Title

A Joint Measurement of ν_{μ}-Disappearance and ν_{e}-Appearance in the NuMI beam using the NOvA Experiment

Author

Nayak, Nitish

Publication Date

2021

UNIVERSITY OF CALIFORNIA,
IRVINE

A Joint Measurement of ν_μ-Disappearance and ν_e-Appearance in the NuMI beam using the NOvA Experiment

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Physics

by

Nitish Nayak

Dissertation Committee:
Prof. Jianming Bian, Chair
Prof. Henry Sobel
Prof. Juan Pedro Ochoa-Ricoux

2021

DEDICATION

To my uncle, Late Sh. Chidambara Jois

TABLE OF CONTENTS

Page

LIST OF FIGURES **vi**

LIST OF TABLES **xiii**

LIST OF ALGORITHMS **xiv**

ACKNOWLEDGMENTS **xv**

VITA **xvii**

ABSTRACT OF THE DISSERTATION **xix**

1 Historical Overview of Neutrino Physics **1**
- 1.1 Neutrinos and the Weak Force 1
- 1.2 The Electroweak Force 5
- 1.3 Astrophysical Neutrino Anomalies and Resolution 10
  - 1.3.1 The Solar Neutrino Anomaly 10
  - 1.3.2 The Atmospheric Neutrino Anomaly 11
  - 1.3.3 Discovery of Neutrino Oscillations 12
- 1.4 Neutrino Oscillation Phenomenology 15
  - 1.4.1 Two Flavor Mixing 15
  - 1.4.2 Neutrino Propagation in Matter 16
- 1.5 Current Paradigm and Open Questions 20
  - 1.5.1 The ν-Standard Model 20
  - 1.5.2 Three Flavor Mixing 23
  - 1.5.3 Open Questions 24
  - 1.5.4 Implications and the Future 29

2 The NOvA Experiment **34**
- 2.1 Introduction 34
- 2.2 The NuMI beam 36
  - 2.2.1 Off-Axis Design 39
  - 2.2.2 Flux Simulation 40
- 2.3 The NOvA Detectors 43
  - 2.3.1 Data Acquisition System 47

2.3.2 Detector Simulation . . . 50
2.3.3 Detector Calibration . . . 54
2.4 Event Reconstruction . . . 60
2.4.1 Event Clustering . . . 62
2.4.2 Vertex Finding . . . 69
2.4.3 Particle Clustering and Tracking . . . 71

3 The Oscillation Analysis **75**
3.1 Particle Identification . . . 75
3.1.1 CVN . . . 76
3.2 Energy Estimation . . . 82
3.2.1 ν_μ-CC Events . . . 82
3.2.2 ν_e-CC Events . . . 85
3.2.3 An Alternate ν_e-CC Energy Estimator . . . 87
3.3 Far Detector Selection and Binning . . . 95
3.3.1 The $\nu_\mu \rightarrow \nu_\mu$ Channel . . . 98
3.3.2 The $\nu_\mu \rightarrow \nu_e$ Channel . . . 104
3.4 Near Detector Constraints . . . 115
3.4.1 Selection . . . 115
3.4.2 The Cross-Section Tune . . . 118
3.4.3 ND → FD Signal Extrapolation . . . 122
3.4.4 Estimation of Beam Backgrounds . . . 127
3.5 Far Detector Prediction . . . 131

4 Systematic Uncertainties **135**
4.1 Flux Uncertainties . . . 137
4.1.1 Hadron Production . . . 137
4.1.2 Beam Transport . . . 139
4.1.3 Principal Component Analysis . . . 140
4.2 Cross-Section Uncertainties . . . 146
4.3 Calibration and Detector Response . . . 149
4.4 Neutron Response . . . 151
4.5 Lepton Reconstruction . . . 152
4.6 ND-FD Acceptance Differences . . . 153
4.7 ND Pile-up . . . 153
4.8 Other Systematics . . . 154
4.8.1 Detector Mass and POT Accounting . . . 154
4.8.2 ν_τ Scale . . . 155
4.8.3 Michel Electron Tagging Efficiency . . . 156
4.8.4 FD Rock Rate . . . 156
4.8.5 Cosmic Scale . . . 156
4.9 Evaluation of Systematic Uncertainties . . . 156

5 **Results** **159**
5.1 Sidebands 160
5.1.1 The $\nu_\mu \rightarrow \nu_\mu$ Channel 160
5.1.2 The $\nu_\mu \rightarrow \nu_e$ Channel 161
5.2 Opening the Box 163
5.2.1 Compatibility with 3-Flavor Mixing 168
5.2.2 Evidence of $\bar{\nu}_e$ Appearance 169

6 **Confidence Intervals** **172**
6.1 Feldman-Cousins 172
6.2 Results 177
6.2.1 Compatibility with Other Experiments 179
6.3 Accelerating Feldman-Cousins 180
6.3.1 Gaussian Process 182
6.3.2 Optimizing the FC Interval Construction 185
6.3.3 Toy Studies 188
6.3.4 Application on NOvA Data 192
6.3.5 Future Prospects for the $\mathcal{GP}$ Algorithm 195
6.4 Future Prospects at NOvA 196

7 **Conclusions** **199**

Bibliography **201**

LIST OF FIGURES

Page

1.1 The Fermi 4-point interaction for neutron decay 2
1.2 The Savannah River neutrino detector and the delayed coincidence signature 3
1.3 Neutron Decay via Massive Intermediate Vector Bosons 5
1.4 The Gargamelle discovery of Weak Neutral Currents. 9
1.5 The Predicted Solar Neutrino Flux based on the Standard Solar Model (SSM) for different fusion processes. Taken from [31] 11
1.6 Observation of Up-Down Asymmetry in the Atmospheric Neutrino Flux by the Super-Kamiokande Experiment 12
1.7 Observed Flux Rates for different neutrino flavors from B^8 solar neutrinos at SNO 13
1.8 The KamLAND Observation of L/E-dependent oscillations in a reactor $\bar{\nu}_e$ flux 14
1.9 The Current Picture with Solar Neutrinos 19
1.10 Measurement of Solar Neutrino Oscillation Parameters 20
1.11 The Neutrino Mass Ordering Picture 24
1.12 Observation of non-zero θ_{13} by the Daya-Bay experiment. [48] 25
1.13 Sensitivities of the DUNE experiment to resolve remaining oscillation parameters 26
1.14 Sensitivities of the Hyper-K experiment to measure CP-Violation and the Neutrino Mass Hierarchy for a planned 10 years of running 26
1.15 The ν_μ Disappearance Spectrum 27
1.16 The Bi-Probability Plot at NOvA 29
1.17 Exclusion of Sterile Neutrinos 30
1.18 The He^3 β decay Kurie plot for measuring the absolute neutrino mass at KATRIN. Taken from [59]. 31
1.19 The current predictions for the effective Majorana neutrino mass $m_{\beta\beta}$ as a function of the lightest neutrino mass and the neutrino mass hierarchy . . . 31
1.20 A neutrino from SN1987A detected in the north-east corner of the IMB detector 32
1.21 The probability of detection of a supernova from a star of given mass by the SKGd and KamLAND experiments as a function of distance from Earth and neutrino mass hierarchy 33

2.1 Illustration of the baselines for different experiments on the NuMI beam . . 35
2.2 Schematic of the proton beam production at NuMI 36
2.3 The NuMI beamline 37
2.4 The NOvA Flux 38

2.5 Event Rates for different flux components at the NOvA Far Detector 38
2.6 Neutrino Flux Dependence on Off-Axis Angle 39
2.7 Schematic of the NuMI Magnetic Horns 41
2.8 Corrections to the hadron production at the beam target from NA49 42
2.9 Hadron Production Uncertainties at NOvA 43
2.10 Schematic of the NOvA detector 44
2.11 Schematic of the NOvA cell 44
2.12 Snapshot of the NOvA Far Detector showing cosmic ray interactions 45
2.13 Snapshot of the NOvA Near Detector showing various neutrino interactions . 46
2.14 Schematic of the NOvA Data Acquisition System 47
2.15 Illustration of multi-point readout 48
2.16 Trigger rates for various NOvA Triggers at the Far Detector 49
2.17 Feynman Diagrams for various important types of neutrino interactions at NOvA 50
2.18 Current Picture of Neutrino and Anti-Neutrino Cross-Sections 51
2.19 Light Yield in a NOvA cell as a function of the particle $\beta\gamma$ 53
2.20 The tricell selection for the attenuation calibration 55
2.21 The raw PE/cm distributions as a function of longitudinal position inside a cell 56
2.22 The raw PE/cm attenuation profile as a function of longitudinal position in a cell for the Near Detector 57
2.23 Threshold and Shadowing Corrections for the Far Detector 58
2.24 The fitted attenuation profiles including the full LOWESS fit 58
2.25 Profiles of the Reconstructed/True Energy for each hit $> 15\,\mathrm{MeV}$ before and after the calibration 59
2.26 The attenuation corrected PE/cm distributions for the stopping muon sample as a function of distance from the track end 60
2.27 Examples of interaction topologies seen at the NOvA detectors for different types of neutrino interactions 62
2.28 Examples of clustering failures seen at the Near Detector due to pile-up . . . 63
2.29 Illustration of the centroid finding algorithm for TDSlicer 65
2.30 Comparison of performance of TDSlicer vs Slicer4D as a function of beam intensity 66
2.31 Example of the same event as before showing TDSlicer is able to separate out interactions compared to Slicer4D 67
2.32 Illustration of the special overlay sample 68
2.33 Illustration of the Elastic Arms algorithm using Hough lines drawn between pairs of points in the hit cluster 70
2.34 Elastic-Arms vertex resolution 71
2.35 A neutrino interaction shown with the reconstructed Fuzzy-K prongs in each view 72
2.36 Performance of the Fuzzy-K reconstruction on electron shower cascades . . . 73
2.37 Performance of the Kalman filter-based reconstruction on muon tracks . . . 74

3.1 Illustration of the convolutional neural network acting on a ν_μ-CC event . . . 77
3.2 The modified MobileNet architecture used for the NOvA CNN PID 79

3.3 The performance of the CVN PID given by the efficiency of the network to identify the true label 81
3.4 The performance of the CVN PID given by the fraction of the predicted labels from the network that are the true label 81
3.5 The fit to the track length of the identified muon for the muon energies . . . 83
3.6 The spline fit to the visible hadronic energy of the event for the hadronic energy estimate 84
3.7 ν_μ-CC energy resolutions for various interaction types 84
3.8 The true ν_e-CC energy as a function of the individual EM and hadronic components weighted by event rates 85
3.9 The ν_e-CC energy resolutions 86
3.10 ν_e-CC energy resolutions as a function of True ν_e-CC energy 86
3.11 The ν_e-CC true energy distributions flattened for input into the training . . 88
3.12 The architecture of the regression CNN for the ν_e-CC energies 89
3.13 Comparison of the performance of the regression CNN based energy estimator for FHC vs traditional approaches as a function of true neutrino energy . . . 91
3.14 Comparison of the performance of the regression CNN based energy estimator for RHC vs traditional approaches as a function of true neutrino energy . . . 92
3.15 Comparison of the performance of the regression CNN based energy estimator for FHC vs traditional approaches as a function of true inelasticity 93
3.16 Comparison of the performance of the regression CNN based energy estimator for RHC vs traditional approaches as a function of true inelasticity 94
3.17 Comparison for the performance of the regression CNN based energy estimator for FHC vs traditional approaches for ν_e-CC quasi-elastic events 94
3.18 Comparison for the performance of the regression CNN based energy estimator for FHC vs traditional approaches for ν_e-CC DIS events 95
3.19 Bias sensitivity of the regression CNN based energy estimator when the pixel inputs are uniformly scaled by ±5% 95
3.20 The different timing windows of the original $550\,\mu$s trigger window used in the analysis 96
3.21 Schematic of the cutflow used for the selection in the disappearance channel 99
3.22 The cosmic rejection BDT used in the disappearance channel 100
3.23 The 4 quantiles of hadronic energy fraction 102
3.24 The ν_μ-CC energy resolutions for different hadronic energy fraction quantiles 104
3.25 Schematic of the cutflow used for the selection in the appearance channel . . 105
3.26 The response of the cosmic rejection BDT used for the core selection 106
3.27 2D distributions of signal purity (color scale) and signal content (box size) in CoreBDT-CVN space used in the core selection 107
3.28 Example of a peripheral ν_e-CC event near the edges of the Far Detector . . . 109
3.29 The response of the cosmic rejection BDT in the peripheral selection for the appearance analysis 110
3.30 2D distributions of signal purity (color scale) and signal content (box size) in PeriBDT-CVN space used in the peripheral selection 111
3.31 Example of a delayed bremsstrahlung shower detached from the parent muon that produced it 113

3.32 The time delay between closest-in-time slices for cosmics and beam events . 114
3.33 Box cuts for events in the time gap peak for closest-in-time slices 114
3.34 The predicted and observed event rates for the ν_μ-CC sample at the Near Detector 116
3.35 The predicted and observed event rates for the ν_e-CC sample at the Near Detector 117
3.36 Comparison of the RFG model used in previous NOvA analyses and the LFG model used now 118
3.37 Corrections to the GENIE hN FSI model based on external pion scattering measurements used in the NOvA tune 119
3.38 The observed distributions along with the out of the box GENIE3 prediction for the visible hadronic energy at the NOvA Near Detector 120
3.39 The phase-space of MEC events in $(q_0, |\vec{q}|)$ space 121
3.40 The distributions of visible hadronic energy at the NOvA Near Detector after the NOvA tune 121
3.41 The distributions of reconstructed transfer momentum at the NOvA Near Detector after the NOvA tune 122
3.42 Illustration of the NOvA signal extrapolation from ND to FD for the disappearance channel 123
3.43 Illustration of the NOvA signal extrapolation from ND to FD for the appearance channel 123
3.44 Selection efficiency as a function of true transverse lepton momentum at the two detectors 125
3.45 Distributions of true transverse lepton momentum for selected events at the two detectors 125
3.46 Quantile boundaries for reconstructed transverse lepton momentum in the appearance channel as a function of the reconstructed neutrino energy at the Far Detector 126
3.47 The ν_μ-CC contained (left) and uncontained (right) samples used for the beam ν_e-CC constraint at the Near Detector 128
3.48 2D phase-space distributions of the parent pion in terms of its $(p_t,\, p_z)$ 128
3.49 The spectrum of the number of Michel e^-s for various components in the ν_e-CC selection at the Near Detector 129
3.50 The uncorrected and corrected distributions for the ν_e-CC sample at the Near Detector 130
3.51 Predicted FD spectra for the appearance channel 132
3.52 Predicted FD spectra for the disappearance channel for FHC using best fit oscillation parameters obtained from a previous analysis 133
3.53 Predicted FD spectra for the disappearance channel for RHC using best fit oscillation parameters obtained from a previous analysis 134

4.1 Illustration of the interpolation procedure for each analysis bin in the appearance channel for an example light model systematic at the Near Detector . . 136
4.2 The correlation matrix based on the hadron production error for the predicted PPFX flux 138

4.3 Different statistical ensembles of the PPFX weights for the predicted ν_μ-CC flux at the Far Detector based on the Hadron production error 141
4.4 The covariance matrix representing the full systematic error in the flux model including hadron production and beam transport 142
4.5 The eigenvalue profile of the flux covariance matrix in the F/N basis 143
4.6 Fraction of bins where the summed error in quadrature from the first N Principle components cover the flux error to $> 99\%$ 144
4.7 Coverage of the overall flux error for the first 5 Principle Components in terms of sensitivities for measuring CP violation or rejecting Inverted Hierarchy based on Asimov data 145
4.8 The total systematic error in the NOvA cross-section model as a function of visible hadronic energy 149
4.9 Distribution of prong energies for a selected neutron sample at the Near Detector broken down into neutron daughters 151
4.10 Total systematic error on the reconstructed muon length based on various effects coming from GEANT or detector mass accounting 152
4.11 Selection efficiencies for the special overlay sample at ND as a function of beam intensity 154
4.12 Comparison of the extrapolated and unextrapolated systematic error on the number of predicted events in the appearance channel for different uncertainties 157
4.13 Overall systematic error on the measurement of δ_{CP} (left) and $\sin^2\theta_{23}$ (right) 157
4.14 Overall systematic error on the measurement of Δm^2_{32} 158

5.1 Exposure in terms of POT for the current analysis 159
5.2 Sideband for the disappearance channel 161
5.3 Core sidebands for the appearance channel 161
5.4 Peripheral sideband for the appearance channel 162
5.5 Oscillated spectra in the appearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error 163
5.6 Oscillated spectra in the disappearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error 163
5.7 Oscillated spectra for each quantile in FHC for the disappearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error 164
5.8 Oscillated spectra for each quantile in RHC for the disappearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error 165
5.9 Total event counts in FHC and RHC in the appearance channel plotted with the best-fit expectations in the 4 quadrants of Hierarchy-Octant space 166
5.10 Total event counts in FHC and RHC in the appearance channel plotted with expectations at a range of oscillation parameters and as a function of δ_{CP} . . 167
5.11 Systematic pulls in the best-fit expectation to the FD data 167
5.12 Goodness of fit $\Delta\chi^2$ for FD data shown along with the $\Delta\chi^2$ distribution for many pseudo-experiments 168
5.13 Significance of $\bar{\nu}_e$-appearance measured via many pseudo-experiments assuming no $\bar{\nu}_e$-appearance 170

6.1 Illustration of the Neyman confidence interval construction 173
6.2 Illustration of the Feldman-Cousins method using a likelihood ratio ($\Delta\chi^2$) distribution generated using pseudo-experiments to find the p-value with respect to data 174
6.3 Significance of inverted hierarchy rejection using the FC method 176
6.4 Confidence intervals in $\sin^2\theta_{23} - \delta_{CP}$ space 177
6.5 Confidence intervals in $\sin^2\theta_{23} - \Delta m^2_{32}$ space 178
6.6 Significances of rejection for each value of the given oscillation parameter for various hypotheses of the profiled parameters. 178
6.7 Significances of rejection for each value of δ_{CP} for various hypotheses of the profiled parameters. 179
6.8 Compatibility of the NOvA measurement with various experiments 180
6.9 Illustration of the Bayesian interpolation procedure using Gaussian Processes for some example data 182
6.10 Illustration of the Bayesian interpolation procedure using Gaussian Processes for the percentile (1-pvalue) as a function of δ_{CP} obtained using FC 184
6.11 The percentile (1-p) surface for some toy oscillation data as a function of $\sin^2\theta_{23} - \delta_{CP}$ 186
6.12 The approximate percentile surfaces in the Inverted Hierarchy along with the acquistion function priority for the toy study 187
6.13 Lineshapes for the flux and cross-section used in the toy oscillation study to mimic NOvA shown along with its systematic error band 188
6.14 Mock datasets generated based on Poisson fluctuations of the expectation at a given set of oscillation parameters in the toy study 188
6.15 Illustration of the binomial proportion confidence interval for a given p-value calculated based on a certain number of pseudo-experiments 189
6.16 Results from the $\mathcal{GP}$ algorithm for approximate contours in $\sin^2\theta_{23} - \delta_{CP}$ space for the Normal Hierarchy 190
6.17 Results from the $\mathcal{GP}$ algorithm for approximate contours in $\sin^2\theta_{23} - \delta_{CP}$ space for the Inverted Hierarchy 190
6.18 Comparison of the $\mathcal{GP}$-based significance curves as a function of δ_{CP} vs the FC result in the toy study 191
6.19 Distribution of accuracies of the $\mathcal{GP}$-based approximated confidence intervals for 200 different mock datasets as a function of fractional number of pseudo-experiments thrown 192
6.20 Comparison of the $\mathcal{GP}$ approximate contours after 4 iterations vs the actual NOvA contours in $\sin^2\theta_{23} - \delta_{CP}$ space 193
6.21 Comparison of $\mathcal{GP}$ approximated significance curves for δ_{CP} vs the actual NOvA result 193
6.22 The approximate p-value surface in the $\mathcal{GP}$ procedure for NH after 4 iterations and the points where the pseudo-experiments were thrown 194
6.23 Errors in the $\mathcal{GP}$ approximate contours for NH 194
6.24 The approximate p-value surface in the $\mathcal{GP}$ procedure for IH after 4 iterations and the points where the pseudo-experiments were thrown 195
6.25 Errors in the $\mathcal{GP}$ approximate contours for IH 195

6.26 Sensitivities for the measurement of CP-violation and Mass Hierarchy at the end of the NOvA lifespan 197
6.27 Potential for rejecting maximal mixing and determination of the θ_{23} Octant based on current analysis techniques at NOvA 197

LIST OF TABLES

Page

2.1 Comparison of Selection Efficiencies at various cut stages for MC on MC spills and MC on Data spills using the special overlay sample reconstructed with TDSlicer . . . 67
2.2 Comparison of Selection Efficiencies at various cut stages for MC on MC spills and MC on Data spills using the special overlay sample reconstructed with Slicer4D . . . 68

3.1 Predicted signal and background components along with the signal selection efficiencies and purities in the disappearance channel at various cut stages for FHC . . . 101
3.2 Predicted signal and background components along with the signal selection efficiencies and purities in the disappearance channel at various cut stages for RHC . . . 101
3.3 Predicted signal and background components with the signal selection efficiencies and purities in the appearance channel at various cut stages for FHC 111
3.4 Predicted signal and background components with the signal selection efficiencies and purities in the appearance channel at various cut stages for RHC 112
3.5 Predicted event rates for various components and PID bins in the appearance channel using best fit oscillation parameters obtained from a previous analysis 133
3.6 Predicted event rates for various components and quantiles in the disappearance channel using best fit oscillation parameters obtained from a previous analysis . . . 134

5.1 Breakdown of individual components in the appearance spectrum at the best fit oscillation parameters . . . 171
5.2 Breakdown of individual components in the disappearance spectrum at the best fit oscillation parameters . . . 171

6.1 Confidence intervals for the individual oscillation parameters in the measurement for each mass hierarchy . . . 179

LIST OF ALGORITHMS

Page
1 $\mathcal{GP}$ iterative confidence interval construction 187

ACKNOWLEDGMENTS

The journey until now has been long-winded and arduous, but I feel extremely fortunate to be in this position currently. There were many points along the way, even before I was in Irvine, where I felt that a career in physics was likely not going to happen. While you never know what's in store later on, it's nice to be able take a step back at this juncture atleast and thank the many people over the years who've given me the opportunity to learn from, contribute towards and experience working in a large particle physics experiment, which can sometimes seem dysfunctional and hectic, but is never not interesting.

First, I'd like to thank my advisor, Jianming Bian, who gave me my start at NOvA and who was always available to answer my dumb questions patiently and with whom I've shared long and fun discussions about our research topics. Thank you as well to other senior people in the Irvine neutrino group – Hank Sobel, Pedro and Michael Smy who between them and Jianming it'd seem, know pretty much all there is to know about neutrino physics. I always felt like I learnt something new and interesting at our weekly meetings. Thank you as well to Jeff Griskevic, our lab overlord at the High Bay who managed to eke out some functional detector R&D work even from a klutz like me.

Thank you to my many many conveners and bosses at NOvA – Chris Backhouse, Alex Himmel, Jeremy Wolcott, Louise Suter, Michael Baird, Evan Niner, Karl Warburton, Fernanda Psihas, Ryan Nichol and Lisa Koerner for the years of mentorship and advice, for putting up with my haphazard talks and generally being the kind of people I'd want to emulate in the future. Thank you as well to other senior members at NOvA – Peter Shanahan, Mark Messier and Patricia Vahle for being approachable and in particular, giving me the opportunity to contribute towards our first anti-neutrino publication, which will remain a professional and personal highlight.

Thank you to the many graduate students and post-docs who've shared this journey with me and who I'm lucky to call friends. From Irvine – Josh, Hasitha, Jingyuan, Yuqi for the many movie nights, Scott, Pierce, Yvonne and Taylor for making office seem bearable, Yiwen for putting up with my rambling explanations for things, Lingge and Lars for being awesome to work with. Ilsoo and Wenjie, whose work ethic as post-docs is something I hope to carry forward myself. Our drinking group – Beda, Ed, Roberto, Ben, Igor, Bri and Olivia for well, making it easy to get through grad school. Thanks as well to Geoff and Steve for the fun pub crawls. To my former Desi roommates – Jawad and Bodhi for making me feel less homesick.

Huge thanks to all the grad students and post-docs from NOvA, especially from the 3-flavor group with whom I've had the pleasure of working with for many years – Shiqi, Liudmila, Abhilash, Micah, Erika, Ashley, Diana, Bruce, Tyler, Alex, Andrew, Derek, Steven, Zoya and many many more. Thanks as well to the Desi group at NOvA – Kuldeep, Biswa, Rijeesh, Reddy, Prabhjot and Sijith for helping me out a ton and making me feel like home on my first visit to Fermilab, which can be awfully intimidating. Big thanks also to a bunch of school and undergrad friends for the many Discord calls, games of Dota and so on, without which the pandemic would have seemed a lot less bearable.

Finally, I would like to thank my family – my parents and my brother, whose constant support and encouragement is the biggest reason that this is even possible. We can be a stubborn bunch but I like to think that can be a good thing sometimes..

This work was funded by a grant from the United States Department of Energy.

VITA

Nitish Nayak

EDUCATION

Doctor of Philosophy in Physics	**2021**
University of California, Irvine	*Irvine, California*
Master of Science in Physics	**2016**
Indian Institute of Technology, Bombay	*Mumbai, India*
Bachelor of Technology, Electrical Engineering	**2012**
Indian Institute of Technology, Bombay	*Mumbai, India*

RESEARCH EXPERIENCE

Graduate Research Assistant	**2017-2021**
University of California, Irvine	*Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant	**2016-2017**
University of California, Irvine	*Irvine, California*

SELECTED PUBLICATIONS

Efficient neutrino oscillation parameter inference using Gaussian processes	**2020**
L. Li, N. Nayak, J. Bian, P. Baldi	PhysRevD.101.012001
First Measurement of Neutrino Oscillation Parameters using Neutrinos and Antineutrinos by NOvA	**2019**
M. Acero et al.	PhysRevLett.123.151803
New constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in the NOvA experiment	**2018**
M. Acero et al.	PhysRevD.98.032012

ABSTRACT OF THE DISSERTATION

A Joint Measurement of ν_μ-Disappearance and ν_e-Appearance in the NuMI beam using the NOvA Experiment

By

Nitish Nayak

Doctor of Philosophy in Physics

University of California, Irvine, 2021

Prof. Jianming Bian, Chair

The discovery of neutrino oscillations provides the first indication of a lepton flavor violating (LFV) process, one that isn't predicted by the Standard Model. As such, NOvA is part of a rich experimental program to constrain unknown parameters in the neutrino oscillation model, described for three neutrino flavors using the PMNS unitary matrix. It is a long-baseline experiment utilizing two detectors, a Near Detector at Fermilab and a Far Detector in Ash River, Minnesota for a total baseline of 810 km. It receives a predominantly $\nu_\mu/\bar{\nu}_\mu$ beam peaking at 1.8 GeV from the NuMI beam facility at Fermilab. There are four oscillation channels used in the analysis, $\nu_\mu \to \nu_\mu$, $\nu_\mu \to \nu_e$, $\bar{\nu}_\mu \to \bar{\nu}_\mu$ and $\bar{\nu}_\mu \to \bar{\nu}_e$. With a total exposure of 13.6×10^{20} and 12.5×10^{20} protons on target for the neutrino and anti-neutrino beam modes respectively, 82 candidates are seen in the $\nu_\mu \to \nu_e$ channel for a total predicted background of 26.8 events. Similarly, 33 candidates are seen in the corresponding anti-neutrino channel for a total predicted background of 14.0 events. In the $\nu_\mu \to \nu_\mu$ ($\bar{\nu}_\mu \to \bar{\nu}_\mu$) channel, 211 (105) candidates are seen with an expectation of 1156.1 (488.1) events at no oscillations.

Consequently, this dissertation reports a measurement for oscillation parameters based on a joint fit for the spectra in these four channels, which is given by :

- $\Delta m^2_{32} = (2.41 \pm 0.07) \times 10^{-3}$ eV^2

- $\sin^2\theta_{23} = 0.57^{+0.04}_{-0.03}$ (UO)

- $\delta_{CP} = 0.82\pi^{+0.27\pi}_{-0.87\pi}$

In addition, a leading 4.2σ confidence level of evidence is seen for $\bar{\nu}_e$ appearance. The oscillation analysis improves upon previous updates in several areas including particle identification, event reconstruction and cosmic background rejection. A principle component analysis (PCA)-based technique is also implemented for decorrelating important flux and cross-section systematics. In addition, new improvements are proposed in areas of energy estimation as well as confidence interval building.

Chapter 1

Historical Overview of Neutrino Physics

1.1 Neutrinos and the Weak Force

The theory of the weak nuclear force is one of the most enigmatic and puzzling aspects of nature. The theory was given its first concrete form in 1934 by Enrico Fermi to explain the instability of the neutron within certain nuclei [1]. A free neutron, discovered just 2 years earlier by James Chadwick [2], would by itself decay with a half-life of roughly 15 minutes. By incorporating Wolfgang Pauli's idea that the decay of these radioactive nuclei ought to be a 3 body process involving a neutrino [3], Fermi was able to write down a 4-point interaction which is described by a Lagrangian density term :

$$\mathcal{L}_{fermi} = -\frac{G_F}{\sqrt{2}}[\bar{p}\gamma_\mu n][\bar{e}\gamma^\mu \nu] \tag{1.1}$$

This interaction has a number of interesting features.

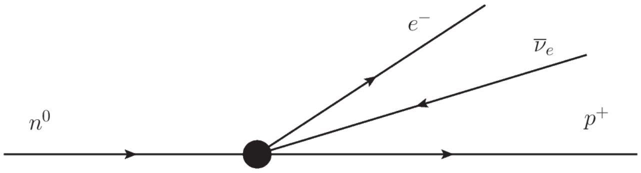


Figure 1.1: The Fermi 4-point interaction for neutron decay

- The neutrino was needed to explain the continuous electron energy spectrum for radioactive nuclei. In particular, a radioactive process described by $N_Z^A \rightarrow N_{Z+1}^A + e^-$ would only be able to emit electrons with discrete energies, which was inconsistent with experimental observations.

- The Fermi constant, G_F is not dimensionless like in the electromagnetic case, ($\alpha = \frac{1}{137}$) and has units of M^{-2}. The value was determined to be $\sim 1.17 \times 10^{-5}$ GeV^{-2} [4]. This is an extremely small number relative to ordinary electrodynamics at the atomic scale. Just for perspective, if we write, $G_F \sim \frac{\alpha}{M_W^2}$, where M_W is the characteristic weak scale and α is the electromagnetic coupling strength, we'd have $M_W \sim 25$ GeV, i.e relevant only at distances ~ 25 times smaller than the proton.

- The theory predicted the inverse beta-decay (IBD) reaction with a cross-section of $\sim \mathcal{O}(10^{-44})$ cm^2 : $\bar{\nu} + p \rightarrow n + e^+$ [5]. This was exploited to prove the existence of the neutrino in 1956 by Clyde Cowan and Frederick Reines [6]. A detector made up of 200 liters of water, acting as a target for the antineutrino flux from the Savannah River nuclear reactor, was surrounded by a liquid scintillator made up of a highly transparent liquid (toluene) and doped with terphenyl. The scintillator would give out flashes in the visible spectrum from the resulting positrons while the neutrons were detected by dissolving Cadmium chloride ($CdCl_2$) in the tank. Cd is a highly effective neutron absorber, releasing a 9 MeV gamma ray in the process ($n + Cd^{108} \rightarrow Cd^{109} + \gamma$). The light flashes from both the positron and the neutron were detected by photomultiplier tubes (PMTs) placed in the scintillation layer. By keeping this detector 12 m underground to shield it from cosmic rays, Cowan and Reines were able to observe the unique and famous “delayed coincidence” signature associated with the IBD interaction where the flashes from the positron and the neutron were observed with a delay associated with the absorption time of the neutron in Cd (~ 5 μs). This neat technique to detect neutrinos (illustrated in Fig. 1.2) is still in wide use today!

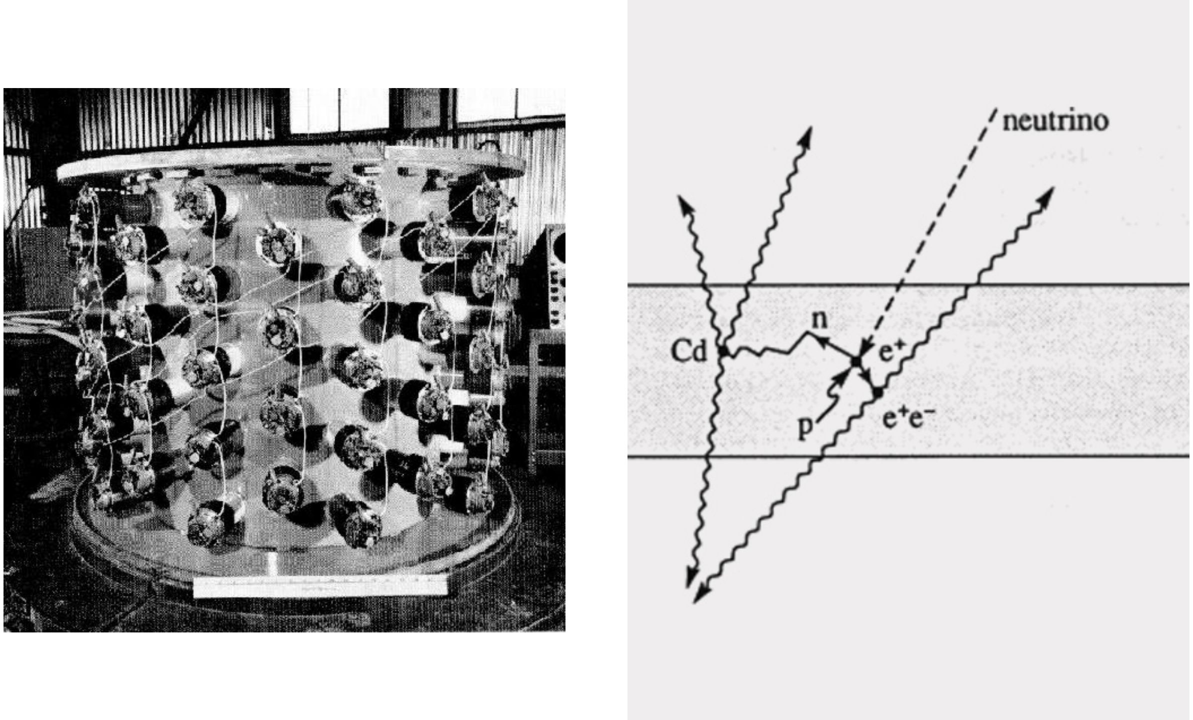


Figure 1.2: The Savannah River neutrino detector (left) and the delayed coincidence signature (right)

Around the same time as Cowan and Reines, Chen-Ning Yang and Tsung-Dao Lee suggested parity is violated in the weak interaction as a way to get rid of the $\tau - \Theta$ puzzle in a class of strange meson decays, now associated with that of the charged kaon, K^+ [7]. Given that two seemingly exactly alike particles with the same mass and overall lifetime (τ, Θ) would decay into 3 and 2 pions (π) respectively, and since the pion has $P = -1$, it'd seem the two decay modes have different parities. If τ and Θ were the same particle then parity must be violated in the weak interaction. This parity violation was verified by Chien-Shiung Wu in a landmark experiment looking at the polarization of β-rays in Co^{60} decays [8]. It was observed that the e^- were emitted in very strong preference to the opposite direction of its spin. The very next year, an experiment developed by Maurice Goldhaber looked at electron captures on Eu^{152} resulting in the emission of a γ-ray and a neutrino moving back to back. By measuring the polarization of the γ-ray, the helicity of the neutrino was determined to be always left-handed [9].

This means that parity is *maximally* violated in the weak interaction and only left-handed (right-handed) neutrinos (anti-neutrinos) are involved. Since neutrinos have only ever been observed from the weak interaction, it was natural to assume that there is no right-handed (left-handed) neutrino (anti-neutrino) and therefore, neutrinos are massless. The Fermi theory of the weak force was also modified to include axial vector currents (V-A theory) that violated parity, i.e

$$\mathcal{L}_W = -\frac{G_F}{\sqrt{2}}[\bar{p}\gamma_\mu(1-\gamma_5)n][\bar{e}\gamma^\mu(1-\gamma_5)\nu] \tag{1.2}$$

In 1962, a different type of neutrino associated with the muon – ν_μ, was discovered by Lederman, Schwartz and Steinberger [10]. In what was a precursor to modern accelerator neutrino experiments, a muon-neutrino "beam" was created by striking accelerated protons onto a fixed Be target producing pions of $\mathcal{O}$(GeV) energies. These would decay into muons and muon neutrinos ($\pi \to \mu + \nu_\mu$), with a small contamination of electron neutrinos arising from the decays of the muons themselves ($\mu \to e + \nu_e + \bar{\nu}_\mu$). The neutrinos were observed by their interaction in a spark chamber, where a ν_μ would produce a long minimum ionizing μ track. A ν_e could be observed with its associated e shower cascade. Modern accelerator experiments still use this technique with the addition of a long cylindrical decay pipe to better collimate the π-decays and hence, the neutrino beam.

The completion of the τ-lepton with the ν_τ discovery would have to wait until 2000, by the DONUT experiment at Fermilab [11].

1.2 The Electroweak Force

The road to a full quantum theory of the weak force started from a symmetry associated with a doublet structure, $SU(2)$, the non-abelian Lie group of rotations in 2 complex dimensions, first developed by C.N Yang and Robert Mills [12] and applied by Sheldon Glashow, Abdus Salam and others [13][14]. The weak doublets,

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_L , \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L , \ldots$$

would be irreducible representations of the Lie group. After the extraordinary success of Quantum Electrodynamics (QED) as a gauge theory associated with the Lie group $U(1)$ of complex phase rotations, it was realized that “gauging” the $SU(2)$ theory, i.e by preserving the theory under local $SU(2)$ transformations – not just global, would give rise to massless intermediate vector bosons that could be interpreted as the carriers of the weak force, since the weak currents were vector or axial vector-like.

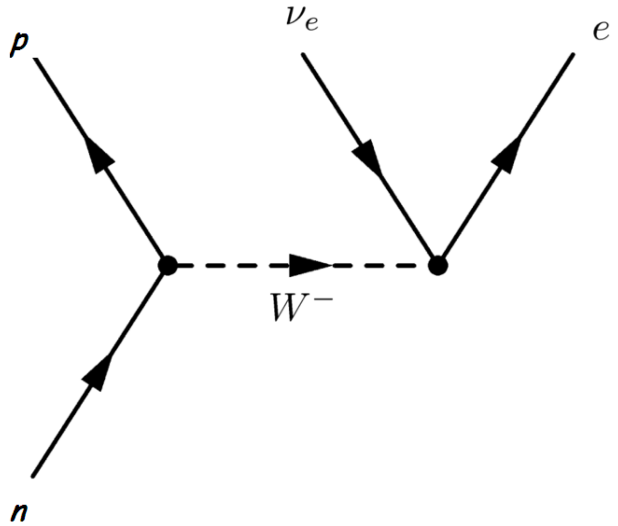


Figure 1.3: Neutron Decay via Massive Intermediate Vector Bosons

However, as discussed above, the weak coupling having a characteristic mass scale meant that the carriers of the weak force must be massive. Any attempts to introduce masses into the Yang-Mills theory however, resulted in bad breakdowns of the gauge theory. Specifically one would expect a large number of massless Goldstone bosons [15] associated with the breaking of this local symmetry, which were nowhere to be seen. In 1964, Peter Higgs, Robert Brout and Francois Englert, along with a few others, reached the conclusion that these Goldstone bosons could be removed by coupling the Yang-Mills fields with a scalar field having a non-vanishing vacuum expectation value [16][17][18]. By the spontaneous and dynamical breaking of the symmetry as the scalar field attains its vacuum expectation energy, the vector bosons would gain a mass through its coupling with the scalar field. The degrees of freedom from the Goldstone bosons would be incorporated into the extra degrees of freedom in the polarization of the vector bosons as they attain a mass.

In 1967, Steven Weinberg used the Higgs mechanism to unify the theories of the weak and electromagnetic forces [19]. In his model of leptons, the underlying gauge symmetry of $SU(2)_L \times U(1)_Y$ denoting the electroweak force was spontaneously broken to $U(1)_Q$ by the Higgs field attaining its vaccum expectation. The leftover $U(1)_Q$ was associated with the electromagnetic force, given by the photon field, A_μ, while the broken $SU(2)_L$ piece gave rise to massive intermediate vector bosons, $W_\mu^+, W_\mu^-, Z_\mu^0$.

Under the electroweak gauge group $SU(2)_L \times U(1)_Y$, the left-handed and right-handed fields would transform in the representation given by :

$$\Psi_L \bigoplus \Psi_R \sim (\mathbf{2},\textbf{-1/2}) \bigoplus (\mathbf{1},\textbf{-1}) \tag{1.3}$$

The Lagrangian would be given by :

$$\mathcal{L}_{EW} = i\bar{\Psi}_L \gamma^\mu D_\mu \Psi_L + i\bar{\Psi}_R \gamma^\mu D_\mu \Psi_R \tag{1.4}$$

where Ψ_L and Ψ_R denote the left-handed and right-handed fermion fields : $\begin{pmatrix} e \\ \nu \end{pmatrix}_L$, (e_R) respectively.

The covariant derivatives D_μ would be given by :

$$D_\mu \Psi_L = (\partial_\mu + ig\tau_a W^a + ig' Y B_\mu) \Psi_L \tag{1.5}$$

$$D_\mu \Psi_R = (\partial_\mu + ig' Y B_\mu) \Psi_R \tag{1.6}$$

where $\tau_a = \sigma_a/2$, the Pauli matrices and W_a, B are the generators of the Lie algebra of $SU(2) \times U(1)$ and associated with the gauge bosons of the theory.

After the symmetry is broken to a $U(1)_Q$, the new gauge bosons can be written as :

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{1.7}$$

$$W_\mu^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}} \tag{1.8}$$

where θ_W is the Weinberg angle denoting the relative strengths of the weak and electromagnetic couplings.

$$\tan\theta_W = \frac{g'}{g} \tag{1.9}$$

$$g_e = g \sin\theta_W \tag{1.10}$$

where $\frac{g_e^2}{4\pi} = \alpha_{em}$. The electroweak interaction for electrons would then be described by the

Lagrangian given by :

$$\mathcal{L}_{EW} \supset g_e \left[\bar{e}\gamma^{\mu} e\right] A_{\mu} - \frac{g}{2\cos\theta_W} \left[\bar{e}\gamma^{\mu} \left(g_V^e - g_A^e \gamma_5\right) e\right] Z_{\mu} \tag{1.11}$$

where $g_V^e = -\frac{1}{2} + 2\sin^2\theta_W$ and $g_A^e = -\frac{1}{2}$ denote the vector-like and axial-vector like couplings respectively and the left and right-handed indices are contracted.

The neutrino interactions written in a chiral form are described by :

$$\mathcal{L}_{EW} \supset -\frac{g}{2\cos\theta_W} \left[\bar{\nu}_R \gamma^{\mu} \nu_L\right] Z_{\mu} - \left\{ \frac{g}{\sqrt{2}} \left[\bar{\nu}_R \gamma^{\mu} e_L\right] W_{\mu}^{+} + h.c \right\} \tag{1.12}$$

These terms are repeated for each of the three lepton generations and consequently the electroweak Feynman rules for scattering amplitudes can be written down.

The electroweak Lagrangian makes evident the chiral V-A structure of the weak force. Only left-handed (right-handed) particles (anti-particles) are involved in the Charged Current (CC) interactions mediated by the W-bosons. Along with the conservation of angular momentum (spin), this results in a suppression of the anti-neutrino cross-sections for large scattering angles compared to the neutrino case. Indeed, the anti-neutrino/neutrino total cross-section ratio for the free scattering case comes out to ~ 0.33.

Significantly, this theory also predicts the existence of weak Neutral Current (NC) interactions mediated by the Z-boson. These were observed experimentally in 1973 by the Gargamelle bubble chamber detector at CERN for both $\nu - e$ and $\nu - p/n$ scattering processes [20].

The development of the Standard Model was completed in the 1970s with the theory of the strong force as a manifestation of a $SU(3)$ gauge symmetry, by David Politzer, Frank Wilczek and David Gross [21][22]. This was inspired by the the experimental discovery of proton sub-structure and the parton model via e-p deep inelastic scattering (DIS) and

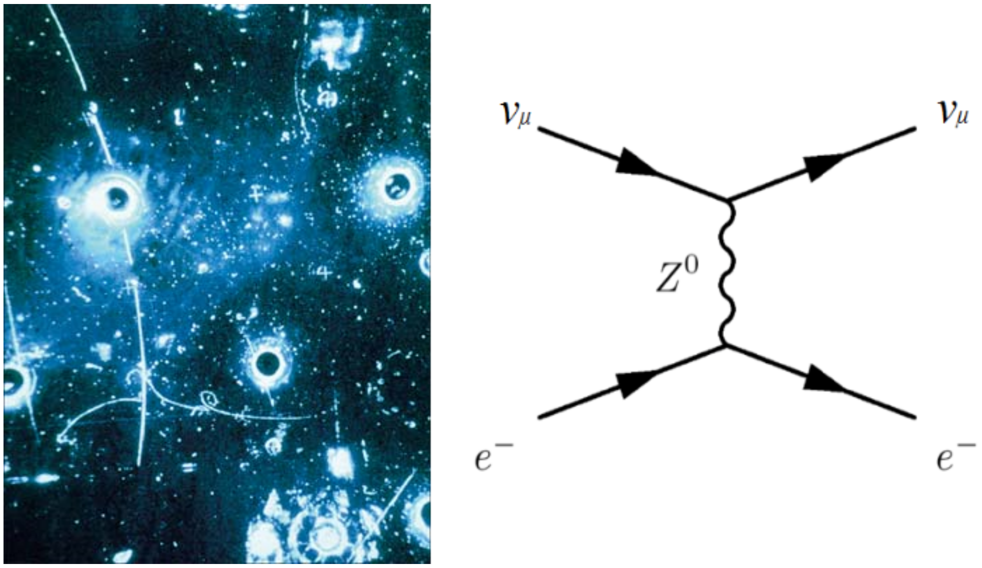


Figure 1.4: The Gargamelle discovery of Weak Neutral Currents. The NC interaction can be seen in the lower half as an electron shower cascading towards the left. [20]

the existence of quarks from the experimental validation of Bjorken scaling in DIS regimes [23][24]. Finally, Gerard 't Hooft showed the renormalizability of Yang-Mills theories, and the self-consistency of the theory was proven [25].

The $W^{\pm}$, Z^0 bosons were discovered at the SPS at CERN in 1983 [26] and their properties measured to high precision at LEP subsequently. Specifically their masses were found to be $M_W = 80.420 \pm 0.003\,\text{GeV}$ and $M_Z = 91.1876 \pm 0.0021\,\text{GeV}$. These masses are related to the Higgs vacuum expectation value and the Weinberg angle as :

$$\begin{aligned} M_W^2 &= \frac{g^2 \nu_{EW}^2}{2} \\ \sin^2 \theta_W &= 1 - \frac{M_Z^2}{M_W^2} \end{aligned}$$

giving $\nu_{EW} \sim 246\,\text{GeV}$ and $\sin^2 \theta_W = 0.2233$. The final piece of the electroweak model, the Higgs boson was discovered at the LHC in 2012 with a mass $M_H = 125.7 \pm 0.4\,\text{GeV}$ [27].

1.3 Astrophysical Neutrino Anomalies and Resolution

As can be appreciated in the previous section, the history of neutrino physics was intimately tied up with the development of the electroweak theory, both experimentally and theoretically. By the end of the 1960s, the picture of a theory necessarily involving chiral fermions coupled to massive vector gauge bosons became clear. Neutrinos, carrying no electrical charge or mass themselves, truly represented this situation as manifestations of chiral interactions.

1.3.1 The Solar Neutrino Anomaly

However, around the same time as Weinberg's model of leptons, a few kinks started developing in this edifice due to experimental observations of an astrophysical nature. An experiment carried out at the Homestake mine in South Dakota by Ray Davis Jr. [28] with theoretical input from John Bahcall [29], set out to observe electron neutrinos from fusion processes within the Sun, which were predicted to be substantial, as shown in Fig. 1.5. The detector made up of $\sim 615\,\mathrm{t}$ of C_2Cl_4 was able to observe these solar neutrinos via the reaction $\nu_e + Cl^{37} \rightarrow e^- + Ar^{37}$. Ar^{37} decays via electron capture from an inner shell of the atom, releasing a characteristic energy which can eject an outer shell electron in the same atom. These electrons were counted to determine the neutrino reaction rate and consequently the solar neutrino flux. The observed capture rate was found to be $\sim 30\%$ of the predicted value from the Standard Solar Model (SSM). The neutrino energy threshold for this reaction was 0.814 MeV which would make it insensitive to the most abundant source of solar neutrinos, namely the pp fusion process.

Subsequent experiments using Gallium, with an energy threshold of 233 keV confirmed this anomaly by measuring a capture rate of $\sim 50\%$ of the predicted value [30]. The Kamiokande and the Super-Kamiokande experiments in Japan, using large cylindrical Cerenkov detectors made up of pure water and surrounded by a large array of PMTs, also observed this deficit for B^8 neutrinos via the Elastic Scattering (ES) reaction $\nu_x + e^- \to \nu_x + e^-$. The measured deficit was $\sim 40\%$.

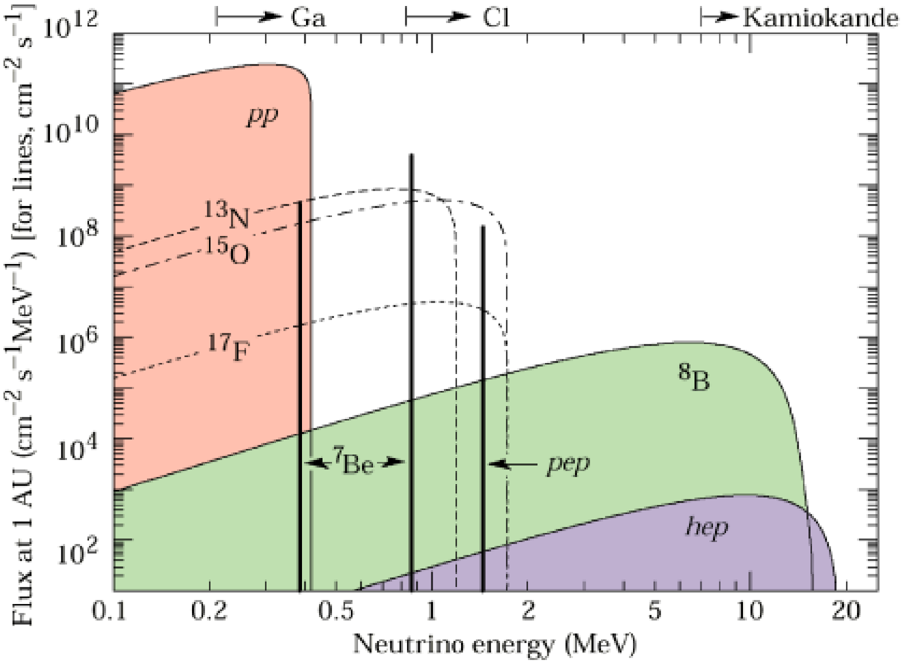


Figure 1.5: The Predicted Solar Neutrino Flux based on the Standard Solar Model (SSM) for different fusion processes. Taken from [31]

1.3.2 The Atmospheric Neutrino Anomaly

Another astrophysical source of neutrinos was cosmic-ray interactions in the upper atmosphere, producing pions and kaons which would decay to primarily ν_μ ($\pi^+ \to \nu_\mu + \mu^+$) of energy $> 0.1\,\text{GeV}$. A small contamination of ν_e was expected from the channel $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$. Therefore, one would roughly expect the flux ratio $\frac{\Phi_{\nu_\mu}+\Phi_{\bar{\nu}_\mu}}{\Phi_{\nu_e}+\Phi_{\bar{\nu}_e}}$ to be ~ 2. Moreover, since neutrinos were assumed massless, this flux ratio would be similar for both upward going and downward going neutrinos, atleast for energies $> 1\,\text{GeV}$, taking into account the asymmetries in the cosmic ray interactions due to Earth's geomagnetic field. For experiments that utilized both upward and downward going atmospheric neutrinos, like Kamiokande, the flux ratio was observed to be ~ 1.2. In 1998, the Super-Kamiokande experiment observed significant asymmetry ($> 6\sigma$) between upward and downward going ν_μ events in their $> 1.33\,\mathrm{GeV}$ μ-like sample, as shown in Fig. 1.6. The measured asymmetry $A = \frac{U-D}{U+D}$ was found to be $-0.296 \pm 0.048 \pm 0.001$. This meant that neutrinos exhibited significant path-dependent effects since $\sim$ half of the upward going ν_μs travelling $\sim 10.000\,\mathrm{km}$ "disappeared" relative to downward ν_μs travelling $\sim 10\,\mathrm{km}$, which could not be explained by the Weinberg model [32].

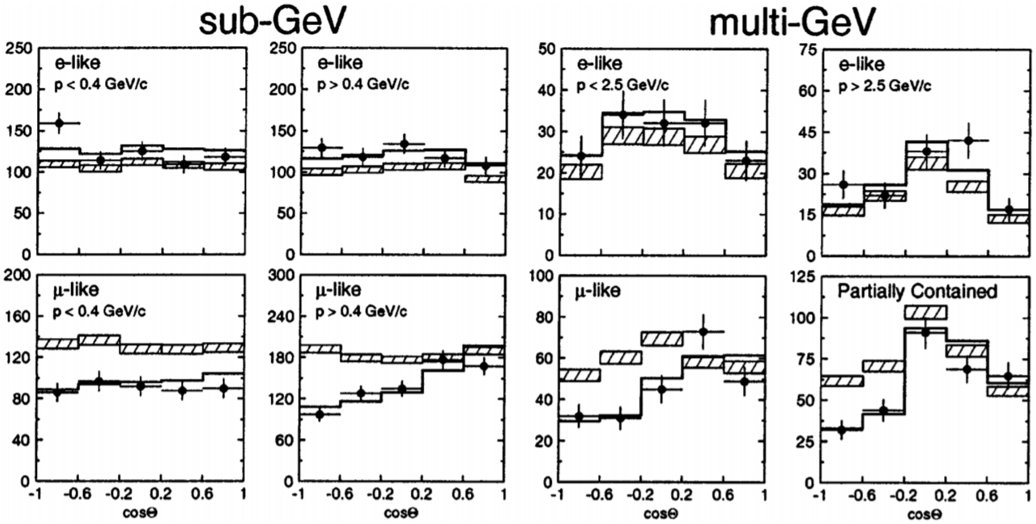


Figure 1.6: Observation of Up-Down Asymmetry in the Atmospheric Neutrino Flux by the Super-Kamiokande Experiment [32]

1.3.3 Discovery of Neutrino Oscillations

By the early 2000s, the picture of neutrino flavor transitions along its propagation length was conclusively established by several experiments, including SNO [33], Super-Kamiokande [32] and KamLAND [34]. As a consequence, the solar and atmospheric neutrino anomalies were explained by realizing that particular neutrino flavors were "disappearing" by mixing into other flavors as they propagated to the detectors on Earth.

The SNO experiment used a spherical detector made up of 1 kt heavy water (D_2O) surrounded by an array of PMTs that were sensitive to B^8 solar neutrinos from three different channels.

- Charged Current (CC) : $\nu_e + d \rightarrow e^- + p + p$
- Neutral Current (NC) : $\nu_x + d \rightarrow \nu_x + p + n$
- Elastic Scattering (ES) : $\nu_x + e^- \rightarrow \nu_x + e^-$

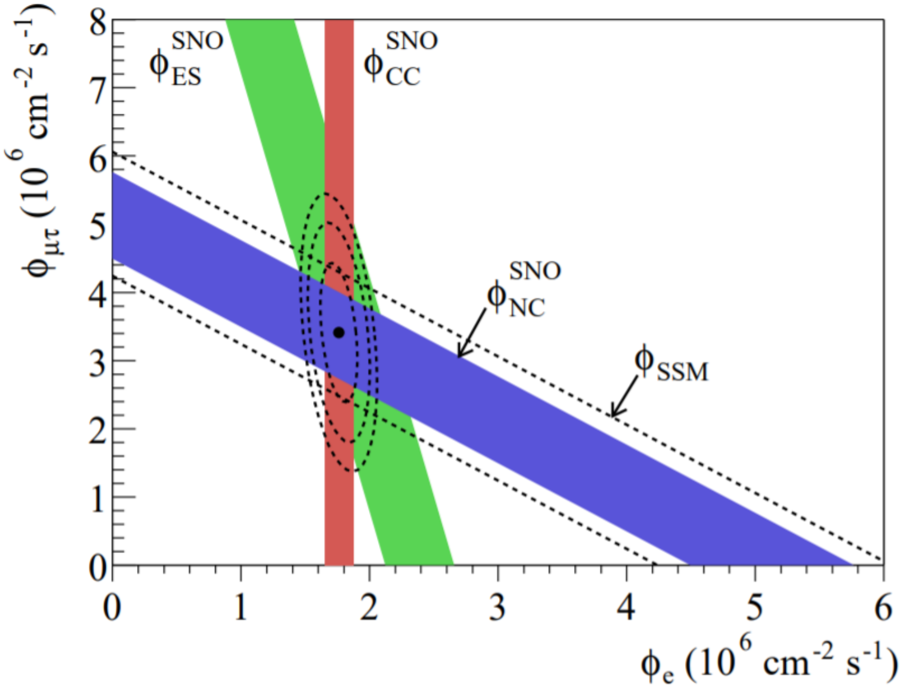


Figure 1.7: Observed Flux Rates for different neutrino flavors from B^8 solar neutrinos at SNO [33]

. The NC and ES channels are flavor insensitive, although the cross-section for ν_e-ES is $\sim 6\times$ that of the other flavors. By 2002, the evidence for a non-ν_e component of solar neutrinos via the NC channel stood at 5.3σ, as can be seen from Fig. 1.7. By adding up the contributions from all these channels, together with the Super-Kamiokande ES measurement, the observed

B^8 neutrino flux was found to match very well with the SSM prediction. In particular, it was found that the CC/NC ratio was ~ 0.5 which is similar to the result from the Homestake experiment.

In addition, the KamLAND experiment in Japan, consisting of 1 kt ultra-pure liquid scintillator, was sensitive to $\bar{\nu}_e$s from nearby nuclear reactors at an average distance of ~ 180 km. In 2002, they reported significant evidence at the $\sim 3\sigma$ level for reactor anti-neutrino disappearance, with a clear oscillatory pattern as a function of L/E where L ~ 180 km and E is the measured energy of the neutrino in a later update, as shown in Fig. 1.8. Together with the observation of the path-dependent asymmetry of atmospheric neutrinos by the Super-Kamiokande experiment, a consistent explanation relating the observed $\nu_\mu, \nu_e, \bar{\nu}_e$ disappearances to L/E-dependent neutrino flavor oscillations was established. This would necessarily involve non-zero neutrino masses and hence require an update to the Standard Model to take into account this phenomenon. .

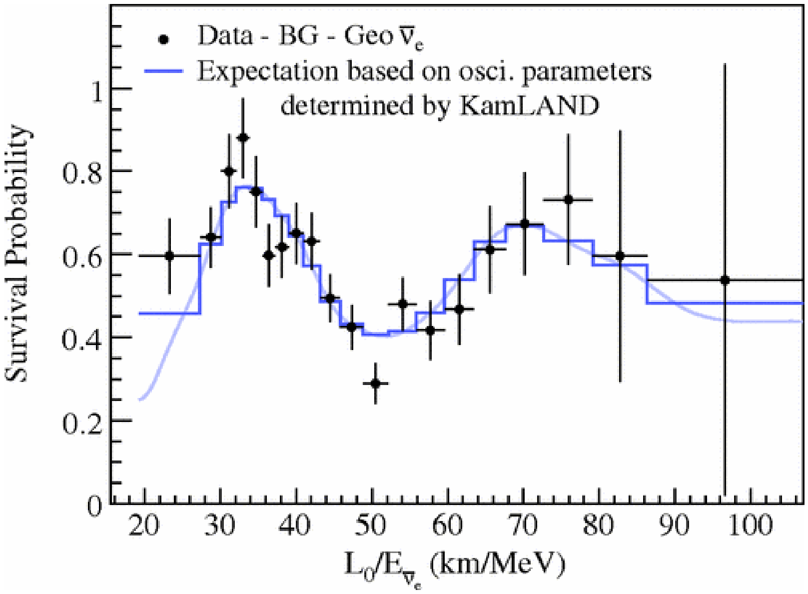


Figure 1.8: The KamLAND Observation of L/E-dependent oscillations in a reactor $\bar{\nu}_e$ flux [35]

1.4 Neutrino Oscillation Phenomenology

To be able to understand neutrino flavor transitions, a couple of key tools are developed.

1.4.1 Two Flavor Mixing

Let us assume two neutrino flavor states : $|\nu_a\rangle$ and $|\nu_b\rangle$ related to its mass-eigenstates : $|\nu_1\rangle$ and $|\nu_2\rangle$ by a basis transformation :

$$\begin{pmatrix} |\nu_a\rangle \\ |\nu_b\rangle \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} \tag{1.13}$$

Assuming ultra-relativistic neutrinos, one can write down the Schrodinger equation of motion in the mass eigenstate basis as :

$$|\nu_i(t)\rangle = e^{-iHt}\,|\nu_i(0)\rangle = e^{-iE_it}\,|\nu_i(0)\rangle \tag{1.14}$$

Identifying t with L, the length of propagation and using the approximation

$$E_i = \sqrt{p_i^2 + m_i^2} \approx p_i + \frac{m_i^2}{2p_i} \approx E_\nu + \frac{m_i^2}{2E_\nu} \tag{1.15}$$

where E_ν is the energy of the neutrino. One then obtains :

$$|\nu_i(L)\rangle = e^{-iE_\nu L} e^{-i\frac{m_i^2 L}{2E_\nu}}\,|\nu_i(0)\rangle \tag{1.16}$$

The disappearance probability, $P_{\nu_a \to \nu_a}(L)$ is then given by :

$$|\,\langle\nu_a|\nu_a(L)\rangle\,|^2 = |\sum_{i=1,2} R_{ai}^2 e^{-i\frac{m_i^2 L}{2E_\nu}}\,|^2 \tag{1.17}$$

where R is the rotation matrix in Eqn. 1.13. This can be written as

$$P_{\nu_a \to \nu_a}(L) = 1 - \sin^2 2\theta \sin^2 \left(\frac{\Delta m^2 L}{4E_\nu} \right) \tag{1.18}$$

$$= 1 - \sin^2 2\theta \sin^2 \left(\frac{1.27 \Delta m^2 [eV^2] L[km]}{E_\nu [GeV]} \right) \tag{1.19}$$

Here, $\Delta m^2 = m_2^2 - m_1^2$. Here θ would denote a constant mixing angle while the oscillations would be L/E dependent. One can immediately apply this to the Super-Kamiokande atmospheric neutrino result by realizing that for the large negative zenith angle case ($\cos\theta_z = -1$) where ν_μ are upward going, the large baseline across the diameter of the earth would imply that the oscillations will be averaged out and the disappearance is given by :

$$\left\langle P_{\nu_\mu \to \nu_\mu} \right\rangle = 1 - \frac{\sin^2 2\theta}{2} \approx \frac{1}{2} \implies \theta \approx \frac{\pi}{4} \tag{1.20}$$

Combined with the fact there wasn't an observed excess for e-like events, the Super-Kamiokande data can be therefore described very well with $\nu_\mu \to \nu_\tau$ oscillations in the 2-flavor approximation with $\Delta m^2 \approx 2.5 \times 10^{-3}$ eV^2 and $\theta \approx \frac{\pi}{4}$.

1.4.2 Neutrino Propagation in Matter

To be able to understand the solar neutrino results, let us first begin by describing how neutrinos propagate in matter using the structure of the electroweak model. We will consider the elastic forward scattering case where neutrino states exist in the final state. The interactions therefore are either NC flavor independent with static e, p, n in the matter medium or CC involving ν_e and e^- only. In what follows, we'll only look at the CC case. As has been shown elsewhere, the NC cancels out completely for e and p if the medium is neutral and only depends on n_n, the neutron number density [36]. Moreover, since the effective matter potential is the same for all flavors in the NC case, it does not affect flavor transitions. For the CC case, one can write down the effective Lagrangian for the neutrino state :

$$L_e \supset 2\sqrt{2}G_F[\bar{\nu}_R\gamma^\mu e_L][\bar{e}_R\gamma_\mu\nu_L] \tag{1.21}$$

Since the electron field is static, the only non-zero component of the above potential is the γ^0 component. The Lagrangian then reduces to :

$$L_e \supset 2\sqrt{2}G_F[\bar{\nu}_R\gamma^0 e_L][\bar{e}_R\gamma_0\nu_L] \tag{1.22}$$

Averaging this operator over the electron polarization states, one obtains an effective matter potential for ν_e states :

$$V = 2\sqrt{2}G_F\left\langle (e_L)^\dagger e_L\right\rangle = \sqrt{2}G_F n_e \tag{1.23}$$

where n_e is the electron number density. The $\bar{\nu}_e$ state potential is given by $V = -\sqrt{2}G_F n_e$.

The Schrodinger equation of motion in the mass basis, taking this matter potential into account is given by :

$$i\frac{d}{dL}\begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} = \left[\begin{pmatrix} \frac{m_1^2}{2E} & 0 \\ 0 & \frac{m_2^2}{2E} \end{pmatrix} + V_{matter}\right]\begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} \tag{1.24}$$

V_{matter} can be written as :

$$V_{matter} = R(\theta)^T V R(\theta) \tag{1.25}$$

where $R(\theta)$ is the rotation matrix from Eqn. 1.13. When ν_e is identified with ν_a from Eqn.

1.13, V is given by :

$$V = \begin{pmatrix} \sqrt{2} G_F n_e & 0 \\ 0 & 0 \end{pmatrix} \tag{1.26}$$

One can then write down :

$$i \frac{d}{dL} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} = \frac{1}{2E} \begin{pmatrix} m_1^2 + A\cos^2\theta & A\cos\theta\sin\theta \\ A\cos\theta\sin\theta & m_2^2 + A\sin^2\theta \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} \tag{1.27}$$

where $A = 2\sqrt{2} E G_F n_e$. In general, this equation is very hard to solve if A is an L-dependent quantity, like in the case of the sun, and one is forced to resort to approximations. If A is constant however, for eg. the case of neutrino propagation in the Earth's crust, one can write down an effective mixing angle and Δm^2 by diagonalizing this matrix and following the procedure in the previous section.

$$\Delta m^2_{eff} = \sqrt{\left(\Delta m^2 \cos 2\theta - A\right)^2 + \left(\Delta m^2 \sin 2\theta\right)^2} \tag{1.28}$$

$$\sin 2\theta_{eff} = \frac{\sin 2\theta}{\sqrt{\left(\cos 2\theta - \frac{A}{\Delta m^2}\right)^2 + \sin^2 2\theta}} \tag{1.29}$$

. The anti-neutrino case is described by $A \to -A$ in the above equations. Fortunately, the solar neutrino data can be described well by an adiabatic approximation, where the matter potential doesn't vary too quickly. It turns out then that higher energy B^8 solar neutrinos are produced in the regime, $A \gg \Delta m^2 \cos 2\theta$, where $\theta_{eff} \sim \frac{\pi}{2}$. Consequently, the produced ν_e is effectively the larger mass eigenstate and the survival probability is just $|\langle \nu_e | \nu_2 \rangle|^2 = \sin^2\theta \approx 0.3$. For smaller energy solar neutrinos ($< 1\,\mathrm{MeV}$) in the pp fusion process, the produced ν_e are in the regime $A \ll \Delta m^2 \cos 2\theta$ and hence propagate as if they

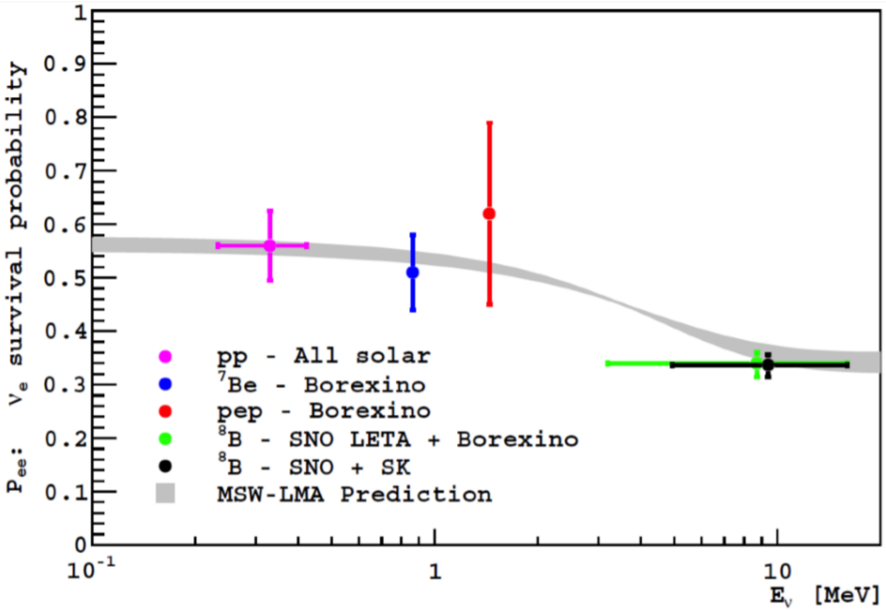


Figure 1.9: Solar Neutrino Survival Probability as a function of Energy conforming to the Adiabatic approximation of Solar densities in the MSW Large Mixing Angle (LMA) solution. [37]

were in vacuum. The survival probability is just given exactly as Eqn. 1.20 :

$$\langle P_{\nu_e \to \nu_e} \rangle = 1 - \frac{\sin^2 2\theta}{2} \approx 0.58 \tag{1.30}$$

which agrees with the measured value. This is known as the MSW effect [38][39]. Fig. 1.9 shows the current fit to the solar neutrino data assuming the MSW-Large Mixing Angle hypothesis. Independent confirmation for the mixing angle was found by the KamLAND experiment. The value of Δm^2 from a combined fit to solar neutrino and KamLAND data was found to be 7.5×10^{-5} eV^2 as shown in Fig. 1.10.

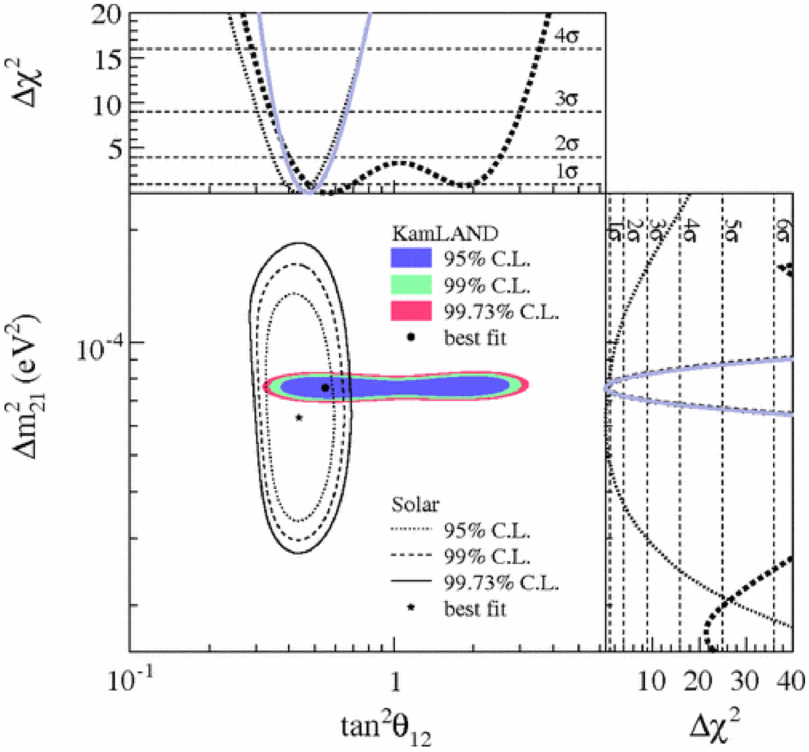


Figure 1.10: Comparing the measurements of solar neutrino oscillation parameters from KamLAND and SNO/SK [35]

1.5 Current Paradigm and Open Questions

We've seen how the phenomenology of two neutrino flavor mixing works quite well in describing the solar and atmospheric anomalies. Let us now try to touch ground with the Standard Model.

1.5.1 The ν-Standard Model

To explain the phenomenon of neutrino masses, one can append to the Standard Model new renormalizable terms in the Lagrangian [40]. This necessarily needs new degrees of freedom as the Standard Model has exhausted all the allowed renormalizable terms already. For neutrino masses, one can add a new right-handed neutrino which transforms under

$SU(2)_L \times U(1)_Y$ as :

$$\nu_R : (\mathbf{1}, \mathbf{0}) \tag{1.31}$$

which is a pure singlet state. We can then write down Dirac masses, which after symmetry breaking looks like :

$$\mathcal{L}_{\nu-D} = -Y \bar{l}_L \phi^* \nu_R + h.c \rightarrow -\frac{y v_{ew}}{\sqrt{2}} \bar{\nu}_L \nu_R + h.c \tag{1.32}$$

where y denotes the neutrino Yukawa mass after the Higgs field attains its vacuum expectation value. One can also write down another renormalizable operator using the new singlet state as :

$$\mathcal{L}_{\nu-M} = -\frac{m}{2} \nu_R^* \nu_R \tag{1.33}$$

where $\nu_R^* = \bar{\nu^{\mathcal{C}}}$, $\mathcal{C}$ being the charge-conjugate operator. This is a Majorana mass term and is unique to neutrinos owing to the fact that they are neutral. This term violates lepton number, which is an accidental symmetry of the Standard Model. In the following discussion, we will not make use of this term as they don't have any relevance for neutrino oscillation physics. The Majorana nature of neutrinos is the focus of many ongoing experiments however, as they have a number of appealing properties in Beyond the Standard Model (BSM) scenarios, including the possibility of explaining light neutrino masses in a more natural way through the see-saw mechanism.

For the Dirac masses, the Yukawa couplings before symmetry breaking, Y relates the doublet left-handed state to the singlet right-handed state and is in general not diagonal in the 3 lepton generations, i.e there are unitary matrices, U_L^f, U_R^f, U_L^ν, U_R^ν such that Y can be diagonalized as :

$$U_L^{f\dagger} Y_f U_R^f = diag(y_e, y_\mu, y_\tau) \tag{1.34}$$

$$U_L^{\nu\dagger} Y_\nu U_R^\nu = diag(y_{\nu_1}, y_{\nu_2}, y_{\nu_3}) \tag{1.35}$$

Here all the Yukawa masses, y are real and positive and f runs over (e, μ, τ) indices. But one cannot choose U_L^f and U_L^ν independently as any redefinition of the left-handed lepton doublet field to make Y_f diagonal in general will not make Y_ν diagonal. If we go to the bases that makes Y_f and Y_ν diagonal separately, we find

$$l_L^{'} = \left(U_L^f\right)^\dagger l_L \tag{1.36}$$

$$l_L^{''} = (U_L^\nu)^\dagger l_L \tag{1.37}$$

That is

$$l_L^{'} = \left(\left(U_L^f\right)^\dagger U_L^\nu\right) l_L^{''} \tag{1.38}$$

which relates the neutrino mass eigenstates to the charged lepton mass eigenstates. We can define a unitary matrix therefore,

$$U_{PMNS} = \left(\left(U_L^f\right)^\dagger U_L^\nu\right) \tag{1.39}$$

It is easy to see that this is the preferred basis to work with experimentally since neutrinos in general are not observed in their mass eigenstates but rather as products of weak interactions, whereas the charged leptons are observed in their mass eigenstates. One can then detect neutrinos in a different flavor to what one starts out as they propagate as a mixture of their mass eigenstates [41][42]. This is very similar to the quark sector, where one sees weak transitions between quarks in different generations. There too, we have a mixing matrix denoted as V_{CKM} [43][44]. The structure of U_{PMNS} can be parameterized similarly by :

$$U_{PMNS} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.40}$$

where we have $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The phase, δ_{CP} is a CP-violating parameter. If $\delta_{CP} \neq 0, \pi$ then there doesn't exist a basis in which all the phases in the Yukawa couplings can be absorbed into the fermion fields. This would imply $\mathcal{CP}$ is violated in the lepton sector along with the already observed violation in the quark sector.

1.5.2 Three Flavor Mixing

As before, we relate the flavor states to a set of mass-eigenstates. The basis transformation is now described by the 3 × 3 unitary matrix, U_{PMNS} defined above.

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \\ |\nu_\tau\rangle \end{pmatrix} = U_{PMNS} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \\ |\nu_3\rangle \end{pmatrix} \tag{1.41}$$

The unitary assumption is to ensure the oscillation probabilities add up to 1. This isn't satisfied if there are more neutrino states involved and testing the unitarity is part of ongoing experimental goals [45][46]. The oscillation probability of a flavor state $|\nu_\alpha\rangle$ transforming into $|\nu_\beta\rangle$ over a propagation length, L is given by :

$$P_{\alpha\rightarrow\beta} = |\,\langle\nu_\beta|\nu_\alpha(L)\rangle\,|^2 = |\sum_{i=1}^{3} U^*_{\alpha i} U_{\beta i} e^{-\frac{m_i^2 L}{2E_\nu}}|^2 \tag{1.42}$$

This reduces to

$$P_{\alpha\to\beta} = \delta_{\alpha\beta} - 4\sum_{i<j}^{3} Re\left[U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j}\right]\sin^2\frac{\Delta m^2_{ij}L}{4E} + 2\sum_{i<j}^{3} Im\left[U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j}\right]\sin\frac{\Delta m^2_{ij}L}{2E} \tag{1.43}$$

The corresponding anti-neutrino probability now is given by $U_{PMNS} \to U^*_{PMNS}$, which changes sign of the imaginary part in the above equation. This equation has two characteristic oscillation frequencies related to parameters Δm^2_{21} and Δm^2_{32}.

1.5.3 Open Questions

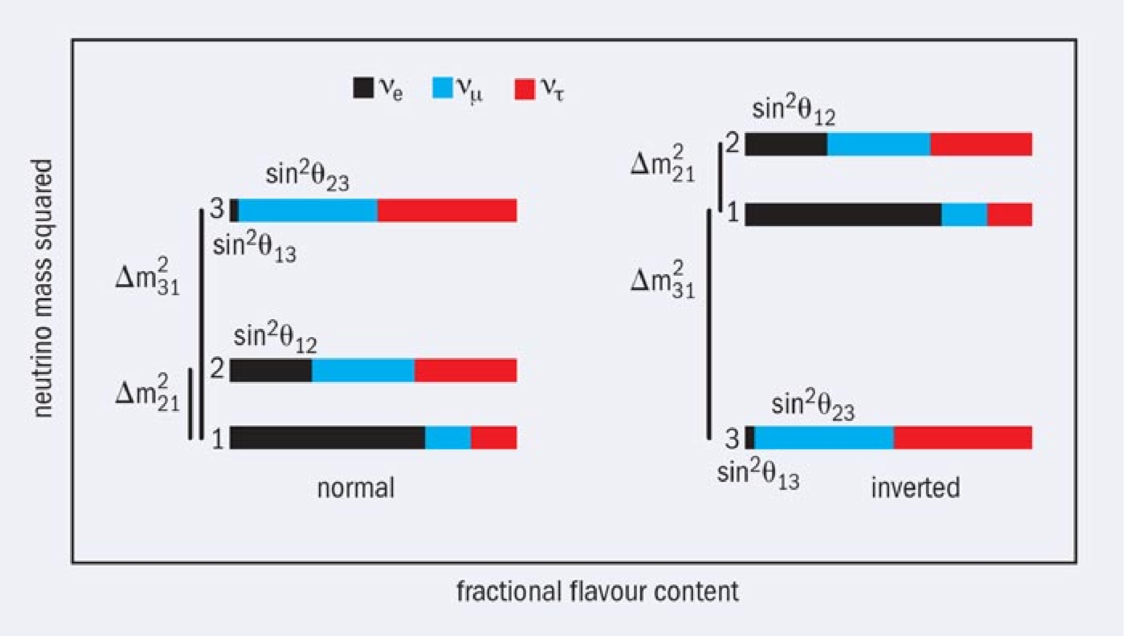


Figure 1.11: The Neutrino Mass Ordering Picture [47]

In looking at applying the above to the neutrino oscillation data, Δm^2_{21} is chosen for the solar sector and is measured to be positive ($\sim 7.5 \times 10^{-5}$ eV^2). This is because the matter effects in the sun constrain the product of $\Delta m^2_{21}\cos 2\theta_{21}$ to be positive. There is still, however, an ambiguity in the sign of Δm^2_{32} in interpreting it as the atmospheric frequency ($\sim 2.5 \times 10^{-3}$ eV^2). This is related to the ordering of neutrino mass state $|\nu_3\rangle$ relative to $|\nu_1\rangle$ where $|\nu_1\rangle$

and $|\nu_2\rangle$ have been identified as being $|\nu_e\rangle$ heavy. In the "Normal Hierarchy" case, $\Delta m^2_{32} > 0$, while the negative case is the "Inverted Hierarchy".

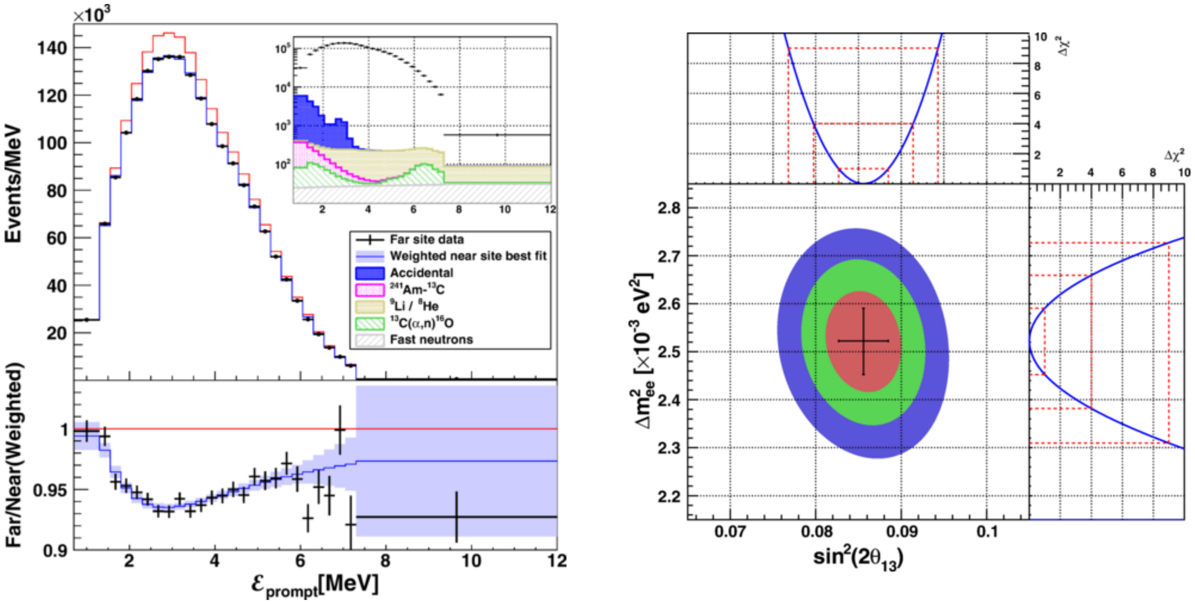


Figure 1.12: Observation of non-zero θ_{13} by the Daya-Bay experiment. [48]

The mixing angles, θ_{12} and θ_{23} are therefore identified with the solar and atmospheric sectors respectively, where $\sin^2\theta_{12} \approx 0.3$ and $\sin^2\theta_{23} \approx 0.5$, although the latter case has large error bars and is not known to exactly conform the maximal mixing scenario, $\theta_{23} = \frac{\pi}{4}$.

Since, in the atmospheric case, ν_μ was predominantly oscillating into ν_τ, one can say that $|\nu_3\rangle$ is fairly light in $|\nu_e\rangle$ and therefore $\theta_{13} \approx 0$. The success of the two-flavor approximation can be explained by the fact that the masses are sufficiently hierarchical, i.e $\Delta m^2_{21} \ll |\Delta m^2_{32}|$.

Experiments like Daya Bay [48], RENO [49] and Double CHOOZ [50], looking at reactor anti-neutrinos at short baselines have proven sufficiently sensitive to the oscillation spectrum of $\bar{\nu}_e$-disappearance that $\sin^2\theta_{13}$ has been precisely measured to 0.022 ± 0.0006 [36]. Fig. 1.12 shows the Daya-Bay result. Unlike the disappearance measurement in KamLAND which is at a relatively long baseline, the shorter baselines are sensitive to the faster oscillations from $|\Delta m^2_{31}|$. This shows definitely that there is a ν_e component in ν_3 which opens up the $\nu_\mu \to \nu_e$ oscillation channel at long baselines. While most of the ν_μ would oscillate to ν_τ, an excess of ν_e can be looked for. This could be sensitive to δ_{CP} by measuring the asymmetry in $P(\nu_\mu \rightarrow \nu_e)$ and $P(\bar{\nu}_\mu \rightarrow \bar{\nu}_e)$, which by Eqn. 1.43 would be related to the imaginary part of the PMNS matrix and hence, δ_{CP}.

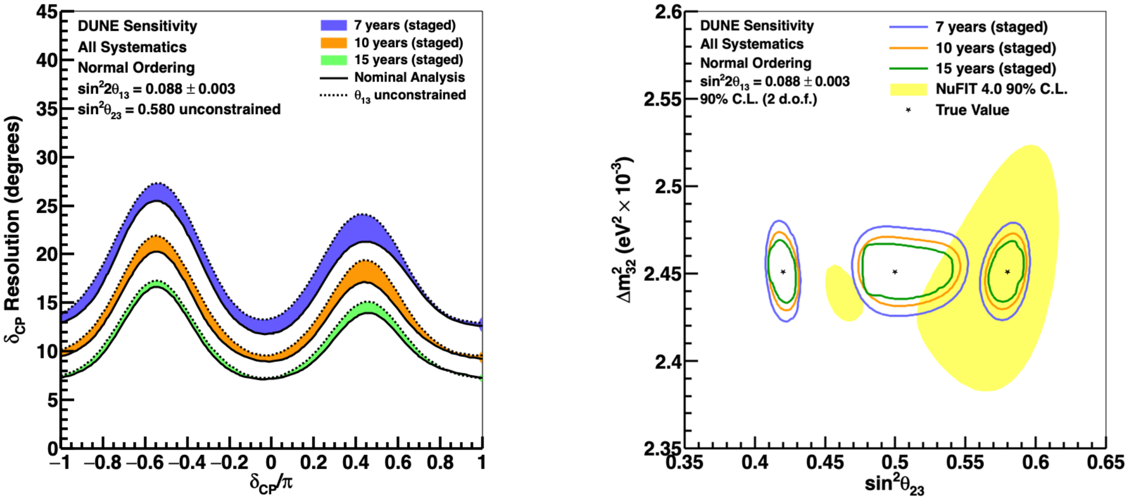


Figure 1.13: Sensitivity of the DUNE experiment to resolve δ_{CP} as a function of its true value (right) and 90% C.L for various true values in $\Delta m^2_{32} - \sin^2\theta_{23}$ space (left) for different years of planned running [51]

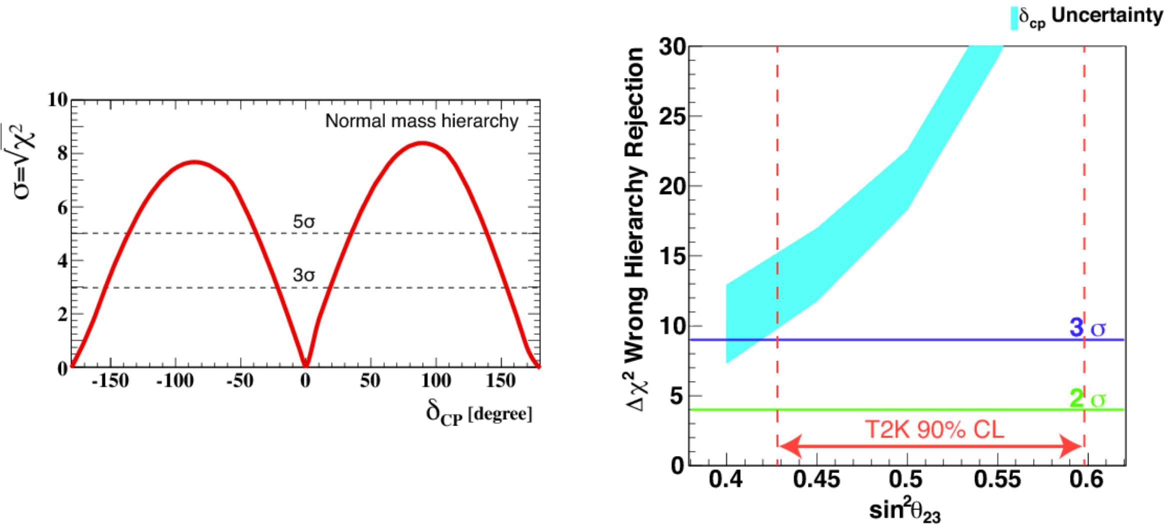


Figure 1.14: Sensitivities of the Hyper-K experiment to measure CP-Violation as a function of true value of δ_{CP} (left) and the Neutrino Mass Hierarchy as a function of $\sin^2\theta_{23}$ for a range of δ_{CP} values (right) after a planned 10 years of running [52]

As we move to the precision era of neutrino oscillation physics, it is therefore imperative that not only do we try to measure the remaining PMNS parameters but as well, overconstrain it as much as possible to be able to test for the unitarity assumption. Accelerator neutrino experiments like NOvA [53] and T2K [54] that look at ν_μ oscillations will be able to constrain these parameters to a large extent while future experiments like DUNE [55] and Hyper-Kamiokande [56] will be able to definitively measure them (illustrated in Figs. 1.13 and 1.14). A predominantly ν_μ beam is studied at baselines ranging from 200 to 1000 km, depending on the experiment. The approximate ν_μ disappearance probability is

$$P_{\nu_\mu \to \nu_\mu} \approx 1 - \sin^2 2\theta_{\mu\mu} \sin^2 \frac{\Delta m^2_{\mu\mu} L}{4E_\nu} \approx 1 - \cos^2\theta_{13} \sin^2 2\theta_{23} \sin^2 \frac{\Delta m^2_{32} L}{4E_\nu} + \mathcal{O}(\alpha, s^2_{13}) \tag{1.44}$$

where

$$\sin^2\theta_{\mu\mu} = \cos^2\theta_{13}\sin^2\theta_{23} \tag{1.45}$$

$$\Delta m^2_{\mu\mu} = \sin^2\theta_{12}\Delta m^2_{31} + \cos^2\theta_{12}\Delta m^2_{32} + \cos\delta_{CP}\sin\theta_{13}\sin 2\theta_{12}\tan\theta_{23}\Delta m^2_{21} \tag{1.46}$$

$$\alpha = \Delta m^2_{21}/\Delta m^2_{32} \tag{1.47}$$

It is easy to see that this channel is agnostic towards the neutrino mass hierarchy as well as

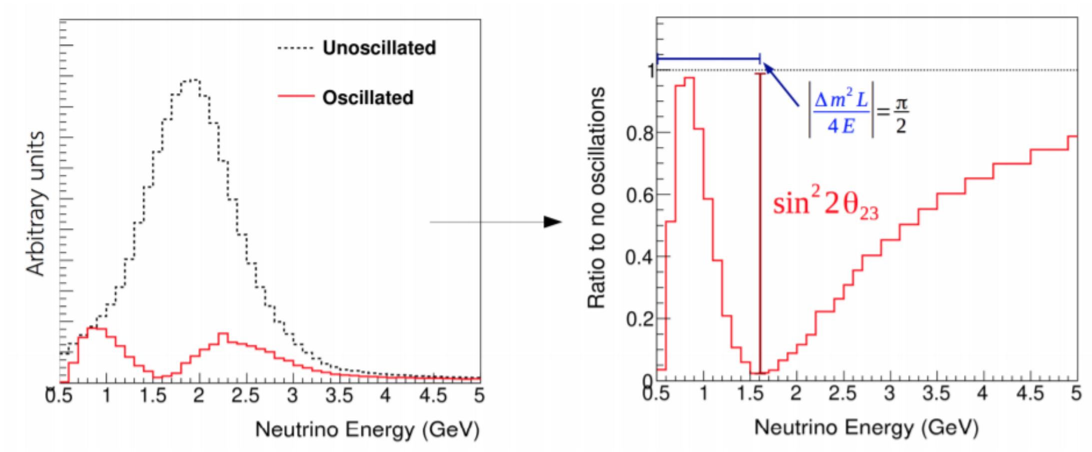


Figure 1.15: The ν_μ Disappearance Spectrum

the octant of θ_{23} (whether $\theta_{23} < \frac{\pi}{4}$ or $> \frac{\pi}{4}$) but does give strong constraints on $|\Delta m^2_{32}|$ and $\sin^2 2\theta_{23}$, as illustrated in Fig. 1.15. For the ν_e-appearance channel, one must also consider effects from interactions in the Earth's crust. For NOvA, which is the experiment we will be concerned with, at energies of $\approx 2\,\mathrm{GeV}$ and a baseline of 810 km, the matter potential, A ($= 2\sqrt{2}G_F n_e E_\nu$) comes out to $4.3 \times 10^{-4}\ \mathrm{eV}^2$ which as we've seen from Eqn. 1.29 modifies the θ_{13} angle in the Normal Hierarchy case as :

$$\sin 2\theta_{eff} \approx 1.17 \sin 2\theta_{13} \tag{1.48}$$

where we've used the mean density of the Earth's crust, $\rho = 2.84\,\mathrm{g/cm^3}$. For the anti-neutrino case, the effect is opposite where

$$\sin 2\theta_{eff} \approx 0.86 \sin 2\theta_{13} \tag{1.49}$$

For the Inverted Hierarchy, this effect is interchanged between $\nu \to \bar{\nu}$, as is illustrated in a bi-probability plot shown in Fig. 1.16. It is evident therefore, that NOvA's baseline has a large enhancement on the $\nu_e/\bar{\nu}_e$-appearance probability depending on the hierarchy, which can be utilized to constrain it. The approximate ν_e-appearance probability can also be written down to first order in α as :

$$P(\nu_\mu \to \nu_e) \approx 4\sin^2\theta_{13}\sin^2\theta_{23}\frac{sin^2\Delta}{(1-M)^2} + 8\alpha J_{CP}^{max}\cos(\Delta + \delta_{CP})\frac{\sin\Delta M}{M}\frac{\sin\Delta(1-M)}{1-M} + .. \tag{1.50}$$

where $M = \frac{A}{\Delta m^2_{31}}$. Here J_{CP}^{max} measures the maximum CP-violation in the Jarlskog's invariant, J_{CP} given by :

$$J_{CP} = \cos\theta_{12}\sin\theta_{12}\cos\theta_{23}\sin\theta_{23}\cos^2\theta_{13}\sin\theta_{13}\sin\delta_{CP} \tag{1.51}$$

and taking $\sin\delta_{CP} \to 1$. For $\bar{\nu}_e$-appearance, the equation is modified as $M \to -M$ and

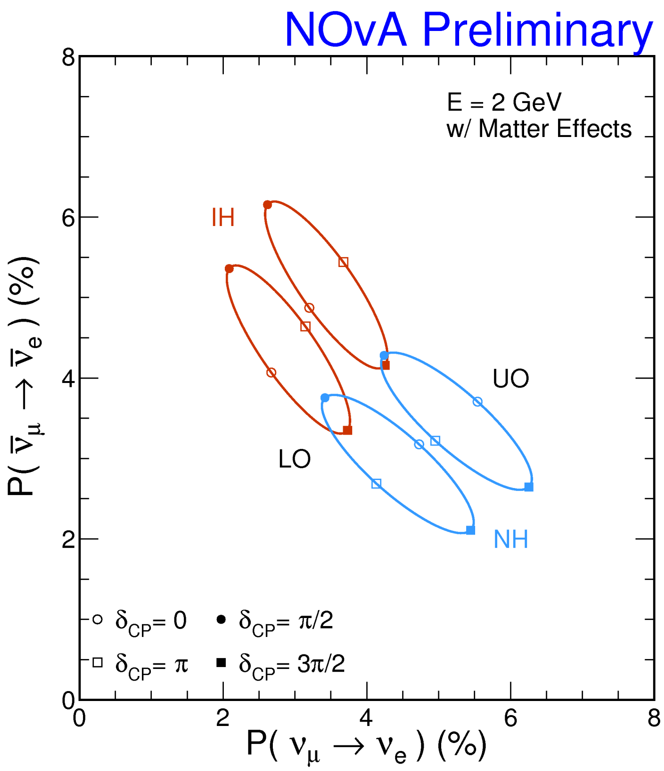


Figure 1.16: Phase-space of measurement of ν_e ($\bar{\nu}_e$) appearance at NOvA for a baseline of 810 km and a neutrino energy of 2 GeV

$\delta_{CP} \to -\delta_{CP}$. This form shows that the appearance probabilities are also sensitive to the octant of θ_{23} as well as δ_{CP} and the mass hierarchy, if one looks at the asymmetries in neutrino and anti-neutrino appearance. As a result, current and future sensitive accelerator experiments can resolve the remaining elements of the PMNS matrix and complete the 3-flavor paradigm.

1.5.4 Implications and the Future

As noted before, it is important to constrain the PMNS matrix as precisely as possible to test for unitarity. The unitarity assumption can be violated by new heavy neutrino states for eg. in some see-saw mechanisms that explain the light neutrino masses. This can therefore act as an indirect probe for the see-saw scale in particular at the electroweak scale ($\sim$ TeV)

[57]. This could also show up as non-standard neutrino interactions in the Earth's crust which experiments like NOvA and DUNE could be sensitive to. In addition, the unitarity is violated more severely if there are light sterile neutrinos and the resulting parameter space can be heavily constrained using measurements of the PMNS matrix [58].

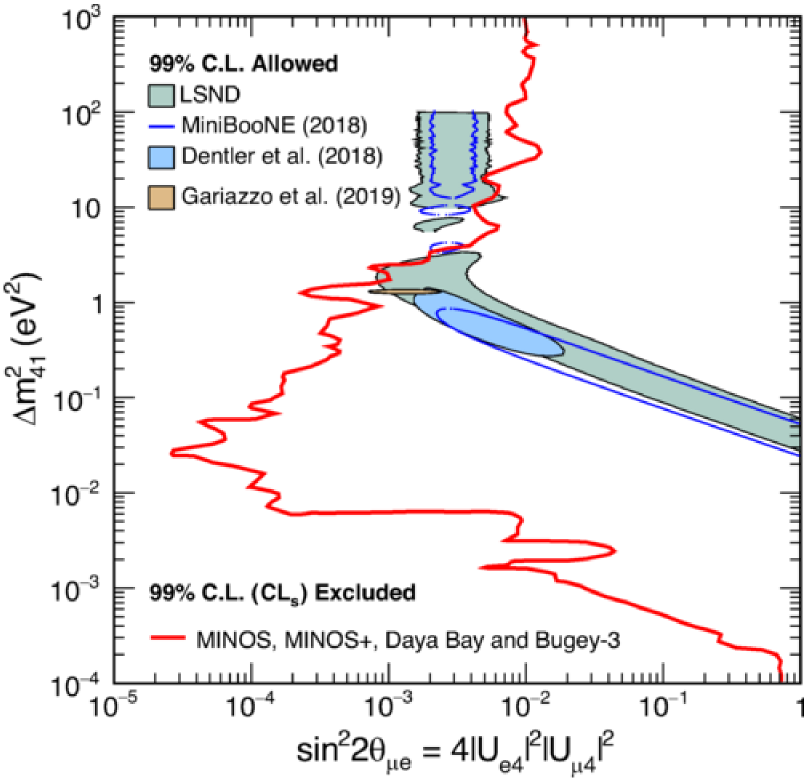


Figure 1.17: Exclusion limits for a 4^{th} sterile neutrino flavor based on a combined measurement of Daya-Bay, MINOS, MINOS+ and Bugey-3. This is compared to allowed regions as seen in the MinoBooNE/LSND anomalies. [58]

In general, the nature of neutrino mass is also unknown. Although there are strong cosmological bounds on the sum of neutrino masses, it is important to measure them in a more controlled setting to keep it model independent. Experiments like KATRIN [60] are involved in directly measuring the neutrino mass by looking for a kink in the end-point β-decay spectrum of He^3 nuclei, as illustrated in Fig. 1.18. It is therefore easy to see that measurements of mass hierarchy, i.e whether $|\nu_e\rangle$ is predominantly in the heavy mass eigenstates or the light eigenstates, will have a direct impact on sensitivity for the mass measurement. Experiments directly probing the existence of Majorana neutrinos by looking for neutrinoless double beta

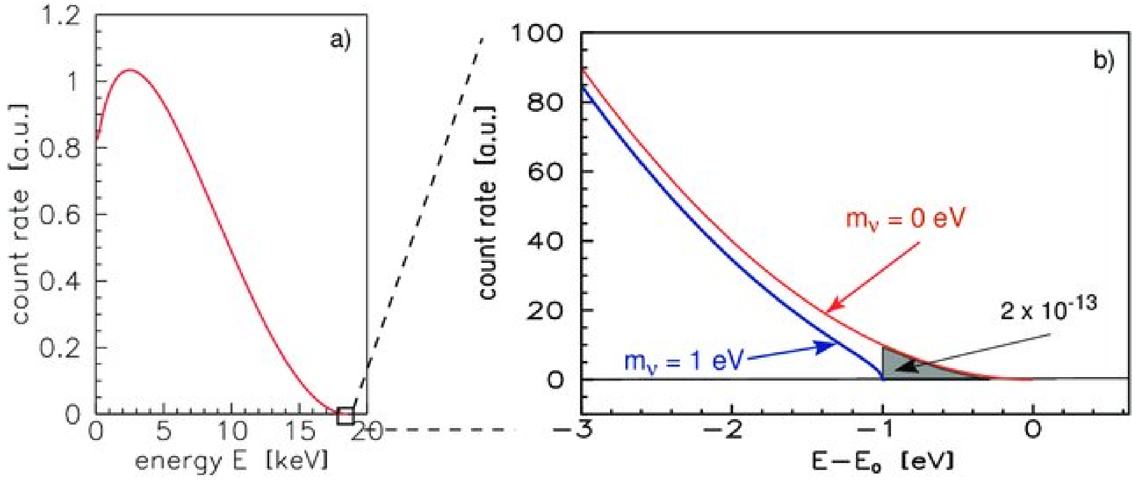


Figure 1.18: The He^3 β decay Kurie plot for measuring the absolute neutrino mass at KATRIN. Taken from [59].

decay ($0\nu\beta\beta$) are also affected by the mass hierarchy in terms of their reach [61] (as shown in Fig. 1.19).

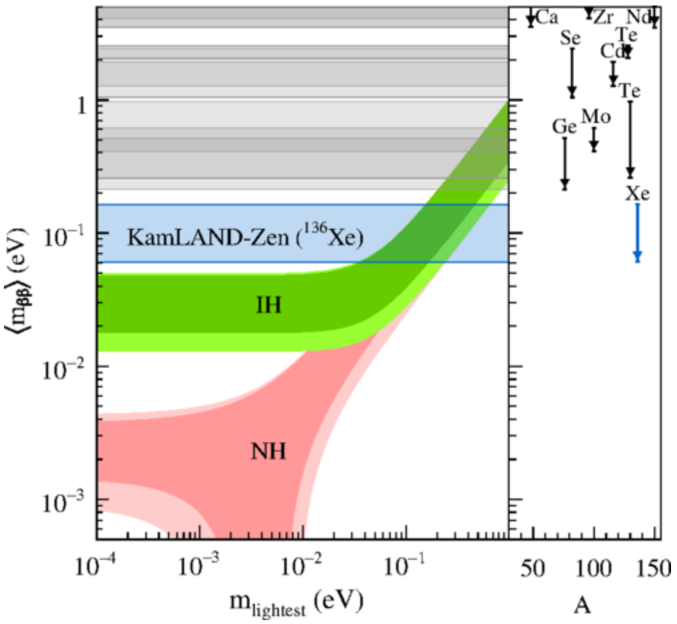


Figure 1.19: The current predictions for the effective Majorana neutrino mass $m_{\beta\beta}$ as a function of the lightest neutrino mass and the mass hierarchy. The dark shaded regions indicate predictions at current best-fit oscillation parameters, while the light shaded regions indicate a 3σ uncertainty on those parameters. Also shown are the current 90% Upper Limits from KamLAND-Zen along with limits from other experiments utilizing different β-decay isotopes on the right. [61]

In addition, CP-violation in the lepton sector can provide an important input for models explaining the observed baryon asymmetry via leptogenesis, especially if the measured CP-violation from δ_{CP} is found to be large [62].

Neutrino oscillations are also the first sign of lepton flavor violation (LFV) in the Standard Model. While the cross-sections of LFV processes for the charged leptons are too tiny to observe within the Standard Model, usually proceeding via two loops suppressed by the W-boson mass [40], there are many BSM scenarios that predict a larger cross-section for these processes, for eg. in $\mu \to e + \gamma$ [63].

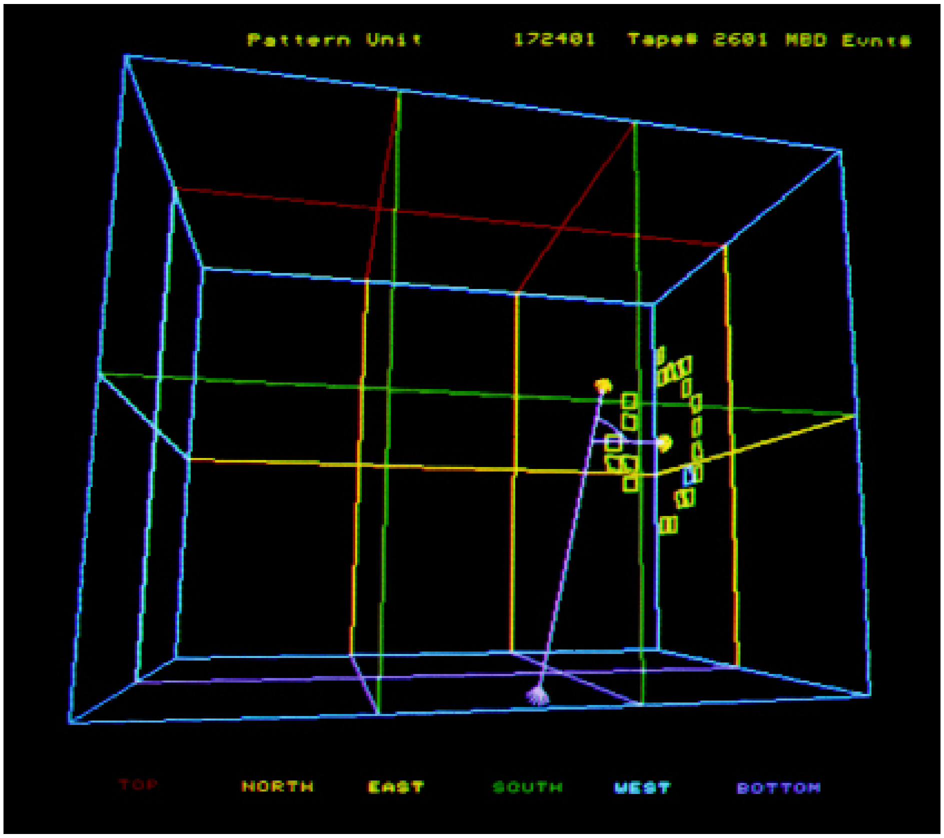


Figure 1.20: A neutrino from SN1987A detected in the north-east corner of the IMB detector. Shown in the purple line is the direction of the neutrino from the Large Magellanic Cloud along with the lit PMTs from the resulting Cerenkov cone after the neutrino interaction. [64]

Finally, neutrino physics is important in an astrophysical context, as neutrinos are copiously produced in some of the most violent events in the universe and can therefore act as probes for those events. Since neutrinos rarely interact, they are able to provide an unprecedented window into objects such as core-collapse supernovae, active galactic nuclei and so on. Ever since the discovery of neutrinos from the supernova SN1987A by the Kamiokande-II [65], IMB [64] and Baksan [66] experiments, a new frontier of neutrino-based astronomy has emerged. Major experimental efforts are therefore underway in order to detect neutrinos from galactic supernovae (predicted within the century) to provide input about the nature of the core-collapse process. In addition, other rapidly developing fronts include detection of the diffuse supernova neutrino background, ultra-high energy neutrinos, relic neutrinos and so on [67].

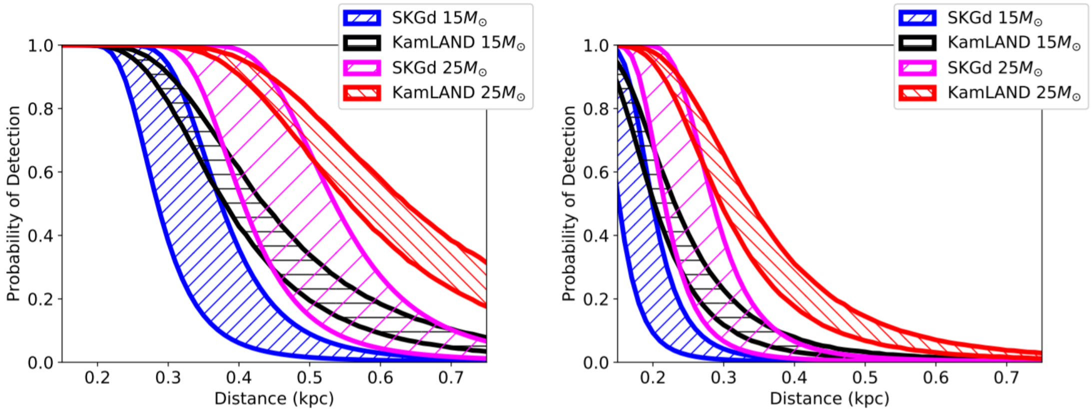


Figure 1.21: The probability of detection of a supernova from a star of given mass by the SKGd and KamLAND experiments as a function of distance from Earth and neutrino mass hierarchy. Normal (left) and Inverted (right). The bands indicate the uncertainty due to background levels. Taken from [68].

A comprehensive understanding of flavor physics and weak interactions is therefore tied up intimately with the nature of neutrino mass and can continue to be the focus of a large experimental program. In addition, the field of neutrino physics has a rich and significant interplay with fundamental questions about the evolution of the universe.

Chapter 2

The NOvA Experiment

2.1 Introduction

NOvA (NuMI Off-Axis ν_e - Appearance) is a long baseline accelerator experiment dedicated to neutrino oscillation measurements using a ν_μ ($\bar{\nu}_\mu$) beam. As discussed previously, the non-zero value of θ_{13} measured by reactor anti-neutrino experiments has opened up the possibility of constraining the remaining elements of the PMNS matrix as well as the neutrino mass hierarchy via the $\nu_\mu \rightarrow \nu_e$ ($\bar{\nu}_\mu \rightarrow \bar{\nu}_e$) channels. The experimental technique that NOvA uses is similar to a range of experiments like T2K and MINOS (a precursor to NOvA which used the same NuMI beam at Fermilab) seeking to answer the same questions but with different sensitivities to different parameters, depending on the baseline and the range of beam energies [69].

NOvA utilizes two detectors. The Near Detector (ND) which is about 1 km away from the beam source, is placed underground at a depth of 100 m and weighs about 300 t. The Far Detector (FD) is a much larger detector, weighing 14 kt but otherwise functionally identical to the ND, and is placed at a baseline of 810 km from the beam. The ND is housed in an

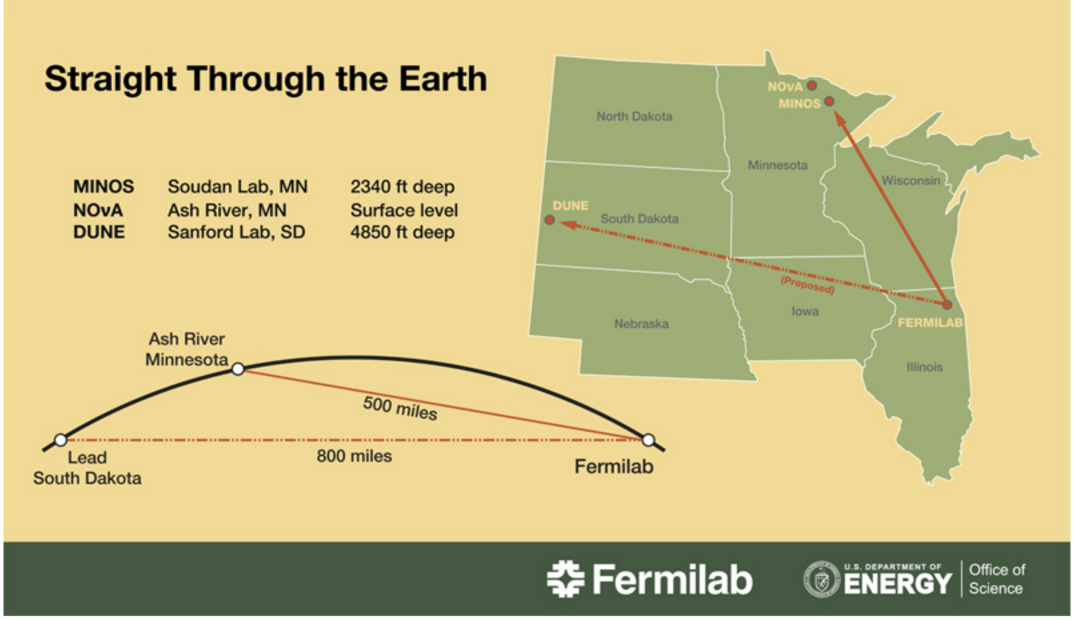


Figure 2.1: Illustration of the baselines for different experiments on the NuMI beam. Taken from [70]

underground cavern dug up near the MINOS Near Detector hall at Fermilab while the FD is on surface in Ash River, MN near the Canadian border, as illustrated in Fig. 2.1.

These detectors are sensitive to neutrinos, primarily made up of ν_μ or $\bar{\nu}_\mu$, albeit with a slight contamination from ν_e and $\bar{\nu}_e$ (for two separate modes of operation). These neutrinos come from a high intensity beam source produced at the Main Injector at Fermilab – the NuMI beamline. In addition the Far Detector seeks to observe oscillated events both in the disappearance channel, where it ascertains how many ν_μ ($\bar{\nu}_\mu$) have disappeared compared to the unoscillated prediction, and the appearance channel, where it tries to observe how many ν_e ($\bar{\nu}_e$) have appeared above the predicted beam contamination. The Near Detector, where neutrinos haven't yet had a chance to oscillate, is primarily used to characterize the beam source in terms of its neutrino flavor components as well as constrain systematic uncertainties for the oscillation measurement.

In what follows, the experimental design and the detector technologies will be described in detail. Aspects of the beam simulation as well as the detection principle and event reconstruction will also be discussed.

2.2 The NuMI beam

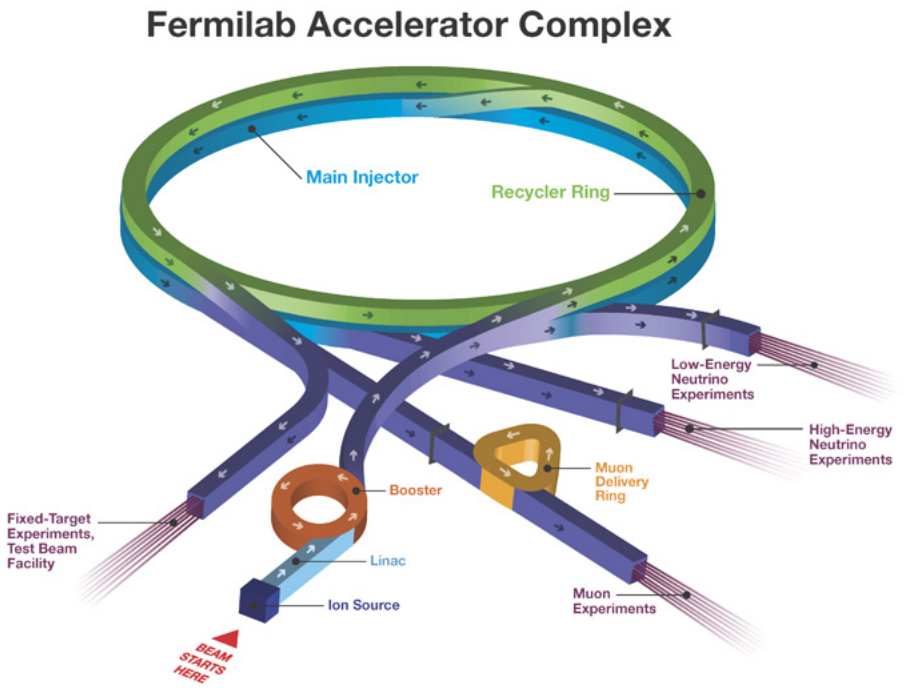


Figure 2.2: Schematic of the proton beam production at NuMI. Taken from [70]

The NuMI beam at Fermilab [71] is designed to produce neutrinos and anti-neutrinos at high intensity by colliding a 120 GeV proton beam onto a fixed graphite target. The proton beam is first accelerated to 8 GeV in a rapid cycling synchrotron called the Booster and then delivered to the Main Injector ring in batches of 1.6 μs with 4.3×10^{12} protons per batch. The Main Injector is a larger synchrotron that accelerates the protons to 120 GeV. The Recycler which sits on top of the Main Injector merges two batches at a time from the Booster in a process called "slip-stacking" and delivers a total of 12 Booster batches or 6 double-intensity batches to the Main Injector. The Main Injector is then able to deliver these batches, made up of $\sim 5 \times 10^{13}$ protons to the graphite target in $10\,\mu$s spills at an average rate of 1.33 s/spill. The total protons on target delivered (POT) is the exposure metric for the neutrino beam. With routine operations at > 700 kW, the Main Injector is able to deliver $\sim 6 \times 10^{20}$ POT every year. Fig. 2.2 shows the schematic of the proton accelerator.

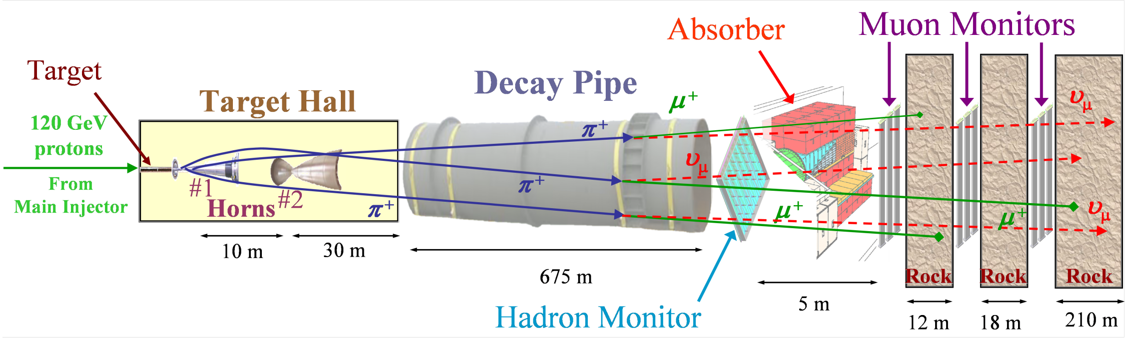


Figure 2.3: The NuMI beamline. Taken from [71]

The proton collision with the 1.2 m graphite target produces mainly $\pi^\pm$ and $K^\pm$. Downstream of the target, there are two magnetic focussing horns with a nominal current of 200 kA and a 1 mm water layer to prevent overheating. The magnetic horns serve to focus the meson beam into a 675 m long, roughly cylindrical decay pipe, where they decay into primarily ν_μ and $\bar{\nu}_\mu$ depending on the sign of the parent meson, i.e $\pi^+/K^+ \rightarrow \mu^+ + \nu_\mu$ or $\pi^-/K^- \rightarrow \mu^- + \bar{\nu}_\mu$. The $\mu^\pm$ can very occasionally decay into ν_e and $\bar{\nu}_e$ and as a result, contaminate the ν_μ beam. This is an irreducible background to the oscillation analysis. Fig. 2.3 illustrates the production of the neutrino beam.

There are primarily two modes of operation we will be concerned with, depending on the sign of the horn current. The forward horn current mode (FHC) primarily selects the positively charged mesons of relatively low momentum by giving it a large transverse momentum (p_T) "kick". The reverse horn current mode (RHC) does the opposite by selecting for the negatively charged mesons. The "wrong-sign" mesons are thus focussed away from the beamline. However, there is a contamination of the wrong-sign mesons at large longitudinal momentum, p_z measured by the Feynman-x variable, $x_F = 2p_z/\sqrt{s}$ that do not receive this kick and are therefore unfocussed. These decay into wrong-sign neutrinos (in RHC) or anti-neutrinos (in FHC).

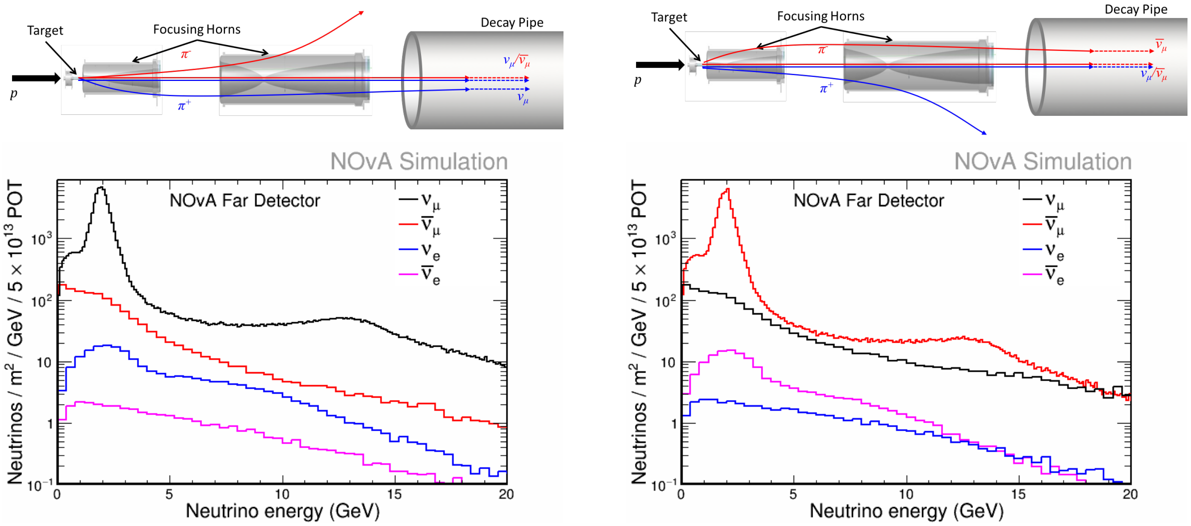


Figure 2.4: The two different neutrino beam modes, with FHC (left) and RHC (right) along with the predicted flux for each neutrino component

As seen in Fig. 2.5, between $1-5\,\mathrm{GeV}$ neutrino energies, which is the relevant range for the oscillation analysis, the beam is 96% pure in ν_μ for FHC and 83% pure in $\bar{\nu}_\mu$ for RHC. The ν_e and $\bar{\nu}_e$ components add up to $< 1\%$. The rest of the beam is made up of wrong-sign neutrinos or anti-neutrinos, where the contamination in RHC is larger due to the larger interaction cross-section for neutrinos relative to anti-neutrinos at the NOvA detectors. The

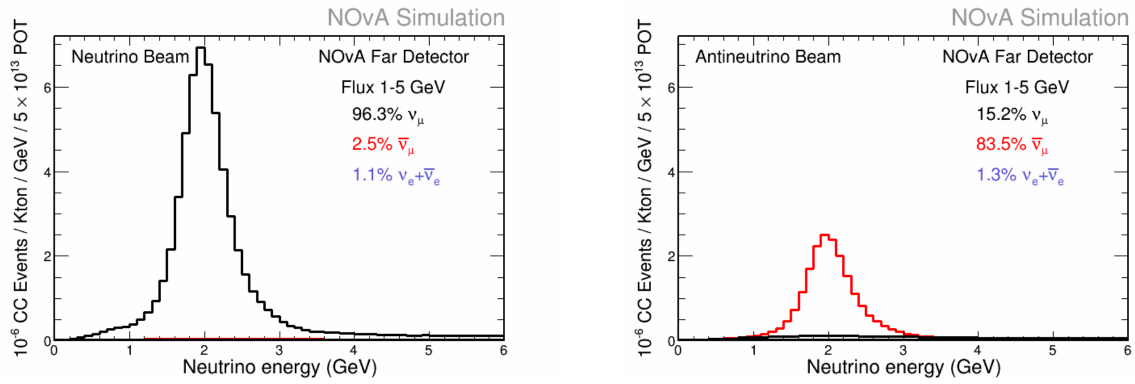


Figure 2.5: The predicted neutrino and antineutrino event rates at the NOvA Far Detectors for FHC (left) and RHC (right)

ν_μ and $\bar{\nu}_\mu$ below 7 GeV predominantly come from $\pi^\pm$ while the $K^\pm$ component dominates at higher energies. The intrinsic beam $\nu_e/\bar{\nu}_e$ component below 4 GeV mostly come from μ decays produced by $\pi^\pm$ although there is a significant contribution from $K^\pm$ decays as well. They also have a relatively larger tail for higher energies dominated by the $K^\pm$ component.

Finally the remaining hadrons are absorbed at a hadron dump just beyond the decay hall. The muons from the meson decays are used to monitor the health of the beam by a set of ionization detectors that interleave the rock layers at the end of the beamline.

2.2.1 Off-Axis Design

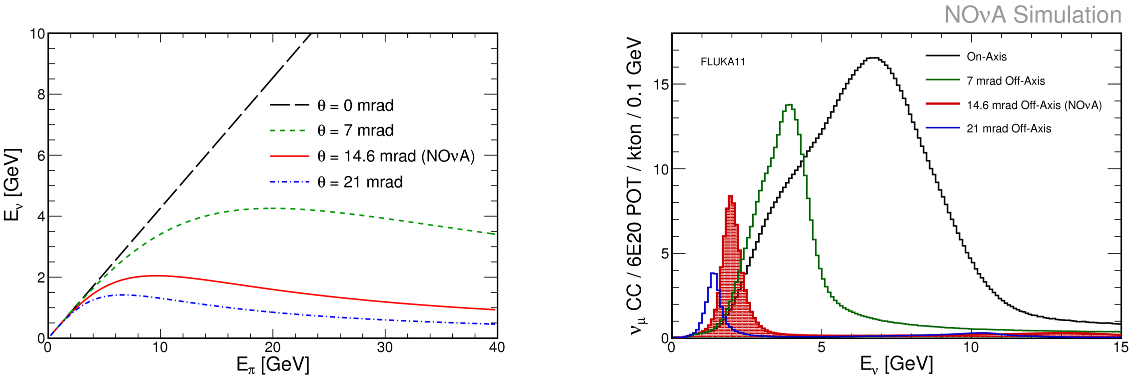


Figure 2.6: Neutrino energies as a function of the parent π energy for different off-axis angles (left) and the overall predicted flux per given POT (right)

As shown in Fig. 2.6, the NuMI beam is a relatively wide band neutrino beam in terms of neutrino energies. In order to get a relatively narrow flux peak at the $\nu_\mu \rightarrow \nu_e$ oscillation peak, NOvA places its detectors slightly off-axis from the beam direction. The neutrino energy as a function of the parent pion energy and off-axis angle is given by :

$$E_\nu = \frac{[1 - m_\mu^2/m_\pi^2]E_\pi}{1 + \gamma^2\theta^2} \tag{2.1}$$

where $\gamma = E_\pi/m_\pi$ [36]. A similar relation holds for the K parents. For $\theta = 0$, the neutrino energies scale linearly as a function of parent energies, while it tails off relatively quicker for

π compared to K as $\theta > 0$, since $\gamma_\pi > \gamma_K$. One can therefore optimize the design based on a non-zero off-axis angle and available baselines to ensure a narrow band flux at the ν_e-appearance oscillation maximum. This is given at the first oscillation maximum in L/E_ν for $\theta = 14.6\,\mathrm{mrad}$ at a baseline of 810 km, resulting in the neutrino flux peaking at $\sim 2\,\mathrm{GeV}$.

This is a similar design to the T2K experiment at a shorter baseline and in contrast to the proposed DUNE experiment at a larger baseline but placed on axis, $\theta = 0$. The resulting wide band beam in DUNE will sample a large portion around the first oscillation maximum but also cover part of the second maximum. The NOvA design allows one to constrain the interplay between different parameters such as the mass hierarchy, the octant of θ_{23} and δ_{CP} and is especially sensitive to the mass hierarchy and the octant at its baseline. The T2K design is sensitive to large asymmetries in the ν_e and $\bar{\nu}_e$ oscillation probabilities and is therefore able to look for $\mathcal{CP}$-violation at the expense of the mass hierarchy. The DUNE experiment will be sensitive to the value of δ_{CP} itself by collecting an enormous amount of statistics in the ν_e and $\bar{\nu}_e$ appearance channels for a wide range of neutrino energies.

2.2.2 Flux Simulation

The simulation of the NuMI beamline [72] and the resulting neutrino flux seeks to take into account each component described above and provide realistic physical uncertainties for the NOvA oscillation analysis. There are two main components associated with the flux, the physics of hadron production at the graphite target and their transport into the decay volume and consequent neutrino production. This is done by an integrated framework built on GEANT [73], called G4NuMI.

G4NuMI was built as part of Minerva, a neutrino experiment dedicated to measuring interaction cross-sections and which has since been decommissioned. Minerva shares the same NuMI beam as NOvA but is placed on axis. The framework tries to simulate each element

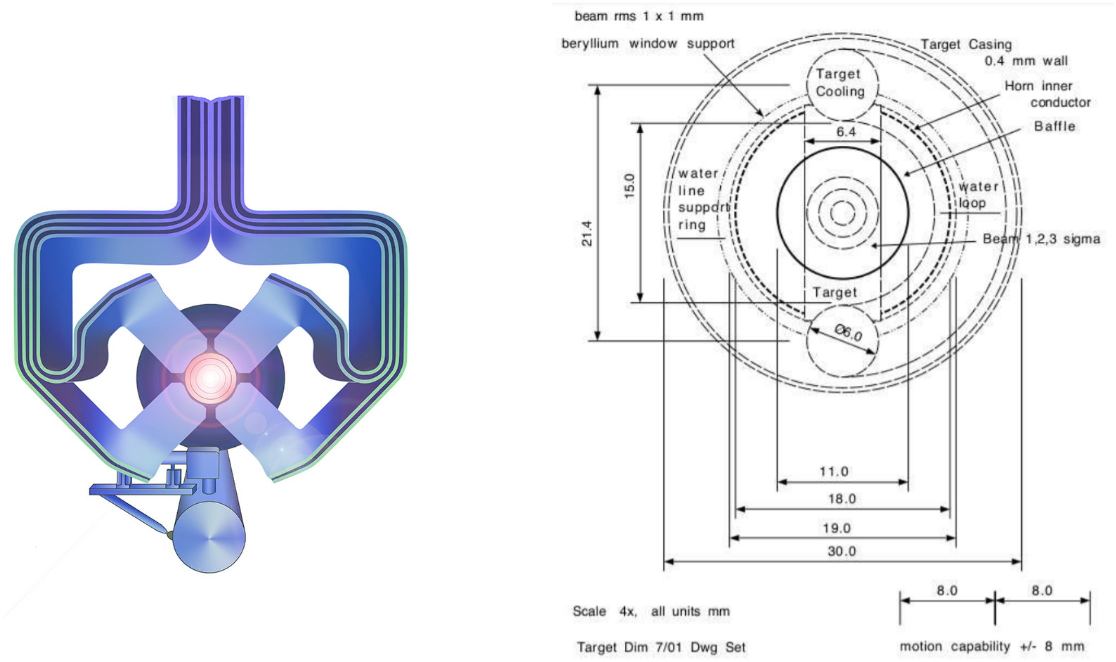


Figure 2.7: Schematic of the NuMI Magnetic Horns. Taken from [70][71]

of the beamline geometry starting from the protons delivered by the Main Injector at the target. Several improvements from the previous NOvA flux simulation built on FLUKA have been made, including a more accurate description of the geometry of the magnetic horns, the water layer on top of the horns which acts as a coolant and the beam spot size.

The hadron production simulation is handled by the Package to Predict the Flux (PPFX), also developed at Minerva but suitably modified for NOvA [74]. The description of hadronic cross-sections for proton collisions on graphite at 120 GeV is theoretically uncertain and turns out to be the biggest source of error in the flux model. As a result it is important to constrain and correct the underlying GEANT model, FTFP_BERT with external measurements made on similar material, which is what PPFX does.

Internally, PPFX does an accounting of the survival probabilities for each hadron produced at the target upto its interaction point and applies a correction based on external measurements for the cross-sections at the interaction point. The external datasets include primarily the thin-target measurements at NA49 experiment at CERN [75] and the MIPP experiment at

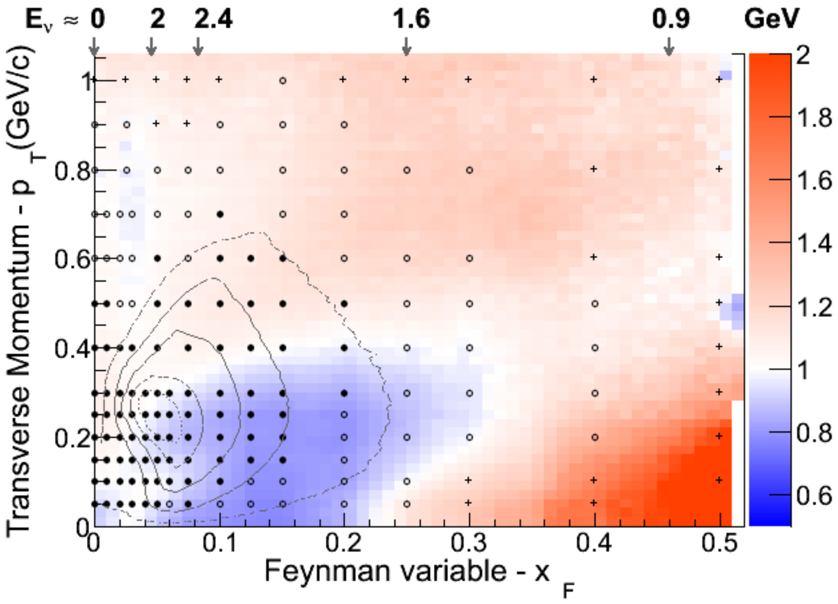


Figure 2.8: Corrections applied to the hadron production at the beam target from NA49 for π^+. Contours represent 10%, 20%, 30%, 50% and 80% of the pion yield for carbon.

Fermilab [76]. The NA49 experiment provides pion and kaon production cross-sections for proton interactions on Carbon at 158 GeV. The corrections are applied for relatively lower longitudinal momentum fractions in bins of the Feynman variable, x_F and p_T, as shown in Fig. 2.8, and are scaled to the appropriate energies relevant for NuMI. For pions at higher x_F, an older dataset from Barton is used and an appropriate scaling is done to the NuMI energy range and target material. For kaons at higher x_F, a ratio of K/π yields from the MIPP experiment is used and multiplied with the NA49 pion yields. These corrections are then applied as a weight to the survival probability and propagated to the neutrino flux prediction. The uncertainties are evaluated from the errors in the external measurements as a statistical ensemble of weights for the neutrino flux prediction in what is called a "multi-universe" technique. The overall size of the uncertainties is roughly $\sim 9\%$ at the 2 GeV peak (as can be seen in Fig. 2.9), although they are highly correlated across neutrino energies, which will be discussed later on.

Uncertainties from other aspects of the beamline simulation including realistic shifts in the

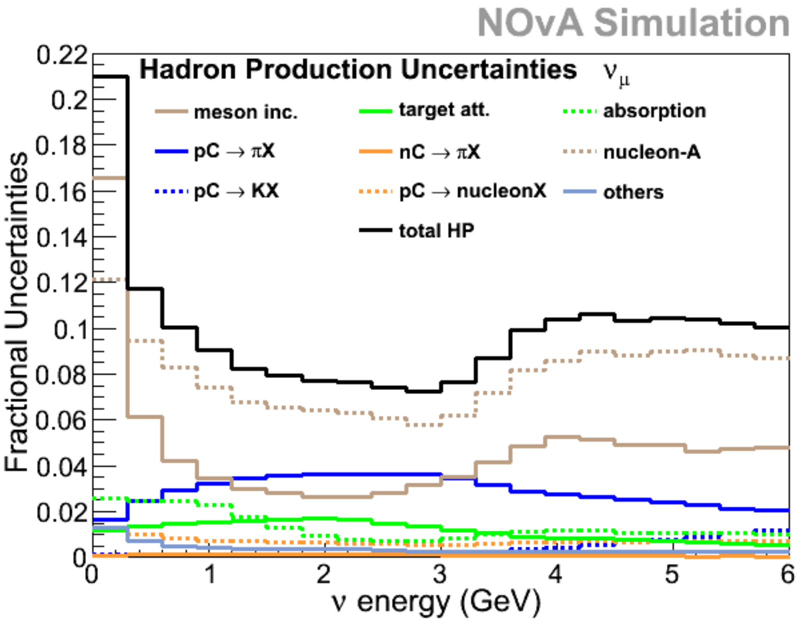


Figure 2.9: The overall hadron production fractional uncertainty as a function of neutrino energies broken down into various components

horn and beam spot positions, the beam spot sizes, the current in the horns and the presence of Earth's geomagnetic field in the decay pipe are also evaluated and propagated as weights for the nominal flux prediction as a function of neutrino energy.

2.3 The NOvA Detectors

NOvA uses a two detector technique to constrain the oscillation signal and backgrounds at the Far Detector. Fig. 2.10 shows a schematic of the NOvA detectors. As such, the design philosophy seeks to keep it functionally identical in order to control the dominant systematic effects that are correlated across the detectors. The primary difference is in their sizes, the Near Detector being 300 t and the Far Detector being 14 kt.

The detectors are finely segmented tracking calorimeters made up of a liquid scintillator (mineral oil doped with ~ 5% pseudo-cumene). The detector unit is an extruded PVC cell that houses the scintillator. For each cell, there is a wavelength-shifting (WLS) fiber that

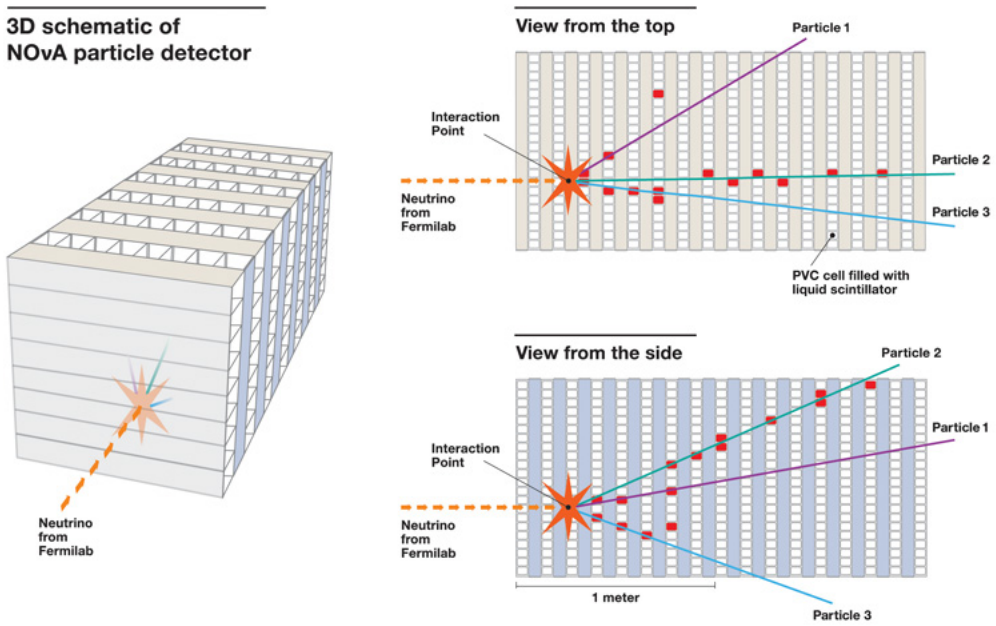


Figure 2.10: Schematic of the NOvA detector [70]

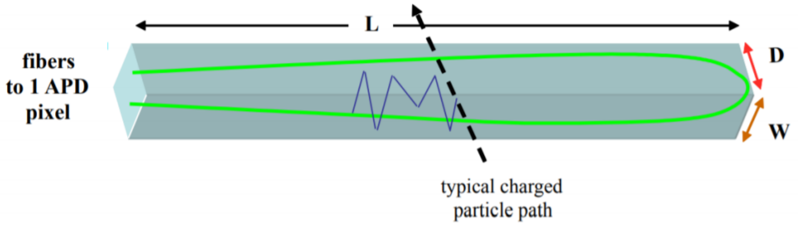


Figure 2.11: Schematic of the NOvA cell [77]

converts the UV scintillation light produced by charged particles passing through the cell into the visible spectrum, primarily in the blue-green range (450 nm – 600 nm). The PVC cells are coated with highly reflective TiO_2 to ensure all the light is reflected back into the cell where it can be captured by the WLS optical fiber [78]. The ends of the WLS fiber are connected to a single channel on an Avalanche Photo Diode (APD), as illustrated in Fig. 2.11 where the emitted photons are converted into photo-electron (PE) units. Each APD has 32 channels in total and reads in signals from 32 cells. These signals are then digitized and readout.

The cells are the same in cross-section for both the detectors, although they differ in length.

The cells are 3.9 cm × 6.6 cm with a small thickness of $\sim 2-5$ mm. The Far Detector cell is extruded to 15.8 m in length and the Near Detector to 3.8 m. Modules of 32 cells each (with a common APD) are glued together in a plane in bunches of 12 and 3 for the FD and ND respectively. This ensures that each plane have the dimensions of a square. These are then stacked horizontally and vertically in alternate planes to allow for 3-dimensional reconstruction using the two projections, the top view (XZ) and the side view (YZ) where Z is identified with the beam direction. Groups of 64 planes are referred to as diblocks with a total of 14 diblocks in the FD and 3 in the ND.

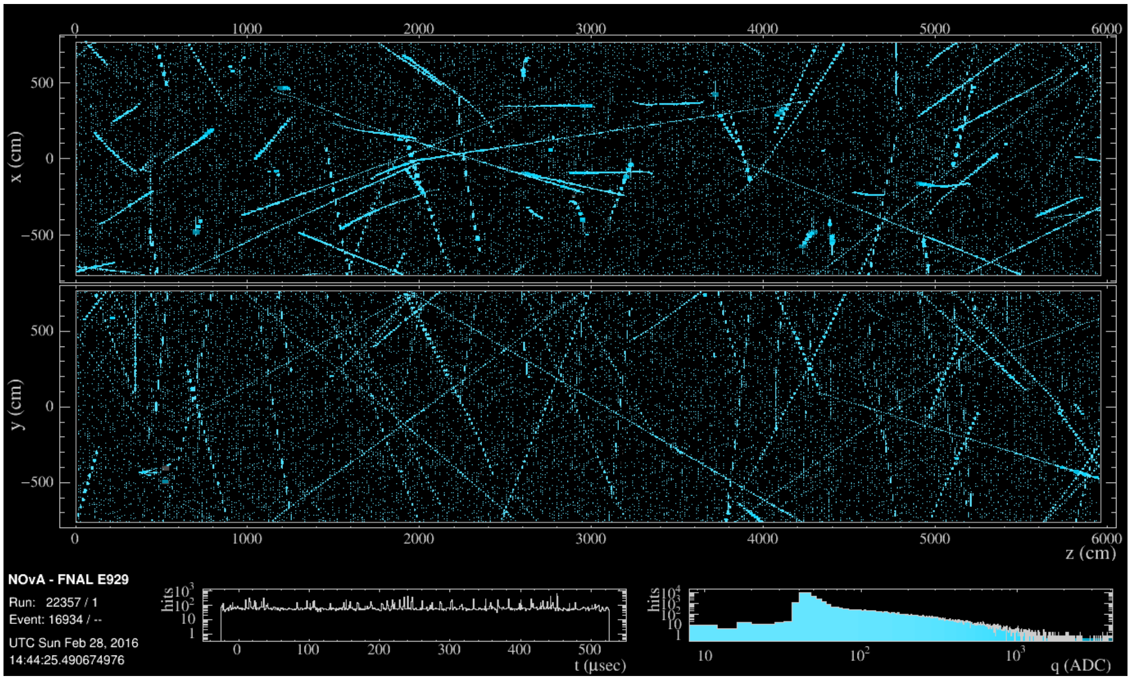


Figure 2.12: Snapshot of the NOvA Far Detector showing cosmic ray interactions

The Far Detector is therefore of dimensions 15.8 m × 15.8 m × 60 m. The Near Detector has an additional component called the “muon-catcher” which acts as a muon ranger for long muon tracks. The planes in the muon catcher are interleaved by 10 cm thick steel planes and are only 2.6 m long compared to 3.8 m in the active region. The Near Detector is then of dimensions 3.8 m × 3.8 m × 15.9 m, the last 3.2 m of which comprises the muon-catcher made up of 22 planes. The Far Detector is therefore made up of 344,064 channels or cells and is $\sim 65\%$ active, while the Near Detector is made up of 20,192 cells, being $\sim 45\%$ active.

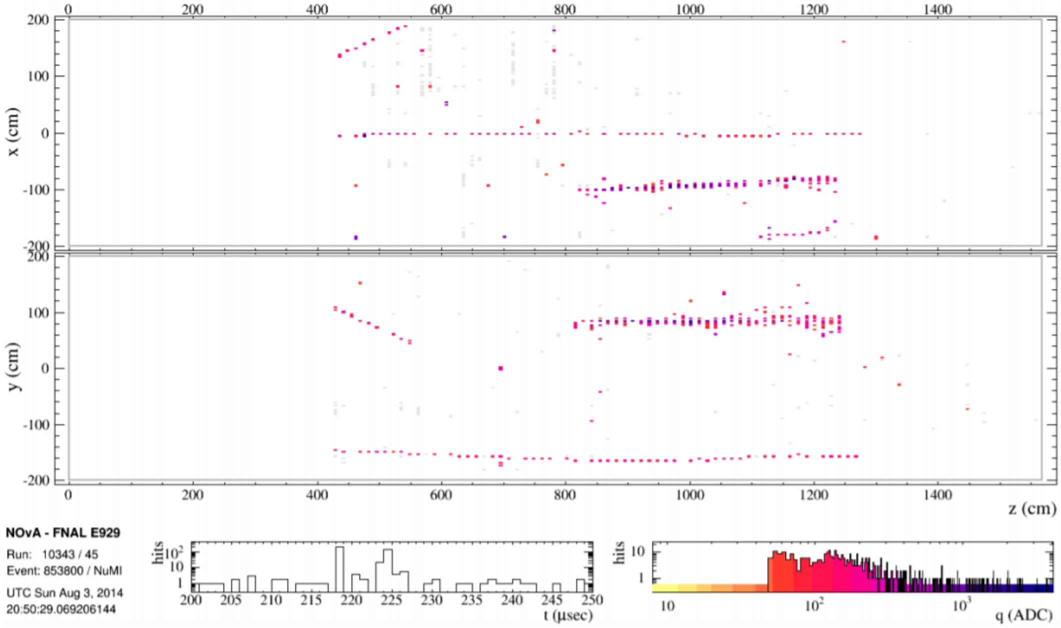


Figure 2.13: Snapshot of the NOvA Near Detector showing various neutrino interactions

The primary element for interaction is carbon, which is Low-Z. As a result, this allows for fine segmentation of shower cascades with a radiation length, $X_0 = 35.8\,\mathrm{cm}$ (6 cell-equivalents) and a Moliere radius of the shower cone, $R_M = 11\,\mathrm{cm}$. This is useful to identify electromagnetic showers from hadronic showers. Muons produce minimum ionizing tracks which deposit $\sim 1.76\,\mathrm{MeV/cm}$, corresponding to roughly 2.2 GeV total energy deposit for a muon travelling horizontally across the length of the active portion of the Near Detector.

Finally, the Far Detector is placed on surface in a detector hall with a modest overburden of 1.2 m concrete and 15 cm barite. Here it collects an enormous number of cosmic rays at a rate $\sim 150\,\mathrm{kHz}$, which acts as a fundamental background to the oscillation analysis. Fig. 2.12 shows a snapshot of the interactions at the Far Detector. The Near Detector is placed 100 m underground, where cosmic rays are negligible. The beam divergence across the 810 m baseline results in smaller statistics at the Far Detector compared to the Near Detector but a tighter flux peak at 2 GeV assuming no oscillations.

2.3.1 Data Acquisition System

As mentioned before, each channel of the APD reads in the wavelength-shifted scintillation photons as a particle deposits energy in a cell. The job of the APD channel is to shape the photon pulse in terms of photo-electron (PE) units and thus provide a measure of the total PE deposited in the cell. Each APD is serviced by a custom Front-End Board (FEB) designed by NOvA that digitizes this signal (in terms of ADC units) after discriminating between different pulses and associates a timestamp to the pulse.

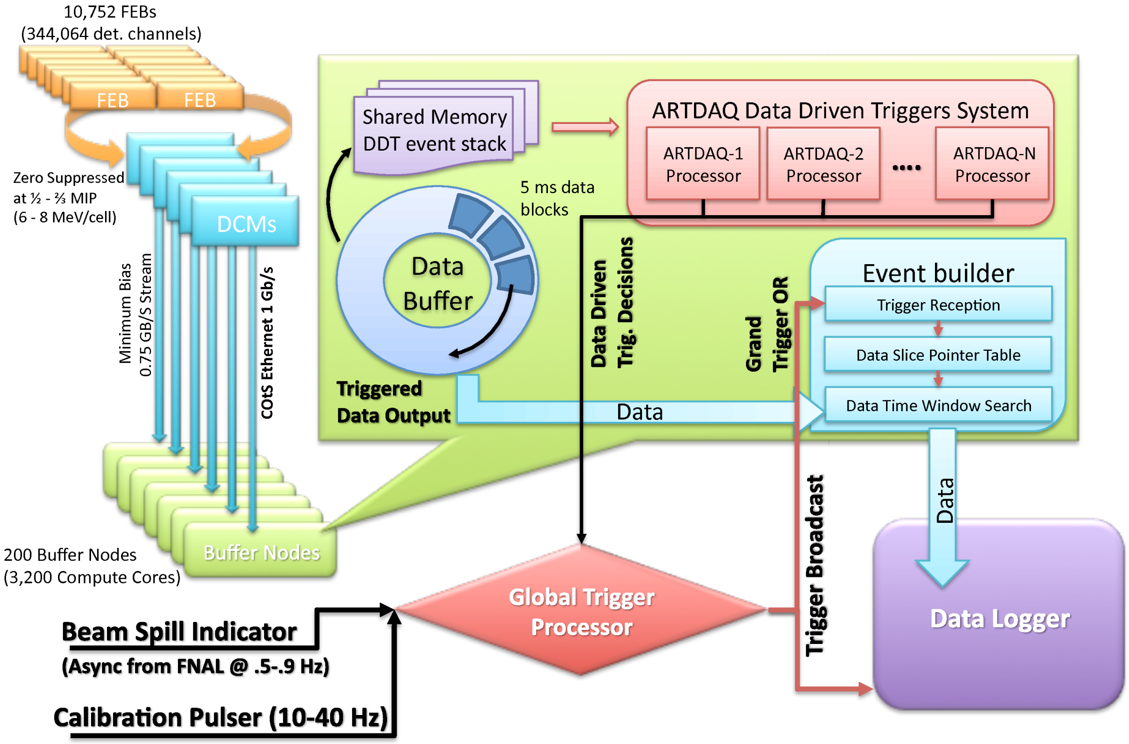


Figure 2.14: Schematic of the NOvA Data Acquisition System

The FEBs run at a clock rate of 16 MHz. In the Far Detector, the 32 channels are multiplexed 8 at a time, so it is able to read out the ADC pulses at an effective rate of 2 MHz, similar to the programmed rise-time of the pulse. At the Near Detector, where the intensity of the beam causes a significant amount of pile-up, the channels are multiplexed 2 at a time, which quadruples the read-out rate and improves pulse discrimination. In addition, the FEBs employ Dual Correlated Sampling (DCS) where the signal is discarded if the measured pulse height from two sample values are below a threshold ($\sim 20\,\mathrm{PE}$). If it is above threshold, the pulse is read out in multi-point mode (although there is a fraction of Far Detector data in single-point mode) where 4 consecutive samples of the rising edge of the pulse are forwarded to the storage element, as shown in Fig. 2.15. The pulse shape is then determined by an offline fit which further improves the timing resolution of the detector, $\sim 10\,\mathrm{ns}$ at FD, $\sim 6\,\mathrm{ns}$ at ND [78].

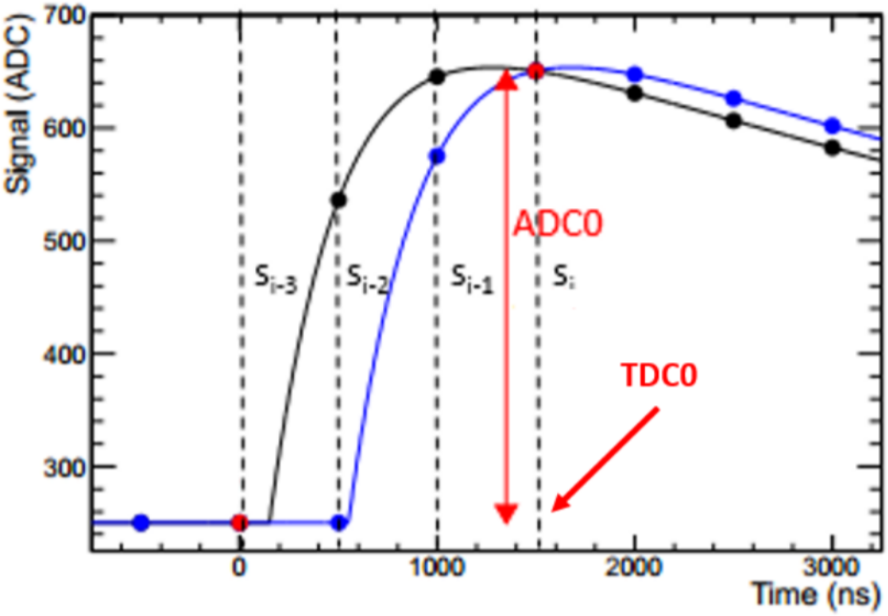


Figure 2.15: Illustration of multi-point readout where intermediate values from the ADC outputs are saved to provide precise pulse discrimination within a short time window [79]

A Data Concentrator Module (DCM) collects signals from 64 FEBs in windows of $50\,\mu\mathrm{s}$ ("microslices"). These are then continuously sent to buffer nodes via a 10 Gbps optical fiber, from which a trigger decides to build events of a certain timespan based on the associated timestamp of the microslices. Given the cosmic ray rate and the number of channels in the Far Detector, the data throughput can be more than 1 GBps. The relevant triggers for the oscillation analysis are :

- ND and FD NuMI Trigger : This builds a $550\,\mu\mathrm{s}$ event centered around a $10\,\mu\mathrm{s}$ beam

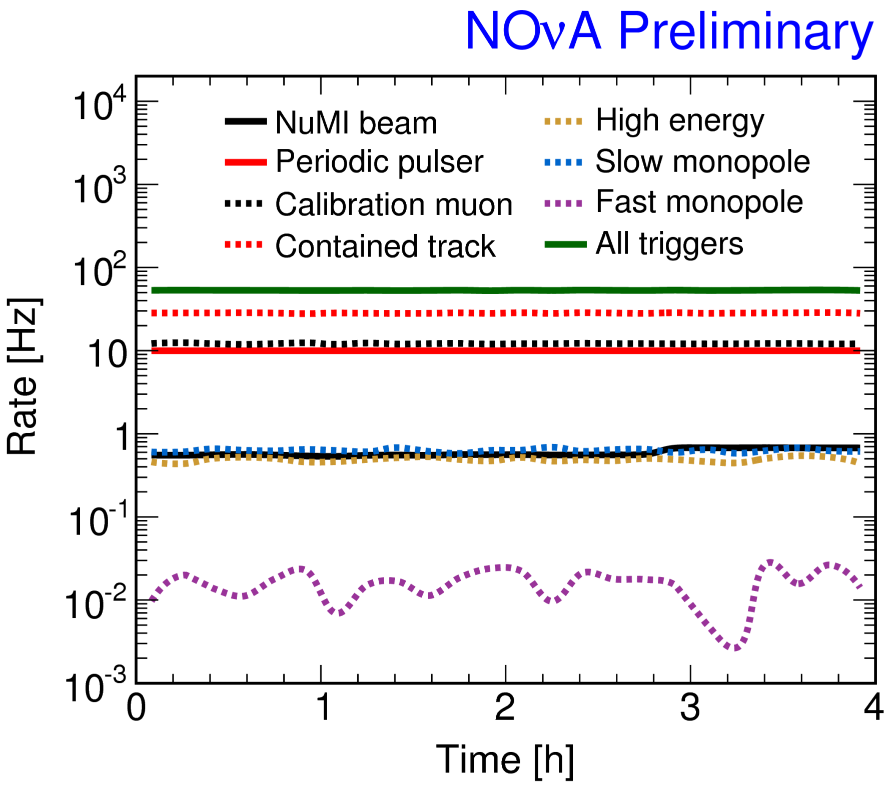


Figure 2.16: Trigger rates for various NOvA Triggers at the Far Detector

spill whenever a signal is received from Fermilab. The rate at which this happens is $\sim 1\,\text{Hz}$. The beam is therefore always in the range of $218 - 228\,\mu\text{s}$ relative to the start of the trigger window. A fraction of the early NuMI data however spans a range $64\,\mu\text{s}$ shifted from this range (i.e $282 - 292\,\mu\text{s}$) due to faulty timing. These are the main events studied in the oscillation analysis.

- ND DDActivity Trigger : This trigger builds events of $50\,\mu\text{s}$ whenever there is activity in the detector outside the NuMI trigger window. These are primarily cosmics at a rate of $\sim 5\,\text{kHz}$. These events are used to calibrate the Near Detector response.

- FD Cosmic Trigger : This builds $550\,\mu\text{s}$ events at a rate of $10\,\text{Hz}$ outside the NuMI trigger window. These serve as minimum-bias cosmic events that can be used to calibrate the Far Detector and provide an estimate for the cosmic backgrounds in the oscillation analysis.

Fig. 2.14 gives a rough schematic of the NOvA data acquisition system, while Fig. 2.16 compares the trigger rates for different NOvA triggers. Finally, there exists a timing system made up of a series of Time Distribution Units (TDUs) that synchronizes all the channels in a detector to the same local time. It also maintains a global time synchronized with a GPS clock given in UTC, which is used to co-ordinate events between the detectors and the accelerator at Fermilab.

2.3.2 Detector Simulation

NOvA uses Monte-Carlo (MC) methods to simulate each step of the physics process of neutrino interactions in the detector environment, which for eg. at the Near Detector could involve a substantial number of interactions in the surrounding rock. The neutrino interaction generator used is GENIE [80] (v3.0.6). As mentioned before the main interaction element is Carbon, although there is in general a variety of elements that the neutrino can interact with, including Hydrogen, Silicon (for rock interactions) etc. GENIE uses a variety of data-driven models that predict the cross-section for interactions with each of these elements (initial-state) as well as the transport of particles produced within the nuclei (final-state).

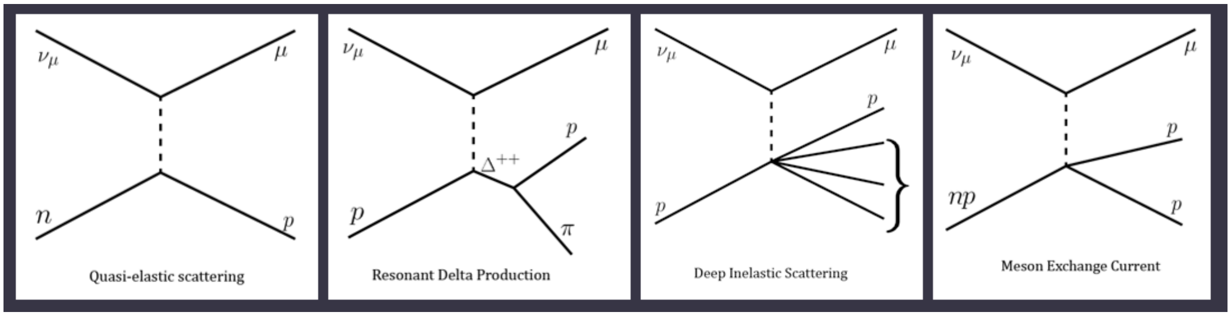


Figure 2.17: Feynman Diagrams for various important types of neutrino interactions at NOvA [81]

Fig. 2.18 shows the current picture of the neutrino and anti-neutrino inclusive cross-sections at energy ranges relevant to NOvA. The NOvA energy range makes it sensitive to many

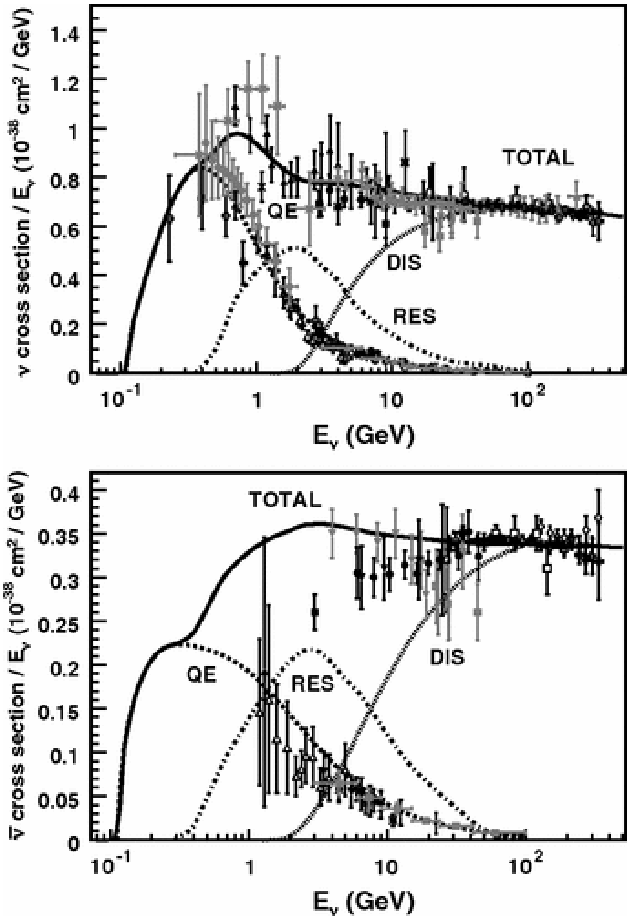


Figure 2.18: The cross-sections for various types of neutrino interactions along with the inclusive cross-section as a function of energy for neutrinos (top) and anti-neutrinos (bottom). Plotted with various measurements over the years. Taken from [82]

different interaction types, as shown by the various Feynman diagrams given in Fig. 2.17, each with a significant cross-section. The main ones are :

- Quasi-Elastic (QE) : Here the neutrino scatters off a single nucleon within the nucleus elastically. In the Charged Current (CC) case, the neutrino (anti-neutrino) interacts with a neutron (proton) producing a proton (neutron) and the relevant lepton flavor associated with the neutrino. The overall $Q^2(= -q^2)$ for this interaction is relatively low and QE events are generally characterized by clean lepton signatures travelling relatively in line with the beam direction and a small hadronic component which usually involves a single proton track for the neutrino case. GENIE uses the Valencia 1p1h model which takes into account various nuclear effects coming from a local Fermi gas nuclear model with some long-range correlations (Random Phase Approximation - RPA) [83]. Additional experimental constraints on the axial form factor are modelled as a z-expansion [84], related to the axial current propagator.

- Meson Exchange Current (MEC) : These are interactions at a slightly higher Q^2 where the neutrino scatters off a correlated nucleon pair instead of a single nucleus. The nucleon pair is predominantly a np pair which produces two protons (neutrons) in the final state for neutrinos (anti-neutrinos). The theory on this is highly uncertain and GENIE uses calculations made at Valencia once again [85].
- Resonant Interactions (RES) : These interactions produce baryon resonances in the final state (Δ) which decay into pions and protons. If a π^0 is produced that usually decays via $\pi^0 \rightarrow \gamma\gamma$. Sometimes a single γ can be produced by the Δ decay as well ($\Delta^+ \rightarrow p\gamma$). These can exhibit a variety of signatures but usually involve either charged pion tracks which can hard-scatter in the detector and produce kinks or a pair of electromagnetic shower cascades coming from π^0 decay. Sometimes multiple proton tracks can be seen as well. This is modelled in GENIE using the Berger-Sehgal model that includes effects from differences in μ and e mass [86].
- Deep Inelastic Scattering (DIS) : These are in general messy interactions at a very high Q^2 where the relevant initial state particles involved are the quark soup within each nucleon. The event signature is a large hadronic component due to the resulting hadronization of the final state quarks. Here the lepton signatures aren't very clean, being relatively small and travelling at a large angle with respect to the beam. These events are modelled in GENIE using the Bodek-Yang calculation [87].

GENIE is also able to model important final-state interactions (FSI) where the particles produced as above, move within the nucleus and produce pions and other nucleons on their own. This is done using a semi-classical framework called the hN-cascade model [88]. NOvA applies a further tune for some of the parameters within these models as they were found to be fairly discrepant with the Near Detector data. Conservative uncertainties are also chosen where appropriate given the state of current knowledge.

The final state particles and their 4-momenta coming from GENIE are propagated within the detector material and the surrounding rock, using GEANT ($v4.2$) and a full-scale model of the detector environment geometry.

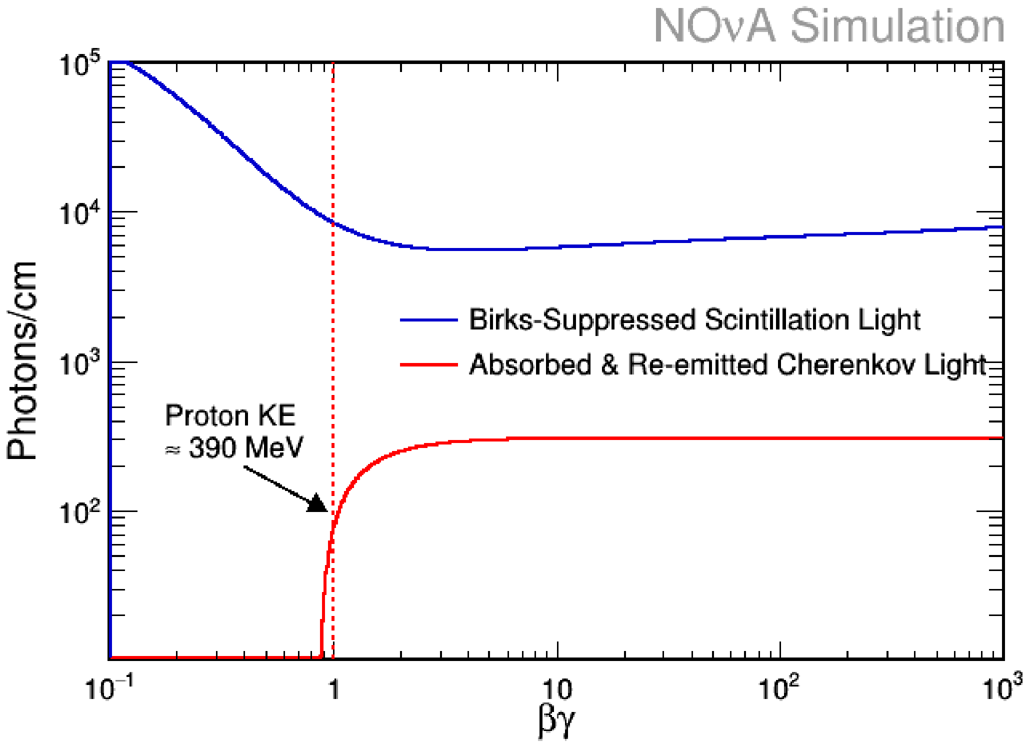


Figure 2.19: Light Yield in a NOvA cell as a function of the particle $\beta\gamma$

Finally, NOvA-specific modules (PhotonTransport and ReadoutSim) are used to simulate the production and the collection of scintillation light within a NOvA cell and their conversion to PE/ADC units for a given physics hit [89]. The scintillation model involves a Birks-Chou parameterization (for quenching and recombination effects) of the number of photons produced as well as an additional term that takes into account Cerenkov emission [90], as illustrated in Fig. 2.19. The parameters in the model are then tuned to bring it into agreement with stopping muon samples developed at the Near Detector.

For cosmic ray interactions, a wealth of data is collected using the cosmic ray trigger and overlaid directly onto the FD simulation. However, where necessary, a separate set of files using the CRY generator are produced to simulate cosmic interactions at FD using the expected flux [91].

2.3.3 Detector Calibration

In addition to a model for the detector response, a calibration is done for both real data events and simulated MC events (coming from ReadoutSim) to convert the PE output of each cell into a GeV estimate. The scintillation photons collected by different points on the wave-length shifting fiber are attenuated along the length of the fiber. This results in an attenuation of the measured PE as the light is collected further away from readout. This is a particularly large effect at the Far Detector because of the length of the cell. In addition, it was found that different fibers have different attenuation lengths and overall responses in terms of the total collected PE. So the first step of the calibration procedure is to do a relative (cell-by-cell) calibration to take into account these effects. This procedure uses a through-going muon sample which are sufficiently minimum ionizing and hence uniform over the extent of the detector. The next step is to perform an absolute calibration that converts the corrected PE response (PECorr) to a GeV estimate. The absolute calibration uses a stopping muon sample in order to get a clean estimate of the total energy deposited by the muon using the well known Bethe-Bloch equation. At the Near Detector, both samples are given by cosmic events using the DDActivity trigger. At the Far Detector, events from the Cosmic trigger are used. CRY is used in both cases for the calibration of simulated events.

For the relative calibration to data, the effect from the non-uniformity of fiber responses is already taken into account by performing the calibration in each cell and each plane of the detectors individually. For the simulation, the procedure is a bit different. This is done so as to not have to simulate each channel of the two detectors individually. The attenuation properties of each fiber were measured during the construction of the detector and a double-exponential fit is performed to the individual fiber quality data. The distribution of attenuation lengths are then binned in 12 quantiles of "fiber-brightness". Cells that stride across the boundaries of these quantiles are then simulated with a fiber of attenuation length given by the boundary of that quantile [92]. The relative calibration is then done in each of these 12 fiber-brightness bins by picking out the relevant cells.

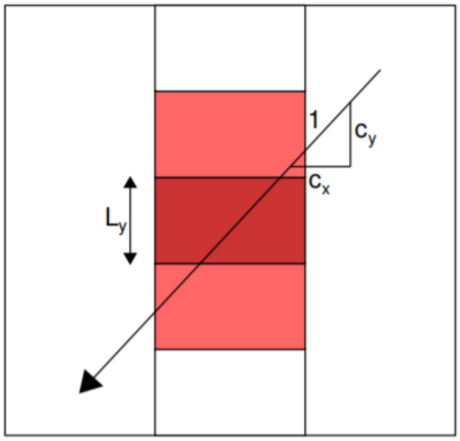


Figure 2.20: The tricell selection for the attenuation calibration. The Path length of the particle is estimated to be L_y/c_y where c_y is the direction cosine along the y-direction

The initial step of the relative calibration is to collect a set of suitable hits from the through going muon sample. The suitability is determined by being able to get a reliable estimate of the path lengths of the muon through any given cell. The cosmic events are characterized by straight-line downward going muons. A minimal and fast reconstruction is performed assuming a straight-line topology (CosmicTrack) in order to cluster hits of a cosmogenic nature. Hits from the cosmic track are selected for calibration if there exists two neighbouring hits in the same plane, below and above the given hit ("Tricell" selection), as illustrated in Fig. 2.20. Hits are also selected if one of the neighbouring hits is a dead channel. The path length of the muon can then be estimated rather precisely by the ratio of the width of the cell and the direction cosine reconstructed using CosmicTrack. The PE output is then normalized by the path length in units of PE/cm, as shown in Fig. 2.21 and 2.22. This is done in order to correct for biases in the reconstruction of muon tracks for different track angles [93].

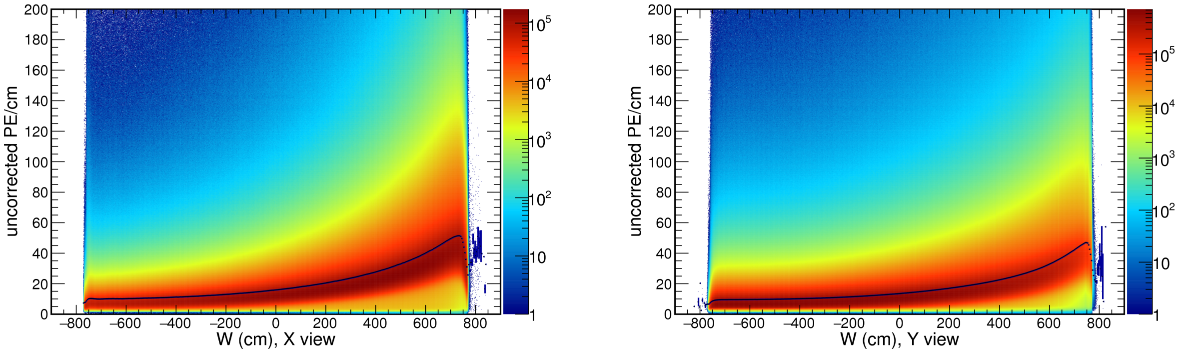


Figure 2.21: The raw PE/cm distributions as a function of longitudinal position (W) inside a Far Detector cell for the X-view (left) and the Y-view (right). The black curves represent the profile of the 2D distributions.

From this, a profile of the mean PE/cm as a function of W (the distance along the cell from its center) for each cell is made. The profile is then corrected for threshold and shadowing effects. Threshold effects arise from the fact that for a hit to be seen, the PE deposited needs to be above a certain threshold which could bias the attenuation calibration. Shadowing effects arise from detector self-shielding where even for a given true energy deposited within the cell for a minimum ionizing particle, there is a certain non-uniformity in the visible energy output across the detector. The corrections for these effects are taken from simulation as

$$T = \frac{PE}{\lambda} \times \frac{E_{true}}{E_{mip}} \tag{2.2}$$

where T is the correction factor, PE is the simulated photo-electrons at the readout, λ is the Poisson expectation for the PE, E_{true} is the true deposited energy in the cell and E_{mip} is the energy deposited by a naive scaling of the path-length. The first factor simulates the

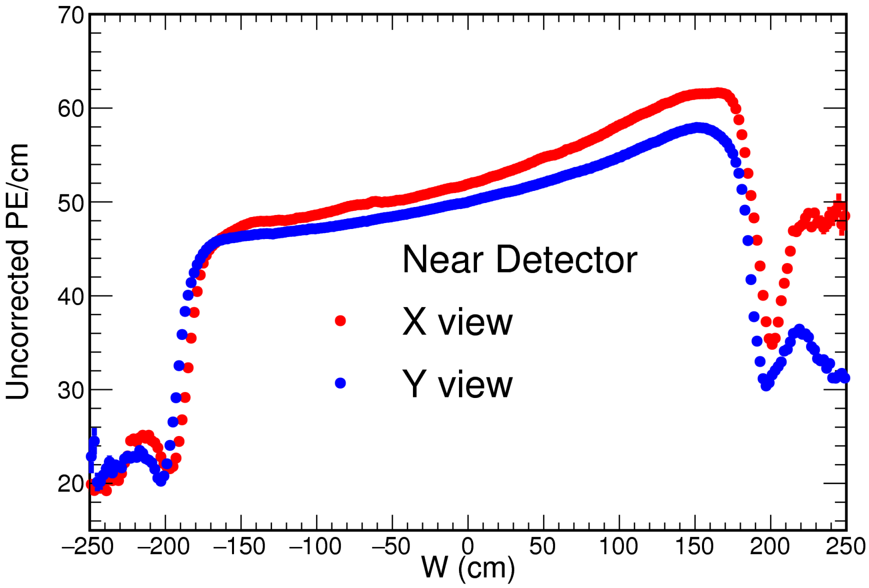


Figure 2.22: The raw PE/cm attenuation profile as a function of longitudinal position in a cell (W) for the Near Detector

threshold effect and the second, the shadowing effect. The corrections are smoothed using a polynomial fit and applied as a function of W, cell, view (X : top view, Y : side view) and for the simulation, fiber brightness. Fig. 2.23 shows the smoothed threshold and shadowing corrections for each cell at the Far Detector. The biggest threshold effects arise at the far end of the cell furthest from readout, where the attenuation is large enough to drop physics hits below threshold. The biggest shadowing effects are at the diblock boundaries or the bottom of the detector as the downward going muons have a different energy deposition profile.

The corrected profiles are then fit to a functional form given as :

$$y = C + A\left(exp\left(\frac{W}{X}\right) + exp\left(-\frac{L+W}{X}\right)\right) \tag{2.3}$$

where C and A are fitted parameters, X is the fitted attenuation length, L is the length of the cell, y is the corrected PE/cm. The functional form is motivated by the fact that the fibers are bent in a U-shape across the cell, so the scintillation photons have two paths to the

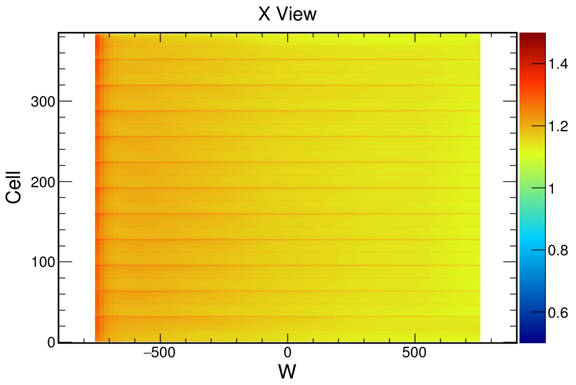


Figure 2.23: Threshold and Shadowing Corrections for the Far Detector for the X-view (left) and Y-view (right). The shadowing can be seen at various diblock boundaries as horizontal features while the thresholding effect is seen at the edges of each cell for negative W.

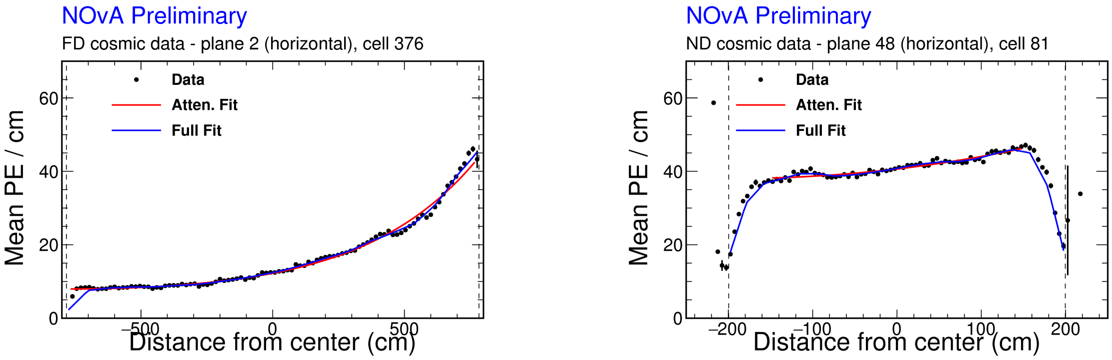


Figure 2.24: The fitted attenuation profiles including the full LOWESS fit for the Far Detector (left) and the Near Detector (right)

readout. The fit is performed in a range of $[-750, 750]$ cm for the Far Detector, $[-150, 150]$ cm in the active portion of the Near Detector and the Y view of the muon catcher, $[-150, 50]$ cm for the X view of the muon catcher. Near the edges of the cell there is a pronounced roll-off observed in data. To capture this and the remaining shape in the residuals to the exponential fit, a LOWESS fit is done to smooth out the variations, as shown in Fig. 2.24. The curve is formed by a weighted mean of the deviations from the exponential fit, where the weighted mean at a given point is taken from nearby points in a range $[-30, 30]$ cm who get a weight :

$$w_i = \left(1 - \left|\frac{W - W_i}{\sigma}\right|^3\right)^3 \tag{2.4}$$

The weight is taken to be 0 if W_i is more than 30 cm away from W.

To take into account effects from changing detector conditions over time, the relative calibration is done separately for different epochs of data taking where the overall detector response is expected to be constant.

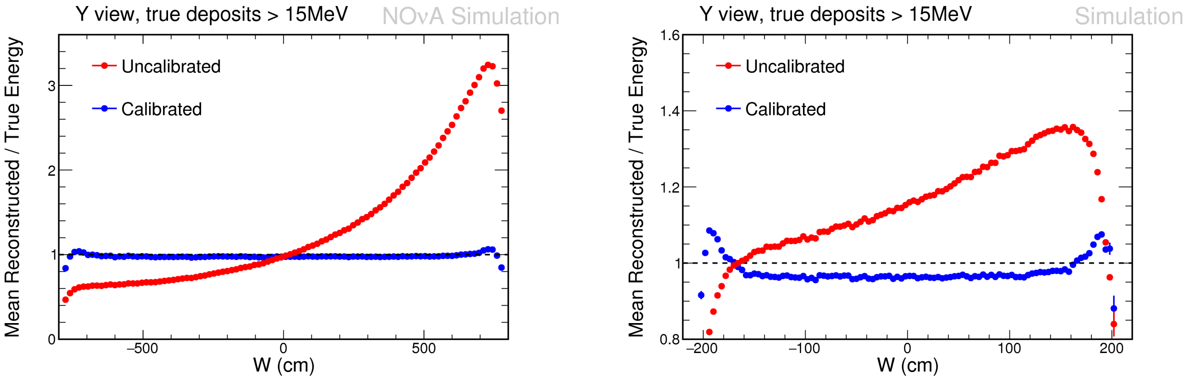


Figure 2.25: Profiles of the Reconstructed/True Energy for each hit $> 15\,\mathrm{MeV}$ before and after the calibration

Finally, as shown in Fig. 2.26, an absolute calibration is performed on the PECorr/cm output from the above procedure using a stopping muon sample that have muon tracks sufficiently far away from detector edges where the attenuation calibration relies on simulation. The absolute calibration selects a $1-2\,\mathrm{m}$ track window from the end of the muon track (where the deposited energy is expected to be MIP-like) and converts the PECorr/cm measurement to a dE/dx estimate (in units of GeV/cm) using the Bethe-Bloch equation. This sets the energy scale of the detector [93].

Fig. 2.25 shows the change in the reconstructed energy due to the overall calibration procedure as a function of position in the cell. The overall calibration procedure is also validated with some standard candles like the π^0 mass peak [94] and the Michel e^- spectrum.

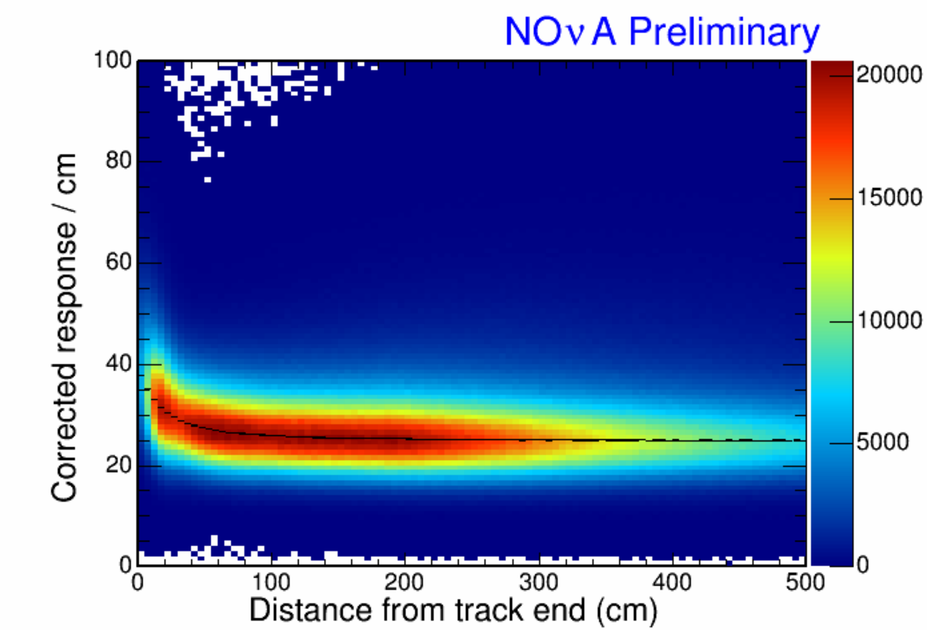


Figure 2.26: The attenuation corrected PE/cm distributions for the stopping muon sample as a function of distance from the track end. The black curve shows the fit for the overall energy scale.

2.4 Event Reconstruction

The reconstruction of interactions in the detector deals with clustering calibrated hits and consequently provides relevant physics information about the interaction in terms of the energies and momenta of the resulting particles. This is vital in order to characterize the topology of the interaction and thus provide means to identify the neutrino flavor and the incoming neutrino energy. There are 4 main interaction classes that are relevant for the oscillation analysis.

- Charged Current ν_μ : These are readily identifiable by a long and straight muon track made up of minimum ionizing hits. The hadronic component depends on the details of the interaction but are in general clustered around the interaction vertex.
- Charged Current ν_e : These are characterized by an electron in the final state which produces a roughly conical shower cascade driven by pair production as the electron

interacts with the dense material. The hadronic component is similar to the CCν_μ case. The radiation length of the material is sufficient to produce a well developed electron shower, which is on average 6 cell-depths. In general it is difficult to separate oscillated CCν_e events and events coming from the intrinsic beam contamination. As a result, the beam ν_eCC component is an irreducible background to the appearance channel. However, a small separation in their energy distributions can be exploited to improve sensitivity.

- Neutral Current (NC) : These are flavor independent interactions where the observed final state particles are only the hadronic component, as the neutrino just scatters off the nuclei and can't be observed. If the hadronic component involves a π^0 then one can observe the resulting pair of γs from their decay. The γs produce shower cascades very similar to electron showers and can therefore mimic CCν_e events at typical energies, especially if the opening angle between the γs isn't very large to begin with. Charged pions can also be produced resulting in track like topologies, which can be misidentified as short muon tracks and thus CCν_μ events. As a result, these events are important backgrounds to both oscillation channels.

- Cosmic Interactions : At the Far Detector, where a large number of cosmic rays are seen, one can in general observe a variety of topologies. Most of them are long muon tracks coming in from the top of the detector and can be differentiated by their directions from a CCν_μ event. Some of them can also emit bremsstrahlung radiation at a variety of angles, which are sufficiently energetic to mimic CCν_e events. Therefore, these are important backgrounds as well to the two oscillation channels.

Three of these are illustrated in Fig. 2.27. The goal of the reconstruction chain is to first isolate separate interactions in the detectors and then provide a full physics description of the event, in terms of their locations, the types of particles produced and their directions. All these steps exploit the fine timing and segmentation offered by the two detectors as

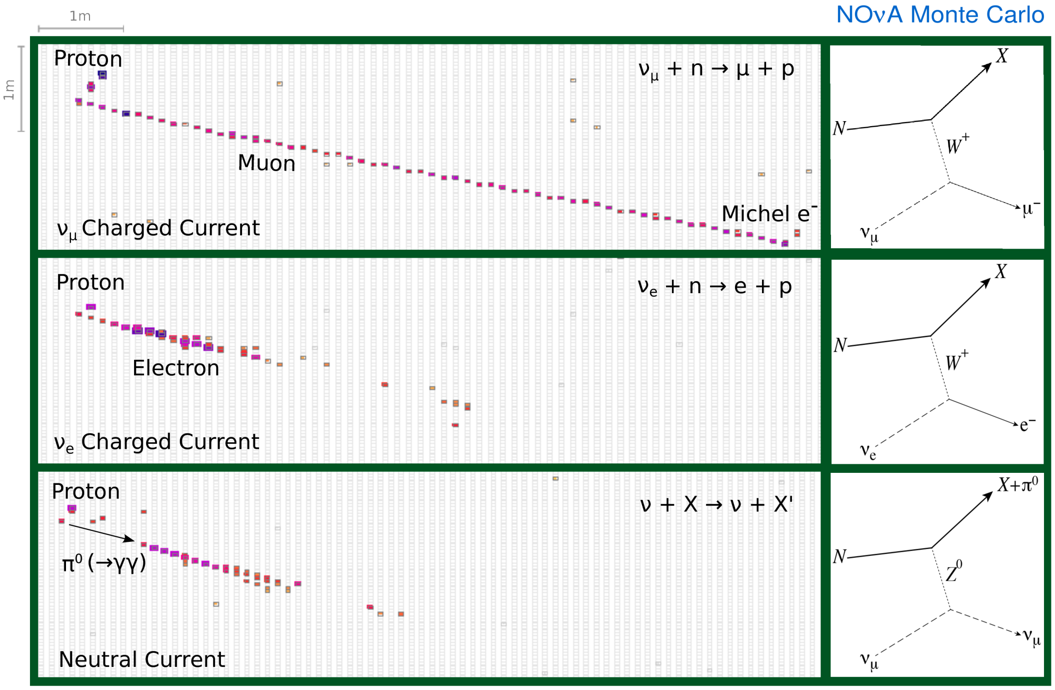


Figure 2.27: Examples of interaction topologies seen at the NOvA detectors for different types of neutrino interactions, with ν_μ-CC (top), ν_e-CC (middle) and NC (bottom)

described in the previous section.

2.4.1 Event Clustering

The original NOvA event clustering algorithm, called Slicer4D, relied on a DBSCAN method [95] in order to group relevant physics hits correlated in space and time. This was reasonably successful that it was used for all of NOvA's oscillation results until the last one [96]. Updates to this method were mainly motivated by pile-up issues related to the high-intensity NuMI beam ($> 700\,\text{kW}$).

The Slicer4D design was built on being able to cluster neighbouring hits across the two projections (XZ : top view, YZ : side view). The algorithm proceeded in a fairly straightforward way where a minimal cluster of neighbouring hits is built, where neighbouring points are defined as points that are above a certain local density threshold, such that the cluster can expand to include all such points and then terminate at a point that falls below the density threshold. The density threshold is tuned according to a neighbour score that penalizes large distances in space and time.

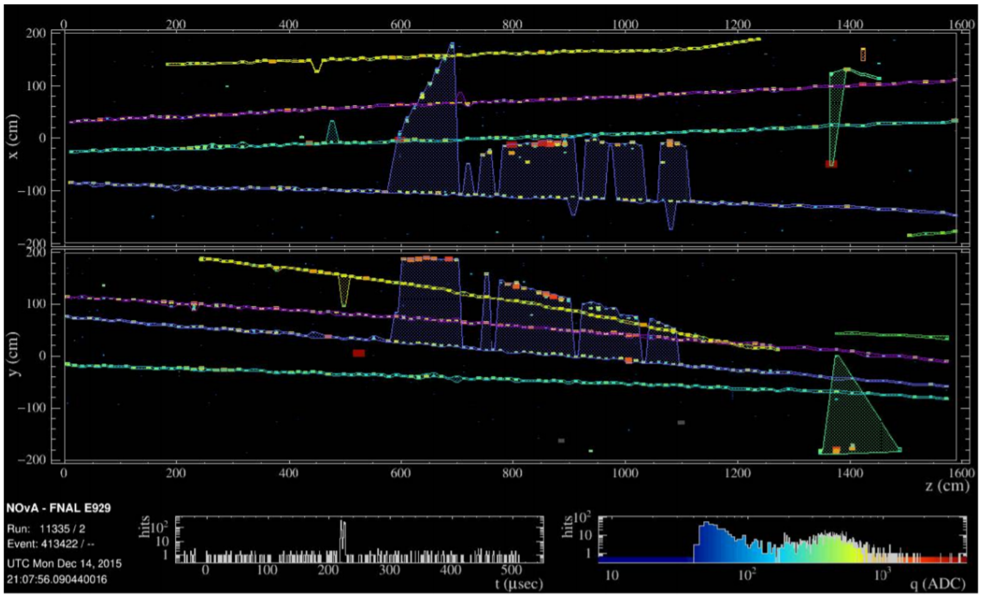


Figure 2.28: Examples of clustering failures seen at the Near Detector due to pile-up where causally separated interactions are nevertheless still clustered together resulting in a loss of reconstruction efficiency

For hits in opposing projections however, a neighbour score cannot include distances between x or y co-ordinates because one of them is ill-defined for each hit. For such points, the metric relies on the z and t co-ordinates only. As such, hits that are causally separated in truth with a large Δxy can be clustered together because of a small Δzt. In addition, Slicer4D can be fairly greedy in grouping up noise hits and also has some failure modes associated with pile-up, especially in the Near Detector but also sometimes in the Far Detector, due to the large number of cosmogenic hits in the environment of a neutrino cluster. The failure mode arises from certain "bridge hits" that can lie sufficiently between distinct physics clusters, that the cluster expansion includes the bridge hit and consequently the distinct physics slice. This is especially apparent at the Near Detector where a large number of rock-muons are seen traversing the length of the detector and a cluster for a contained neutrino interaction *in* the detector expands to include the rock-muon as well, as can be seen in Fig. 2.31. This pile-up related failure mode results in an attenuation of slicing efficiency with respect to the intensity of the beam, which can be $\sim 12\%$ at 700 kW.

To address these issues, a new slicing algorithm called the Time Density Slicer (TDSlicer) has been implemented for the latest set of oscillation results [97]. TDSlicer relies on finding a centroid for a grouping of hits, upon which a cluster is built *separately* for each projection [98]. The two 3D clusters thus do not contain causally separated hits and as a consequence, a merging operation can be performed that associates clusters in one projection to a cluster in another.

To estimate the centroid, a local density (ρ) and isolation score (δ) are defined for each hit as :

$$\rho_i = \sum_j e^{-d_{ij}^2/\tau^2} \tag{2.5}$$

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}) \tag{2.6}$$

where d_{ij} is the causal distance between hits, i and j, defined as :

$$d_{ij} = \frac{||\Delta t| - R/c|}{\tau} \tag{2.7}$$

Here, τ is a constant parameter related to the timing resolutions of the detectors, 16 ns and 80 ns for ND and FD respectively. R is the spacial distance between the two hits. This is done separately for each projection. A large density score identifies a reasonable cluster in the environment while requiring a large isolation score provides the centroid estimate within the cluster. Centroids are thus found by requiring $\rho > 3(10)$ and $\delta > 8(6)$ for the ND (FD). This is illustrated in Fig. 2.29. The cluster is then built by grouping hits with the centroid at its base and accepting points that are within a certain distance threshold ($d_{ij} = 8(5)\tau$ for

ND(FD)) to a point already in the cluster.

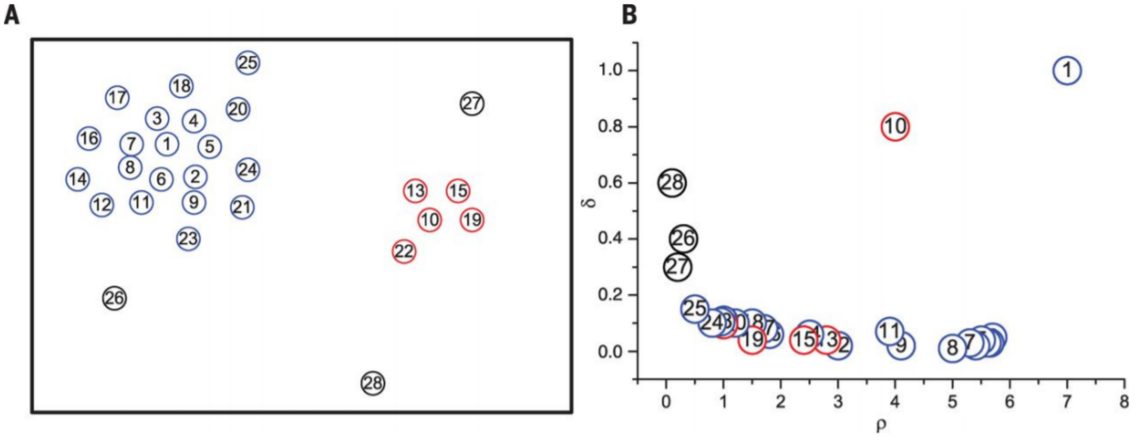


Figure 2.29: Illustration of the centroid finding algorithm for TDSlicer. Here, points labelled 10 and 1 have a high local density (ρ) and isolation score (δ) and as a result are tagged as centroids to build clusters around.

Finally a view merging step is performed that penalizes pairs of 3D clusters based on differences in their average zt compared to length and time scales, 200(60) cm and 6(60) ns for ND(FD). The 4D cluster is then given by the best such pairing across all possible pairs. The remaining unmatched clusters are tagged as noise.

TDSlicer thus achieves a trade-off which is to improve upon the “greediness” of Slicer4D by ignoring the bridge-hits at the cost of a modest loss of low-energy physics hits at the edges of the interaction in some cases. This results in an improved attenuation profile as a function of intensity (shown in Fig. 4.11) as well as a larger overall slicing efficiency and increased slice purity, defined as the fraction of the energy in the cluster coming from the true particle. The overall completeness, defined as the fraction of deposited energy in the cluster by the true particle, is slightly reduced.

One way to validate the performance of the new slicer is to study it in an artificially intense environment using a special “overlay” sample. In this method, a MC neutrino interaction is overlaid into a 10 μs nominal beam spill made up of real data events at the Near Detector,

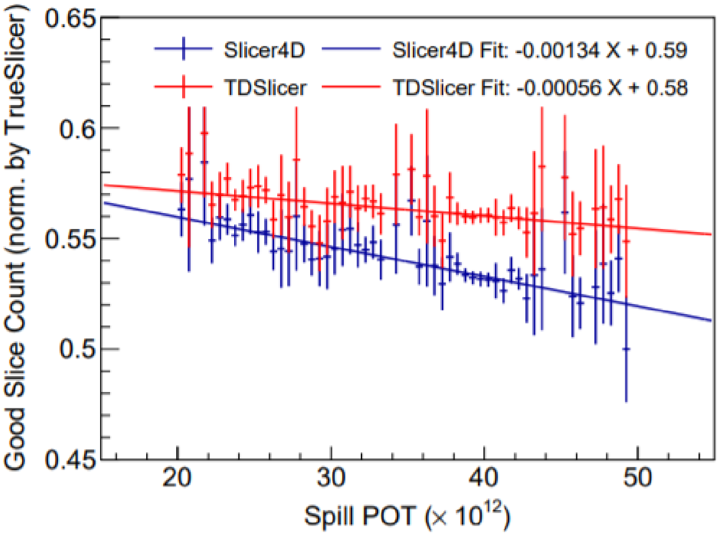


Figure 2.30: Comparison of performance of TDSlicer vs Slicer4D as a function of beam intensity. TDSlicer shows smaller attenuation at higher intensities and as a result is less sensitive to pile-up

as illustrated in Fig. 2.32. The slicer can then be studied using the truth information given in the overlaid MC event. One can also compare behavioural differences between the two slices on the same MC event [99].

This study validated expectations that TDSlicer was less sensitive to pile-up with larger beam intensity. As a result, the efficiency of making good slices (defined as slices with $> 90\%$ completeness and purity) that are contained in the detector increased by $\sim 5\%$. The actual number will be slightly lower as the effective intensity in the overlay sample is higher due to more having neutrinos per beam spill. The validation also showed that TDSlicer was in general less prone to the failure mode associated with clustering rock-muons and the neutrino event in the detector. At the Far Detector, where an overlay sample is unncessary and instead real cosmic data events are overlaid on top of the beam MC events, the improvement in slicing efficiency for good contained slices was found to be $\sim 9\%$ and the fraction of events with a neutrino slice was improved by $\sim 6\%$.

Using the overlay sample at ND, the mean completeness of good contained slices degraded

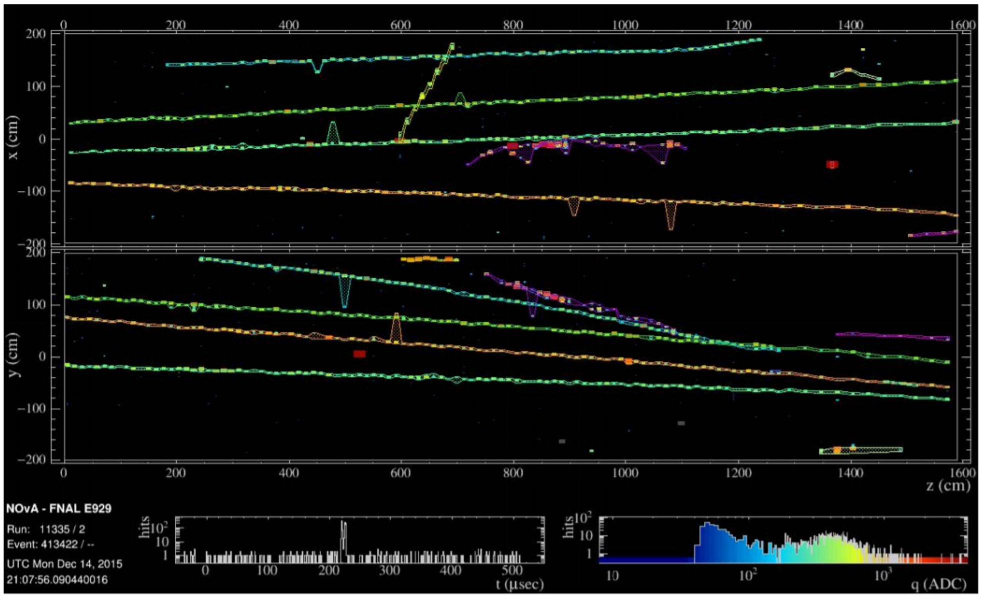


Figure 2.31: Example of the same event as before showing TDSlicer is able to separate out interactions compared to Slicer4D

TDSlicer	MC-MC		Data-MC		
Stage	Events (10^3)	Efficiency (%)	Events (10^3)	Efficiency (%)	Difference (%)
Overlaid	138.50	100	211.70	100	0.000
Sliced	137.00	98.92 ± 0.028	209.43	98.93 ± 0.022	0.010 ± 0.0360
Quality	129.54	93.53 ± 0.066	197.71	93.39 ± 0.054	-0.153 ± 0.0912
Contained	32.61	23.54 ± 0.114	50.04	23.64 ± 0.092	**0.414 ± 0.6232**

Table 2.1: Comparison of Selection Efficiencies at various cut stages for MC on MC spills and MC on Data spills using the special overlay sample reconstructed with TDSlicer

by ∼ 1% as the resulting new slices made were found to be less complete in general. A loss of ∼ 2% in capture efficiency of neutron hits was found, where the largest differences were seen for delayed neutrons which zoom around in the detector for a fair amount of time. This had negligible effects on the neutrino energy resolution once the energy estimators were re-tuned with the clusters from the new slicer.

One can also develop an overlay sample for MC events on MC simulated spills. By comparing analysis selection efficiencies between the MC-on-data and MC-on-MC overlay samples, one can get a handle on normalization effects that creep in to the simulated clusters, mainly related to detector electronics noise. The normalization effects were in general found to be

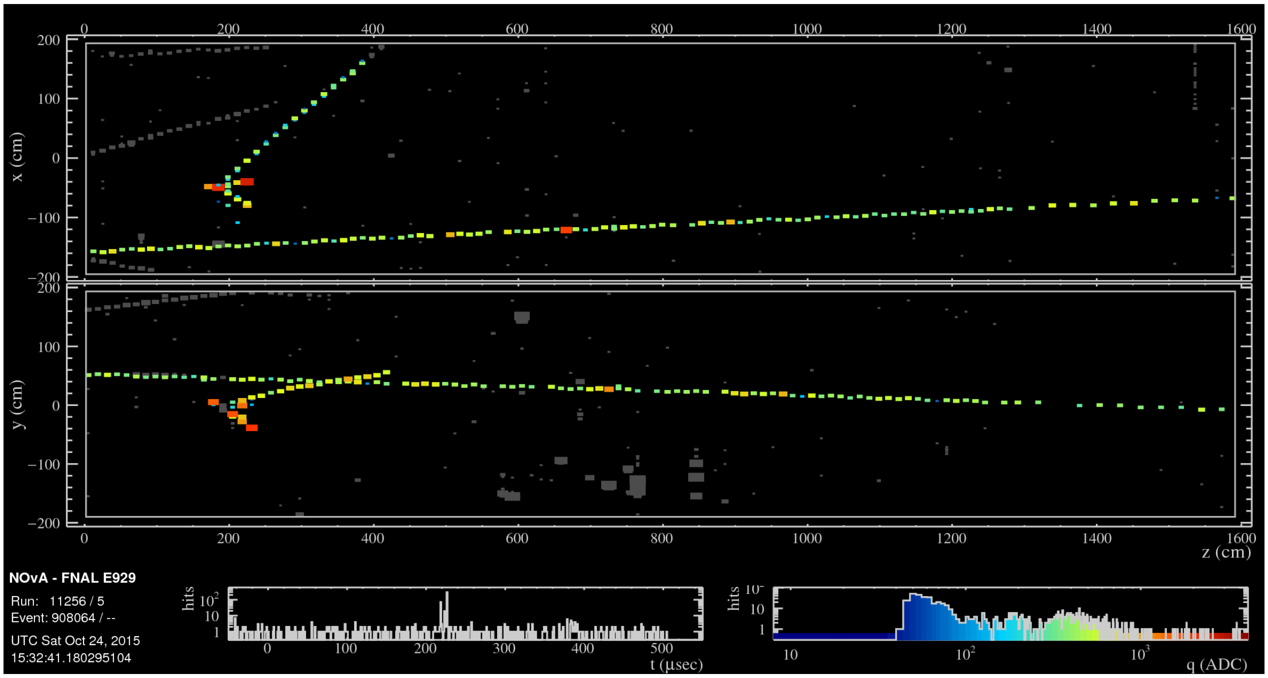


Figure 2.32: Illustration of the special overlay sample where a MC ν_μ event is overlaid onto a real data spill at the Near Detector

Slicer4D	MC-MC		Data-MC		
Stage	Events (10^3)	Efficiency (%)	Events (10^3)	Efficiency (%)	Difference (%)
Overlaid	82.10	100	184.51	100	0.000
Sliced	81.83	99.67 ± 0.020	183.83	99.63 ± 0.014	-0.046 ± 0.0245
Quality	76.37	93.03 ± 0.090	172.32	93.39 ± 0.058	0.390 ± 0.1140
Contained	18.14	22.09 ± 0.145	41.49	22.48 ± 0.097	**1.781 ± 0.7895**

Table 2.2: Comparison of Selection Efficiencies at various cut stages for MC on MC spills and MC on Data spills using the special overlay sample reconstructed with Slicer4D

smaller when looking at TDSlicer clusters vs Slicer4D clusters, indicating that the new slicer was able to reject noise hits better. Tables 2.1 and 2.2 compares the performance between the two slicers in this regard.

As a result, TDSlicer was chosen to be the main clustering algorithm for the latest oscillation analysis.

2.4.2 Vertex Finding

The next downstream step of the reconstruction chain [100] is to identify the interaction vertex of the neutrino interaction, once the cluster associated with the neutrino has been identified as above. In each view, lines are drawn between every pair of hits in the cluster and mapped to Hough space [101], parameterized by the 2D polar co-ordinates (ρ, ϕ). This is a projective transformation where lines are mapped to points. The Hough lines are illustrated in Fig. 2.33. To save computation, horizontal lines are ignored if the hit pairs are more than a certain distance threshold apart. The points in Hough space are then smeared by a Gaussian, where the weight of the smear is given by a "vote" that penalizes a large distance between hits and large angles from the beam direction. As a result, if many hit pairs are collinear to each other, a gaussian peak is formed in the vote distribution in Hough space. The peaks are larger if more hits are collinear. Dominant lines are thus found by finding the largest peak and its associated line, removing the hits associated with that peak and iterating the process.

These dominant lines are used as input to an Elastic-Arms algorithm that identifies a vertex candidate. Vertex "seeds" are chosen as intersection points of the dominant lines and an energy cost function is minimized with respect to the seeds and the lines via simulated annealing. The energy cost function is given by :

$$E = \sum_{i=1}^{N}\sum_{a=1}^{M} V_{ia}M_{ia} + \lambda \sum_{i=1}^{N}\left(\sum_{a=1}^{M} V_{ia} - 1\right)^2 + \frac{2}{\lambda_v}\sum_{a=1}^{M} D_a \tag{2.8}$$

Here, M_{ia} is the distance of hit i to arm a. V_{ia} is defined as the likelihood of hit i belonging to arm a and is given by a partition function :

$$V_{ia} = \frac{e^{-\beta M_{ia}}}{e^{-\beta\lambda} + \sum\limits_{b=1}^{M} e^{-\beta M_{ib}}} \tag{2.9}$$

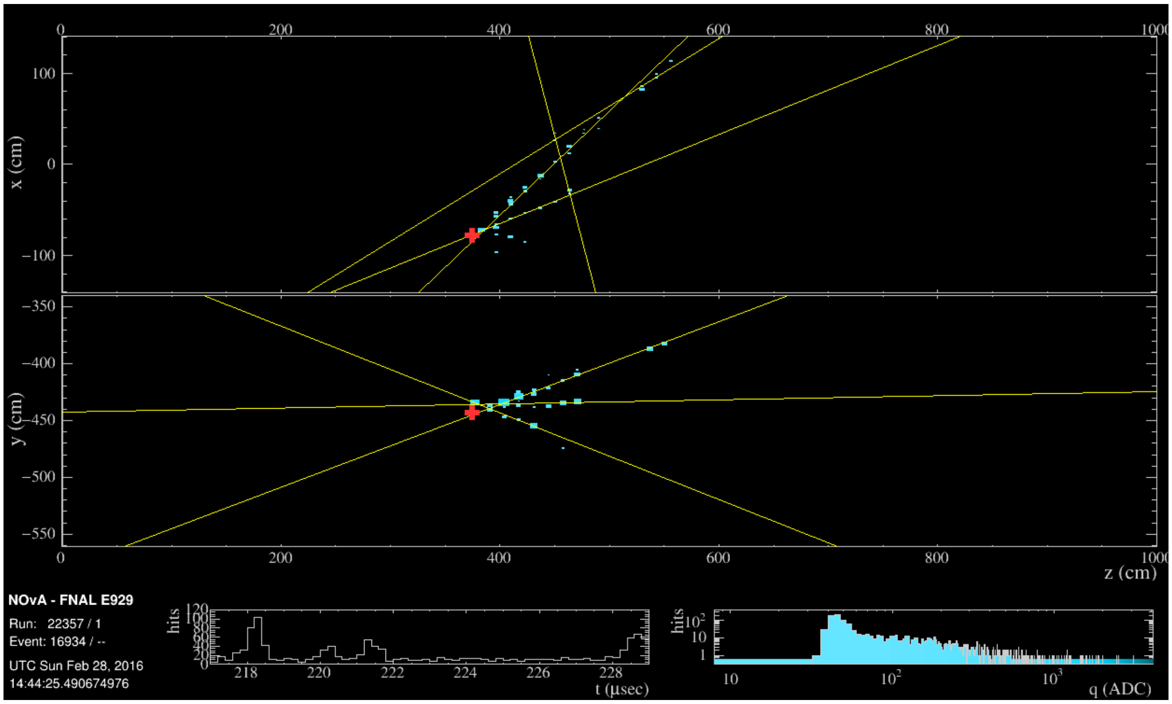


Figure 2.33: Illustration of the Elastic Arms algorithm using Hough lines drawn between pairs of points in the hit cluster

β measures a "temperature" ($\beta = 1/T$), where higher temperature arms have more hits associated with them. The first term therefore minimizes the distance of hits to the arms weighted by the likelihood that the hit belongs to the arm. The 2$^{\text{nd}}$ term in the cost function is a noise penalty term with strength parameter, λ that penalizes hits that aren't associated with any arm. The 3$^{\text{rd}}$ term is a displacement penalty term associated with the parameter λ_v which is tuned to the photon conversion length in the detector. Here D_a refers to the distance of the vertex seed from the first hit in arm, a. As is evident, the term penalizes vertex points and Hough arms that are more than λ_v away from each other. This takes into account the possibility that the first hit in the interaction can be different from the vertex.

As seen in Fig. 2.34, the resulting vertex resolution from the co-ordinates in the two views is found to be $\sim$ 10 cm ($\sim$ 2 cell depths) for ν_eCC and ν_μCC events. For NC events, the resolution is $\sim$ 25 cm ($\sim$ 4 cell depths).

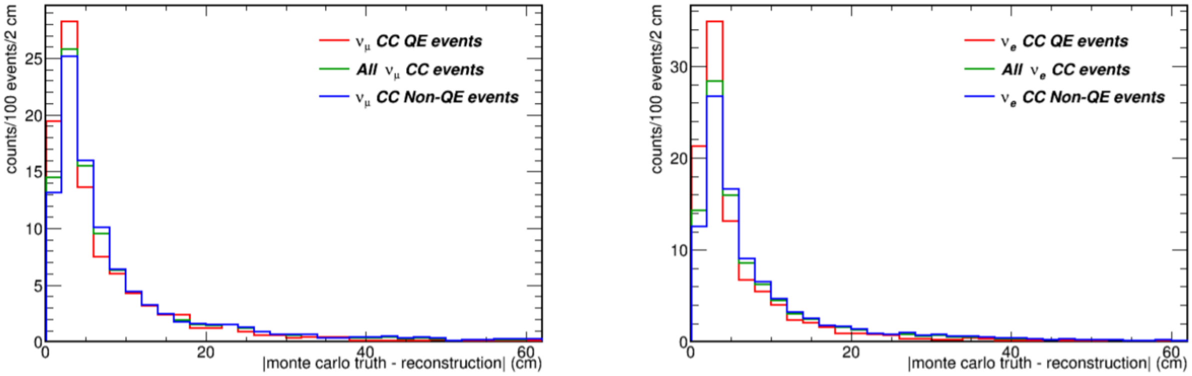


Figure 2.34: The vertex resolution for ν_μ-CC events (left) and ν_e-CC events (right)

2.4.3 Particle Clustering and Tracking

Once the vertex has been identified, multiple clusters associated with each interaction product can be built. This is done by a “Fuzzy K”-means algorithm. The clusters are referred to as “prongs” to distinguish it from “slices” made by the slicing algorithm. The philosophy here is to allow hits to belong to multiple clusters (“fuzzy”) or no cluster (“noise”). Since the number of particles are unknown apriori, the clustering needs to be unsupervised and is done by a k-means based approach [100].

First, the hits in the slice are mapped to an angular space, defined by an angle about the vertex, θ with respect to the z axis. The hit energy distribution in this θ space determine the number of prong centres around which the cluster can be built. This is an iterative process where the highest peak is first chosen, thereby assuming 1 prong centre. A membership score for each hit to the prong centre is then calculated as :

$$U_{ij} = \exp\left(\frac{m\sqrt{c}d_{ij}}{\beta}\right) \tag{2.10}$$

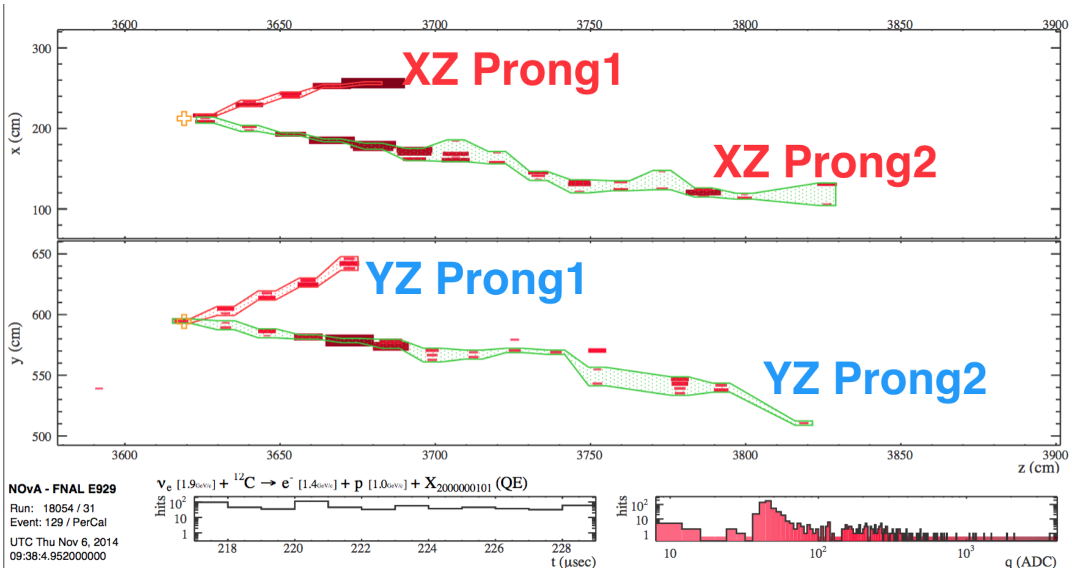


Figure 2.35: A neutrino interaction shown with the reconstructed Fuzzy-K prongs in each view

where d_{ij} is the angular separation of hit j from prong centre i, given by :

$$d_{ij} = \left(\frac{\theta_j - \theta_i}{\sigma_j} \right)^2 \tag{2.11}$$

where σ_j is an angular uncertainty determined by the multiple scattering of muons ($\sim$ 5.4 mrad for a 2 GeV muon), to allow for scattering further away from the vertex, for example with charged pions. The parameters in the membership score are taken to be the allowed threshold membership of hits to prongs (m), the total number of clusters (c) and the allowed width of the clusters (β). The prong centers are updated by the equation :

$$\theta_i' = \theta_i + \frac{\sum_{j=1}^{n} \mathcal{A}_{ij} \left(\theta_j - \theta_i \right)}{\sum_{j=1}^{n} \mathcal{A}_{ij}} \tag{2.12}$$

until the prong centre angles converge with a tolerance of 1×10^{-7} rad. The process is then iterated by generating a new angular distribution with low membership and so on until all hits are clustered or a maximum number of prongs are made. This is done separately for the two views and the prongs are merged by a Kolmogorov-Smirnov test. This process captures $> 90\%$ of the lepton energy in the energy range of interest for NOvA with $> 80\%$ purity, as seen in Fig. 2.36. The performance improves significantly for Quasi-Elastic events, as expected.

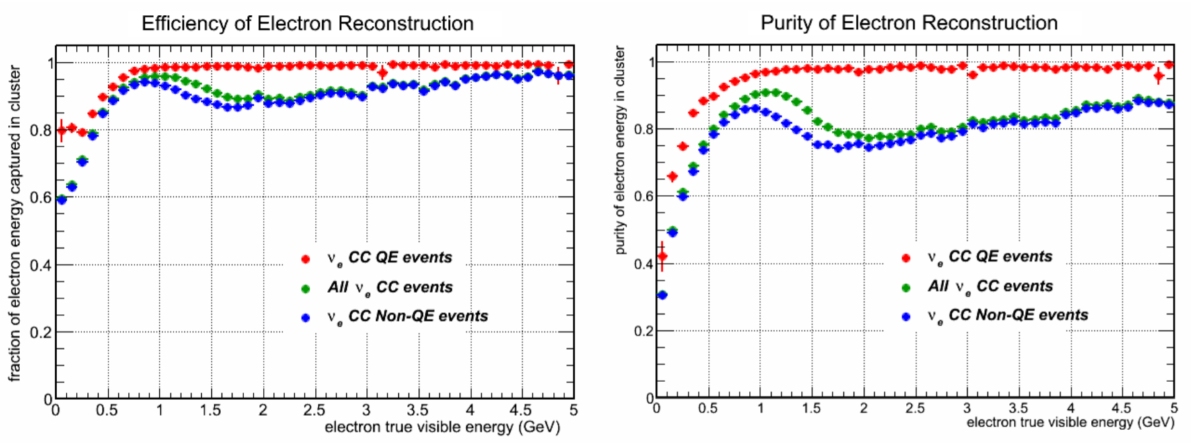


Figure 2.36: Performance of the Fuzzy-K reconstruction on electron shower cascades with the efficiency (left) and purity (right) of the reconstructed shower

This can be applied to muons as well, but since muon tracks are relatively straight and long with deviations from straight tracks given by a multiple Coulomb scattering process (MCS), a traditional Kalman tracking algorithm is used to accurately measure the length of the muon track and its angle. This is important since the muon energy resolution is mainly driven by how well its length can be estimated, rather than the total calorimetric energy deposited. The Kalman tracking algorithm builds tracks by adding nearby hits that satisfy the MCS model within statistical noise and estimating the position and slope for the next hit. These tracks are back-propagated to refine the process iteratively. Finally, tracks made in separate views are merged based on an overlap score between the z-start and z-stop co-ordinates for each track pair. The completeness of the muon track for QE (non-QE) is $\sim 95\%$ ($\sim 92\%$), as seen in Fig. 2.37. The mean difference between the reconstructed and true muon direction is $< 0.1\%$ in cosine units.

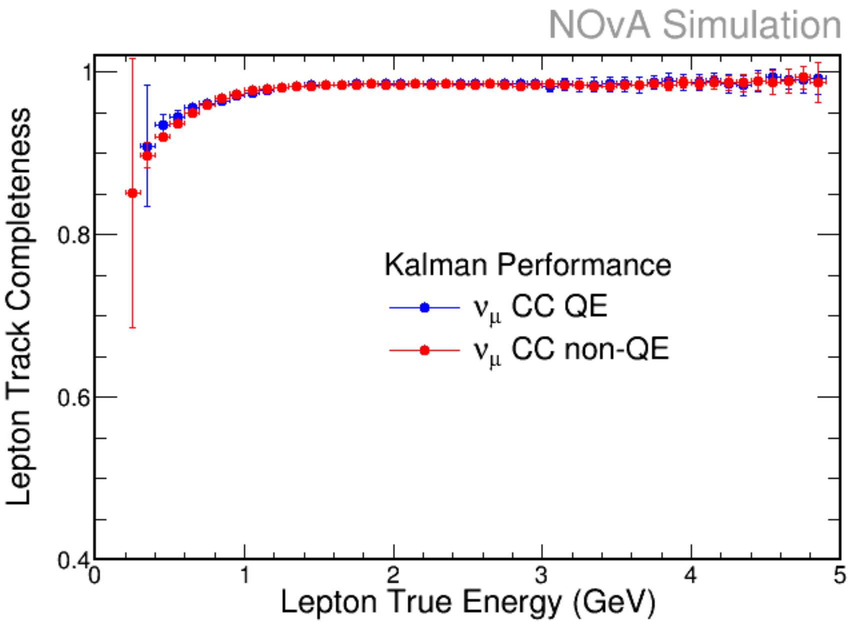


Figure 2.37: Performance of the Kalman filter-based reconstruction on muon tracks. Lepton Track Completeness is defined as the fraction of true muon hits reconstructed in the given kalman track

Chapter 3

The Oscillation Analysis

We can finally start to describe the components of the oscillation analysis. There are many moving parts, so in this chapter we will be concerned mainly with the fundamental elements. These are related to neutrino flavor identification (PID), neutrino energy estimation and the selections developed to maximize signal content in the relevant oscillation channels at the Far Detector. In addition, constraints on the oscillated prediction using Near Detector data will be described.

3.1 Particle Identification

As mentioned before, the two main backgrounds for the oscillation channels are NC interactions and cosmic rays. In both the ν_μ-disappearance and ν_e-appearance channels, these backgrounds have the potential to mimic the signal events. The task of identifying ν_μ-CC events however is significantly easier due to the readily identifiable muon track especially at peak flux energies. As a result, the ν_μ-disappearance channel is largely background free. This channel is agnostic towards neutrinos and anti-neutrinos, since by the $\mathcal{CPT}$-theorem, they should oscillate at the same rate if the initial and final particles before and after oscillation are the same. Therefore, this channel does not try to discriminate between neutrinos and anti-neutrinos and the signal is defined as the total ν_μ-CC and $\bar{\nu}_\mu$-CC that have remained unoscillated over the baseline of the experiment.

For the appearance channel, the oscillation probabilities are different for neutrinos and anti-neutrinos and are related to the measurement of the parameters we are interested in, such as δ_{CP}, the mass hierarchy and the octant of θ_{23}. Therefore, the signals are ν_e-CC ($\bar{\nu}_e$-CC) events for the neutrino (anti-neutrino) mode. One therefore needs to discriminate between neutrinos and anti-neutrinos in this case. These are especially important in the anti-neutrino mode where the contamination of "wrong-sign" neutrinos in the beam is substantial due to the higher interaction cross-section. It turns out however that due to the lack of a magnetised detector, this background is sufficiently hard to reject and techniques developed to separate wrong-sign events based on their kinematic differences do not improve sensitivity to the oscillation parameter measurement significantly. In addition, the cross-section ratio of anti-neutrinos to neutrinos increases as a function of energy. This means that these backgrounds are less of a problem for NOvA at energies of $\sim 2\,\text{GeV}$ than for example, T2K at energies of $\sim 0.9\,\text{GeV}$. As a result, no effort is made to identify them in the particle identification and appropriate systematic uncertainties are taken in the simulation.

3.1.1 CVN

NOvA uses a common multi-label PID to classify the 4 types of interactions, ν_e-CC, ν_μ-CC, NC and Cosmics. Within particle physics, NOvA has pioneered a deep learning approach to perform this task. This has been aided by recent advances in computational power, making these deep learning techniques more viable. In particular, NOvA utilizes a convolutional neural network (CNN) architecture to perform deep inferences about the type of interaction from its topology [102][103].

CNNs are fairly estabilished techniques for 2D image classification purposes and have found use in a variety of different contexts. For our case, it is important to realize that such techniques are readily applicable due to the fine segmentation and pixelated nature of our detectors. Unlike traditional machine learning techniques, CNNs are able to extract the most important features that aid in classification by themselves, instead of relying on features developed by hand.

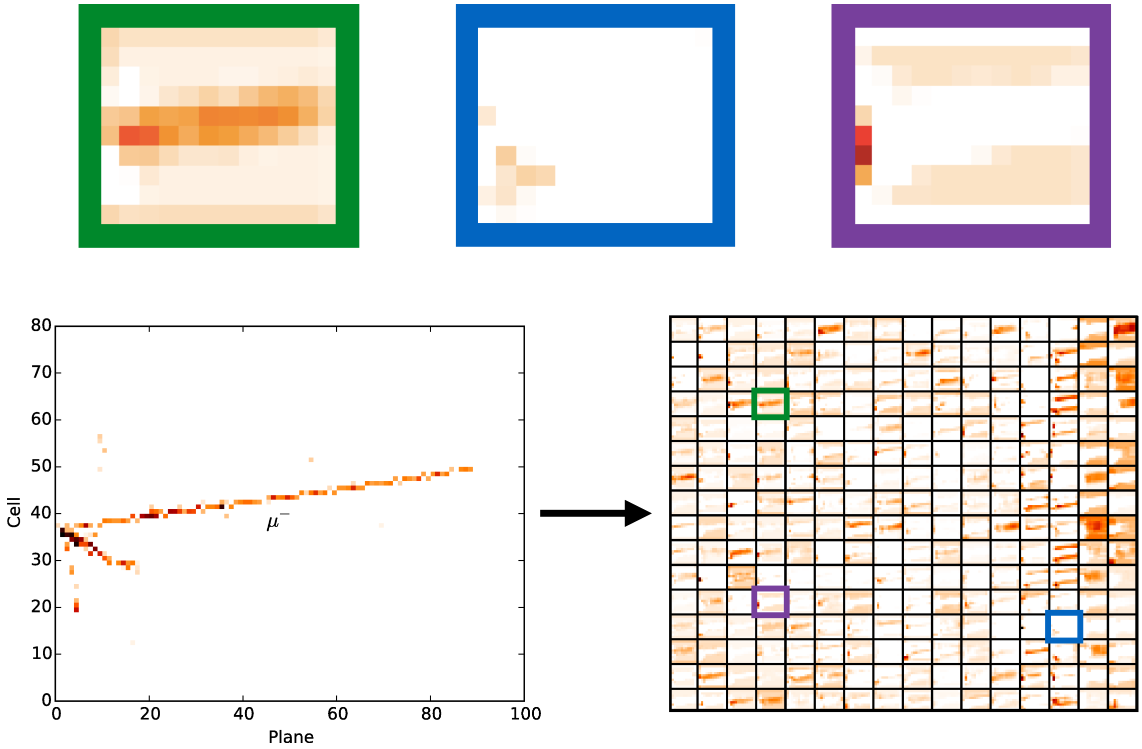


Figure 3.1: Illustration of the convolutional neural network acting on a ν_μ-CC event. The features highlighted in the feature map correspond to various topological features of the event including on the track and hadronic activity.

The fundamental component of a CNN is the convolutional layer that takes an input 2D image, represented by a $m \times n$ matrix of pixel values ("pixel-map") and performs convolutions on smaller pixel blocks (in our case, this will be 5×5) in that image. These convolutions are performed using a kernel matrix in a sliding window fashion. The kernel is of the size of the smaller pixel block and aggregates information within a group of pixels through a set of weights. The output of these convolutions can thus be interpreted as abstract 2D matrices with each element representing various local features within a certain region of the original image, as illustrated in Fig. 3.1. Within image processing, this can be used to extract object edges, sharpen images and many other things. Once this abstract feature-map has been suitably transformed into even lower dimensionality, often to reduce computational complexity, they are then used as inputs to a traditional fully-connected neural network architecture. The elements of the feature map are thus linearly transformed using a set of weights for each neuron in the first layer and forwarded to the next layer and so on until it is identified with its classification target, which in our case is the type of interaction. The identification step is usually preceded by a non-linear activation function (for eg. a sigmoid) that pins the output of the neural network to a specific range of values for eg. $(0, 1)$ for classification. This can then be identified as a probability measure of classifying the input data to a given target label.

During the training process, a cost-function is built for mislabelling any event in the training dataset. This cost-function is then minimized with respect to all the weights used throughout this architecture. Viewed from far above, this is a convex non-linear transformation of the training data into a high dimensional space where the cost-function is minimized to set the boundary between different clusters representing different target labels. But in contrast to traditional machine learning, this space is more abstract (represented by outputs of the convolutions) and is of much larger dimensionality, which results in very accurate classifications. To make sure there is no overtraining happening, this process is of course tested on an entirely separate dataset.

In practice, modern CNN architectures are large and complicated beasts that involve multiple layers of convolutions and pooling (aggregating high dimensional feature maps into smaller dimensional feature maps), as shown in Fig. 3.2. Many different architectures have found reliable use for NOvA. More recently, a modified MobileNet architecture has been used as it is considered one of the more light-weight ones ($\sim 3\times$ inference speed compared to previous NOvA architectures), with an ability to perform fast computations without losing accuracy. In general, the use of CNNs at NOvA have resulted in significant gain in classification accuracies ($\sim 30\%$ improvement in equivalent exposure) compared to more traditional machine learning techniques, including artificial neural networks (ANNs).

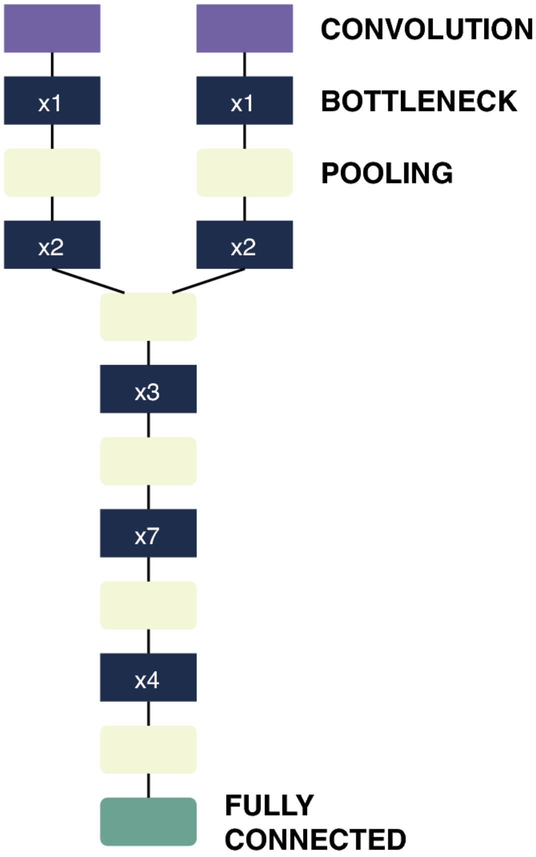
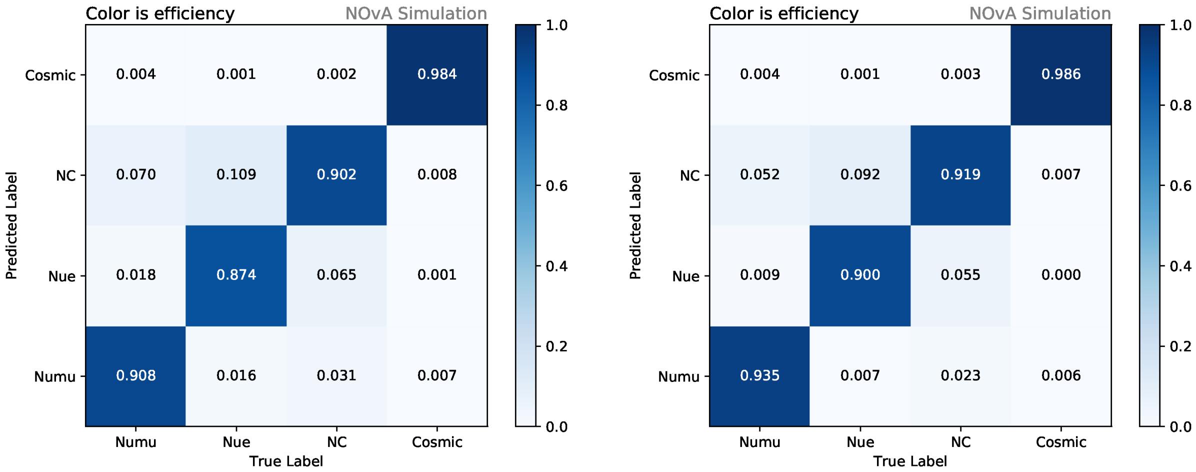


Figure 3.2: The modified MobileNet architecture used for the NOvA CNN PID

The input data to the NOvA CNN (dubbed the Convolutional Visual Network or CVN), is the pixel-map of the interaction clustered using the slicer and nothing else. The pixels are the calibrated energy deposits in each hit of the interaction, given in GeV and recast into a 8-bit integer (between 0 and 255) for faster computations. Each of the two projections in the top view and the side view are fed separately into the CNN in a Siamese-like architecture that combines features from the two images downstream. The pixel map is of dimensions, 100planes $\times$ 80cells in each projection starting from the first hit in z, which is able to cover all relevant features of the interaction. The muon track usually overshoots the image but the missed portion of the track is not important for identification. The rest of the interactions are covered to a large extent ($> 99\%$). This form of the data has the advantage that the classification does not depend on the details of the reconstruction chain apart from the slicer and is therefore blind to any reconstruction inefficiency that crops up for various events, downstream of the slicer. The fully connected neural network is made up of a single layer of 1024 neurons and are connected to the target layer, made up of 4 nodes for each type of classification label, ν_e-CC, ν_μ-CC, NC and Cosmics. The activation function is a simple softmax function that normalizes the scores between (0, 1) [104][105].

The training dataset is made up of an equal mixture of the three types of neutrino interactions ($\sim$ 2.5million events). Cosmics are down-sampled to 10% of the dataset so as to let the network concentrate more on the neutrino interactions. No significant selections are made on the input training dataset to keep the inference general. A simple cosmic veto that removes obvious cosmics is applied. The veto removes events that touch opposite walls of the Far Detector as well as events that have a measured transverse momentum fraction (pT/p) less than 0.95. pT/p measures how aligned the event is with the beam direction and since cosmics are mostly vertical, a significant number of cosmics can be rejected this way and the classification is done for tougher backgrounds. The pixel maps are themselves uniformly scaled by a random number distributed as $\mathcal{N}(1, 0.1)$ which makes the training more robust to energy scale uncertainties in the calibration. The training is performed using Keras on a Tensorflow backend and is done separately for neutrino and anti-neutrino modes. This is to ensure that specific variations in topology across the beam modes are taken into account, which results in an overall $\sim$ 10% gain in accuracy compared to the case when the FHC network is directly used on RHC events.

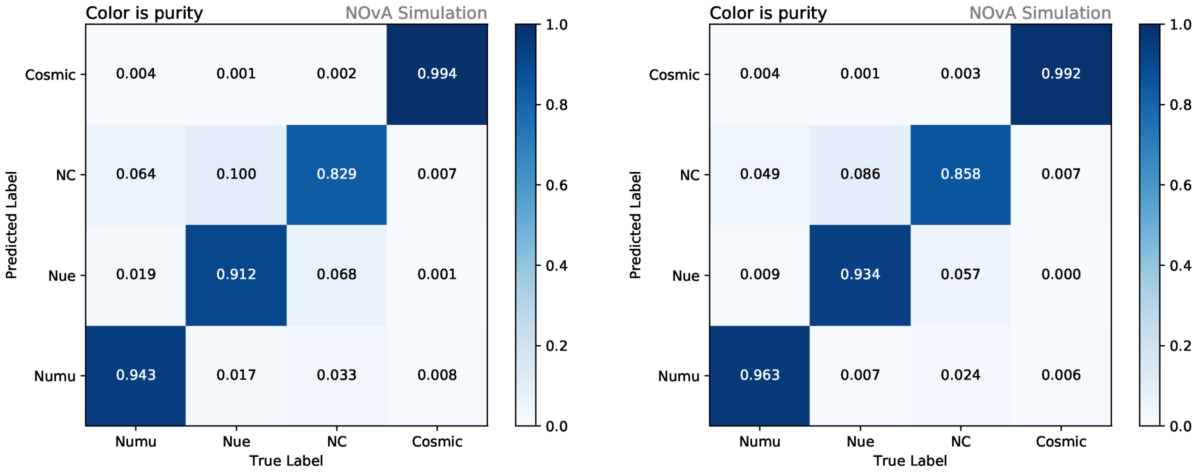
The results of the training can be described by classification matrices in terms of the efficiency of getting the right answer or the purity, which is the fraction of predicted labels that are right in truth. These are shown in Figs. 3.3 and 3.4. In general, both FHC and RHC networks show ∼ 90% efficiency for the relevant signal events, with most of the misclassifications related to NC.



Predicted Label	Numu	Nue	NC	Cosmic
Cosmic	0.004	0.001	0.002	0.984
NC	0.070	0.109	0.902	0.008
Nue	0.018	0.874	0.065	0.001
Numu	0.908	0.016	0.031	0.007

Predicted Label	Numu	Nue	NC	Cosmic
Cosmic	0.004	0.001	0.003	0.986
NC	0.052	0.092	0.919	0.007
Nue	0.009	0.900	0.055	0.000
Numu	0.935	0.007	0.023	0.006

Figure 3.3: The performance of the CVN PID given by the efficiency of the network to identify the true label for the neutrino mode (left) and the anti-neutrino mode (right)



Predicted Label	Numu	Nue	NC	Cosmic
Cosmic	0.004	0.001	0.002	0.994
NC	0.064	0.100	0.829	0.007
Nue	0.019	0.912	0.068	0.001
Numu	0.943	0.017	0.033	0.008

Predicted Label	Numu	Nue	NC	Cosmic
Cosmic	0.004	0.001	0.003	0.992
NC	0.049	0.086	0.858	0.007
Nue	0.009	0.934	0.057	0.000
Numu	0.963	0.007	0.024	0.006

Figure 3.4: The performance of the CVN PID given by the fraction of the predicted labels from the network that are the true label for the neutrino mode (left) and the anti-neutrino mode (right)

3.2 Energy Estimation

Neutrino energy estimation is fundamental to the sensitivity of the oscillation analysis as the oscillation probabilities are a function of neutrino energy. In the disappearance channel, this is important especially in the so-called "dip-region" (between $1 - 2\,\mathrm{GeV}$ for NOvA) where the shape and extent of the dip directly corresponds to the wavelength and amplitude of the oscillation, i.e Δm^2_{32} and $\sin^2 2\theta_{23}$ respectively. For the appearance channel, the neutrino energy provides a small amount of discrimination power with the irreducible beam-ν_e contamination as well as an improvement in sensitivity for the oscillation parameters. We will therefore look at the techniques used by NOvA to estimate neutrino energies for ν_μ-CC and ν_e-CC events separately as well as explore possible improvements.

3.2.1 ν_μ-CC Events

The energy estimator for ν_μ-CC events is built by summing contributions from the leptonic and hadronic portions of the event [106].

$$E_\nu = E_\mu + E_{had} \tag{3.1}$$

As discussed before, E_μ can be estimated accurately using the length of the muon track as there is a significant correlation between these observables due to the Bethe-Bloch relation. E_{had} is then estimated calorimetrically and includes both off-track and on-track components. True ν_μ-CC events are selected that pass a preselection requiring them to be well-contained and having a reconstructed kalman track. The kalman track describing the muon is selected by using a special Boosted Decision Tree (BDT) trained on the following characteristics [107].

- log-likelihood that the dE/dx profile follows a MIP-like distribution

- scattering log-likelihood for deviations from straight lines
- length of the track
- estimated hadronic fraction on track which is smaller for muons away from the vertex.

The muon energy is then estimated by a piecewise-linear spline fit to the true muon energy using the reconstructed track length. In the Near Detector, the fit is done separately for portions inside the active region and the muon catcher and summed. The fit is done separately in three regions of track length to take into account variations due to multiple Coulomb scattering behaviour, as shown in Fig. 3.5. The energy resolution for the muon comes out to $\sim 2.5\%$

The hadronic energy is estimated calorimetrically by fitting to the true hadronic energy (true neutrino energy - true muon energy) using an estimate of the visible hadronic energy in the interaction. The visible hadronic energy is estimated by summing up off-track energies and the on-track contamination. A piecewise-linear spline fit is then performed in three regions of visible energy so as to take into account different regions in interaction phase-space, as shown in Fig. 3.6. The energy resolution for the hadronic component comes out to $\sim 25\%$, since this is a much more diffused dataset.

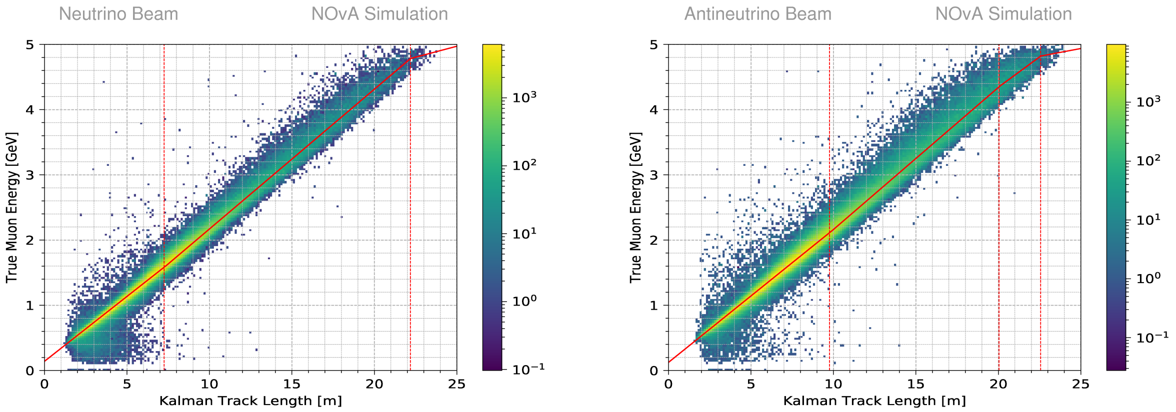


Figure 3.5: The fit to the track length of the identified muon for the muon energies. FHC is given on the left, while RHC on the right

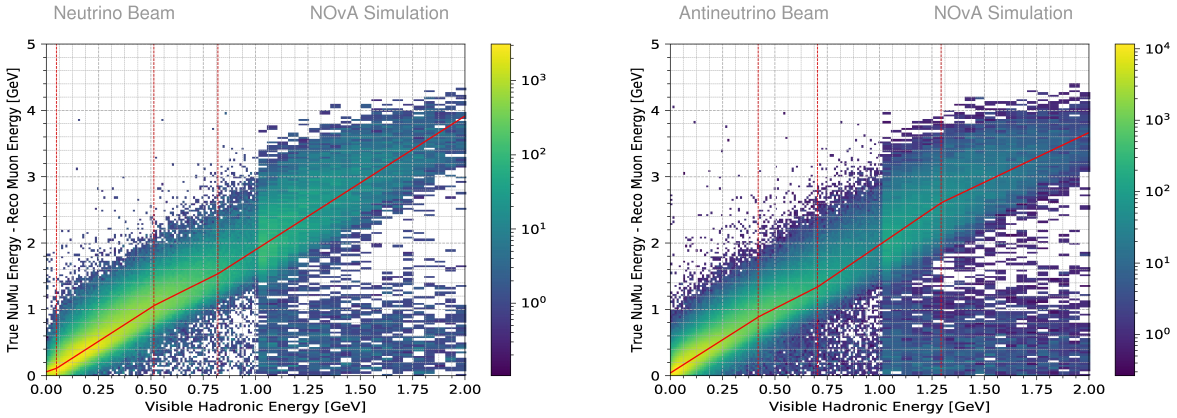


Figure 3.6: The spline fit to the visible hadronic energy of the event for the hadronic energy estimate in 3 different regions. FHC is given on the left, while RHC on the right

The fits are performed separately for ND and FD as well as FHC and RHC. The overall neutrino energy resolution is $\sim 6\%$ for both FHC and RHC, with a slight improvement for RHC events since they are less inelastic. As can be seen in Fig. 3.7, the resolution has a significant dependence on the type of interaction, with QE events exhibiting better estimator performance than DIS events, due to the fact that the muon carries more of the neutrino energy. The estimator performance is also measured by the bias in the resolution peak (deviation from 0) and the estimator is fairly flat across neutrino energies of $(0.5, 4.5)$ GeV and especially so in the dip-region. This is important to ensure data events don't move out of the dip region as a consequence of the estimator which can significantly bias the oscillation parameter measurement.

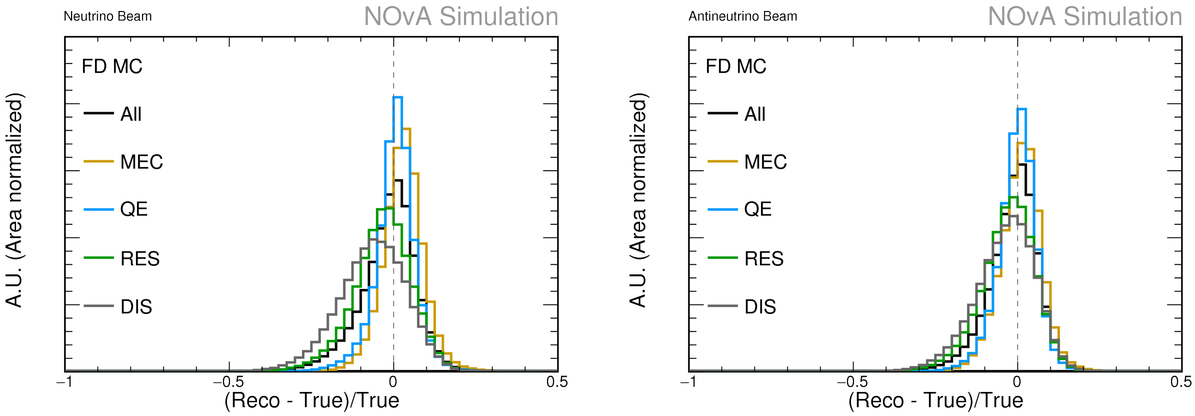


Figure 3.7: ν_μ-CC energy resolutions for various interaction types

3.2.2 ν_e-CC Events

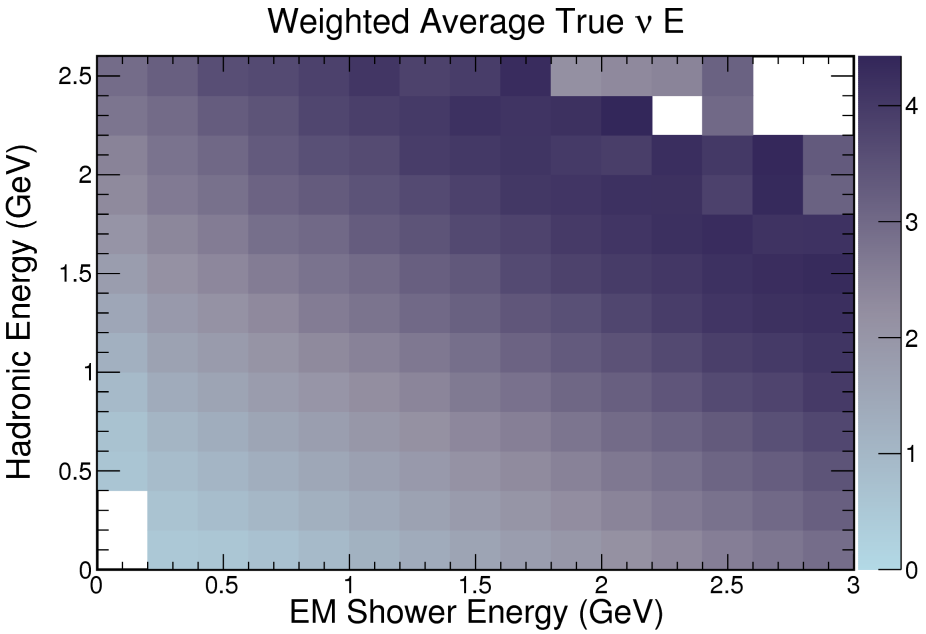


Figure 3.8: The true ν_e-CC energy as a function of the individual EM and hadronic components weighted by event rates

The ν_e-CC energy estimation is significantly more difficult due to the nature of the electron shower cascade process and the overlap between the leptonic and hadronic components. Given that a significant fraction of the true energy of the particle isn't seen in the detector (due to dead materials), a simple calorimetric approach by summing up all reconstructed hit energies and rescaling them to match the true energy will only work well for events at the flux peak where most of the events exist. The resulting purely calorimetric estimator will then exhibit a significant bias for events in the flux tails. As a result, the estimator tries to exploit differences between the electromagnetic and hadronic components to improve performance, where the electromagnetic component includes contributions from the electron shower as well as γ cascades coming from π^0 decays. This is motivated by the fact that these components have a different detector response, i.e roughly 26% more energy is deposited on average for a given true energy if the source is electromagnetic compared to hadronic. This comes from the fact that the hadronic component includes neutrons, quasi-MIP-like charged pions and so on. The differences in detector response are taken into account by estimating the electromagnetic and hadronic contributions separately and then performing a fit to the true neutrino energy using a quadratic functional form given as :

$$E_\nu = \alpha(p_0 E_{EM} + p_1 E_{had} + p_2 E_{EM}^2 + p_3 E_{had}^2) \tag{3.2}$$

Fig. 3.8 shows the relationship between the hadronic and the electromagnetic energies. The quadratic form is motivated by the shape of the dependence between the EM energy fraction and the average detector response, measured by the fraction of true visible neutrino energy that is deposited in the detector. The linear coefficients, p_0 and p_1 take into account the different detector response between the respective components [108].

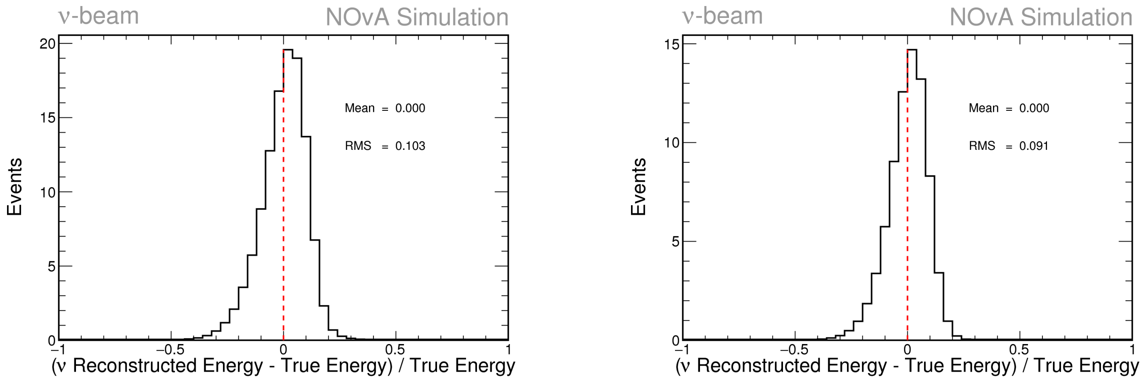


Figure 3.9: The ν_e-CC energy resolutions for FHC (left) and RHC (right)

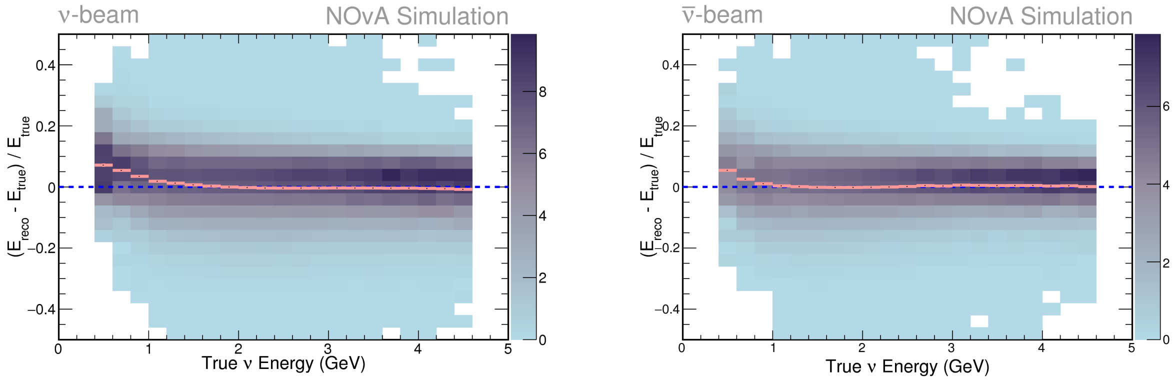


Figure 3.10: ν_e-CC energy resolutions as a function of True ν_e-CC energy for FHC (left) and RHC (right). The colored curves represent the profile of the respective distributions

The electromagnetic and hadronic components are identified using a CNN-based particle ID [109] that is trained on individual reconstructed FuzzyK prong images with an additional image input of the entire neutrino interaction slice to provide it information about the context in which that prong is produced. The particle ID scores are summed up for electrons and photons to generate an EM score. Prongs are tagged as EM if their EM score is > 0.5 and the rest of the hits are included in the hadronic component. The fit to the true neutrino energy gives an overall energy resolution of $\sim 10\%$ ($\sim 9\%$) for FHC (RHC), as shown in Fig. 3.9. Importantly, the quadratic functional form allows the estimator bias to be flat between $(1, 5)$ GeV, which was the original motivation to extend the purely calorimetric estimator, as shown in Fig. 3.10.

3.2.3 An Alternate ν_e-CC Energy Estimator

A big motivation for improving upon the current estimator design is to handle different interaction topologies equally well. We've seen how the current ν_e-CC estimator builds upon the purely calorimetric approach which is good only for the flux peak so as to ensure a flat bias across all neutrino energies. Similarly for ν_μ-CC events, the spline fits are done in separate regions of muon track length or visible hadronic energy to keep the bias flat especially in the dip-region. However, these approaches are not able to handle the additional complication that arises from having different interaction types included in a sample of given neutrino energy.

It turns out that for the ν_e-CC estimator, the energy for QE events is over-estimated while for DIS events, it is underestimated. This is understandable since the estimator is tuned to be correct for the entire sample on average and high-Q^2 events have a large missing energy component, resulting in an underestimation and consequently an overestimation for low-Q^2 events. The estimator does not have enough degrees of freedom to handle all the complexities in the interaction phase-space.

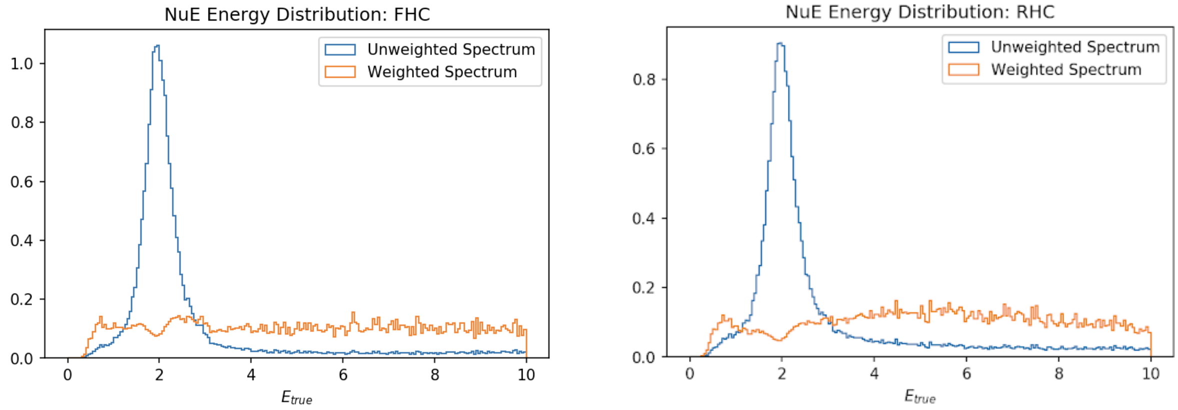
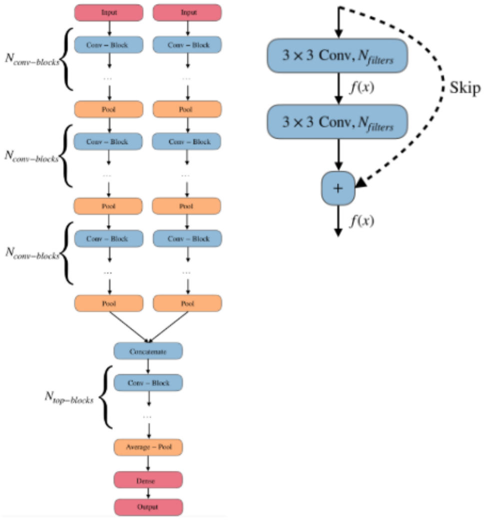


Figure 3.11: The ν_e-CC true energy distributions flattened for input into the training for FHC (left) and RHC (right)

To take this into account, we build upon the deep-learning approach for the PID (CVN) by utilizing a CNN to handle the regression task of estimating the neutrino energy as well [110]. We've seen how CNN architectures have the capability of performing deep inference directly from image representations of the interaction and this provides us with the degrees of freedom required to handle the different types separately. We keep a similar philosophy as the CNN PID by utilizing a Siamese-style architecture trained on true ν_e-CC events at the FD using the two projections. The training sample is made up of events passing a cosmic veto and are required to be contained. The pixel maps are the same size as the CNN PID ($100\text{planes} \times 80\text{cells}$ in each projection). To ensure the training sample is not biased towards the flux peak, we reweight the true energy spectrum in the training so as to ensure all events with energies between $0.5 - 10\,\text{GeV}$ get the same importance, as shown in Fig. 3.11. The training is carried out separately for FHC and RHC simulated events [111].

The same architecture was used for the FHC and RHC trainings, illustrated in Fig. 3.12. A ResNet architecture was used with a Siamese network structure. The Siamese network structure consists of two identical sub-networks whose outputs are merged to produce the final output. Each sub-network processes an image from one view. Weights are not shared between the sub-networks to provide independent information aggregation in each view. The sub-networks are constructed from convolutional layers and pooling layers.



Architecture:
Pool: Average
Activation f(x): ReLU
Kernel Initialization: He Normal
N conv blocks between pooling: 1
N top blocks: 0
N filters: 64
Dense: 512

Training:
Optimizer: Adam
Learning Rate: 1e-3
Reductions: multiply learning rate by 0.1
if no improvement >0.4 for 4 epochs

Figure 3.12: The architecture of the regression CNN for the ν_e-CC energies

As described in previous sections, convolutional layers apply a weight matrix in a sliding window fashion to the input image. This allows computing the same feature at different locations in the input image producing an output referred to as a feature map. Using multiple weight matrices allows learning a variety of features from the input images. Stacking these convolutional layers makes it possible to learn higher level features with each additional layer. Pooling layers take a feature map and reduce its dimensionality. This is done by tiling the feature map and reducing each tile to its maximum or average. Pooling layers are normally used between convolution layers.

In our case, a single convolution block consisting of two 3×3 convolution filters was used before each pooling layer. A total of 64 filters were used, as is typical of a ResNet architecture, which has been shown to be highly performant. A ReLU activation function was used across the filters. These are all summarized below.

Neural networks for supervised learning are trained by defining a differentiable loss function L between neural network outputs $f_{\mathbf{W}}(\mathbf{x}_i)$ and target values y_i. Here, $\mathbf{W}$ represents the weights of the neural network and $\mathbf{x}_i$ is the neural network input. The loss function represents the metric by which the neural network accuracy is assessed. During training the neural network weights $\mathbf{W}$ are iteratively updated to minimize L using its gradient and a step size α.

Typically these updates are computed over mini-batches of size n which make up a partition of the total training dataset. Iteration over all mini-batches constitutes one epoch. Training parameters such as the step size α and the batch size n are referred to as hyperparameters which must be selected before training and possibly tuned using the loss function on the validation dataset. We utilize the hyperparameter optimization software SHERPA which implements a number of hyperparameter optimization strategies and visualization tools. By automating the task, this software significantly speeds up the computationally expensive task of finding optimal hyperparameters.

Unlike classification problems such as image recognition and particle identification, the target value y_i for the output (energy) of our regression neural network is a continuous variable varying over events. This requires the definition of an appropriate loss function for the task of the regression neural network.

The goal is to minimize the standard deviation of a Gaussian fit to the peak of the histogram given by the energy resolution $\frac{E_{reco}-E_{true}}{E_{true}}$ on the test set. While this quantity cannot be directly optimized the absolute scaled error loss provides an appropriate surrogate for the task. The training loss function is then given by:

$$L(\mathbf{W}, \{\mathbf{x}_i, y_i\}_{i=1}^n) = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{f_{\mathbf{W}}(\mathbf{x}_i) - y_i}{y_i}\right| \tag{3.3}$$

Traditional loss functions for regression problems are the mean squared error $\frac{1}{n}\sum_{i=1}^{n}(f_{\mathbf{W}}(\mathbf{x}_i) - y_i)^2$ or the mean absolute error $\frac{1}{n}\sum_{i=1}^{n}|f_{\mathbf{W}}(\mathbf{x}_i) - y_i|$. The former is often used due to its relationship to the log-likelihood when the data distribution is assumed to be Normal, its

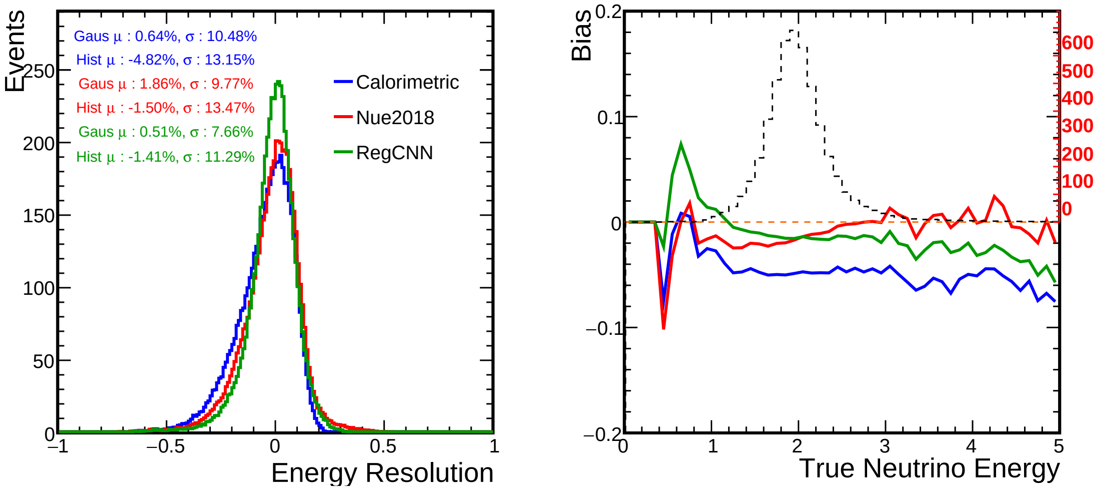


Figure 3.13: Comparison of the performance of the regression CNN based energy estimator for FHC vs traditional approaches both for the overall resolution (left) and profiled resolution as a function of true neutrino energy (right)

strict convexity, the fact it can be decomposed into variance and bias, and many other desirable properties. In our case, however, the mean squared error is suboptimal, because its derivative with respect to $\mathbf{W}$ is $\frac{1}{n}\sum_{i=1}^{n} 2(f_{\mathbf{W}}(\mathbf{x}_i) - y_i)\frac{\delta f_{\mathbf{W}}(\mathbf{x}_i)}{\delta \mathbf{W}}$. This increases proportionally with the distance of the predicted value from the truth. In other words, outliers will have increased impacts during gradient descent. In the training for neutrino energy, the events with large invisible energy due to dead material and hadronic interactions shouldn't have much larger impacts than those whose visible energy is close to the true energy, so we choose the absolute error instead of the squared error in the loss function.

A collection of the results are described in Figs. 3.13 - 3.18. The FD selections include standard preselections and containment cuts w/o any energy cuts. Simple oscillation weights are also applied using an approximate appearance probability formula for $\delta_{CP} = 0$ and utilizing the atmospheric and solar measurements ("dumb" oscillations). Comparisons are made to the current estimator (Nue2018) as well as the purely calorimetric approach, which is a raw sum of the calorimetric energies scaled by a constant factor of 1.78 (as a dead material correction).

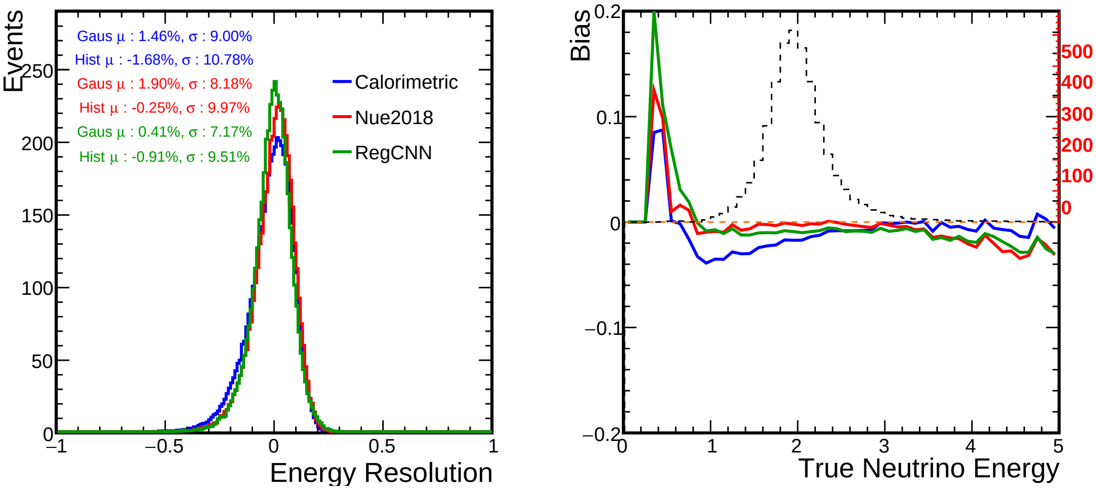


Figure 3.14: Comparison of the performance of the regression CNN based energy estimator for RHC vs traditional approaches both for the overall resolution (left) and profiled resolution as a function of true neutrino energy (right)

RMS of the resolutions both for a gaussian fit to the peak of the resolution distribution as well as raw RMS of the histograms are provided. The overall conclusions are that the resolutions are significantly better compared to the traditional approach with a better bias at higher inelasticities and for different interaction modes.

In summary, the performance of the neutrino energy estimator is better in almost every metric looked at so far. The FHC neutrino energy resolution improves from 9.8% to 7.7% while the RHC resolution improves from 8.2% to 7.2%. The overall bias vs true energy is equivalent to the previous estimator which shows that the reweighting does its job. In fact, the bias is very similar across all interaction modes and is flatter vs true inelasticity as well. This is in contrast to the old estimator, where it overcorrects for QE and MEC events while undercorrecting for RES and DIS events.

All of these conclusions support the idea that not only we can obtain significantly superior performance in terms of energy resolution for ν_e-CCs using this approach, we can reduce the systematic dependence of the interaction model on the energy estimator. To first order this changes the composition of the different interaction types in the simulation and this

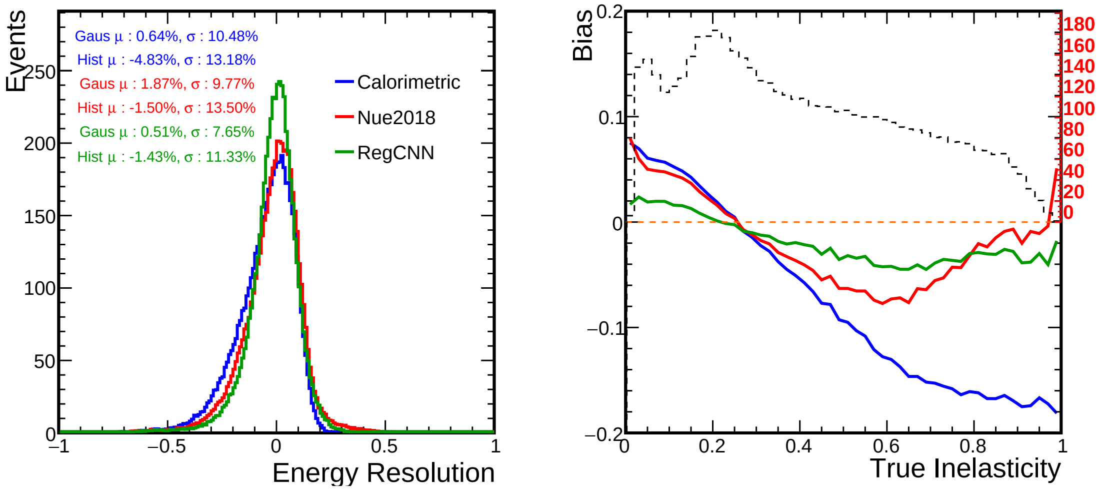


Figure 3.15: Comparison of the performance of the regression CNN based energy estimator for FHC vs traditional approaches both for the overall resolution (left) and profiled resolution as a function of true inelasticity of the event (right)

approach ensures that there are few knock-on effects on the estimator itself.

One other point is with regards to the response of the estimator w.r.t the absolute energy scale. One way to gauge this is to shift the energy deposits of each cell in the pixel map by ±5%. This doesn't take into account all the downstream changes in the reconstruction that could take place, for example subtle threshold effects but is a useful exercise nonetheless. The results from the pixel map shifts are shown below in Fig. 3.19.

This indicates that the response of the estimator is essentially linearly related to the sum of the energies in each cell in the pixel map and is therefore much more dependent on the underlying calorimetry rather than the topological features in the interaction. One way to square this with the observation of similar performance across different topologies is to think of the CNN as parameterizing different topologies separately but in a linear fashion.

To ensure that this is a workable estimator in the oscillation analysis, future studies will validate this approach on more complicated systematics related to particle transport within the detector and the dependence of the estimator on other details such as secondary interactions

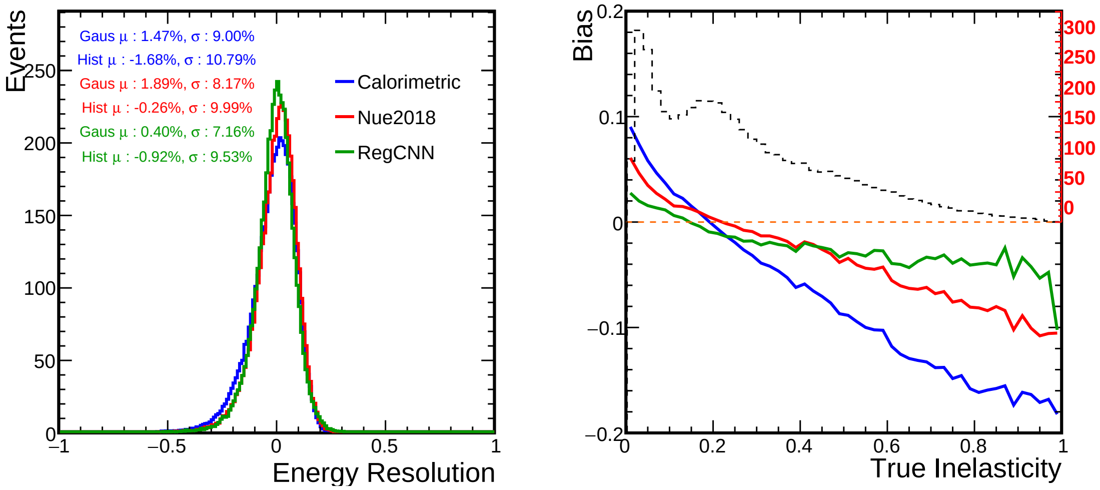


Figure 3.16: Comparison of the performance of the regression CNN based energy estimator for RHC vs traditional approaches both for the overall resolution (left) and profiled resolution as a function of true neutrino energy (right)

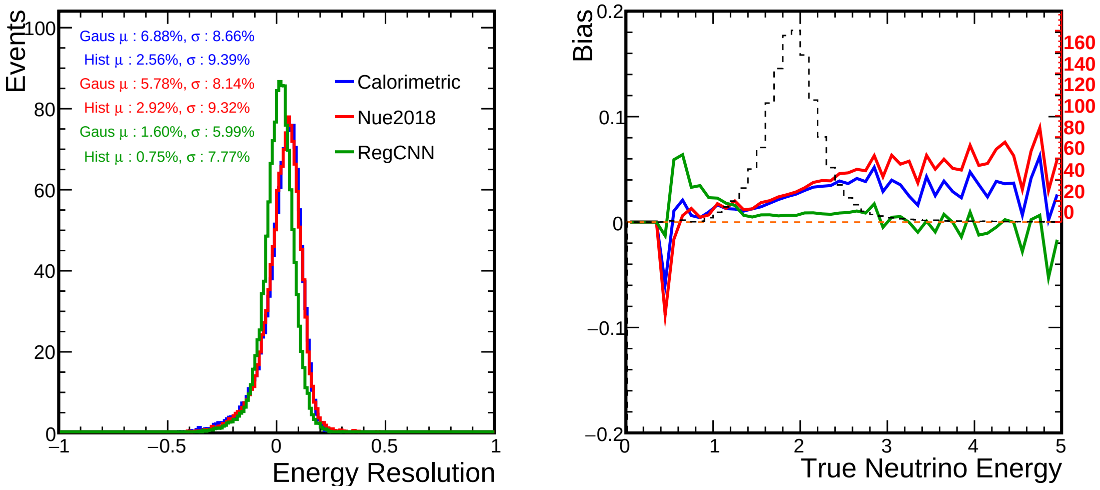


Figure 3.17: Comparison for the performance of the regression CNN based energy estimator for FHC vs traditional approaches for ν_e-CC quasi-elastic events

of the hadronic shower and so on.

For the ν_μ-CC estimator, a LSTM-based approach with a sequence of reconstructed inputs rather than a pixel map is being studied. The pixel-map approach is complicated by the fact that one needs the muon track to be contained within the pixel map for energy estimation and this blows up the computational requirements due to the size of the pixel map needed.

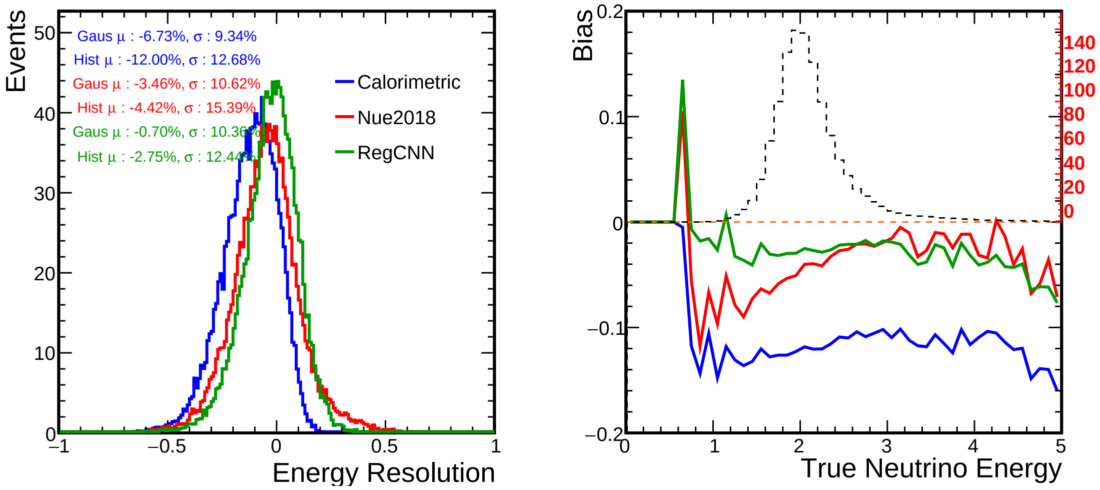


Figure 3.18: Comparison for the performance of the regression CNN based energy estimator for FHC vs traditional approaches for ν_e-CC DIS events

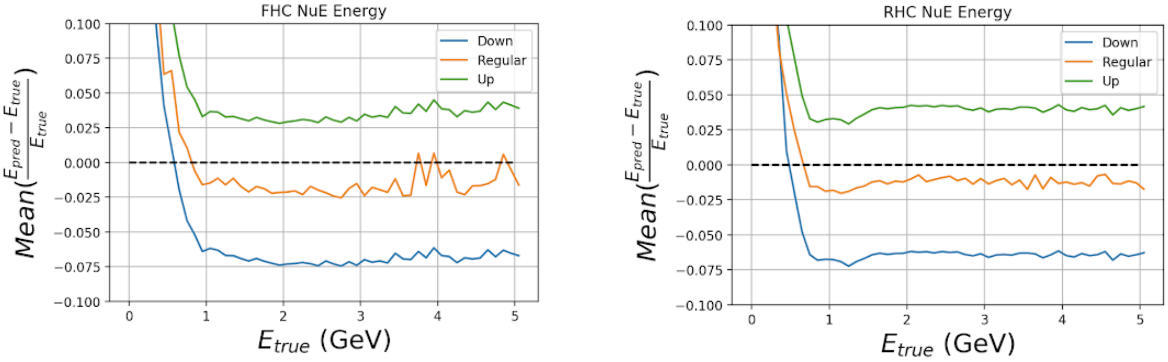


Figure 3.19: Bias sensitivity of the regression CNN based energy estimator when the pixel inputs are uniformly scaled by ±5% for FHC (left) and RHC (right)

A new approach involving a similar pixel-map size for ν_μ-CC events but targeted for the true hadronic energy instead (which is well contained within the pixel map) is also being studied and has been shown to be promising [112].

3.3 Far Detector Selection and Binning

With the PID and Energy Estimation in place, we can begin to describe the details of the FD oscillation analysis. We will first describe common selections used for both the oscillation channels and then discuss the individual pieces specific to the disappearance and the appearance channel.

The selection flow is designed using a consistent philosophy for both the channels. A large part of the design involves cosmic rejection, since we collect around 150 kHz cosmic data at the Far Detector. Cosmic backgrounds are estimated using the minimum bias 10 Hz cosmic trigger, which ensures that the selection tuning is data-driven for this particular background. A wide range of techniques are utilized to pare down the cosmic background to $\mathcal{O}(1)$ events [113].

Events are required to pass certain spill quality cuts to ensure that the data quality is good. These cuts require acceptable ranges in beam parameters such as the horn current, the beam position on target and the beam width. In addition, the POT per spill is required to be larger than 2×10^{12}. Problematic spills due to DAQ issues arising from the Data Concentrator Modules (DCMs) in sections of the Far Detector are also removed. Atleast 4 contiguous diblocks are required to be active during these spills. This is to remove certain spills during early data-taking periods when the full 14 diblock detector was not active [114].

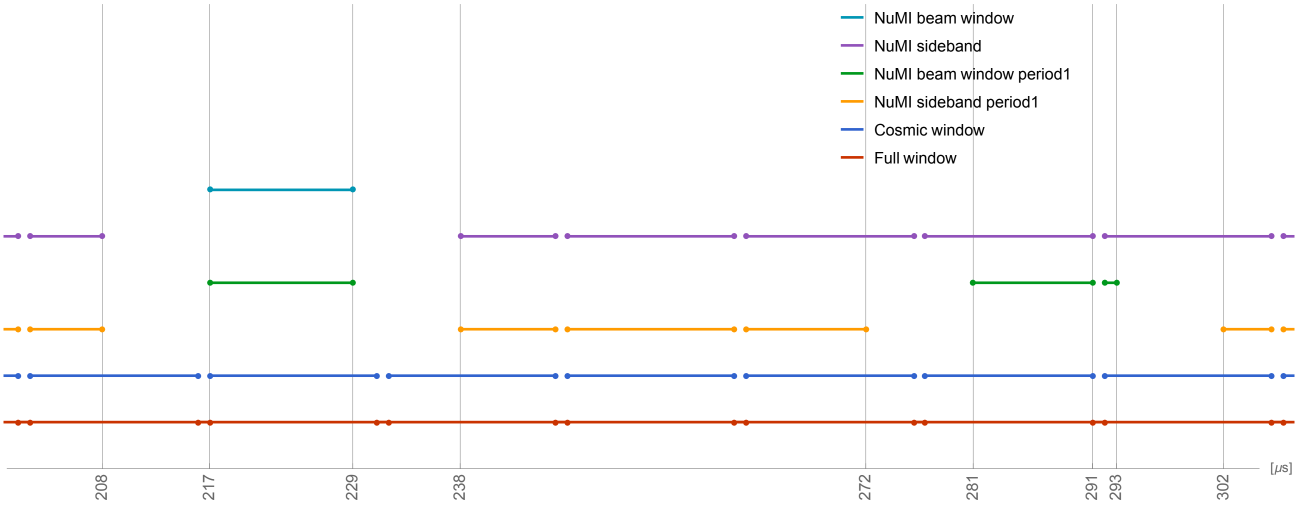


Figure 3.20: The different timing windows of the original 550 μs trigger window used in the analysis

Since the NuMI trigger captures events within a 550 μs window centered around the beam spill ($218 - 228\,\mu$s, with a fraction of early runs having this beam spill shifted by $64\,\mu$s), it becomes prohibitive to reconstruct every single event within this $550\,\mu$s trigger window. Therefore, the trigger window is split into $16\,\mu$s separate windows, each having a $1\,\mu$s overlap, as shown in Fig. 3.20. The splitting is done in such a way so as to ensure the beam spill is fully contained within the smaller time window. Then, a CNN-based cosmic veto algorithm [115] is trained on snapshots of the entire detector during these time windows in order to reject windows having no neutrino-like events. The resulting CNN score is then used to veto reconstruction for these windows. All events within the beam spill are reconstructed. However, the CNN score is still saved for this special window and applied downstream in the analysis files. Selections are tuned using simulated neutrino events within this beam spill along with an estimate of cosmic backgrounds using separate files containing only cosmic events. The cosmic events are derived from the minimum bias 10 Hz cosmic trigger and rescaled to the appropriate livetime of the detector. It is important to note that the reconstruction proceeds for simulated neutrino events within a cosmogenic environment even in the simulation to ensure no biases crop up with respect to real data. These cosmic events in the simulation files also come from the cosmic trigger and are overlaid on top of the simulated neutrino interactions. However, due to computational requirements there is not enough of a cosmic sample overlaid in the sideband of the beam spill window to be able to use them for selection tuning. Therefore, separate cosmic trigger files are used for the selection tuning. In actual FD data spills using the NuMI trigger, this is not so much of a problem and there are enough cosmic events in the beam spill sideband even after veto to provide a data-driven cosmic background prediction within the beam spill after all the tuned selections are applied. This is done by rescaling cosmic events passing all selections in the sideband to the length of the beam spill window.

All this is done to ensure that the analysis design is completely blind to the FD data before everything is finalized. The cosmic veto, selection tuning using separate cosmic trigger events, the cosmic overlay process are designed without looking at any data collected from the FD NuMI trigger. These are then applied appropriately to actual data spills, both as a way to save reconstruction time and estimate the actual cosmic background in the oscillated data within the NuMI beam spill. Data-driven estimates of the beam neutrino background are also developed, but these require utilizing the Near Detector data which will be described later on. The selections themselves are tuned assuming "dumb" oscillations (described in the previous section), which basically use approximate formulae for the oscillation probability and are designed to ensure that the tuning doesn't hinge on a particular choice of oscillation parameters.

3.3.1 The $\nu_\mu \rightarrow \nu_\mu$ Channel

Fig. 3.21 illustrates the selection flow for the disappearance channel. For this channel, a series of basic quality cuts requiring less than 5 GeV reconstructed energies and having atleast 20 hits and a reconstructed kalman track are applied. The CNN-based cosmic veto cuts are also applied.

Events are required to be fully contained in the Far Detector. This gets rid of a large portion of cosmogenic background that travel the full extent of the detector, mainly coming in from the top wall. The important reconstructed variables deciding the containment criteria are :

- Projected number of cells to the detector boundary using the kalman track direction, both forward and backward
- The number of planes to the front and back of the detector from the interaction slice
- The minimum distance of each FuzzyK prong to the end walls of the detector (6 in total)

The containment cuts are tuned using a loose preselection involving the BDT identifying

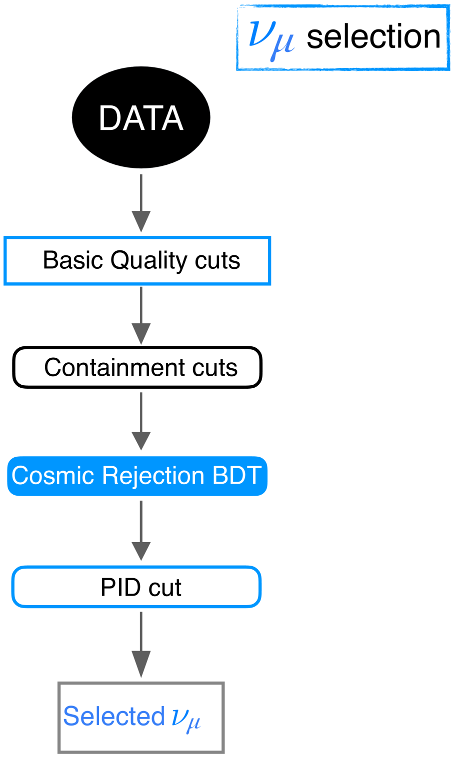


Figure 3.21: Schematic of the cutflow used for the selection in the disappearance channel

the kalman track (ReMID > 0.75) and the cosine of the track angle (> 0.5). Different metrics were used for tuning and were assessed using the resulting signal/background ratio. In addition the tuned cuts were seen to be similar for both FHC and RHC, so a common set of cuts are applied to both beam modes. The metrics used were the fraction of escaping energy in the detector and the energy resolution of the estimator. These were found to be similar in effect for the signal/background ratio. The idea is that events closer to the edge of the detector have a higher fraction of escaping energy and consequently a lower energy resolution. The tuning was then done to find points in the detector where the spectrum of the metric was flat and unchanged across the fiducial volume. Since a lot of cosmic background comes from the top wall, the tuning also took this feature into account by tightening the minimum distance required of the prongs to the top wall. As a result the fiducial volume considered in the analysis involves regions 60 cm away from the top wall and ~ 15 cm away from the other walls (The actual cuts are 18 cm from the front and back and 12 cm from the rest of the walls). In addition, atleast 1 plane is required behind and ahead of the interaction slice and atleast 6 cells are required projected forwards and backwards from the kalman track direction to any detector boundary. The quality and containment criteria serve to reduce $\mathcal{O}(10^6)$ cosmic background events to $\mathcal{O}(10^3)$ events. The signal efficiency at this stage is $\sim 40\%(\sim 45\%)$ for FHC (RHC).

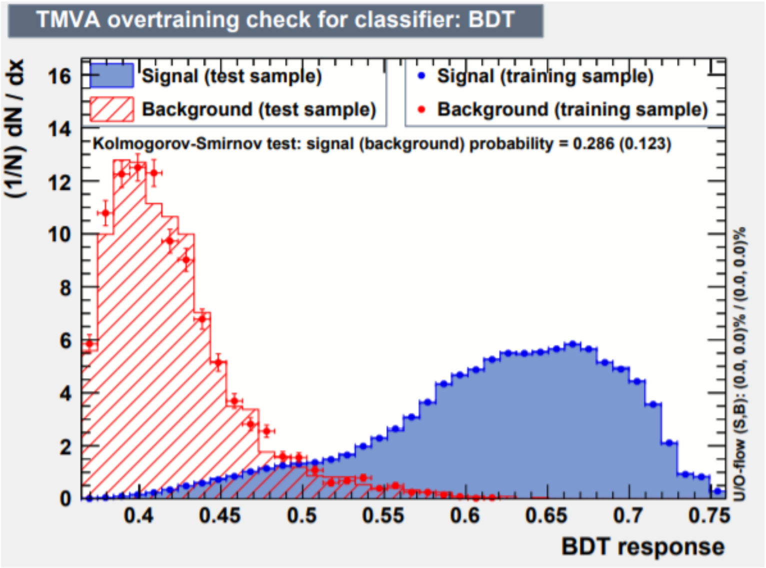


Figure 3.22: The cosmic rejection BDT used in the disappearance channel

Downstream of these selections, a dedicated cosmic rejection BDT is developed to further reduce cosmic events in conjunction with the CVN ν_μ score and the muon PID - ReMID. The cosmic rejection BDT is developed on the following kalman track inputs. The kalman track is identified as the track having the largest muon PID.

- cosine of the kalman track angle with the beam direction and the Y direction
- the length of the kalman track
- the maximum Y-coordinate of the track
- the distance of the track from each end wall of the detector
- fraction of total slice hits in the kalman track

- the transverse momentum fraction of the kalman track

Cut Level	ν_μ	$\bar{\nu}_\mu$	Total Beam Bkg	NC	Cosmics	Effic	Purity
No Cut	602	43.1	124	563	2.52×10^6	-	-
Quality	569	41.0	121	490	4.72×10^5	0.95	1.2×10^{-3}
Containment	249	14.1	83.9	336	6.99×10^3	0.41	0.03
Cosmic Rej.	239	13.8	78.2	293	1.54×10^3	0.40	0.11
Full PID	201	12.5	0.72	2.51	6.02	0.33	0.96

Table 3.1: Predicted signal and background components along with the signal selection efficiencies and purities in the disappearance channel at various cut stages for FHC

Cut Level	$\bar{\nu}_\mu$	ν_μ	Total Beam Bkg	NC	Cosmics	Effic	Purity
No Cut	193	111	45.8	225	1.63×10^6	-	-
Quality	186	106	45.0	195	2.99×10^5	0.97	6.2×10^{-4}
Containment	86.7	35.5	32.0	132	4.38×10^3	0.45	0.02
Cosmic Rej.	84.9	32.8	28.9	107	635	0.44	0.1
Full PID	79.5	25.7	0.34	0.79	1.65	0.41	0.97

Table 3.2: Predicted signal and background components along with the signal selection efficiencies and purities in the disappearance channel at various cut stages for RHC

Each of these inputs exhibit significant separation power between signal ν_μ-CC events and cosmic events. The BDT technique allows us to exploit the correlations between these inputs to further separate them. The tunings were done in 3D PID space (BDT vs CVN ν_μ vs ReMID) to further exploit correlations between the different PIDs. These are tuned according to a figure of merit (FOM) derived as :

$$FOM^2 = \frac{S^2}{S+B} \tag{3.4}$$

The FOM is calculated for both events between $(0, 5)$ GeV and events only in the dip region (between $(1, 2)$ GeV). The performance was found to be similar for both metrics. As a result, the PID cuts are Cosmic BDT > 0.45, CVN ν_μ score > 0.8, ReMID > 0.3. The CVN ν_μ score was found to be the most performant related to cosmic rejection and as a result of these cuts, the cosmic contamination was found to be only 6 (1.7) events for FHC (RHC) after the application of these PID cuts. The overall signal efficiency after all selections stands at 33% (41%) and the signal purity is $> 96\%$ for FHC (RHC), as shown in Tables 3.1 and 3.2.

The oscillation spectrum is binned in reconstructed neutrino energy and optimized to maximize the sensitivity of maximal mixing rejection. Since most of the power here comes from the dip region $(1, 2)$ GeV (more pronounced dip $\sim$ closer to maximal mixing), 10 bins of 0.1 GeV are chosen in this region, limited by the neutrino energy resolution. Events are binned in progressively increasing sizes beyond 2 GeV with $(2, 3)$ GeV in 0.25 GeV bins, $(3, 4)$ GeV in 0.5 GeV bins and a single bin for $(4, 5)$ GeV. The bin boundaries before the dip region are 0 GeV and 0.75 GeV.

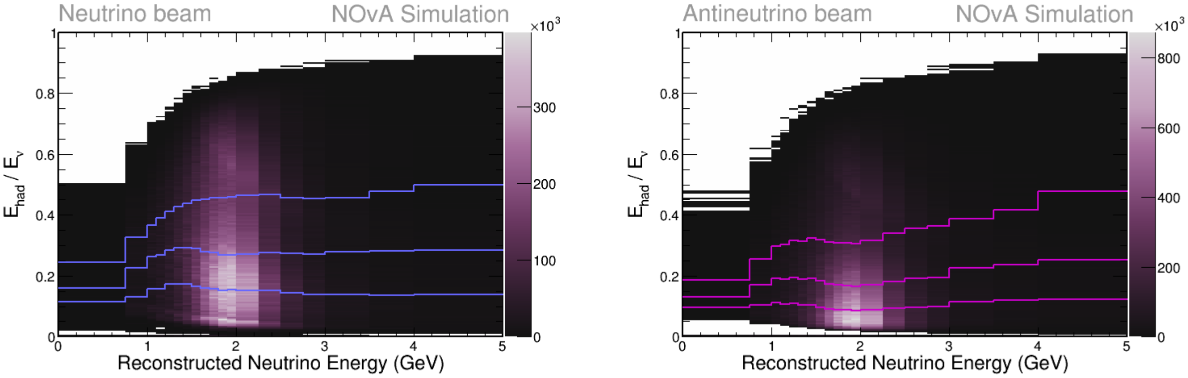


Figure 3.23: The 4 quantiles of hadronic energy fraction for FHC (left) and RHC (right) shown in terms of the chosen analysis binning scheme

A significant improvement in sensitivity to the oscillation parameters relevant in the disappearance channel can be obtained by splitting the FD oscillation spectrum into 4 quantiles of reconstructed hadronic energy fraction, E_{had}/E_{ν} [116], as shown in Fig. 3.23. The quantiles are built by selecting 4 equally populated regions E_{had}/E_{ν} for each E_{ν} bin described above. The quantile boundaries are chosen using true ν_{μ}-CC events without any oscillations. As we've seen, the neutrino energy resolution for ν_{μ}-CC events is largely driven by the hadronic energy resolution, $\sim$ 26% as opposed to the muon energy resolution, $\sim$ 3%. In the full oscillated spectrum in the dip region, there is also a significant danger for events to move across bins or in and out of the dip region because of the estimator which can cause a large bias in the measurement. These events between $(1, 2)$ GeV are a mixture of many different interaction types, QE → DIS as the Q^2 of the interaction increases. The interaction Q^2 is highly correlated to the inelasticity of the event, i.e the hadronic energy fraction E_{had}/E_{ν}.

As a result, QE event energies are easier to resolve compared to DIS events. Consequently, the reconstructed observable E_{had}/E_ν is a good predictor for the interaction type as well as the neutrino energy resolution. The energy resolutions per quantile vary from 4.2% (4.0%) in the first quantile to 10.5% (9.3%) for the last quantile for FHC (RHC) events, as seen in Fig. 3.24. By splitting up the spectrum into quantiles of this observable, one can therefore ensure that

- Events within a given quantile are handled separately from events in another quantile in the fit to the oscillation data. This means that events with better energy resolution are handled separately and consequently reduces the amount of movement of events in the dip region for that quantile, since they are well resolved by the estimator. The problem of events moving into and outside the dip region in the full spectrum is turned into a problem of events moving only in higher hadronic energy fraction quantiles. This improves the sensitivity for say, maximal mixing rejection.
- Events are largely separated by their interaction types. QE events dominate the first quantile, QE and MEC events populate the second quantile, MEC and RES events populate the third and DIS events dominate the last quantile. This ensures that any uncertainty involving the relative composition of the different interaction types has smaller knock-on effects on the spectrum in the dip region and consequently a smaller systematic effect on the oscillation measurement.

In addition, the small background seen in the disappearance channel are seen to pretty much all lie in the last quantile. This has an additional improvement for sensitivity due to the signal and background separation that is a consequence of this procedure. This is due to the fact that most of the background events are either NC which mimic ν_μ-CC events only if the lepton track is quite small and appears mixed in with the hadronic component, or they are energetic cosmic events whose energies cause the muon track to exhibit significant non

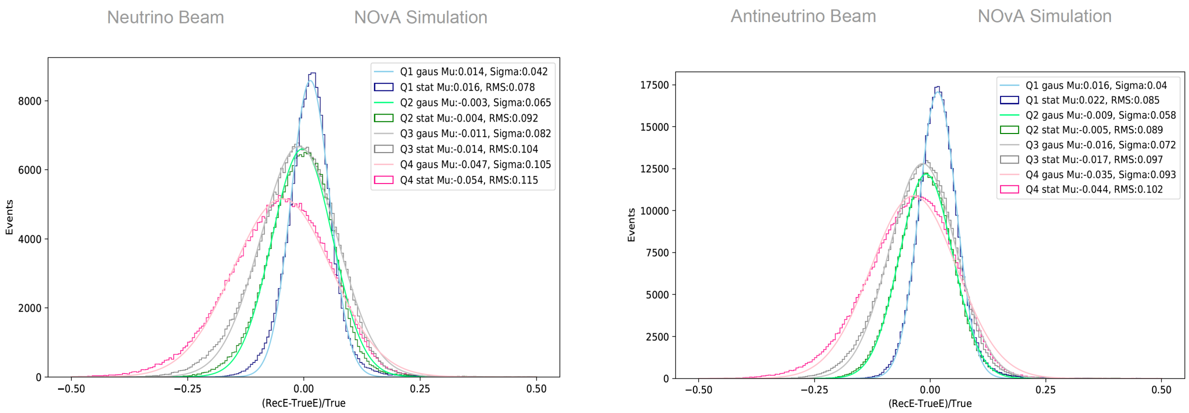


Figure 3.24: The ν_μ-CC energy resolutions for different hadronic energy fraction quantiles shown for FHC (left) and RHC (right)

MIP-like behavior which is picked up the energy estimator as a hadronic component.

All told, the hadronic energy fraction splitting and the varying bin sizes have an effect equivalent to an improvement of 20% in sensitivity for maximal mixing rejection. It was also seen that splitting the spectrum into more than 4 quantiles had diminishing improvements.

3.3.2 The $\nu_\mu \to \nu_e$ Channel

Fig. 3.25 illustrates the selection process in the appearance channel. In this channel, after the cosmic veto, a series of quality cuts are applied to ensure a clean sample of events and even get rid of a bunch of junk cosmic events. Events are rejected if they have more than 8 hits in any given plane to cut out FEB flashers. They are also required to have a reconstructed vertex and atleast one reconstructed FuzzyK prong. A basic preselection is applied to moderate the fraction of ν_μ-CCs and NCs in the sample. Interaction slices are required to have atleast 30 hits and atmost 150 hits. Events are selected with reconstructed ν_e-CC energies between 1 GeV and 4 GeV. The longest prong in the slice is required to be atleast 100 cm and atmost 500 cm. Longest prongs of more than 500 cm are dominated by ν_μ-CC events.

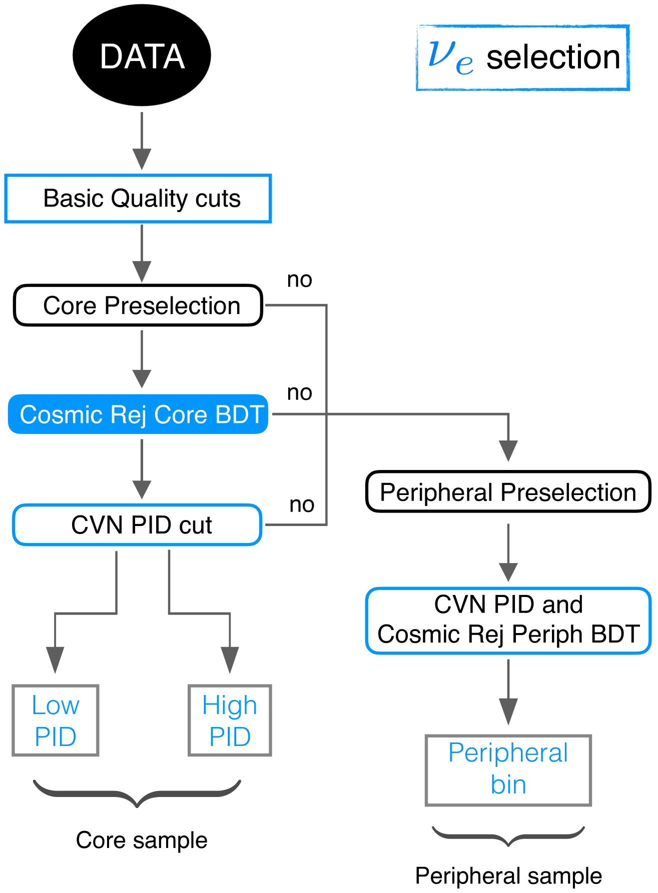


Figure 3.25: Schematic of the cutflow used for the selection in the appearance channel

As for the disappearance channel, the containment cuts are the primary set of criteria to get rid of a large portion of cosmic interactions. The containment cuts are tuned using the minimum distance of all prongs to each of the end walls of the detector, 6 in total. As before these are tuned by looking at the escaping energy fraction and the signal energy resolution metrics. And similar to the disappearance channel again, the distance to the top wall is tightened relative to the other walls since this is where cosmics dominate. As a result, the containment criteria are given by requiring events atleast 63 cm from the top wall, 18 cm from the front and back and 12 cm from the rest.

As with the disappearance channel, a dedicated cosmic rejection BDT [117] (shown in Fig. 3.26) is trained for both FHC and RHC and is tuned in conjunction with the CVN ν_e PID

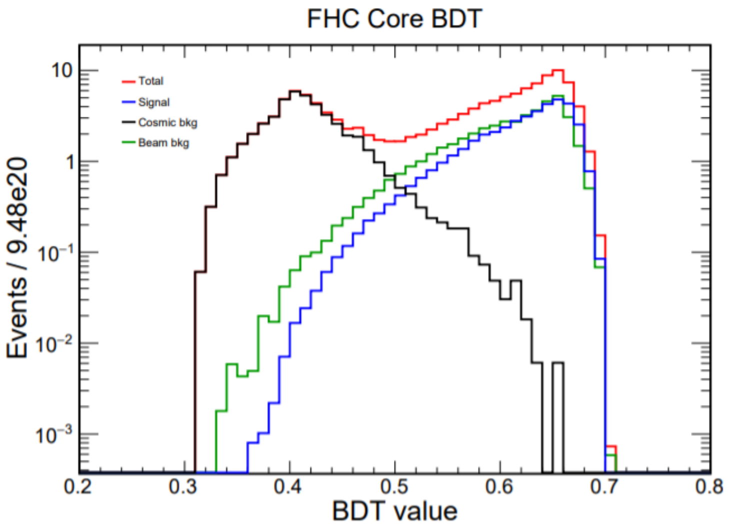


Figure 3.26: The response of the cosmic rejection BDT used for the core selection

score to get rid of a lot of cosmic events that make it into the inner regions of the detector as well as maximize the amount of signal with respect to beam backgrounds, mainly the beam-ν_e-CC contamination but also NCs and ν_μ-CCs. It is important to note that the CVN ν_e PID was trained to be agnostic towards beam ν_e-CC events vs oscillated ν_e-CC events as they're the same topologically. However the tuning treats them differently. The cosmic rejection BDT was trained on the following reconstructed inputs :

- number of hits in the slice
- the asymmetry in the hits at the start and the end of the event
- the transverse momentum fraction of the entire event with respect to the beam direction
- the containment variables, (distance of prongs to each end wall) to see if one can exploit any correlations downstream after applying the containment
- the fraction of energy in the identified electron (most energetic prong)
- the reconstructed width of the electromagnetic shower cascade

One of the most powerful variables in the BDT is the transverse momentum fraction of the event, which by momentum conservation should be 0 for neutrino events if all directions are reconstructed perfectly and close to 1 for downward going cosmic events. In addition, the distance to the top wall of the detector is also seen to be good at rejecting cosmic events by using its separation power in conjunction with the other variables in the BDT.

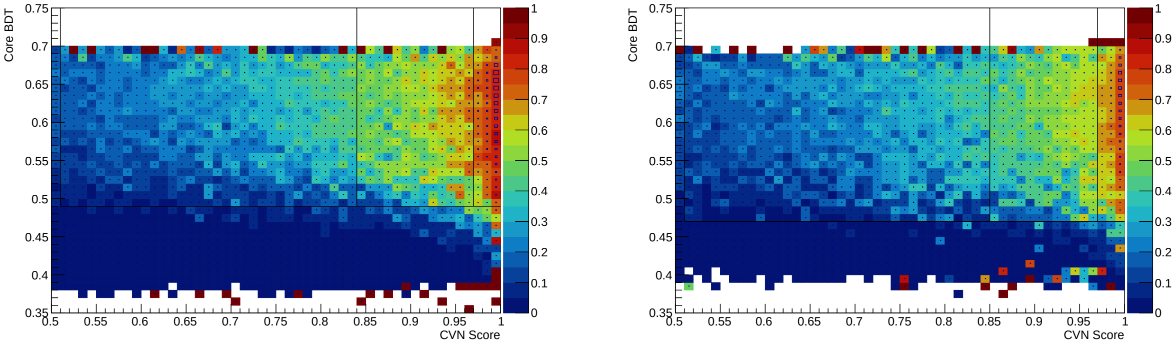


Figure 3.27: 2D distributions of signal purity (color scale) and signal content (box size) in CoreBDT-CVN space for FHC (left) and RHC (right) used in the core selection. Overlaid are the tuned core analysis bins at low, mid and high CVN respectively. Only the latter two are used in the analysis

The tuning is carried out in 2D PID space (BDT vs CVN ν_e), as illustrated in Fig. 3.27. It is important to note that each PID has specific strengths, the BDT is good at getting rid of cosmic events specifically while the CVN can take up a large portion of the slack for rejecting beam backgrounds that are not ν_e-CC like. In fact, the CVN score is also good at getting rid of cosmics but it turns out that there is a correlation between the CVN and the BDT where even at high CVN scores, one can see a significant cosmic contamination which can be gotten rid of by the BDT. One can study this behaviour by looking at a 2D “box-plot” in PID space, where the color palette indicates the signal purity and the box sizes indicates the amount of signal contained in that bin. It is evident from this plot that there are three important regions in this PID space, one at very high CVN scores where the signal purity is great ($> 80\%$) and where most of the signal actually lives, one at an intermediate CVN score where the signal purity is in the range of $\sim 50\%$ to $\sim 80\%$ and one at lower CVN scores where the signal purity is $< 50\%$. The intermediate regions are fairly large in terms of their overall signal content to justify including them in the final sample.

Therefore, the tuning is carried out to find 3 CVN boundaries and a single BDT boundary to maximize the figure of merit given as :

$$FOM^2 = \sum_{i=1}^{3} \frac{S_i^2}{S_i + B_i} \tag{3.5}$$

where i runs over the different regions. One can allow the BDT cut to roam within each region separately as well but it was found to have a tiny effect on the overall FOM. The tuning was carried out with a loose CVN ν_e cut of > 0.5. Different figures of merit were also tried by including an error component in the amount of background selected to mimic a systematic uncertainty, but were found to have little effect on the results of the tune. The tuning ended up pulling the lowest CVN boundary right to the edge of 0.5, suggesting that in this region, the FOM depended much less on the signal purity, given by $\frac{S_i}{S_i+B_i}$ but rather the actual signal content in this CVN bin, S_i. Because the aggregate purity was terrible in this lowest CVN bin ($\sim 20\%$) and the actual signal content itself was also much lower than the other regions, the FOM gain from this bin was small enough ($\sim 3\%$) to justify dropping this sample to ensure our measurement doesn't rely on a CVN bin heavily swamped by backgrounds where the systematic uncertainties would dominate. This selection flow results in a signal efficiency of about $\sim 53\%(\sim 63\%)$ and an overall purity of $\sim 74\%$ ($\sim 64\%$) with CVN boundaries defined by 0.84 (0.85) and 0.97 (0.97) for the FHC (RHC) beam mode . Most of the remaining background is from the beam with cosmics pared down to $\sim 1.5(0.6)$ events. The background is dominated by beam ν_e-CC events which mostly lie in the high CVN bin. The next largest background is NC and most of the non-ν_e-CC background have an electromagnetic component as a result of π^0 decays where the second γ cascade is too small to identify or buried within the first one. The two CVN samples are split into energy bins of size 0.5 GeV from $(1, 4)$ GeV. This spectrum shows a slight separation power between the irreducible beam ν_e-CC background and the signal of oscillated ν_e-CCs which is due to the different flux spectra for these components.

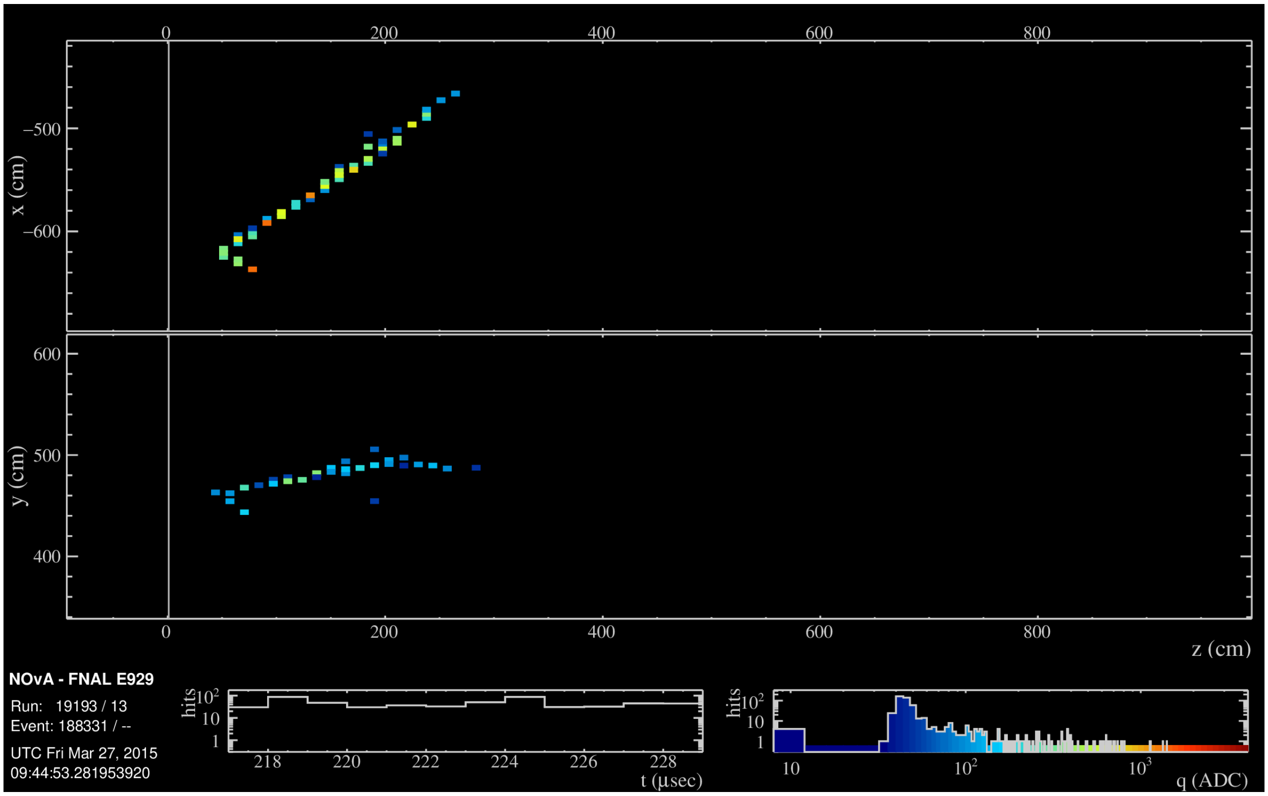


Figure 3.28: Example of a peripheral ν_e-CC event near the edges of the Far Detector

In addition to this sample, a neat idea to improve the sensitivity to the oscillated signal events is by trying to reclaim some of the events that have been rejected by the above set of selections. It turns out that there is a fair chunk of signal events that are lost primarily due to the containment definition. One example of such an event is in Fig. 3.28. The containment criteria described above defines a "core" fiducial volume in the Far Detector within which one can expect our BDT input variables like the transverse momentum fraction and the energy estimator to work well. However, in the periphery of the detector, one finds a fair amount of easily identifiable ν_e-CC events that would add to the sensitivity of the oscillation measurement. Even though this region is dominated by cosmic backgrounds, one can train an additional BDT with inputs specialized to this region and in conjunction with a high CVN ν_e score, moderate this background to an extent where an appreciably pure sample can be added to the original “core” sample. The estimator is not as trust-worthy in this region however, since the signal events might not be fully contained, so this “peripheral” sample is a single-bin sample that is added on to the “core” spectrum binned in energy. The peripheral BDT (shown in Fig. 3.29) is therefore trained with inputs given as :

- the number of hits in the slice (to ensure there are events with enough hits in them)
- the asymmetry in the hits between the start and the end of the cluster
- the transverse momentum fraction, now split into two separate variables, p_x/p and p_y/p, to exploit any differences in the individual momentum fractions
- the distance to the top of the detector
- the vertex co-ordinates in X, Y and Z.

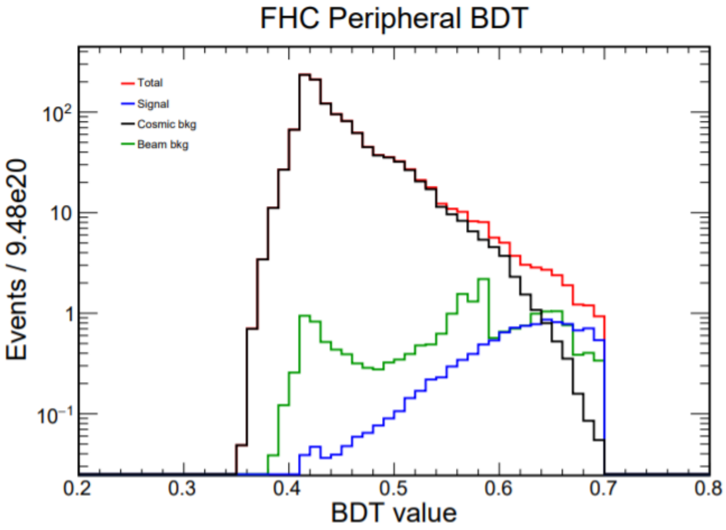


Figure 3.29: The response of the cosmic rejection BDT in the peripheral selection for the appearance analysis

The BDT is trained on events that fail the above containment criteria with a loose CVN cut applied to ensure the signal fraction is high. The primary background here are cosmic events as expected. As is evident, the peripheral BDT is mainly made up of variables that describe

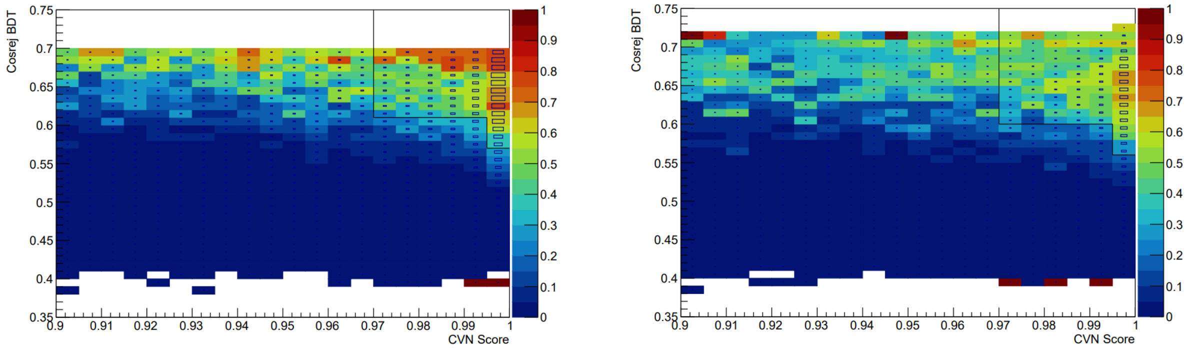


Figure 3.30: 2D distributions of signal purity (color scale) and signal content (box size) in PeriBDT-CVN space for FHC (left) and RHC (right) used in the peripheral selection. The box cut represents the tuned analysis bin

Cut Level	ν_e	$\bar{\nu}_e$	Total Beam Bkg	NC	Cosmics	Effic	Purity
No Cut	96.8	1.84	1.58×10^3	820	2.95×10^6	-	-
Data Qual.	93.2	1.78	1.5×10^3	680	5.12×10^5	0.96	1.81×10^{-4}
Containment	71.4	1.39	560	510	2.58×10^4	0.74	2.65×10^{-3}
Cosmic Rej.	63.5	1.02	91.5	169	353	0.66	0.09
Full Core PID	52.0	0.86	11.4	4.73	1.51	0.54	0.74
Basic Peri Cuts	38.7	0.69	549	508	4.53×10^5	0.4	8.53×10^{-5}
Full Peri PID	8.63	0.19	3.84	0.25	1.63	0.09	0.59

Table 3.3: Predicted signal and background components with the signal selection efficiencies and purities in the appearance channel at various cut stages for FHC

where the slice lies in the detector and other general slice characteristics, rather than looking for electromagnetic shower properties. This improves upon the containment chosen for the core sample since now the BDT can exploit correlations between these variables rather than assuming the containment variables are independent of each other. One can then tune in 2D PID space (Peripheral BDT vs CVN ν_e) as before to guarantee a relatively pure sample of signal events, shown in Fig. 3.30. The tuning is carried out in a similar fashion to the core sample PID tuning but on events that fail any of the selections downstream of the basic quality cuts. A hard constraint is also applied on the CVN boundary, i.e an initial CVN cut defined by the high CVN cut (0.97) in the core sample is chosen. One then finds that at even higher CVN values compared to this, we can loosen the requirements on the BDT for cosmic rejection, since CVN can take up a bit more of the slack here. Therefore, one arrives

Cut Level	$bar\nu_e$	ν_e	Total Beam Bkg	NC	Cosmics	Effic	Purity
No Cut	27.0	5.14	723	328	1.92×10^6	-	-
Data Qual.	26.4	4.92	693	266	3.25×10^5	0.98	8.12×10^{-5}
Containment	20.9	3.72	252	199	1.61×10^4	0.77	1.26×10^{-3}
Cosmic Rej.	19.3	2.57	27.4	57.8	296	0.72	0.05
Full Core PID	17.3	1.96	5.43	1.6	0.59	0.64	0.64
Basic Peri Cuts	8.6	2.25	259	200	2.88×10^5	0.32	2.97×10^{-5}
Full Peri PID	3.01	0.42	2.11	0.12	1.0	0.11	0.45

Table 3.4: Predicted signal and background components with the signal selection efficiencies and purities in the appearance channel at various cut stages for RHC

at a “double-box” like cut that select the purest bins in this space where most of the signal lies. At a CVN cut of 0.995, the BDT cut is looser with a value of 0.56 while at the previous CVN boundary of 0.97, the BDT cut is tighter with a value of 0.6. This is similar for both FHC and RHC. Looseing the CVN constraint of 0.97 would require further tightening the BDT to very high values which doesn’t buy us a lot of signal. In summary, the peripheral tuning exploits a further correlation between the peripheral cosmic rejection BDT and high values of CVN to ensure a relatively pure sample. This sample improves the sensitivity to the oscillation measurement by adding an equivalent of 20% more data.

Since this sample involves very high CVN scores and is in the periphery of the detector, the backgrounds are dominated by beam ν_e-CC and cosmic events. The purity in this sample is $\sim$ 60% ($\sim$ 45%) for FHC (RHC), described in Tables 3.3 and 3.4.

We’ve seen therefore that the selections are capable of selecting as much signal as possible in a detector environment that is flooded by cosmics. They utilize BDTs that incorporate a fuzzier definition of contained events in order to capture the signal wherever they might arise in the detector. We’ve also seen how most of the backgrounds are neutrino events that have electromagnetic characteristics with only a tiny fraction of them given by cosmic events.

The cosmic backgrounds that remain are mostly small clusters of hits that arise from a muon track releasing some bremsstrahlung radiation which are clustered separately by the slicer

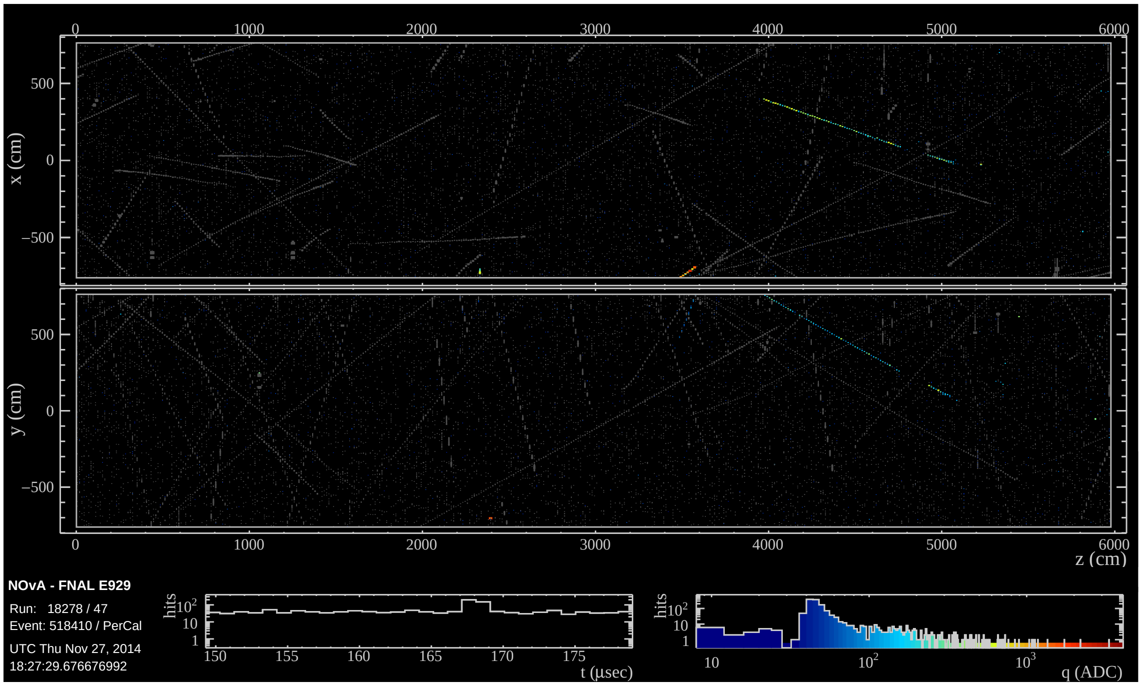


Figure 3.31: Example of a delayed bremsstrahlung shower detached from the parent muon that produced it

and the PIDs are unable to reject as not being ν_e-CC like. In rare cases, these bremsstrahlung γs are sufficiently detached from the muon track and *also* energetic enough to look like a well developed shower cascade, as shown in Fig. 3.31. Since this is a completely separate slice, there are no well-defined characteristics that can be used to remove such events which can contaminate the high purity sample as a result of having high PID scores. One can get rid of a fraction of these events that could exist in the sample by looking for a muon track close in time and space to the bremsstrahlung slice [118]. First, a selection is made on the difference in mean slice times between a given slice and its closest-in-time neighbour slice, as shown in Fig. 3.32. Beam events in general have a much wider Δt distribution as defined above and cosmics exhibit a larger peak. Within this peak, events are rejected from the sample if they're both close in space to its closest-in-time neighbour and if its closest-in-time neighbour is near the top wall of the detector, as shown in Fig. 3.33. The cuts in these variables are tuned separately for each of the three bins (two core and 1 peripheral) after all the regular selections defined above are applied. These set of cuts get rid of $\sim 13\%$ of the

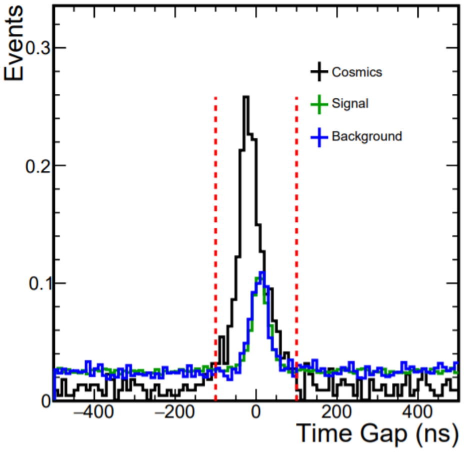


Figure 3.32: The time delay between closest-in-time slices for cosmics and beam events. Events near the peak are considered for rejection using further cuts below

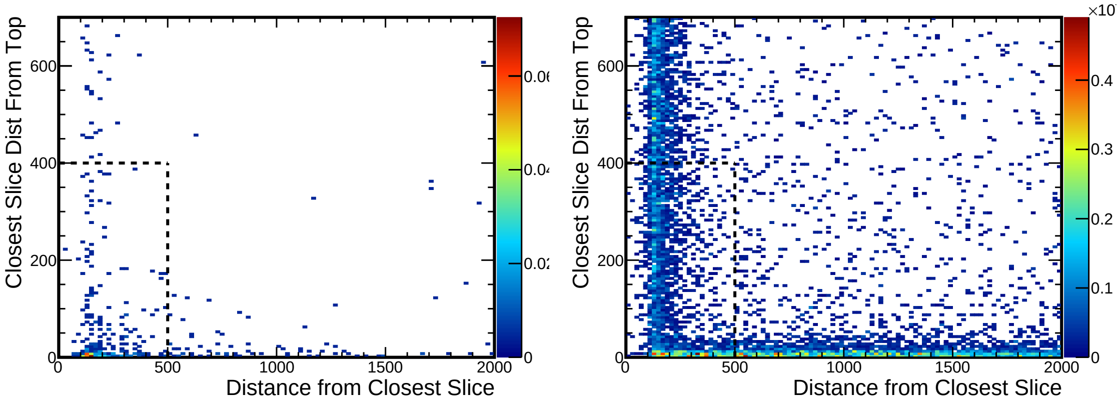


Figure 3.33: Box cuts for events in the time gap peak shown above for cosmics (left) and signal events (right). Events within the box are rejected from the analysis

remaining cosmic background with no loss of signal or beam events. As such, this doesn't add a lot of sensitivity to the measurement since there aren't a whole lot of cosmics left at the end but it does make doubly sure that a certain kind of cosmic event will not contaminate the final sample. Events are required to be atleast 100 ns away from its neighbour or atleast 50 cm (400 cm) from the top wall for the high CVN bins (low CVN bin) or atleast 500 cm away from its neighbour. Here the top wall requirement is tightened for the low CVN bin since that's where most of the cosmics in the final sample are expected to lie. These cuts are also useful for separate analyses looking for NC disappearance where neutrinos oscillate into a sterile neutrino. These detached bremsstrahlung showers are a prominent background for the sterile neutrino analysis and are tuned appropriately but are also carefully studied in the 3-flavor ν_e-appearance channel.

3.4 Near Detector Constraints

The Near Detector data is used to develop constraints on the Far Detector prediction and reduce the systematic burden of the experiment. This is vital since many systematics mimic the effect of oscillations and are in some cases significantly large due to theoretical uncertainties for cross-sections and flux. The Near Detector data is also used to characterize the beam composition and the breakdown of the different beam components in order to ensure a data-driven estimate for the beam backgrounds at the Far Detector, which are especially important in the appearance channel.

The signal component in the two oscillation channels therefore utilize ν_μ-CC data at the Near Detector, while the beam backgrounds for the appearance channel utilize a ν_e-CC rich sample at the Near Detector. The backgrounds for the disappearance channel are tiny enough so one can use the backgrounds selected in the ND ν_μ-CC sample to provide the necessary constraint.

3.4.1 Selection

The selections used for the ND ν_μ-CC sample for the signal constraints mimic the selections used in the Far Detector except for the cosmic rejection [113]. This is done to ensure the Near Detector constraints are carried out using events with a similar kinematic phase space, since the detectors are identical except in size. Important acceptance differences that arise due to the difference in sizes will be described in later sections.

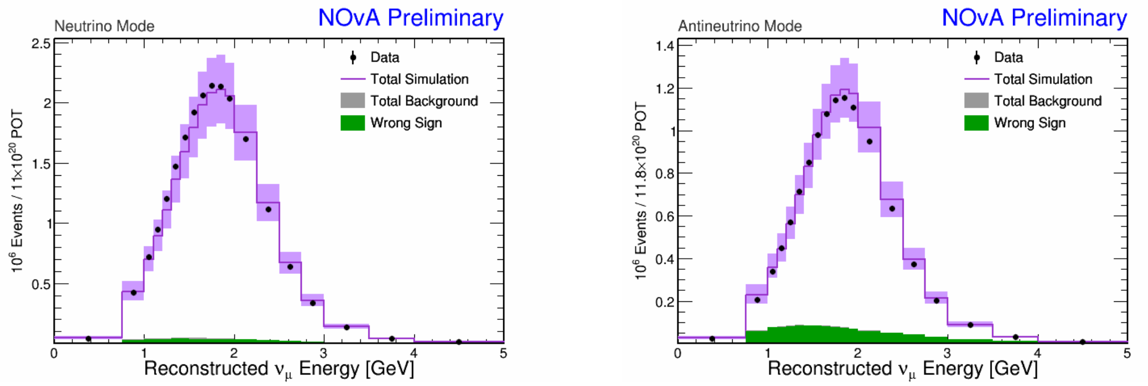


Figure 3.34: The predicted and observed event rates for the ν_μ-CC sample at the Near Detector for FHC (left) and RHC (right). The shaded regions represent the 1-σ systematic error described in the next chapter.

The same basic quality and PID cuts are used as that in the Far Detector disappearance channel. The PID cuts involve the CVN and the Muon ID (ReMID). No cosmic rejection cuts are applied since the cosmic backgrounds in the Near Detector are negligible. The containment cuts are defined by requiring all prongs to be within the regions $(-180, 180)$ cm in X and Y and between $(40, 1525)$ cm in Z, which includes the muon catcher. However, only events with the kalman track identified to be the muon are allowed to extend to the muon catcher region. Events with any other tracks extending to the muon catcher are removed. This is done to ensure good energy estimation since the cells are interleaved by steel rather than PVC in the muon catcher and are much harder to calibrate. The muons extending to this region is not an issue since their energies are estimated by range anyway. Further containment cuts use geometric criteria to ensure a good energy resolution and utilize cuts on projections of the kalman track onto the front and back walls of the Near Detector and so on. In addition, the binning used is similar to the disappearance channel. The wrong sign contribution is $\sim 10\%$ ($\sim 3\%$) for the RHC (FHC) beam mode and is folded in as signal for the disappearance channel but not for the appearance channel. The sample is otherwise relatively background-free, as shown in Fig. 3.34.

For the ν_e-CC rich sample, the selections try to mimic the appearance channel since this sample is used to predict the beam backgrounds in that channel. The same quality cuts on

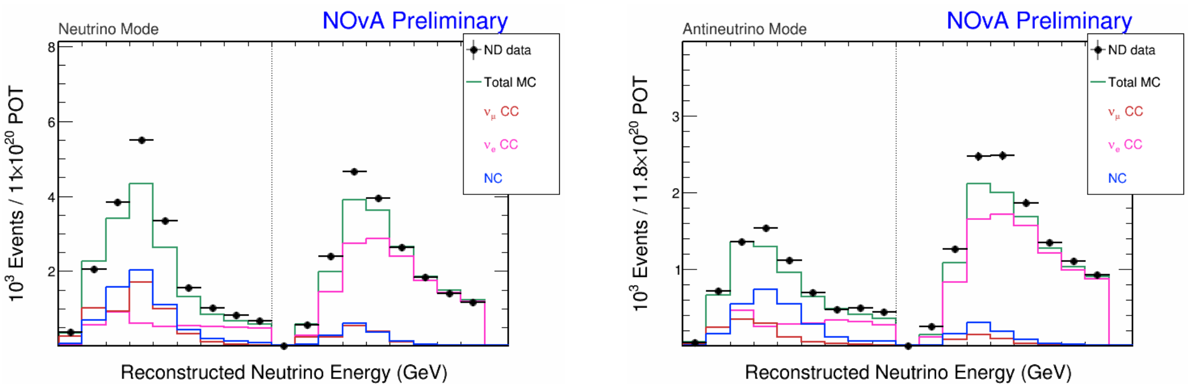


Figure 3.35: The predicted and observed event rates for the ν_e-CC sample at the Near Detector for FHC (left) and RHC (right)

existence of a reconstructed vertex and a FuzzyK prong as well as the maximum allowed number of hits per plane are used. The reconstructed vertex is required to be within the fiducial volume defined by the region $(-100, 160)$ cm in X, $(-160, 100)$ cm in Y and $(150, 900)$ cm in Z. The event is required to be atleast 10 planes away from the front wall. In a similar fashion to the ND ν_μ-CC sample, the containment cuts are defined over the start and stop co-ordinates of each reconstructed FuzzyK prong. These are required to be within $(-150, 170)$ cm in X, $(-170, 150)$ cm in Y and $(100, 1225)$ cm in Z. The pre-selections also include cuts on the number of hits in the interaction cluster (between 20 and 200) and the length of longest prong (between 100 and 500 cm). Events above reconstructed energies of 4.5 GeV are discarded. The same PID cuts on the CVN ν_e-CC ID are applied to mimic the Far Detector sample. They are also binned in a similar fashion, with a low purity and high purity sample defined by low and high CVN scores. There is no analog of the peripheral sample at the Near Detector but the FD peripheral beam backgrounds are described by the ND high PID sample directly, since apart from cosmic rejection the PID cuts between the high PID and the peripheral sample at FD are the same. This sample has large contributions from ν_μ-CC and NC interactions even though it is relatively ν_e-CC rich. These are roughly described by $\sim 25\%$ ν_μ-CC and NC and $\sim 50\%$ ν_e-CC events coming from the beam for the FHC beam mode. In the RHC beam mode, there are contributions from 5 components in this sample due to the larger wrong-sign contamination in the beam, both in ν_e-CC and

ν_μ-CC (neutrinos are wrong-sign in the anti-neutrino beam). There are $\sim 30\%$ wrong-sign events here, mostly coming from ν_e-CC events. This is shown in Fig. 3.35

3.4.2 The Cross-Section Tune

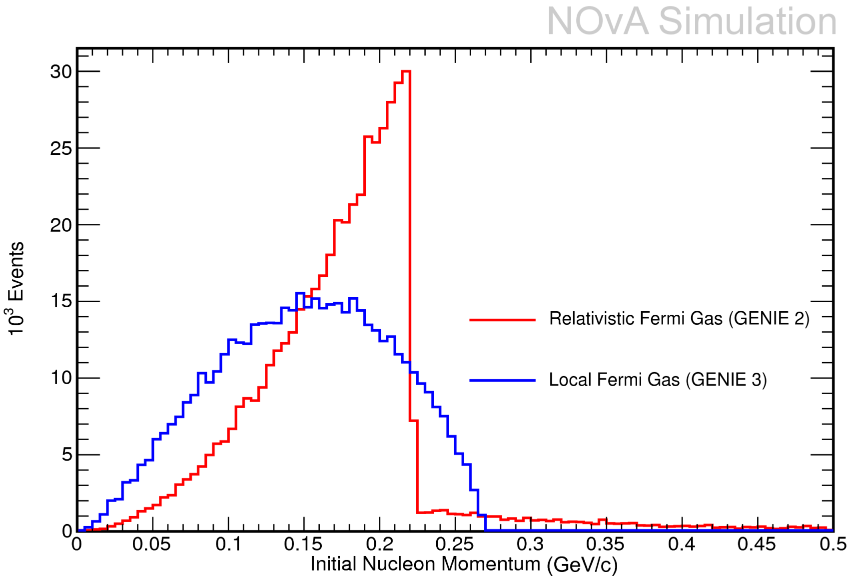


Figure 3.36: Comparison of the RFG model used in previous NOvA analyses and the LFG model used now in terms of the momentum of the struck nucleon

An important step in the analysis is the development of a tune for the cross-section model using the Near Detector data. As mentioned before, the cross section model coming from GENIE v3 has many theoretically uncertain aspects that are still being studied and are therefore incomplete. The Near Detector data can therefore provide an important input to the cross-section model separate from just using it to correct for the Far Detector prediction. Indeed, the out-of-the-box GENIE model does not describe the Near Detector data very well, especially in interaction phase-space variables sensitive to competing interaction effects, such as the 4−momentum transfer, $(q_0, \vec{q})$ and other variables relevant to the hadronic sector.

As mentioned before, the out-of-the box model contains the Valencia 1p1h calculation taking into account long-range correlations within a local Fermi gas nuclear model to describe

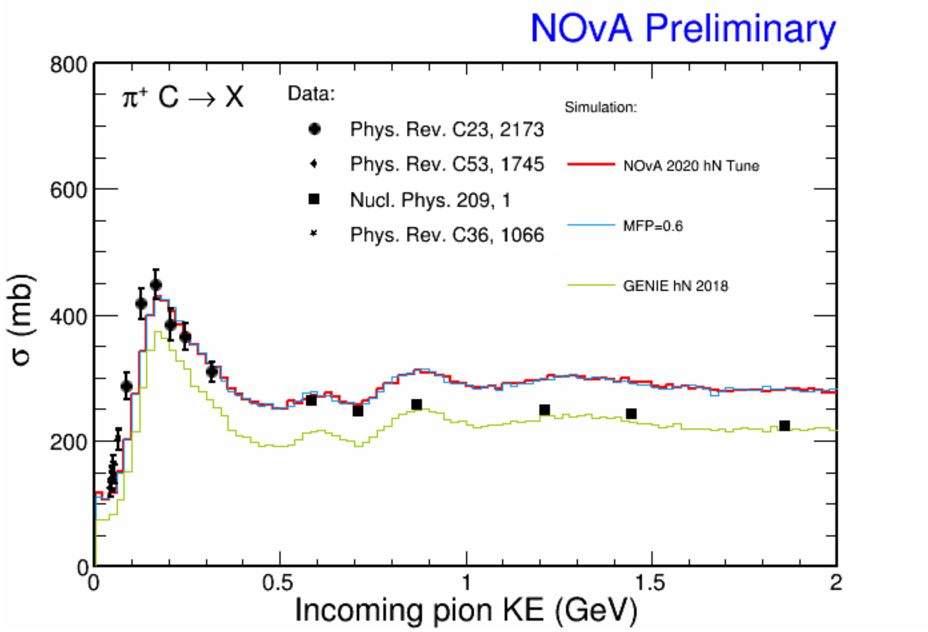


Figure 3.37: Corrections to the GENIE hN FSI model based on external pion scattering measurements used in the NOvA tune. Data shown represents the inclusive pion scattering cross-section measurements on Carbon

quasi-elastic interactions, illustrated in Fig. 3.36. It uses a Valencia calculation again for Meson-Exchange Current (MEC) type interactions. For RES and DIS events, GENIE uses the Berger-Sehgal and the Bodek-Yang models respectively. The final-state interactions (FSI) are modelled by a hN semi-classical cascade model. Of these, the ND data is used only to tune the MEC model since this is where the theory is the most lacking [119][120]. Additionally, since GENIE does not provide reweightable knobs for the FSI model unlike before and it has been known to disagree with external pion-scattering data, the FSI model is also tweaked to agree with these external measurements using a BDT-based technique, shown in Fig. 3.37.

MEC events lie right in between QE and RES events in terms of the magnitude of the 4−momentum transfer, Q^2 $(= -q^2)$. As a result, the tune is designed to model contributions from both QE-like and RES-like MEC events. Shown in Fig. 3.38, the Valencia model is therefore useful as a starting point since it provides enough spread in Q^2 in order to develop the tune. The MEC tune is similar in philosophy to Minerva and is done in the interaction

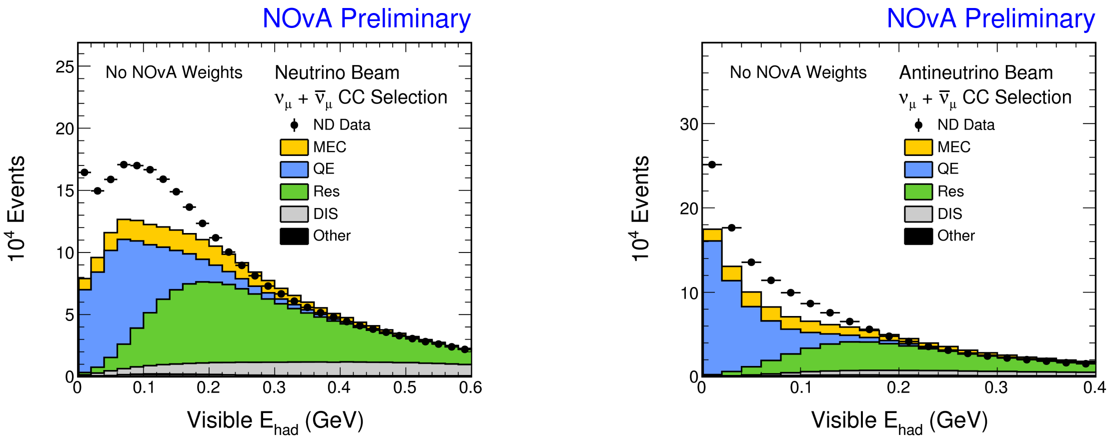


Figure 3.38: The observed distributions along with the out of the box GENIE3 prediction for the visible hadronic energy at the NOvA Near Detector for FHC (left) and RHC (right)

phase-space variables, $(q_0, |q|)$. Different from Minerva, the tune is parameterized using a double-gaussian like model with a total of 13 free parameters as :

$$\mathbf{wgt_{MEC}} = Baseline + a_1 \times gauss_1(q_0, |q|) + a_2 \times gauss_2(q_0, |q|) \tag{3.6}$$

$Baseline$, a_1 and a_2 adjust the normalizations of the base Valencia model and the two gaussians respectively. The other 10 parameters describe the two 2D gaussians in this space with the means and standard-deviations for individual variables q_0 and $|q|$ and a correlation parameter between the two variables. The tuned MEC model therefore shows a double-peaked structure in these variables, as shown in Fig. 3.39, similar to observations from other experiments like T2K and Minerva. The resulting data/MC is much improved especially in variables like the reconstructed visible hadronic energy, hadronic energy fraction, reconstructed hadronic invariant mass (W) and so on, as shown in Figs. 3.40 and 3.41. It is found that the FHC tune works quite well for RHC events as well suggesting that the base Valencia model is able to take differences between neutrinos and anti-neutrinos into account.

It is interesting to note that the overall cross-section for MEC events after the tune is similar to other components, even though one might expect an ab-initio calculation to have a significant drop-off vs the other components, since it requires interacting with a correlated

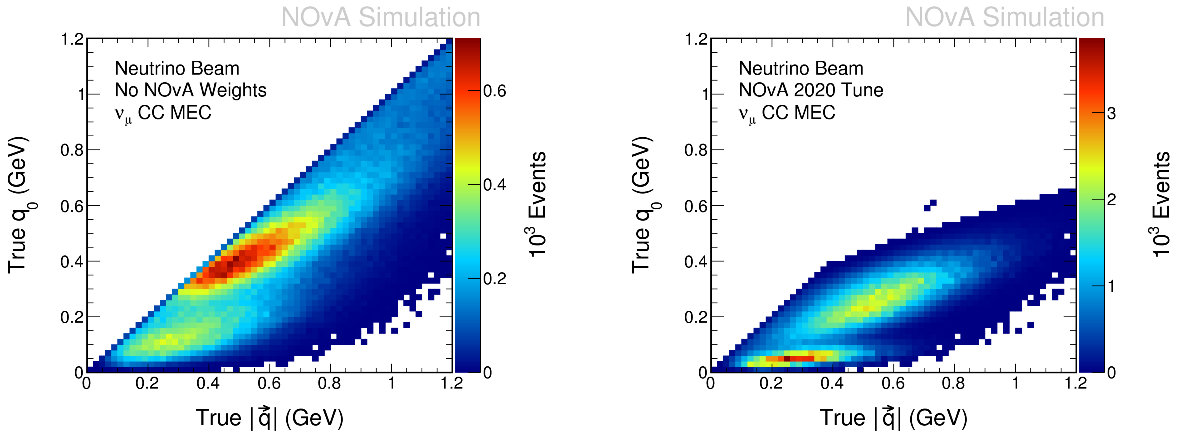


Figure 3.39: The phase-space of MEC events in $(q_0, |\vec{q}|)$ space for uncorrected (left) and corrected using the NOvA tune (right)

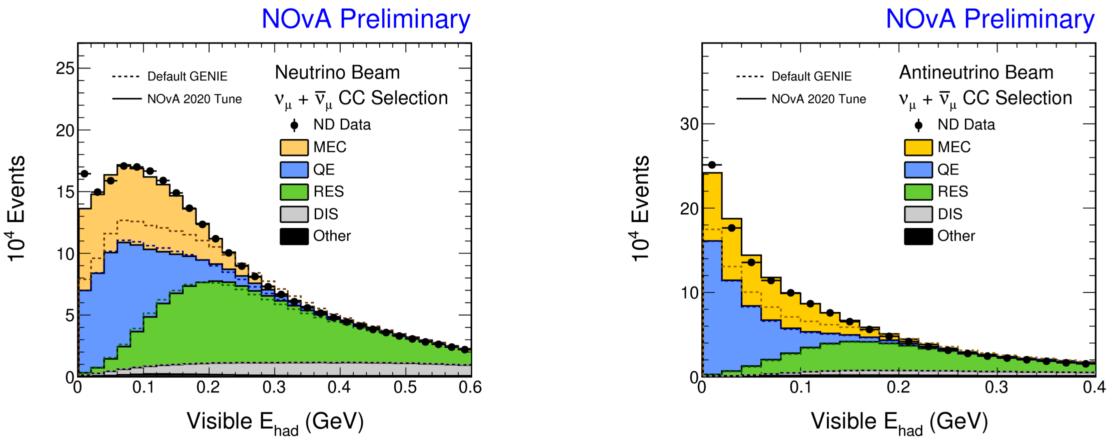


Figure 3.40: The distributions of visible hadronic energy at the NOvA Near Detector after the NOvA tune for FHC (left) and RHC (right)

pair of nucleons rather than a single nucleon. This essentially comes from the fact that there is a large discrepancy between the out-of-the box interaction model and the Near Detector data in a phase-space between QE and RES events and the consequent assumption that it can be described by MEC events. We will discuss the systematic uncertainties in the interaction model in later sections but in general they aim to be fairly conservative given this observation, essentially varying the MEC component significantly between QE-like and RES-like phase space. It is also important to note that since the Near Detector is a calorimetric detector, biases in the composition of the different interaction types in the tuned model results in an overall bias in the energy estimator, since we don't utilize

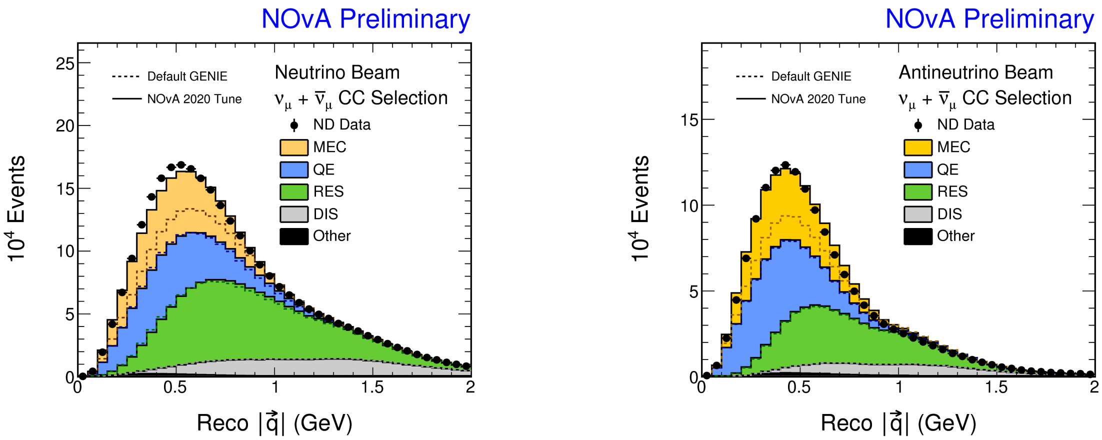


Figure 3.41: The distributions of reconstructed transfer momentum at the NOvA Near Detector after the NOvA tune for FHC (left) and RHC (right)

and indeed find it hard to resolve the interaction phase space in terms of the final state particles. This is mitigated by splitting up the ND ν_μ-sample into 4 hadronic energy fraction quantiles as described in the FD disappearance analysis and constraining each of the FD quantiles separately using the ND quantiles. We will also discuss ways to mitigate other biases later on but this confirms even more the importance of hadronic energy fraction quantile splitting to constrain separately regions where energy resolutions are different. In addition, the development of energy estimators while ensuring the energy estimator bias is flat across neutrino energies and interaction types is very much a worthwile endeavour.

3.4.3 ND → FD Signal Extrapolation

To constrain the Far Detector prediction using the Near Detector data, NOvA uses a F/N-ratio based extrapolation method on Near Detector samples [121]. For the signals to the appearance and disappearance channels at the Far Detector, the ν_μ-CC Near Detector samples are used using the selections defined in previous sections. Let us first describe the extrapolation methodology, illustrated in Figs. 3.42 and 3.43.

The first important point that the extrapolation takes into account is the fact that neutrino

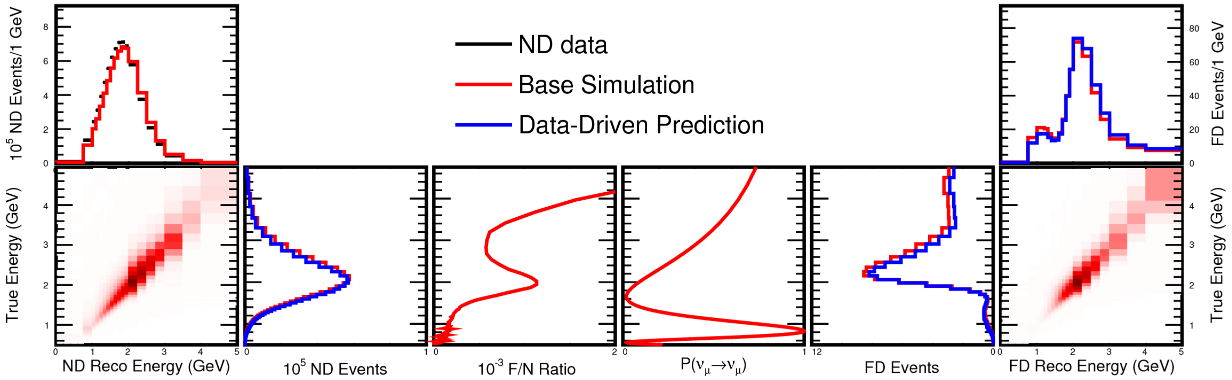


Figure 3.42: Illustration of the NOvA signal extrapolation from ND to FD for the disappearance channel

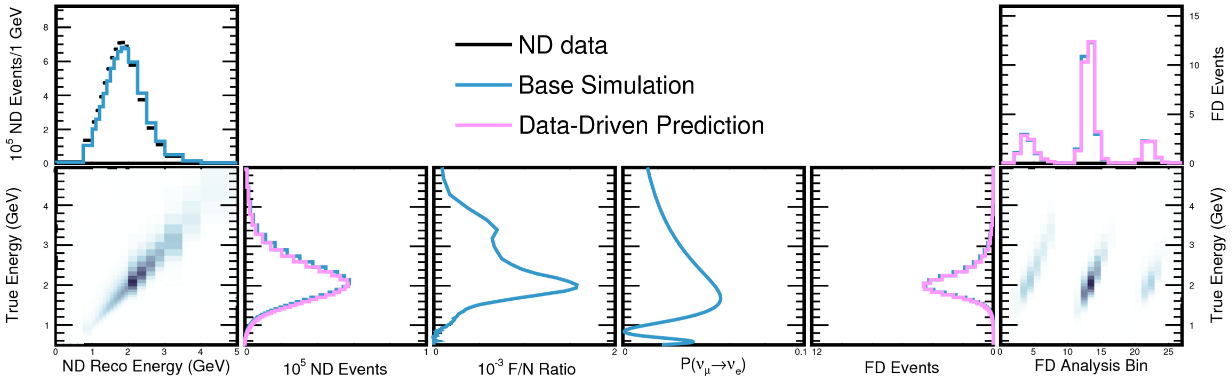


Figure 3.43: Illustration of the NOvA signal extrapolation from ND to FD for the appearance channel

oscillation probabilities are defined over true neutrino energies, while the ND samples are binned in reconstructed energies, which are therefore smeared representations. Consequently, the Near Detector data is unfolded into true energy space by applying a reco-true energy matrix. Since there are some backgrounds in the ND samples, especially in the RHC beam mode due to wrong-sign contamination, the ND signal spectrum is derived by “decomposing” the ND data into the different components proportionally. This is done by taking the ND data and simulated spectrum into true energy space and reweighting the various components of the simulated spectrum proportionally to match the ND data spectrum, i.e the residual data/MC difference is split proportionally between the different components.

$$N_j^{S,T,corr} = \sum_i N_i^{Data,R} \frac{N_{ij}^S}{N_i^{Tot,R}} \tag{3.7}$$

where N denotes the number of events in a particular bin at ND, i and j denote the i^{th} and j^{th} bin of the reconstructed energy and true energy spectra respectively. T and R denote True and Reconstructed labels, while S and Tot denote the signal component and the total simulated ND prediction respectively. N_{ij}^S denotes the element of the reco-true mixing matrix and the sum is taken over all reconstructed bins. A similar equation holds for the background components as well. Once this is done, the true space spectrum is multiplied with the predicted Far/Near ratio assuming a set of oscillation parameters and then re-folded and re-smeared back into the reconstructed energy space to obtain a constrained FD prediction :

$$F_i^{S,R,pred}(\vec{\theta}) = \sum_j N_j^{S,T,corr} \frac{F_{ij}^{S,unosc}}{N_j^{S,T}} P_j^{osc}(\vec{\theta}) \tag{3.8}$$

where $F_{ij}^{S,unosc}$ denotes the reco-true matrix at the Far Detector assuming no oscillations and the prediction is derived for some set of oscillation parameters, $\vec{\theta}$. In the disappearance channel, the wrong-sign contamination is treated as signal while in the appearance channel, the contamination is treated as background, since the respective oscillation probabilities are either identical or different for the two channels respectively. A similar equation holds for the WS backgrounds in the appearance channel.

Systematic uncertainties in the simulation are handled by varying the simulated spectra appropriately as defined by the systematic (both in the ND and FD) and carrying it through this chain individually. Thus, since one is comparing the varying ND spectra to the same data spectrum, the variation at FD is constrained. This method works especially well for systematics that are correlated in effect in both the detectors since there's a cancellation effect and is helped by utilizing two functionally identical detectors where uncorrelated effects are in general small. We will discuss the systematics more thoroughly next chapter but it is important to note that this set-up with two functionally identical detectors already mitigates biases coming from "unknown" unknowns as well as "known" unknowns, since we expect most mismodelling to be correlated across the detectors.

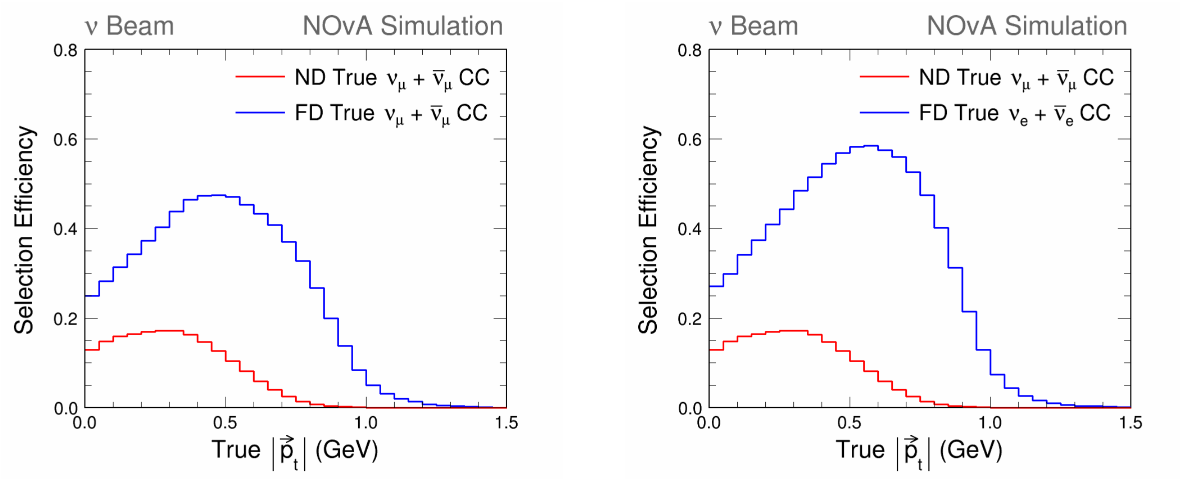


Figure 3.44: Selection efficiency as a function of true transverse lepton momentum at the two detectors for FHC (left) and RHC (right)

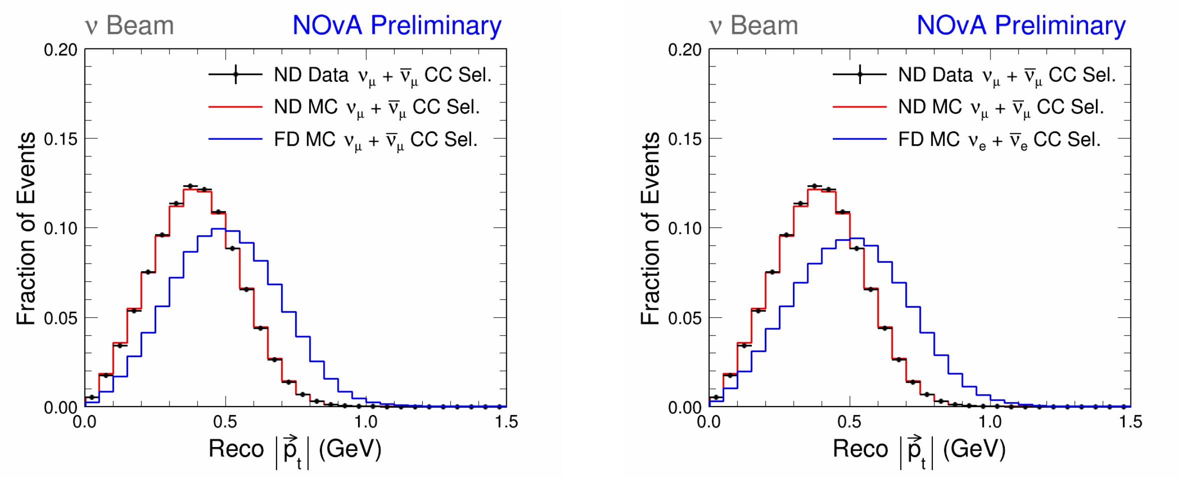


Figure 3.45: Distributions of true transverse lepton momentum for selected events at the two detectors for FHC (left) and RHC (right)

However, there are known biases that need to be corrected for. We already touched upon the fact that the disappearance channel uses 4 samples split into quantiles of reconstructed hadronic energy fraction. Separating these samples correct for differences in energy resolution for given neutrino energies, especially in the dip region between $(1, 2)$ GeV. Each of these samples at the FD have an analogue at the ND by just requiring the same ranges of hadronic energy fraction at the ND. Therefore, the signal prediction in the disappearance channel uses separate extrapolations for each of the 4 hadronic energy fraction quantiles. Thus, we are able to constrain the FD samples by also taking into account different energy resolutions in different interaction phase-space and therefore able to explain the FD prediction more comprehensively using ND data. This also has an additional reduction in systematic uncertainties coming from the cross-section model since the quantiles act as a good separator for the different interaction types. This is done only for the disappearance channel since the splitting into hadronic energy fraction quantiles are self-consistent here. For the appearance channel, it is harder to match the hadronic energy fraction for a ν_μ-CC event at ND to a ν_e-CC event at FD. Therefore, for the appearance channel, the combined ν_μ-CC spectra is used in the extrapolation.

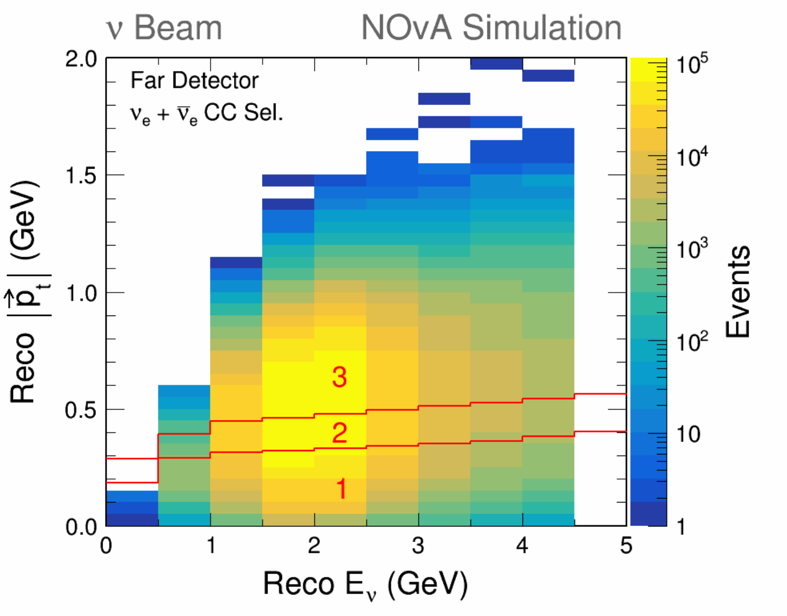


Figure 3.46: Quantile boundaries for reconstructed transverse lepton momentum in the appearance channel as a function of the reconstructed neutrino energy at the Far Detector

Another source of bias is the acceptance between the two detectors, which while being functionally identical, are nevertheless of different sizes and therefore allow different ranges of interaction phase-space for events that are fully contained within each detector, as can be seen in Figs. 3.44 and 3.45. This is evident when looking at the reconstructed transverse momentum of the lepton (p_t) in both the detectors. For a long muon track especially, the containment criteria will select for a tighter p_t spectra in the Near Detector compared to the Far Detector, since the muon can still come off at larger angles while being contained at the Far Detector. There is also an effect when extrapolating the ND ν_μ-CC sample to the FD ν_e-CC sample for the appearance channel, where the electrons at FD are almost always contained. To take this into account, the ND spectra are split into 3 quantiles of p_t and extrapolated separately to the FD, where they're summed to give the FD prediction. This is done for both the disappearance and the appearance channels. For the disappearance channel, the p_t splitting is done for each hadronic energy fraction quantile, resulting in the extrapolation of 12 samples in total at the ND to predict 4 samples at the FD. For the appearance channel, the number of samples extrapolated are 3, as shown in Fig. 3.46. As a result, one can mitigate biases related to the acceptance by constraining FD events at a particular region of phase-space using ND data at a similar phase-space, defined by p_t [122].

3.4.4 Estimation of Beam Backgrounds

The beam backgrounds in the appearance channel consist of misidentified ν_μ-CC and NC events and the ν_e-CC contamination in the beam. Each of these are estimated from ND data using the ν_e-CC selections described above. Here the ND data is designed to correct for the relative composition of each of these components. This is done using separate methodologies for the FHC and RHC beam modes, since the WS contamination in the RHC beam mode is much higher and there are a lot more components to constrain [123].

For FHC, the beam ν_e-CC sample is first constrained using 2 high statistics ND ν_μ-CC samples, a fully contained sample and an uncontained sample, as seen in Fig. 3.47. This is referred to as the Beam Electron Neutrino Decomposition (BEN). The contained sample is

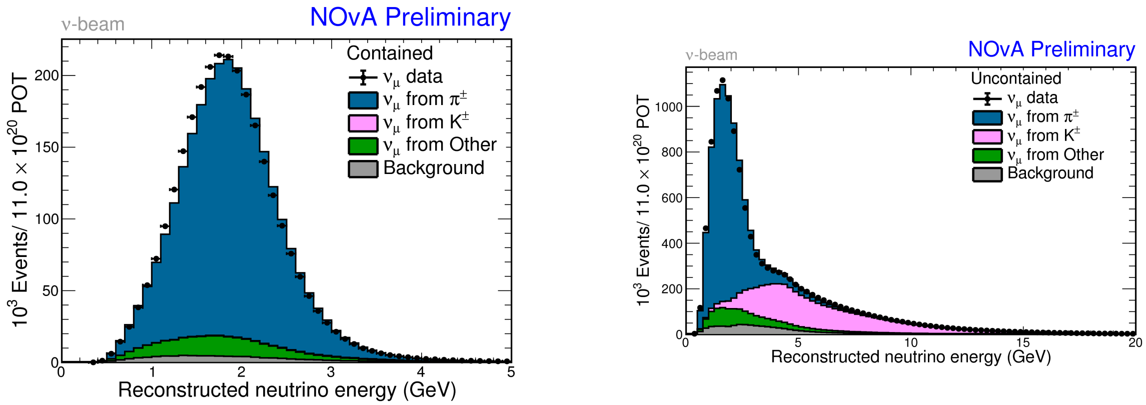


Figure 3.47: The ν_μ-CC contained (left) and uncontained (right) samples used for the beam ν_e-CC constraint at the Near Detector showing relative contributions of the parent mesons

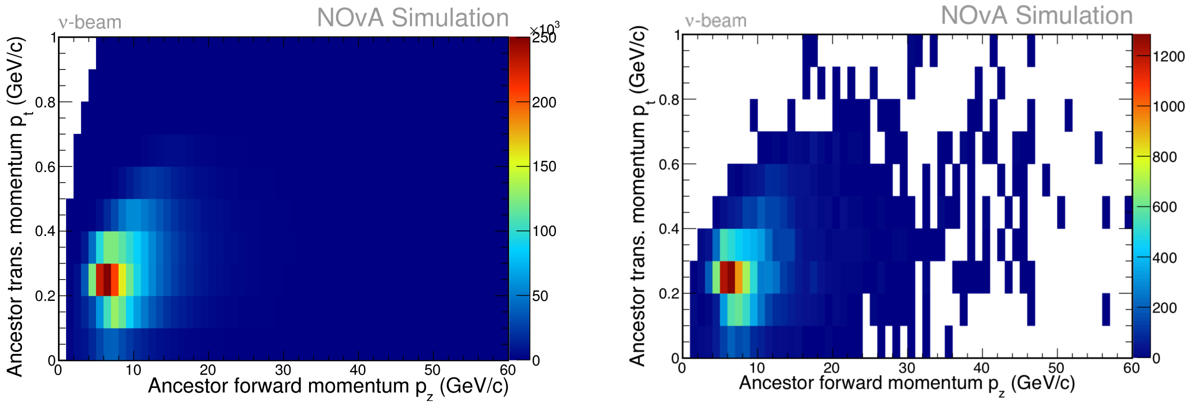


Figure 3.48: 2D phase-space distributions of the parent pion in terms of its (p_t, p_z) shown for ν_μ-CC events in the contained sample (left) and the selected ν_e-CC events at the ND in the analysis (right)

selected using similar criteria as the ν_μ-CC sample for signal extrapolation. The uncontained sample is selected by reversing these containment cuts. The contained sample is dominated by neutrinos from $\pi^\pm$ while the uncontained sample has a larger neutrino energy range and is therefore dominated by $K^\pm$. The ND data in the contained sample is used to derive constraints on the $\pi^\pm$ contribution in $p_t - p_z$ space (shown in Fig. 3.48) , denoting the transverse and longitudinal momenta of the parent $\pi^\pm$. These weights are then transferred over to the $\pi^\pm$ contribution in the uncontained sample and the $K^\pm$ normalization is then fixed to match the ND data in this sample, in the energy range where it is dominant. These weights are then individually used to fix the ND beam ν_e-CC component in the ND selection depending on the individual parents that produced those ν_e-CC events. The ND beam ν_e-CC events coming from $\pi^{\pm}$ are reweighted in $p_t - p_z$ space, while the events coming from kaons are modified by a normalization. The kaon weights are important here since there is quite a significant contribution of the beam ν_e-CCs coming from kaons, even at analysis energy ranges of (0.4.5) GeV. This is mostly in the low CVN score bins. BEN is therefore designed to measure the amount of beam ν_e-CCs more precisely using information from their parents. The uncertainties in the ν_μ-CC/ν_e-CC cross-section ratio are handled by propagating them through the decomposition procedure separately.

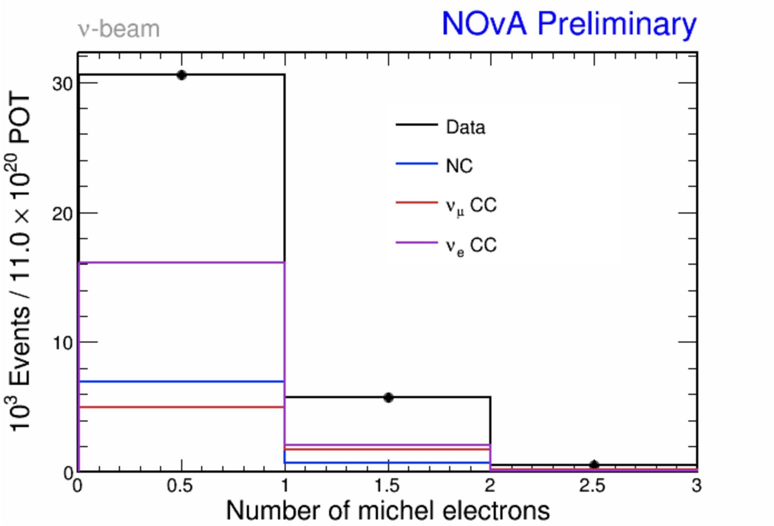


Figure 3.49: The spectrum of the number of Michel e^-s for various components in the ν_e-CC selection at the Near Detector

Once the beam ν_e-CC component is constrained, the other components, ν_μ-CC and NC in the ν_e-CC sample, are constrained by looking at the reconstructed number of Michel electrons in this sample, as shown in Fig. 3.49. The idea is that ν_μ-CCs on average have one extra Michel electron than the NC events, where the Michel electrons in NC come from π decay. Therefore, this observable is a good separator for the two components. The Michel electrons are reconstructed by a separate reconstruction algorithm, called MEFinder, which acts on clusters tagged as noise by the original slicer. A separate PID (MID) is developed to tag some of these clusters as Michel like based on the Δt relative to its parent slice and the energy deposits in the Michel cluster. Based on this PID, the spectra for the number of Michel electrons per event in the ND ν_e-CC sample is drawn in each of the analysis bins in energy × PID space. This is only done for bins where the ν_μ-CC and NC contribution is significant, which is mostly in the Low CVN score region. A likelihood fit to the number of Michel electrons is then done keeping the beam ν_e-CC contribution fixed based on the BEN constraint and modifying the normalizations of the two other components to match the ND data. Thus one obtains a data-driven way to decompose the ND data into the 3 components and any residual data/MC difference is split proportionally between all the components, including small WS contaminations.

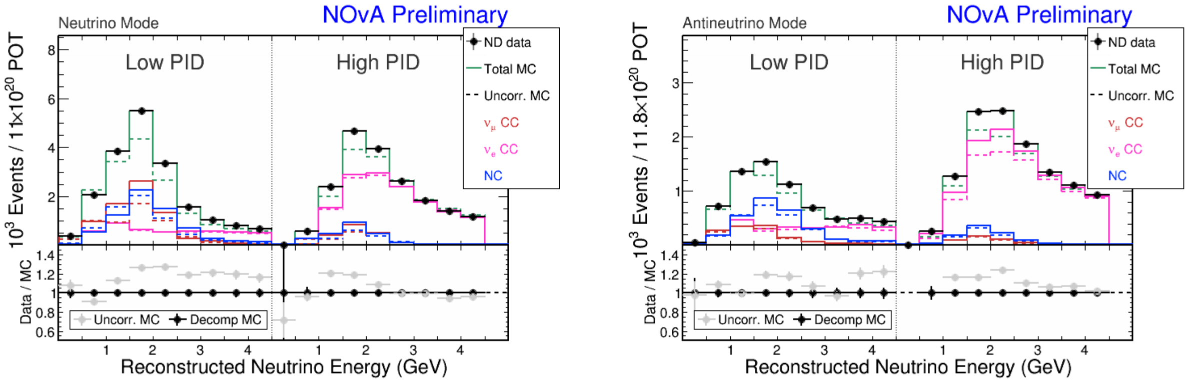


Figure 3.50: The uncorrected and corrected distributions for the ν_e-CC sample at the Near Detector for FHC (left) and RHC (right)

For RHC, since the degrees of freedom are larger and the WS contribution is ∼ 30%, these methods don't work. As a result, the data/MC difference is just split proportionally between all the components, which is called proportional decomposition. Fig. 3.50 illustrates this procedure for both FHC and RHC.

These components are then individually extrapolated to the Far Detector, except the extrapolation doesn't proceed by unfolding into true energy space. This is because the extrapolation of these backgrounds are designed to correct for their relative composition and the ν_e-CC estimator is not designed to work for misidentified components such as ν_μ-CC and NC. Moreover, since these components are expected to oscillate very little, one can be comfortable about making the extrapolated prediction as :

$$F_i^{B,R,pred}(\vec{\theta}) = \sum_j N_i^{B,R,corr} \frac{F_{ij}^{B,unosc}}{N_i^{B,R}} P_j^{osc}(\vec{\theta}) \tag{3.9}$$

where $N_i^{B,R,corr}$ denotes the decomposed result of the particular background component for the i^{th} analysis bin and the sum is taken over bins in true energy space using the FD reco-true matrix. For the peripheral bin in the FD disappearance channel, the constraints from the high CVN bin at ND are directly used since there is no other analogue of the peripheral events at ND. This entire procedure is also done separately in the 3 p_t quantiles defined before of the lepton transverse momentum and summed up to obtain the Far Detector prediction. The p_t quantile splitting is not done for the peripheral bin at FD since the p_t reconstruction is unreliable in this region at the edges of the detector.

3.5 Far Detector Prediction

Once all the neutrino components are extrapolated to the Far Detector, the remaining aspect is the cosmic backgrounds. The cosmic backgrounds in the NuMI trigger within the beam window are estimated using the NuMI timing sideband as mentioned before. These cosmics in the sideband are rescaled to the beam window livetime of $10\,\mu$s per spill. The statistical uncertainty is taken into account as a separate systematic. The shape of the cosmic distribution cannot be predicted from this sideband as there aren't enough statistics. Therefore, the shape is predicted from the minimum-bias 10 Hz cosmic trigger data stream by reweighting the NuMI sideband spectra shape to the cosmic trigger spectra shape. The final predicted events and their respective components are shown in Figs. 3.51, 3.52 and 3.53. The predicted event counts are shown in Tables 3.5 and 3.6.

This prediction is fit to the NuMI data using the traditional Poisson log-likelihood for binned

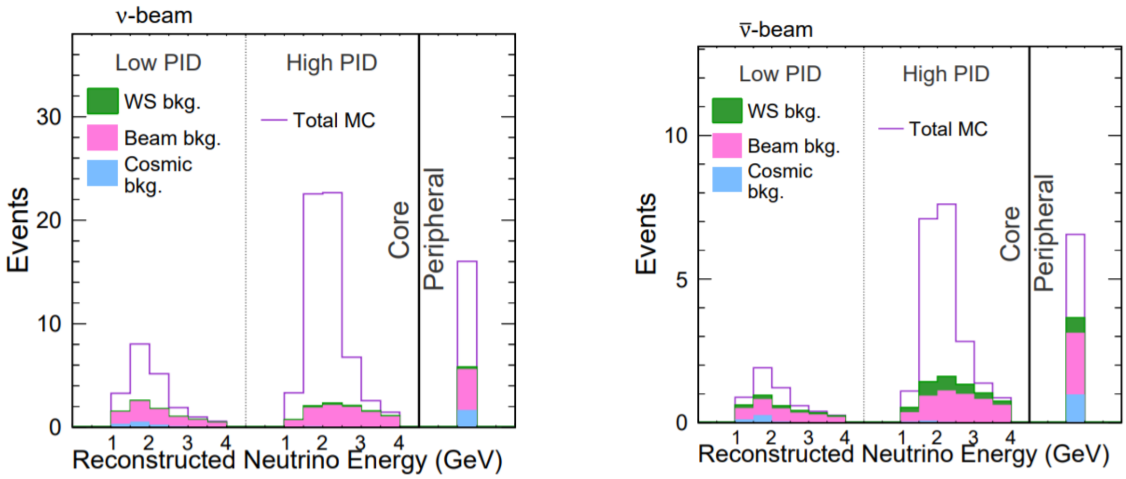


Figure 3.51: Predicted FD spectra for the appearance channel for FHC (left) and RHC (right) using best fit oscillation parameters obtained from a previous analysis

spectra [124]. This is given as :

$$-2\ln\mathcal{L}(\vec{\theta},\vec{\eta}) = -2\sum_{i=1}^{N}\left[v_i(\vec{\theta},\vec{\eta}) - o_i + o_i\ln\frac{o_i}{v_i(\vec{\theta},\vec{\eta})}\right] + \sum_{k=1}^{M}\eta_k^2 \tag{3.10}$$

where v_i denotes the predicted number of events in each analysis bin as a function of the oscillation parameters $\vec{\theta}$ and nuisance parameters $\vec{\eta}$. o_i is the observed number of events in that bin. The nuisance parameters, $\vec{\eta}$ come from systematic uncertainties in the prediction, derived as 1σ, 2σ etc deviations in the central value prediction. The last term denotes the penalty given to favoring a non-zero systematic variation for M set of systematic uncertainties. We will describe how the nuisance parameters are derived in the following chapter.

The best-fit oscillation parameters are derived by minimizing this equation with respect to these parameters and is done using a MIGRAD based algorithm [125]. The likelihood takes into account the Poisson deviations of the prediction from data in each analysis bin (given by the ln term) as well as an overall Poisson fluctuation on the total number of predicted events.

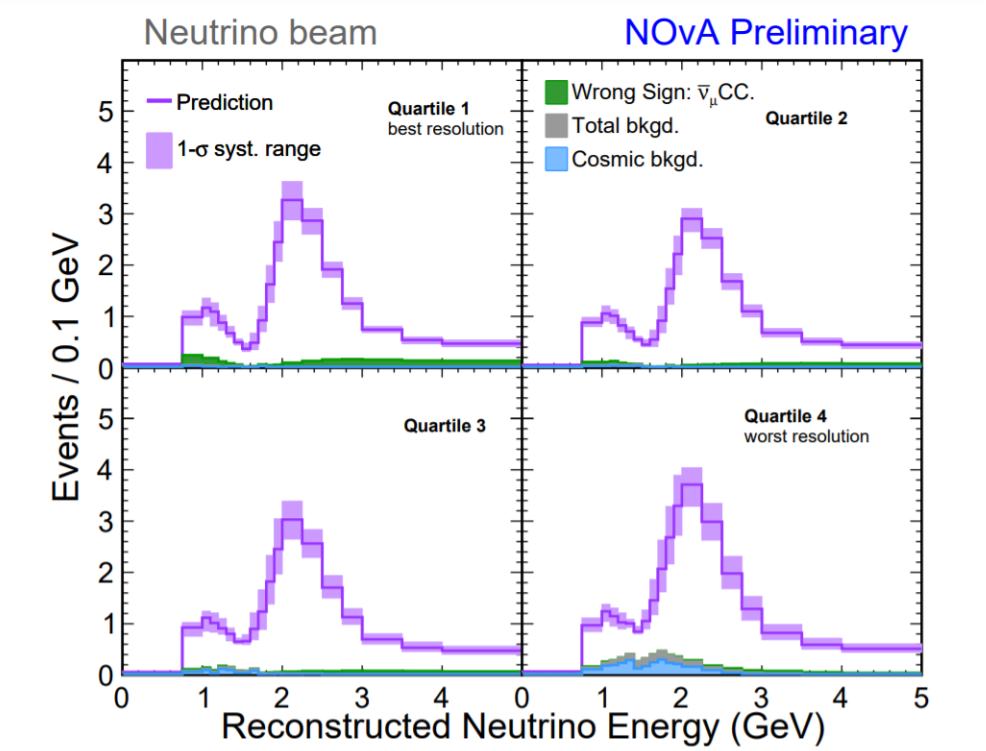


Figure 3.52: Predicted FD spectra for the disappearance channel for FHC using best fit oscillation parameters obtained from a previous analysis

		Low PID	High PID	Peripheral	Total
FHC	ν_e CC Signal	11.73	49.37	10.20	71.31
	$\bar{\nu}_e$ CC WS	0.18	0.65	0.20	1.03
	Beam $\nu_e + \bar{\nu}_e$ CC	2.29	7.59	3.50	13.39
	$\nu_\mu + \bar{\nu}_\mu$ CC	0.57	0.16	0.12	0.85
	$\nu_\tau + \bar{\nu}_\tau$ CC	0.18	0.31	0.12	0.85
	NC	3.66	1.01	0.22	4.88
	Cosmic	1.32	0.22	1.68	3.22
	Total	19.93	59.31	16.04	95.27
RHC	$\bar{\nu}_e$ CC Signal	2.10	14.20	2.90	19.20
	ν_e CC WS	0.48	1.79	0.52	2.80
	Beam $\nu_e + \bar{\nu}_e$ CC	0.88	4.07	1.92	6.86
	$\nu_\mu + \bar{\nu}_\mu$ CC	0.10	0.04	0.05	0.20
	$\nu_\tau + \bar{\nu}_\tau$ CC	0.11	0.19	0.07	0.36
	NC	1.15	0.43	0.10	1.68
	Cosmic	0.42	0.14	0.97	1.53
	Total	5.24	20.86	6.54	32.64

Table 3.5: Predicted event rates for various components and PID bins in the appearance channel using best fit oscillation parameters obtained from a previous analysis

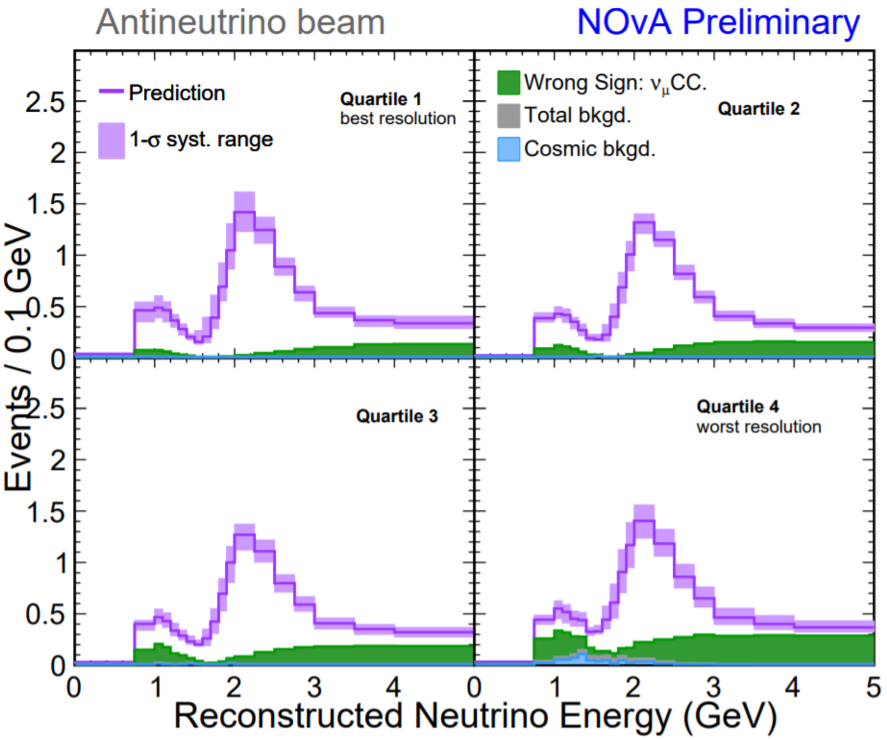


Figure 3.53: Predicted FD spectra for the disappearance channel for RHC using best fit oscillation parameters obtained from a previous analysis

		Q1	Q2	Q3	Q4	Total
FHC	Signal: $\nu_\mu \to \nu_\mu$	41.60	39.96	42.55	48.99	173.10
	$\bar{\nu}_\mu \to \bar{\nu}_\mu$	5.31	2.72	2.27	1.26	11.56
	Total Signal	46.91	42.67	44.82	50.25	184.66
	NC	0.10	0.16	0.40	1.81	2.47
	Cosmic	0.36	0.36	0.89	3.50	5.09
	Other	0.14	0.14	0.16	0.33	0.77
	Total	47.52	43.34	46.27	55.88	193.00
RHC	Signal: $\bar{\nu}_\mu \to \bar{\nu}_\mu$	19.98	16.78	15.60	14.11	66.47
	$\nu_\mu \to \nu_\mu$	3.50	4.74	6.09	9.76	24.08
	Total Signal	23.48	21.52	21.69	23.87	90.56
	NC	0.02	0.04	0.09	0.63	0.78
	Cosmic	0.06	0.03	0.11	0.73	0.94
	Other	0.06	0.08	0.09	0.12	0.35
	Total	23.62	21.66	21.98	25.36	92.62

Table 3.6: Predicted event rates for various components and quantiles in the disappearance channel using best fit oscillation parameters obtained from a previous analysis

Chapter 4

Systematic Uncertainties

The systematic variations in the simulated model are designed to capture many different types of uncertainties coming from the various components that go into making a prediction at the FD, for example in the flux model, the cross-section model, calibration etc. These enter the oscillation fit as penalty terms, where the penalty is calculated for the amount of σ deviation favored in the oscillation fit, i.e if the fit favors a 1.2σ deviation from the central value prediction then the penalty to $-2\ln\mathcal{L}$ is given by $1.2^2 = 1.44$.

However, given this parameterisation it is often hard to define what constitutes a numerical σ of deviation. Many uncertainties in the simulated model which we will discuss in more detail are heavily non-Gaussian and can even be correlated to each other in their effect on the binned spectra. Some of the uncertainties are even purely "turn-on" in the sense that only a positive σ deviation is defined and it turns off at negative σ.

As a result, the parameterisation is derived by defining the variation in the prediction at specific σ values, for example : $-2, -1, 0, 1, 2$ σ and interpolating the deviations in each analysis bin using a cubic spline-based method, as illustrated in Fig. 4.1. This is sufficiently general for all the systematics considered in the analysis, including those with non-Gaussian

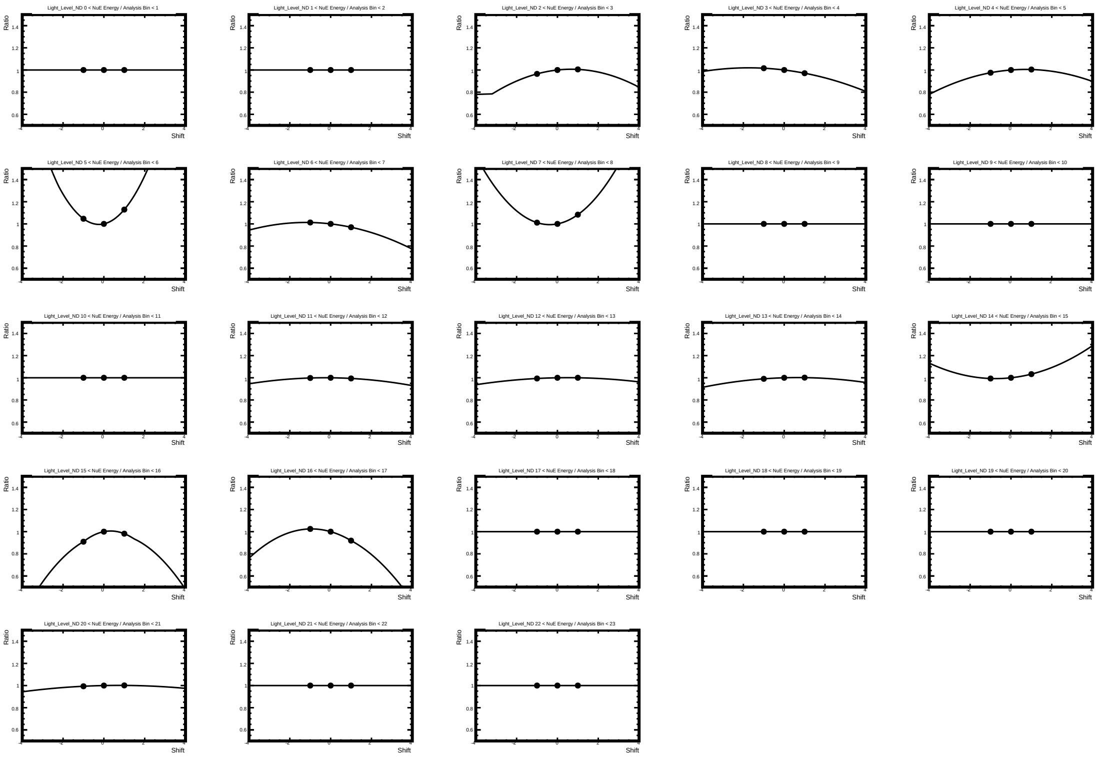


Figure 4.1: Illustration of the interpolation procedure for each analysis bin in the appearance channel for an example light model systematic at the Near Detector. The curves represent the fractional shift in the prediction based on given shifts at (-1, 0, 1) σ

uncertainties. For example, for a purely "turn-on" systematic, the $-2, -1$ σ deviations don't change the central value prediction at all, while the $1, 2$ σ deviations change the prediction as defined in the systematic uncertainty taken in the model. Therefore, the interpolation parameterizes the overall deviation for any given σ by fitting a flat line at 0 across negative σ values and an appropriate curve for positive σ values through the "bridge-points", 1σ and 2σ. This is also done at each analysis bin separately. The variations at the specific σ values are taken by modifying the central value prediction at ND appropriately and extrapolating through the entire prediction chain, including for the beam backgrounds, and arriving at a modified and constrained prediction at FD for those specific σ values. The fit then favors a particular amount of σ deviation given by η to match the FD data spectra after taking into account the penalty given by η^2. Defining the variations in the model for given σ values is therefore the primary function of the systematic analysis of the oscillation parameter measurement.

The functional form of the log-likelihood assumes a separation in $\vec{\theta}$-$\vec{\eta}$ space which is necessary since there are > 50 systematic uncertainties considered in the analysis and navigating such a high-dimensional likelihood is not feasible. This procedure can be justified by noting that the systematic uncertainties are often much smaller than oscillation effects even in regions where they mimic the shape of the oscillation effect. This is why the ND is so important in oscillation parameter measurements at long baselines since otherwise the measurements would be completely drowned out by systematic uncertainties, which for example can be as large as $\sim 9\%$ in the case of the flux model at 2 GeV but because of the ND data constraints and helped by having two functionally identical detectors, it is mostly cancelled out.

The log-likelihood form also requires that individual systematic uncertainties are not correlated to each other. This is somewhat justifiable for uncertainties coming from completely separate parts of the simulation for example the flux vs cross-section. However, when dealing with uncertainties from just the flux model for example, one must take into account these correlations that could crop up within the errors in the hadron production cross-sections for different neutrino energies. This is done using a Principal Component Analysis (PCA) to decompose an ensemble of systematic possibilities in the model into individual uncorrelated pieces.

4.1 Flux Uncertainties

4.1.1 Hadron Production

As mentioned before in the flux simulation, NOvA uses a Geant-based PPFX model to predict the flux at the Near Detector and the Far Detector. The cross-sections for the hadron production at the target, mainly π and K are constrained using external measurements at NA49 and MIPP. The relevant uncertainties are derived from the errors in the external measurements themselves and include errors for producing the individual hadrons at a given neutrino energy. These errors are evaluated in $x_F - p_t$ space of the hadron parent. Based on these errors, the PPFX model provides 100 separate ensembles of flux predictions at the two detectors for each flux component in two beam modes, $(\nu_\mu, \nu_e, \bar{\nu}_\mu, \bar{\nu}_e) \times$ (FHC, RHC). These ensembles are derived by reweighting the central value flux prediction coming from PPFX for each event at a given neutrino energy. The 68% envelope of this ensemble results in $\sim 9\%$ deviation from the central value flux prediction at 2 GeV. It is evident however from Fig. 4.2, that these ensembles are highly correlated across neutrino energies. In particular the correlation matrix is highly non-diagonal with two major pieces, one coming from the π at lower neutrino energies and one from the K at higher neutrino energies [74].

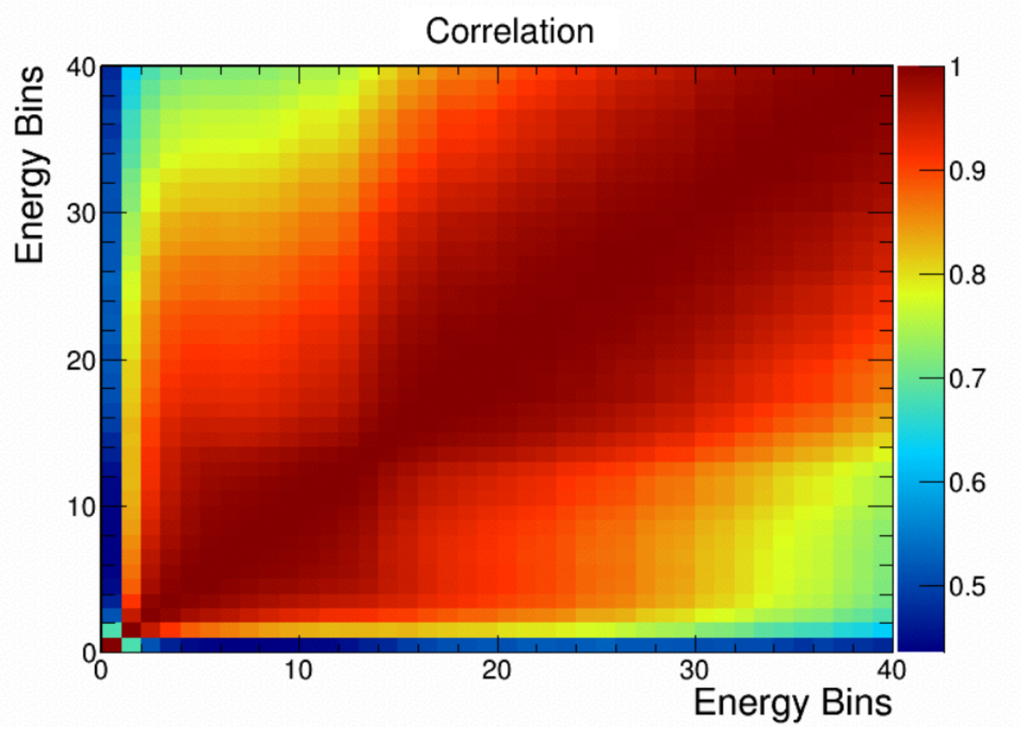


Figure 4.2: The correlation matrix based on the hadron production error for the predicted PPFX flux from (0, 10) GeV in bins of 0.25 GeV

4.1.2 Beam Transport

The other major component of the flux uncertainties come from variations in the parameters involved in the transport of the meson beam using the focussing horns, the beam spot size and so on [126]. The uncertainties considered in the analysis are given by :

- Horn positions in X and Y, transverse to the beam : A ±3 mm deviation is considered in each of these directions.
- Position of the graphite target : A ±7 mm deviation along the direction of the beam.
- Horn current (nominal 200 kA) : A ±1% deviation from the normal horn current for each beam mode.
- Size of the beam spot on the target : A ±0.2 mm deviation from nominal.
- Beam position on target in X and Y : A ±1 mm deviation from nominal. This uncertainty is asymmetric in Y since the NOvA detectors are off-axis.
- Thickness of the water cooling layer in the magnetic horns : A ±0.5 mm deviation from nominal.
- Earth's Magnetic Field in the Decay Pipe : A "turn-on" systematic from the nominal case assuming no magnetic field. The magnetic field was measured to be $(0.1, -0.3, 0.07)$ Gauss in X, Y, Z respectively.

These uncertainties are derived as $\pm 1\sigma$ deviations in the nominal prediction for each of the beam modes and flux components, $(\nu_\mu,\ \nu_e,\ \bar{\nu}_\mu,\ \bar{\nu}_e) \times$ (FHC, RHC).

4.1.3 Principal Component Analysis

The uncertainties on the flux model for the oscillation analysis utilize a PCA-based approach to de-correlate the individual pieces of the flux systematics. In rough terms, this means deriving a covariance matrix in true neutrino energy space by varying all the components of the flux uncertainties, and then diagonalizing it to obtain the de-correlated and orthogonal eigenvectors [127][128].

The covariance matrix is derived from many ensembles of flux predictions in each of the 16 components, (ν_μ, ν_e, $\bar{\nu}_\mu$, $\bar{\nu}_e$) × (FHC, RHC) × (ND, FD). This is done in 40 bins of size 0.25 GeV from $(0, 10)$ GeV of neutrino energies. This range is the relevant neutrino energy range for the oscillation analysis and is larger than the reconstructed neutrino energy range because of NC events whose true energies are in general much higher than their reconstructed energies. There are 2000 ensembles taken in total, where for each of the 100 hadron production ensembles, 20 different ensembles are derived by varying the beam transport parameters randomly using a Gaussian distribution with width defined by the $\pm 1\sigma$ variation in those parameters.

Before deriving the covariance matrix, the hadron production ensembles are smoothed, as shown in Fig. 4.3, since the 100 ensembles don't sample the $x_F - p_t$ variation of the hadron parent sufficiently smoothly. This is a prominent effect especially at FD where the overall flux is small and therefore the smoothening removes any statistical fluctuations in the ensembles.

Due to the fact that many of the beam transport uncertainties are asymmetric around the central value flux prediction, the covariance matrix is derived by taking the variation around the central value rather than the mean of the ensemble. The covariance matrix is also defined in the (ND, FD/ND)-basis rather than the (ND, FD)-basis since this is the relevant basis

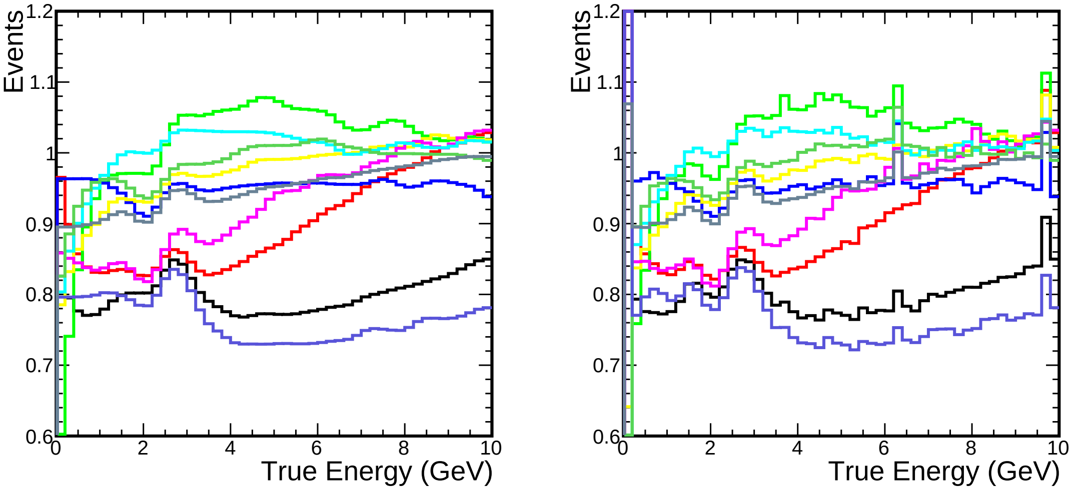


Figure 4.3: Different statistical ensembles of the PPFX weights for the predicted ν_μ-CC flux at the Far Detector based on the Hadron production error. Shown are smoothed (left) and unsmoothed (right) versions

for the extrapolated uncertainty. This is given as :

$$C = \left(\frac{1}{N_u - 1}\right) \sum_{u=1}^{N_u} (\vec{X}_u - \vec{X}_{cv})(\vec{X}_u - \vec{X}_{cv})^T \tag{4.1}$$

where $\vec{X}_u$ and $\vec{X}_{cv}$ represent the binned histograms for a particular ensemble, u and the central value respectively. $\vec{X}$ is therefore of size 40×16 given by 40 bins of neutrino energy $\times$ 4 flux components, (ν_μ, ν_e, $\bar{\nu}_\mu$, $\bar{\nu}_e$), $\times$ 2 beam modes, (FHC, RHC) $\times$ the basis, (ND, FD/ND). In addition, the ensembles are normalized such that the central value spectra is of area 1 for each component, beam mode and detector combination. Note that the ensembles are not of area 1 themselves, the ratio of areas of the ensemble vs the central value are preserved. The normalization to area 1 is done so each component gets equal importance in the covariance matrix. However, the ND spectra are further downweighted by a factor of 10^{-6} to reduce their importance in the covariance matrix. This is done to ensure the diagonalisation concentrates on the block matrix defined over FD/ND space rather than the entire (ND, FD/ND) space. The ND part is still carried through the diagonalisation to ensure a complete basis.

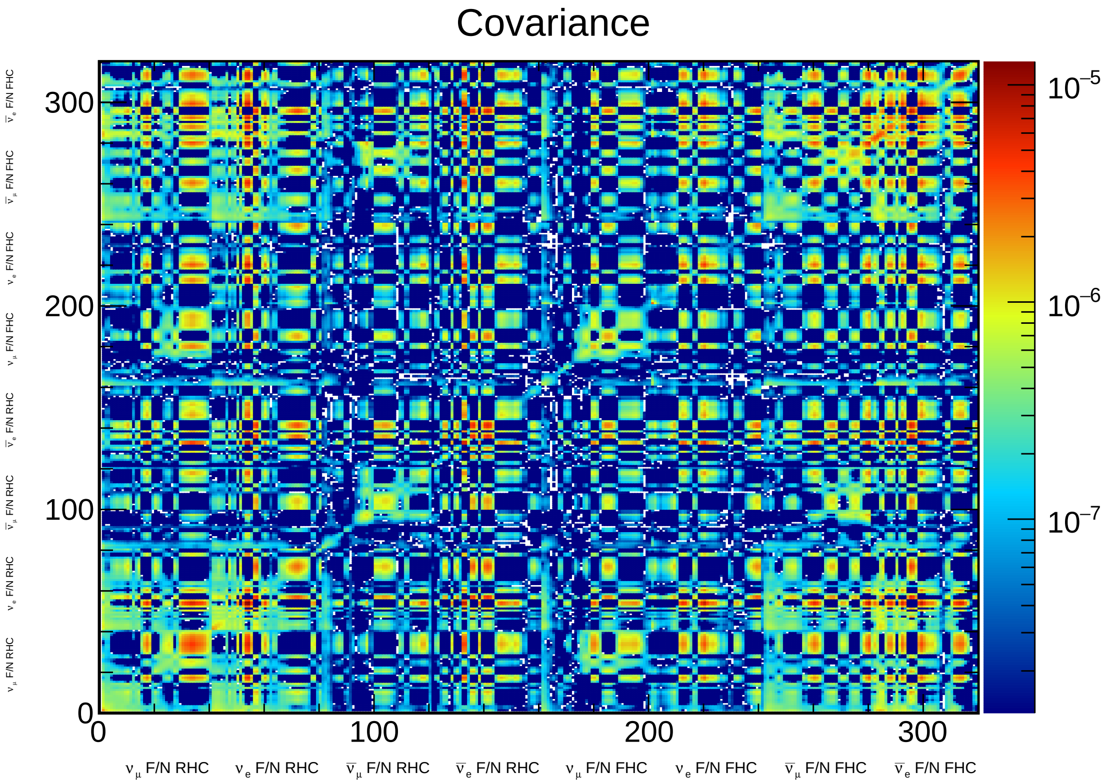


Figure 4.4: The covariance matrix representing the full systematic error in the flux model including hadron production and beam transport. Shown in the F/N basis for different neutrino components in FHC and RHC

The PCA method involves diagonalizing the covariance matrix, shown in Fig. 4.4 into its orthogonal eigenvector pieces. These individual pieces then parameterize the individual flux uncertainties in the fit. Since the covariance matrix is real and symmetric, they are always diagonalizable and moreover, their eigenvectors are orthonormal with positive eigenvalues, i.e they can be put into the form :

$$C = P\Lambda P^T \tag{4.2}$$

where P denotes the matrix of eigenvectors $[v_1, v_2, v_3, ..]$ and $\Lambda = diag(\lambda_1, \lambda_2, ..)$ denotes the diagonal matrix of eigenvalues. The pull terms for the covariance matrix in the log-likelihood

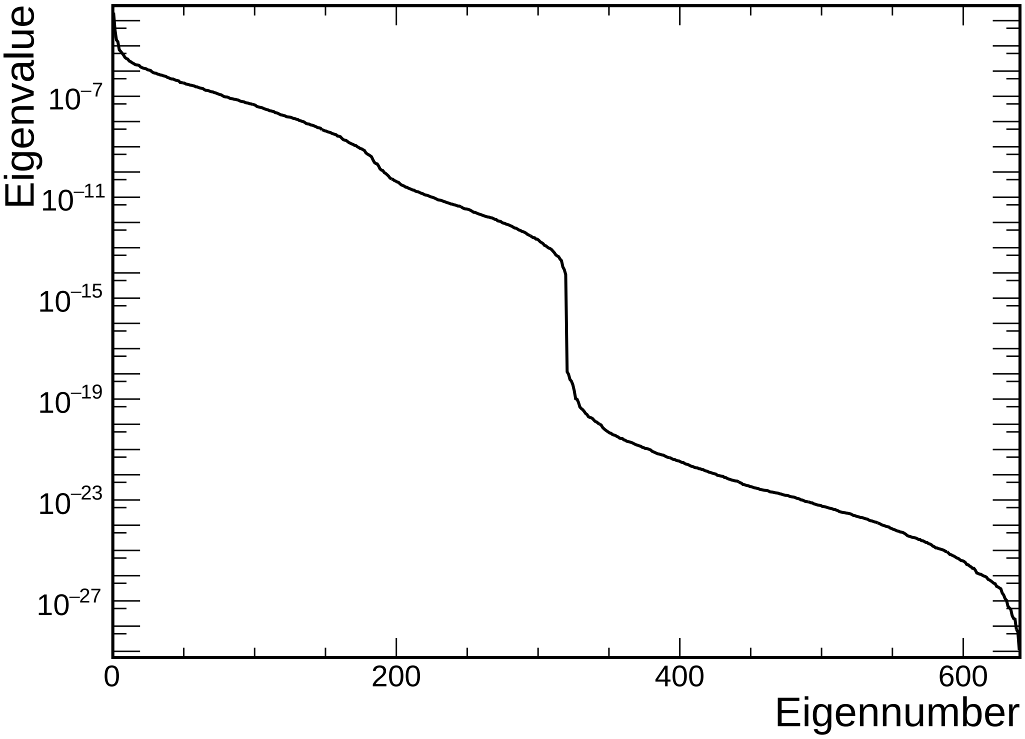


Figure 4.5: The eigenvalue profile of the flux covariance matrix in the F/N basis

fit would be given by :

$$\chi^2 = (\vec{N}_{syst} - \vec{N}_{cv})^T C^{-1} (\vec{N}_{syst} - \vec{N}_{cv}) = (P^{-1}\delta\vec{N})^T \Lambda^{-1} (P^{-1}\delta\vec{N}) \tag{4.3}$$

By taking $\delta\vec{N} = \sqrt{\lambda_i} v_i$ one gets a total pull, $\chi^2 = 1$ for any index i, since the eigenvectors are orthonormal, i.e

$$v_i^T v_j = \delta_{ij} \tag{4.4}$$

Therefore, the individual uncorrelated ±1σ systematic uncertainties in the flux model are given as $\pm\sqrt{\lambda_i} v_i$, which are called the Principal Components. The systematic variation is then derived by adding to this the central value prediction and removing the rescaling done to the area of the spectra. As can be seen from Fig. 4.5, the strength of the PCA lies in the

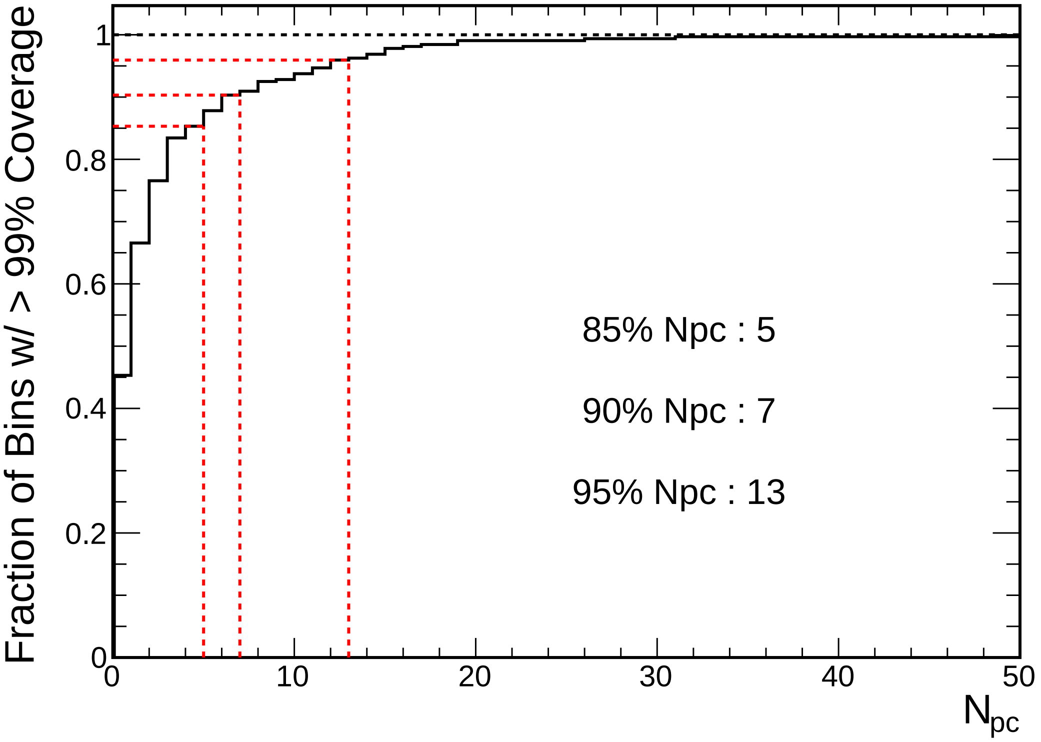


Figure 4.6: Fraction of bins where the summed error in quadrature from the first N Principle components cover the flux error to $> 99\%$

fact that not all indices i have to be used in the analysis since only a few eigenvectors have a significant strength given by their eigenvalues, λ_i. This is akin to fitting a N-dimensional hyper-ellipse to the variate data and ignoring the eigenvectors (defined by the axes directions) which are tiny in magnitude.

The number of Principal Components (PCs) are then chosen by first ordering them in terms of their eigenvalue and discarding all but a few of them by a variety of coverage checks. The first metric involves comparing a set of N PCs added in quadrature to the 68% envelope of the 2000 ensembles. One can then derive a metric based on the fraction of bins that go $> 99\%$ of the way towards the envelope as a function of the number of PCs considered, N. The coverage is seen to be pretty good with just 5 PCs, with $\sim 85\%$ of the bins having $> 99\%$ coverage, as shown in Fig. 4.6. The remaining bins are concentrated in the low energy

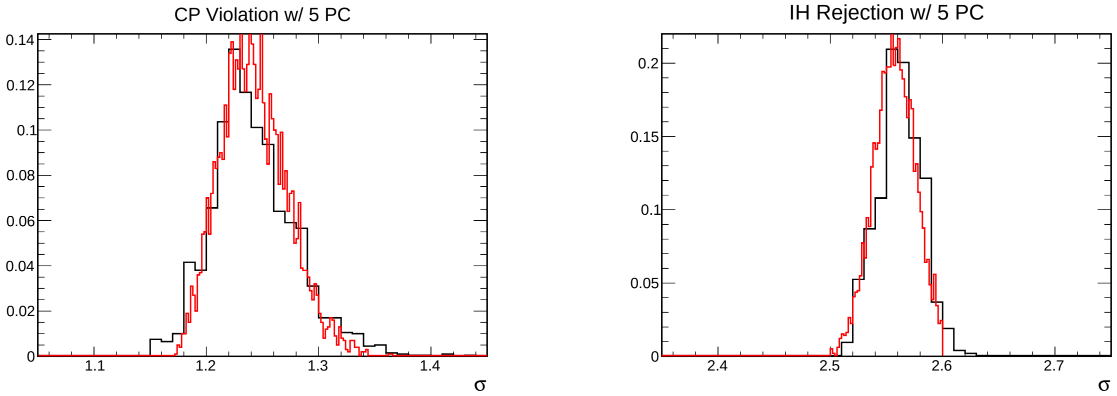


Figure 4.7: Coverage of the overall flux error for the first 5 Principle Components in terms of sensitivities for measuring CP violation or rejecting Inverted Hierarchy based on Asimov data

region, especially for beam ν_e and $\bar{\nu}_e$, which are expected to have a small overall effect on the oscillation measurement. If one only looks at the hadron production error, more than 95% of the bins are captured with $>$ 99% coverage. To cover the remaining portion, the 5 PCs are increased by a factor of 25%.

The choice of 5 PCs is confirmed by evaluating the sensitivities to the important oscillation measurement observables, such as the significance of CP-Violation and Mass Hierarchy Rejection, shown in Fig. 4.7. This is done using a fit to Asimov data which is just the central value prediction and comparing it to systematically varied predictions. The first step involves carrying the 2000 flux ensembles all the way through the analysis chain, including selection, extrapolation and so on and measuring how much these ensembles differ in their measurements with respect to the central value prediction taken at no CP-violation and the Normal Hierarchy at some chosen oscillation parameters. The next is to carry an ensemble derived from the 5 PCs themselves through the analysis chain and doing the same procedure as before. A joint fit to both beam modes and the disappearance and appearance channels is done as in the oscillation analysis. It is evident that the range of sensitivities to the two measurements is captured by the range coming from 5 PCs.

In this way, a first principles based approach can be used to capture the entire error in the flux model while ensuring that the individual pieces of the systematic uncertainty is uncorrelated. The PCA method does hurt interpretability of the individual PCs however, since they're represented in a more abstract form. But intuitively, one can think of them as being a set of Fourier modes of the original uncertainty, with one of the largest components being the normalization uncertainty in the flux and other components having nodes at different points in neutrino energy, i.e the uncertainties go to 0.

4.2 Cross-Section Uncertainties

The uncertainties for the neutrino interaction cross-sections are derived from both the underlying GENIE model and the custom NOvA tune done on top of it. These take the form :

- QE Axial Form Factor : These are derived from the underlying Z-expansion formalism. The knobs provided by GENIE vary the shape and normalization of the axial form factor. The normalization is changed from the nominal by a factor of $-20\%/+15\%$. This is similar to varying m_A, the axial mass in the CCQE Lagrangian derived from the Llewelyn-Smith calculation [129]. The other uncertainties on the Z-expansion include the coefficients of the next 4 terms. These are derived as a function of Q^2 and a PCA technique as described above is done to take into account the correlations in the uncertainties in each of the 4 terms.
- QE RPA : Long range correlations in the nucleus using a local Fermi gas model result in an overall suppression of the CCQE cross-section at low-Q^2 and a slight enhancement at high Q^2. This is taken into account by the GENIE prediction, but the uncertainties are derived from a study done at MINERvA.
- RES Q^2 Suppression : This is motivated by previous NOvA analyses that saw a low-Q^2

Suppression for RES events too, similar to the RPA effect for QE events [130]. The origin of a possible suppression is unclear and the amount of suppression seen in this version of the analysis is much smaller. There are indications that this could arise from lepton mass effects taken into account in the RES model but it is still being studied. Instead, an uncertainty is derived from a similar suppression seen at MINOS and applied to the previous GENIE model v2.12 used in the previous NOvA analysis. This uncertainty is only applied to non-Hydrogen targets in the NOvA detector since this is expected to come from deviations in the nuclear model with multiple nucleons.

- FSI : These are derived by constructing knobs on FSI parameters such as the mean free path of the outgoing particle in the interaction (±30%), the fate of the outgoing particles in terms of its absorption (±40%), possible pion charge exchange (±20%) and quasi-elastic scattering (±30%). The knobs are based on measurements done at T2K/DUET on top of the hN semi-classical cascade model and use the same BDT-based technique for reweighting the nominal model as was done for the central value tune [131].

- DIS Non-Resonant Nπ : These involve 50% normalization uncertainties based on the π multiplicity coming from DIS interactions upto a hadronic mass, W of 3 GeV. Beyond that, the uncertainty decreases linearly upto 5% at W of 5 GeV.

- DIS Formation Zone : These involve uncertainties on the hadronization length of the outgoing partons in the nuclear medium upto ±50%.

- MEC ($q_0, |q|$) dependence : The MEC tune derived using ND data is redone under different assumptions of the base non-MEC model. The non-MEC model is tuned to make it more QE-like or RES-like using uncertainties on the CCQE normalization and shape, the RPA effect, the RES uncertainties on the low-Q^2 suppression and the axial and vector masses, m_A and m_V on the CC RES model. The MEC component is refit after varying these components and the uncertainties for the central value MEC component are taken in this $(q_0, |q|)$ space based on the new fits.

- MEC Nucleon Pair Fraction : The Valencia model assumes the fraction of the struck pair of nucleons, i.e $np/(np + nn)$ $(np/(np + pp))$for the MEC component to be 0.69 (0.68) for neutrinos (antineutrinos). An uncertainty of $-5/15\%$ is taken on this number. This is asymmetric since other MEC models have higher fractions of struck nucleons and the uncertainty is designed to cover that range.

- MEC Absolute Cross-Section : This is done by taking an envelope over different predictions for the MEC cross-section (SuSA and Martini models) using the base model from Valencia. This is parameterized as a function of true neutrino energies.

- ν_μ/ν_e Cross-Section Ratio : These are derived from the base cross-section ratio by taking into account corrections from radiative processes (2%) and possible second-class currents (2% anticorrelated for ν-$\bar{\nu}$).

- Others : These are mainly knobs provided by GENIE itself on various processes such as on the NC/CC cross-sections in the Coherent interaction channel, the NC Elastic scattering parameters, the NC/CC RES axial and vector mass parameters and underlying knobs on the Bodek-Yang and Berger-Sehgal models.

The largest 20 of the above systematics are used directly in the fit, while the rest of the "small" uncertainties are handled using a similar PCA approach as defined above [132]. The PCA is done using the analysis selections and binning and uses the samples, ν_μ-CCQE, ν_μ-CC Non QE, NC in the disappearance selection and ν_e-CCQE, ν_e-CC NonQE, ν_μ-CC and NC in the appearance selection. Similar samples are chosen for anti-neutrinos. The largest 12 PCs are chosen and rescaled by 35% to cover the effect of these small uncertainties. Fig. 4.8 shows the total systematic error taken on the cross-section model.

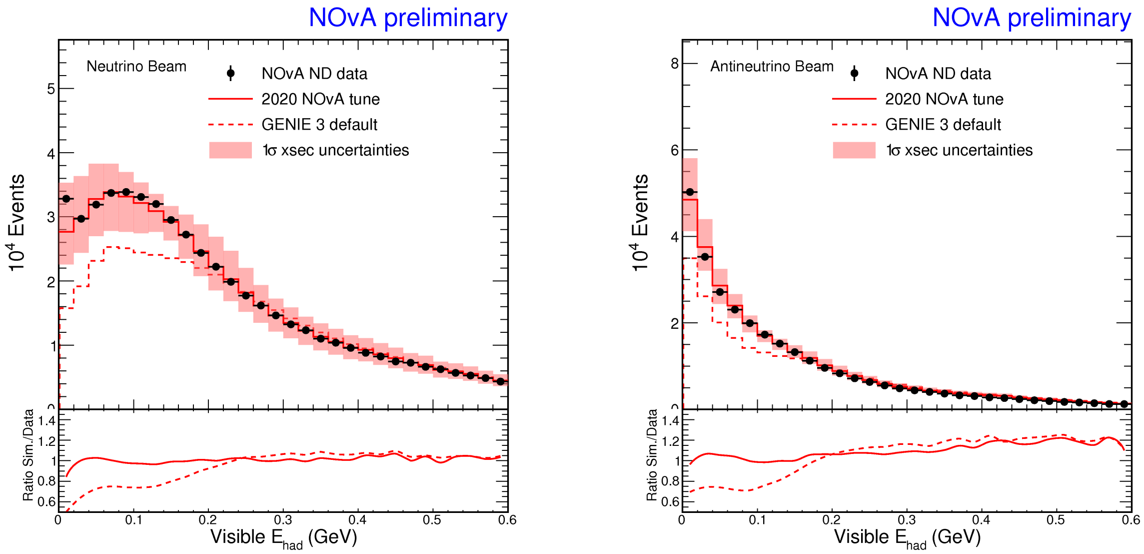


Figure 4.8: The total systematic error in the NOvA cross-section model as a function of visible hadronic energy for FHC (left) and RHC (right)

4.3 Calibration and Detector Response

The uncertainties correspond to the calibration of the energy response and the parameters in the scintillation light model [133]. These are :

- Energy Scale : These come from the absolute calibration done on the stopping muon sample at ND and FD separately. This is motivated by a discrepancy seen in the data/MC dE/dx for protons at ND at the level of $\sim 5\%$. The origin of this discrepancy is not yet known since other standard candle samples such as beam muons, Michel electrons, π^0 reconstructed mass exhibit smaller discrepancies. The uncertainties are taken as both fully correlated for ND and FD and anti-correlated for ND and FD.
- Calibration Shape : A correlated uncertainty across the two detectors is constructed based on the imperfection of the calibration to take into account differences in energy response at the ends of the cells compared to the middle. The calibration is changed to cover the discrepancy between reconstructed and true energies at the edges of the

detector and the difference is folded into the uncertainty. This defines a $+1\sigma$ shift in the calibration, but the shape of the reco/true energies at the edges is reflected around 0 to give the -1σ shift. Overall these are not very well-motivated systematics as their origin is not known but is still used to be conservative.

- Detector Drift : There is an observed attenuation in the reconstructed number of hits as a function of time, presumably due to detector aging effects either from the scintillator or the electronics. This is covered by varying the scintillation model with an overall light output decreasing at a rate of 4.5% per year, offset by the relative calibration done for each different period of data taking. This is a one-sided fully correlated systematic where negative shifts are not allowed.

- Light Level : The light model is modified using an improved tune taking into account problems in the fit from the original light model tune. This is taken as an uncertainty since it was too late to change the original tune. Here the uncertainty is constructed by generating a new sample after varying the light model using the new tune and inverting the calibration built on the previous light tune. The uncertainties come out to $\pm 10\%$ for both X and Y views at ND and $\pm 16\%$ in the X view and $\mp 6\%$ in the Y view for FD. This is therefore uncorrelated across the two detectors.

- Cerenkov Emission : This is an additional uncertainty taken on the observed proton dE/dx data/MC discrepancy where the cause is assumed to be due to the parameters in the scintillation model that describe the amount of Cerenkov photons emitted in a given cell. The light model tune is varied by adjusting these parameters to cover the discrepancy without changing the muon response.

In addition, NOvA is utilizing a dedicated test beam program with a beam of π, p, e, μ towards a small-scale version of the NOvA detectors to study the cause and constrain further some of the above uncertainties. The test beam data is not used in this analysis but will be important in future iterations.

4.4 Neutron Response

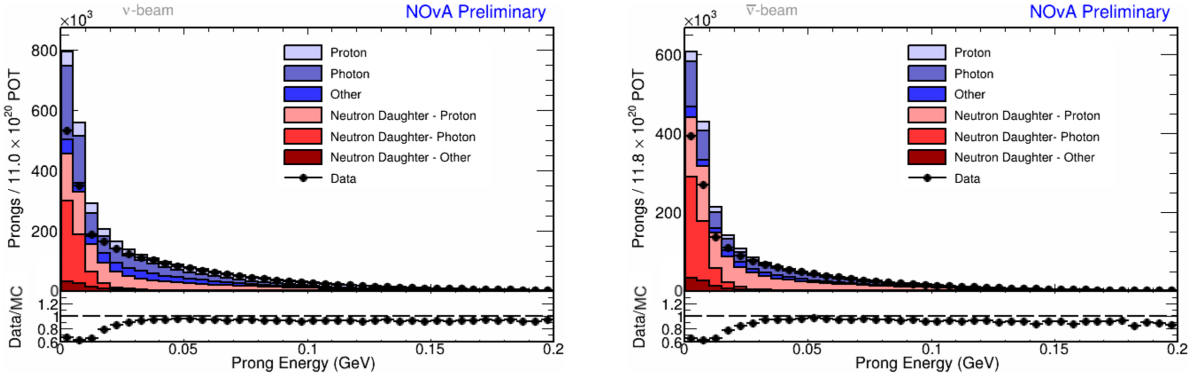


Figure 4.9: Distribution of prong energies for a selected neutron sample at the Near Detector broken down into neutron daughters for FHC (left) and RHC (right)

This uncertainty is motivated by another discrepancy seen at ND for fast neutrons of $>$ 1 MeV energies, as shown in Fig. 4.9. These neutrons either capture producing photons observed in the detector or scatter protons from detector nuclei which then produce light in the detector. The origin of the discrepancy in the simulation, i.e whether it's coming from the generator or the particle transport is not yet known, although indications are that most of it is related to the neutron transport models used in GEANT. Studies are underway to utilize a different neutron model like MCNP, however, in this analysis, an uncertainty is constructed to correct for this discrepancy. This is done by subtracting out a certain amount (47%) of reconstructed visible energy for all FuzzyK prongs coming from a neutron of true energies < 20 MeV. The estimator is then recalculated for such events and the change in the analysis spectra is taken as the 1σ uncertainty [134].

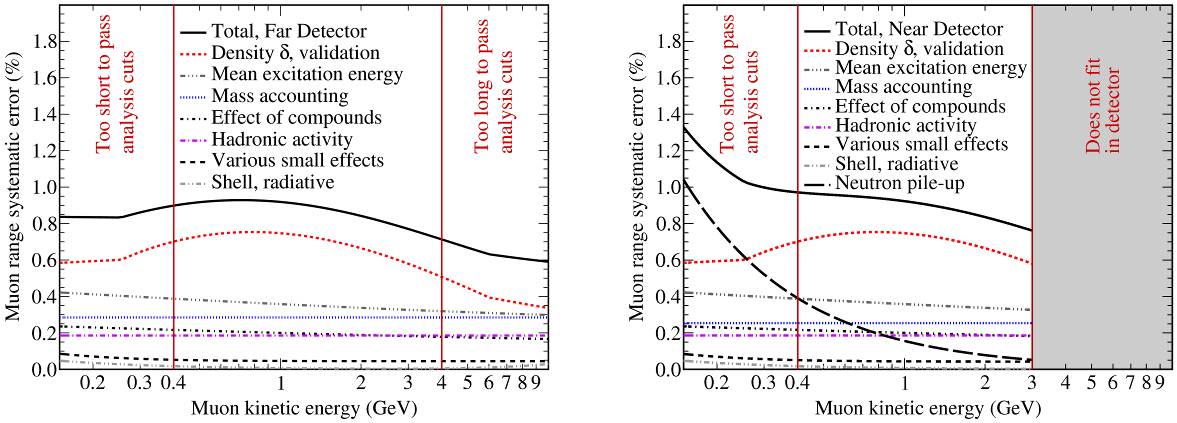


Figure 4.10: Total systematic error on the reconstructed muon length based on various effects coming from GEANT or detector mass accounting at the Far Detector (left) and the Near Detector (right)

4.5 Lepton Reconstruction

This uncertainty is motivated by assumptions used in the reconstruction of the muon track, its length and angle. There are primarily two pieces to the uncertainty. The first piece is an absolute fully-correlated part that comes from uncertainties in the Fermi density effect and GEANT calculation of the muon range compared to a different calculation from Groom et al. These are of the size 0.74% for ND and FD while the error at the muon-catcher is derived to be 0.13%. The second piece is a relative uncorrelated part that mainly comes from calculations of muon range due to mass accounting differences in the two detectors. They come out to 0.15% for FD, 0.13% for ND, 0.48% for the muon catcher. An additional piece coming from neutron pile-up at the ND and its consequent effect on the reconstructed length of the muon track is derived and taken to be 0.46 cm in the active portion of the ND and 1.3 cm in the muon catcher [135]. This is illustrated in Fig. 4.10.

4.6 ND-FD Acceptance Differences

Even though acceptance differences between ND and FD is mitigated by the p_t-based extrapolation, this uncertainty is constructed to bracket the size of "unknown" unknowns and is done only for the signal in the appearance channel. This is motivated by the fact that not only are the kinematic regions for the events different compared to ND vs FD, the corresponding PIDs are also different in the appearance channel and therefore has potential to sculpt the FD distribution differently compared to ND [136].

This uncertainty is derived by reweighting the ND ν_μ-CC data and MC samples used for signal extrapolation to match the phase space of the FD ν_e-CC signal in observables, Q^2, $\cos\theta_{beam}$, p_t^{sum} where p_t^{sum} is the sum of the reconstructed transverse momenta of all prongs in the event (~ 0 for neutrino events reconstructed perfectly). The reweighted samples are then propagated to the FD using the extrapolation procedure and the largest variation between the three observables is chosen to be the uncertainty. This has a marginal effect on the overall signal contribution at FD with a size given by 0.4% for both FHC and RHC.

4.7 ND Pile-up

This uncertainty is motivated by the fact that a significant portion of the data at ND is collecting using a high intensity neutrino beam which can cause an attenuation in the number of selected interaction clusters as a function of intensity, i.e pile-up. To bracket the effect of pile-up, special "overlay" samples are developed where a single MC neutrino is overlaid onto nominal data and MC spills separately and the selection efficiencies are compared. This is similar to the studies done for validating the slicer and the differences turn out to be small when using TDSlicer compared to the older Slicer4D algorithm. The differences in the selection efficiency for MC-on-data vs MC-on-MC samples is taken as an

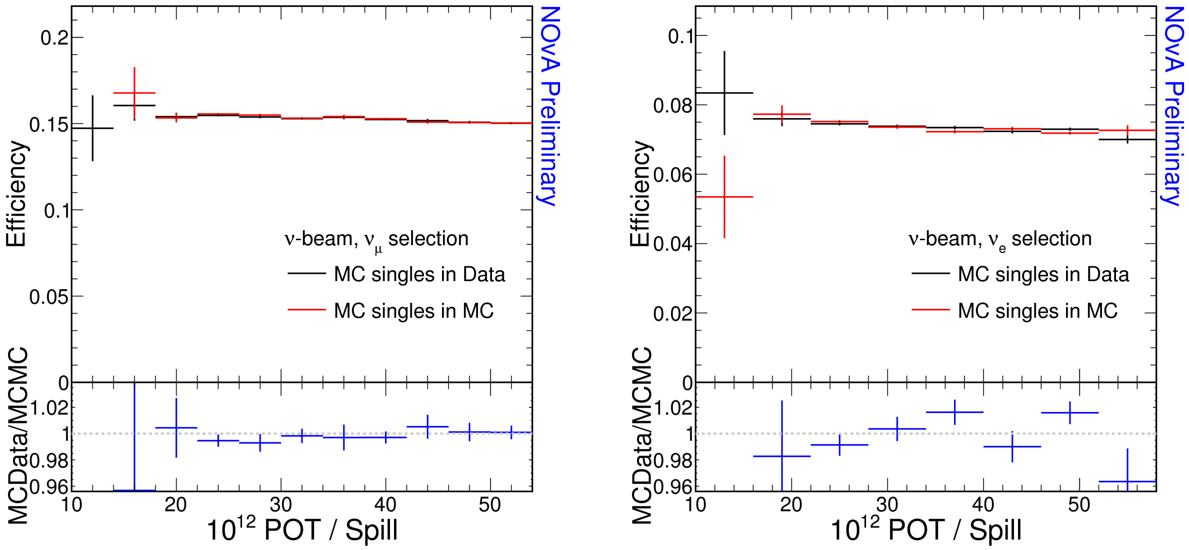


Figure 4.11: Selection efficiencies for the special overlay sample at ND for ν_μ-CC (left) and ν_e-CC (right) as a function of beam intensity. Small differences in the efficiency profile between MC on MC and MC on Data spills is taken as the pile-up systematic

overall normalization uncertainty, to the tune of 0.21% (0.48%) for FHC (RHC) ν_μ ($\bar{\nu}_\mu$) events and 0.41% (0.53%) for FHC (RHC) ν_e ($\bar{\nu}_e$) events. These numbers are calculated after reweighting the MC spills to match the intensity of the data spills, as described in Fig. 4.11 [137].

4.8 Other Systematics

4.8.1 Detector Mass and POT Accounting

A small uncertainty is used on the amount of POT used in the MC when comparing to the observed data. This comes from the fact that fractions of the data at FD are taken in different diblock configurations, especially early on when the full detector wasn't active. Correspondingly, the MC samples the POT for different diblock configurations based on the data POT but could get the overall POT wrong due to statistical uncertainties. Therefore, first an effective mass for each diblock configuration is derived by fitting to the selection efficiencies as a function of number of active contiguous diblocks. Then a weighted sum of POTs is calculated over the different diblock configurations where the weights come from the effective mass. This is done for both data and MC and the residual difference in this weighted sum comes out to $\sim 0.23\%$. To be conservative, a 0.5% normalization uncertainty is chosen taking into account this effect. In addition, it turns out that a small fraction of the early data when the beam spill window was shifted by $+64\,\mu$sec needed to be cut from the analysis due to the way the CNN-based cosmic veto was defined. This comes out to a loss of 0.2% of the total FHC POT. This was added in quadrature to the 0.5% and taken as an overall POT uncertainty of 0.5%.

In addition, differences in mass calculations between the ND and FD by doing a full accounting of the detectors is taken as another source of normalization uncertainty, which comes out to 0.19%.

Further, a bug in GEANT was discovered that caused it to drop particles from the list of primary particles coming from GENIE. This happened predominantly at FD and had no dependence on the interaction phase-space, such as in the inelasticity, y or the invariant hadronic mass, W. The effect comes out to $\sim 0.8\%$ which is taken as a systematic rather than an overall correction since the origin of this is mysterious.

Combined with the pile-up systematics defined above, these are split up into FD-ND correlated and uncorrelated pieces as appropriate and the individual pieces are added in quadrature to derive an overall normalization uncertainty on the analysis [138].

4.8.2 ν_τ Scale

This is derived from uncertainties coming from the cross-section of ν_τ-CC events which make up a tiny background component of the oscillation analysis, especially in the appearance channel [138]. An overall scale uncertainty of 60% is taken motivated by measurements at OPERA.

4.8.3 Michel Electron Tagging Efficiency

A conservative 5% uncertainty on the tagging efficiency of Michel electrons is used and applied to the FHC decomposition procedure for the appearance channel [138].

4.8.4 FD Rock Rate

A small amount of FD events in the appearance channel can come from rock and other materials in the detector hall surrounding the FD detector. A special sample is generated for such events and is added to the central value prediction. An overall 100% scale uncertainty is taken for such events to be conservative. The contribution of such events in the disappearance channel is negligible to completely ignore, therefore no uncertainty is taken here [138].

4.8.5 Cosmic Scale

The statistical error in the unscaled cosmic trigger spectra used for the prediction of the shape of the cosmic background seen in the two oscillation channels is taken as an overall uncertainty on the cosmic spectrum shape itself [138].

4.9 Evaluation of Systematic Uncertainties

As mentioned before, the nuisance parameters in the likelihood fit are derived by interpolating between shifts at given numerical σ values for the above uncertainties for each analysis

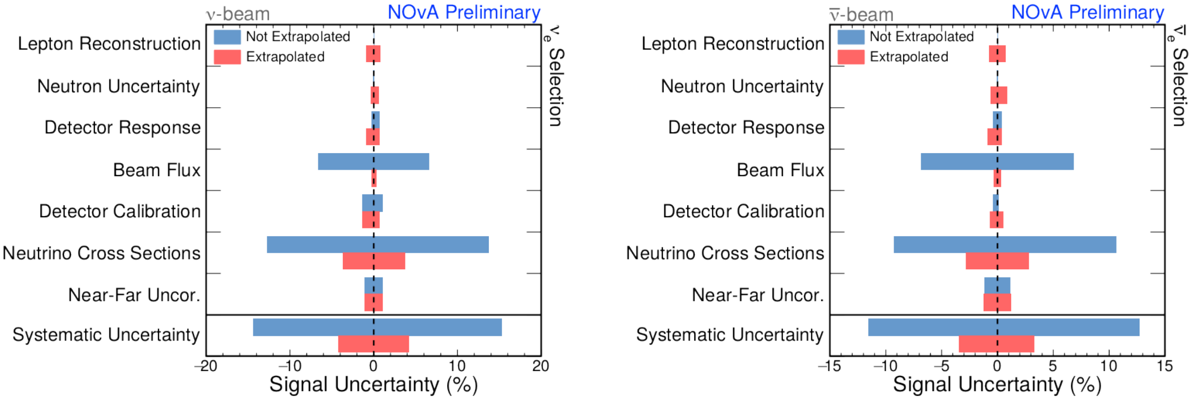


Figure 4.12: Comparison of the extrapolated and unextrapolated systematic error on the number of predicted events in the appearance channel for different uncertainties

bin. One can then evaluate the size of the uncertainties in many ways. For the appearance channel, one can compare different systematics based on the effect it has on the total number of predicted signal and background events when varied by $\pm1\sigma$, as shown in Fig. 4.12. One can also estimate its effect by fitting to Asimov data using the central value prediction and varying the MC expectation based on the systematic variation. This takes the form of the effect of the systematic on the individual oscillation parameters [79], as shown in Figs. 4.13 and 4.14. All of these metrics show that the analysis is statistically limited, i.e the systematic error is smaller than the statistical error in the prediction.

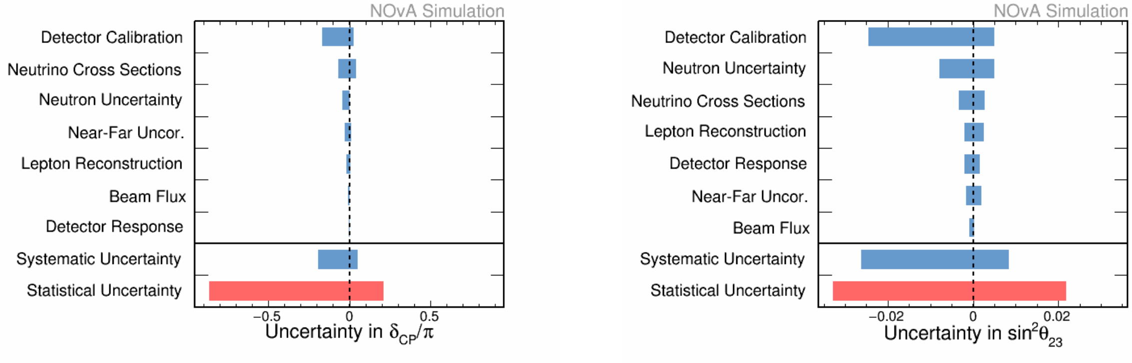


Figure 4.13: Overall systematic error on the measurement of δ_{CP} (left) and $\sin^2\theta_{23}$ (right)

The largest uncertainties are those in the Calibration, mainly coming from the energy scale uncertainty of 5%. The flux errors are mostly cancelled out, while neutron response, the cross-section and the lepton reconstruction systematics have relatively significant effects on

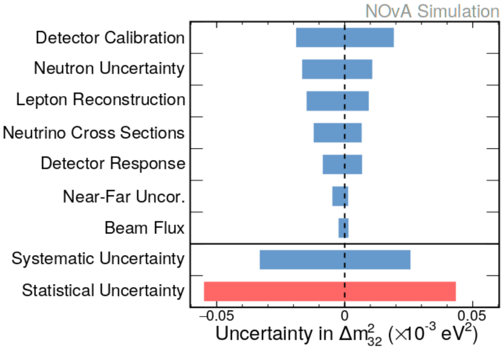


Figure 4.14: Overall systematic error on the measurement of Δm^2_{32}

the oscillation parameter measurement.

The effect of the extrapolation can be seen as well and shows that it is designed to correct for mainly correlated effects between the ND and FD. Uncorrelated effects increase slightly after extrapolation, but overall the method works quite well in reducing the total systematic error by more than half.

Chapter 5

Results

The results here are derived with a total exposure of 13.6×10^{20} POT in FHC and 12.5×10^{20} POT in RHC, as shown in Fig. 5.1. This was first presented in July 2020 [139].

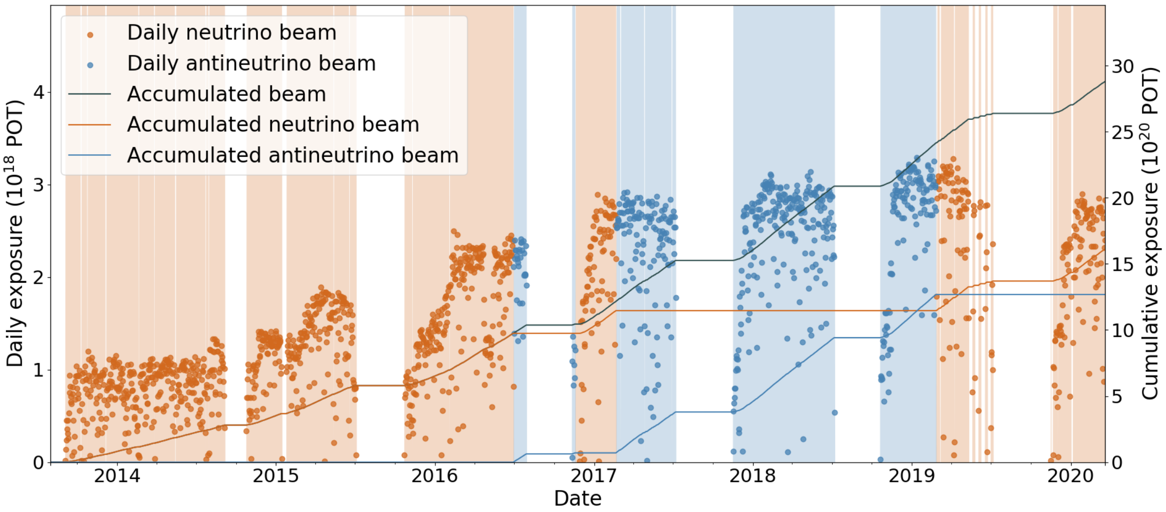


Figure 5.1: Exposure in terms of POT for the current analysis

5.1 Sidebands

The oscillation analysis in the two channels is carried out in a blind procedure where all the analysis components including the FD prediction with the ND constraint, the budgeting of systematic uncertainties and selection procedures are decided before “opening the box” on FD oscillated data. Before this step, a set of 3 selection sidebands are developed in the appearance channel and a high energy sideband in the disappearance channel. The sidebands are designed to be insensitive to any oscillations in the appearance channel, while in the disappearance channel it accesses only a high energy region beyond the oscillation dip, in order to ensure that the FD data matches the predictions well and there are no egregious discrepancies.

5.1.1 The $\nu_\mu \to \nu_\mu$ Channel

For the disappearance channel, a sample of ν_μ candidates at FD is developed by looking at regions $> 3\,\text{GeV}$ of total calorimetric energy in the interaction. The oscillation parameters chosen for this sideband are based mostly on global fits and involve the parameters :

- $\Delta m^2_{32} = 2.67 \times 10^{-3}$ eV^2
- $\Delta m^2_{31} = 7.53 \times 10^{-5}$ eV^2
- $\sin^2 2\theta_{12} = 0.851$
- $\sin^2 2\theta_{13} = 0.082$
- $\sin^2 \theta_{23} = 0.404$
- $\delta_{CP} = 1.48\pi$

The predictions all agree within statistical errors for both FHC and RHC [140], as illustrated in Fig. 5.2.

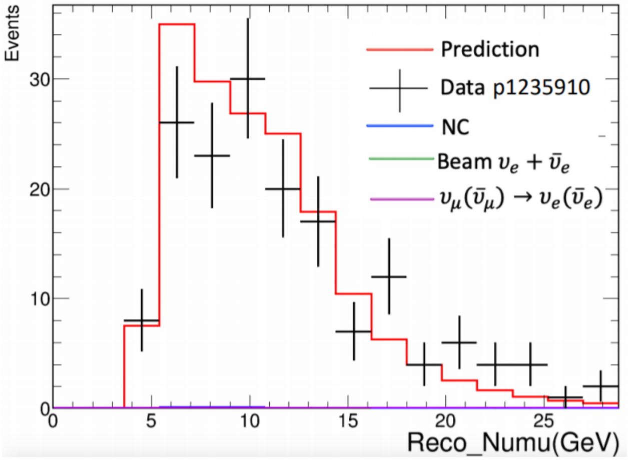


Figure 5.2: Sideband for the disappearance channel

5.1.2 The $\nu_\mu \rightarrow \nu_e$ Channel

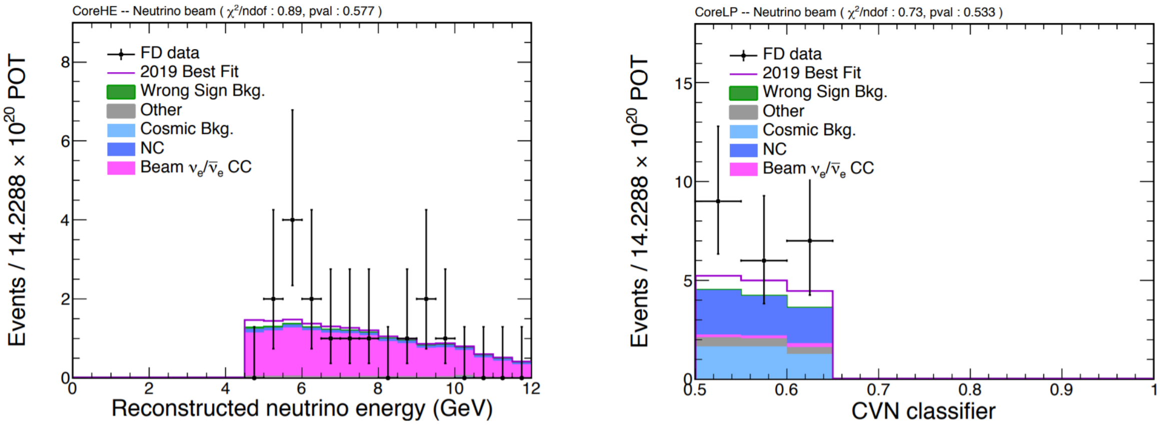


Figure 5.3: Sidebands for the appearance channel. Shown are the Core high energy sideband (left) and the Core low PID sideband (right)

For the appearance channel, since there are primarily three main backgrounds, the irreducible ν_e-CC contamination in the beam, the cosmic background and misidentified beam backgrounds such as ν_μ-CC events and NC events, 3 separate sidebands are designed dominated by each of these backgrounds. The sidebands utilize the selections :

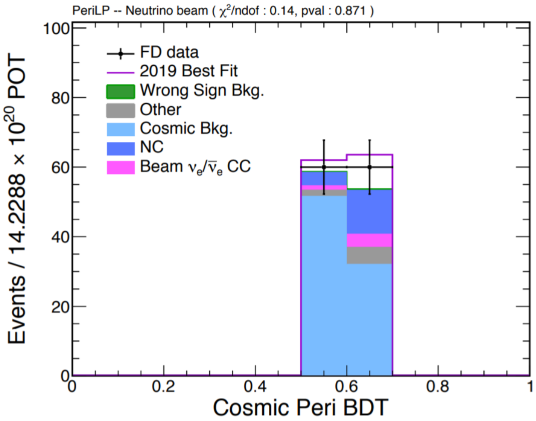


Figure 5.4: Peripheral low PID sideband for the appearance channel

- Core Selection with a Low PID cut (CoreLP) : The same core selections in the analysis are applied but with the CVN PID scores ranging from $(0.5, 0.65)$. This sideband is dominated by NC backgrounds.

- Core Selection at High Energy (CoreHE) : The same core selections in the analysis are applied including the PID score but only events $> 4.5\,\text{GeV}$ are chosen. This sideband is dominated by beam ν_e-CC events due to the fact that the flux of this component has a fatter tail compared to the oscillated ν_e-CC events.

- Peripheral Selection at Low PID : The same peripheral selections in the analysis are applied but with the CVN-BDT scores ranging from > 0.5 and > 0.56 (> 0.57) for FHC (RHC) and excluding the region in the analysis. This sideband is dominated by cosmics which are estimated from the NuMI trigger timing sideband anyway and are hence much less interesting but nevertheless, allow us to verify if there's something egregiously wrong in the predictions for the other beam background components.

A number of distributions in many observables including the various PIDs, the interaction locations and energies are looked at and verified to be within statistical errors, both for FHC and RHC [141], as can be seen from Figs. 5.3 and 5.4.

5.2 Opening the Box

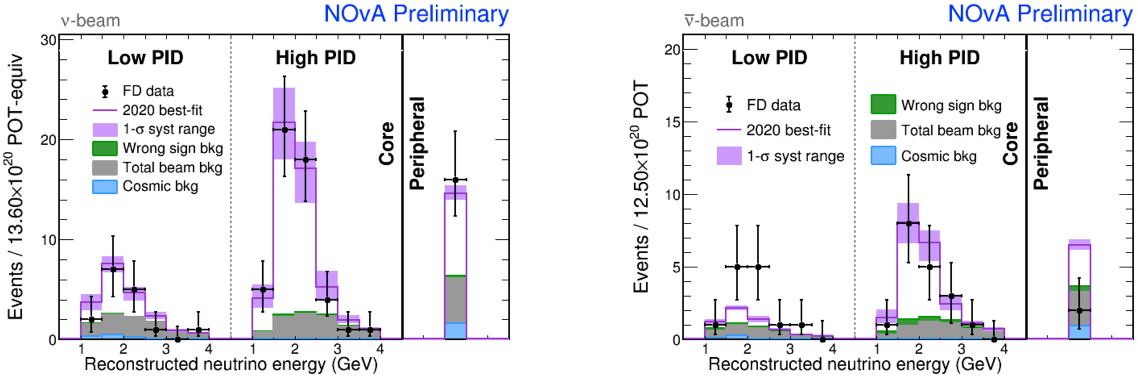


Figure 5.5: Oscillated spectra in the appearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error

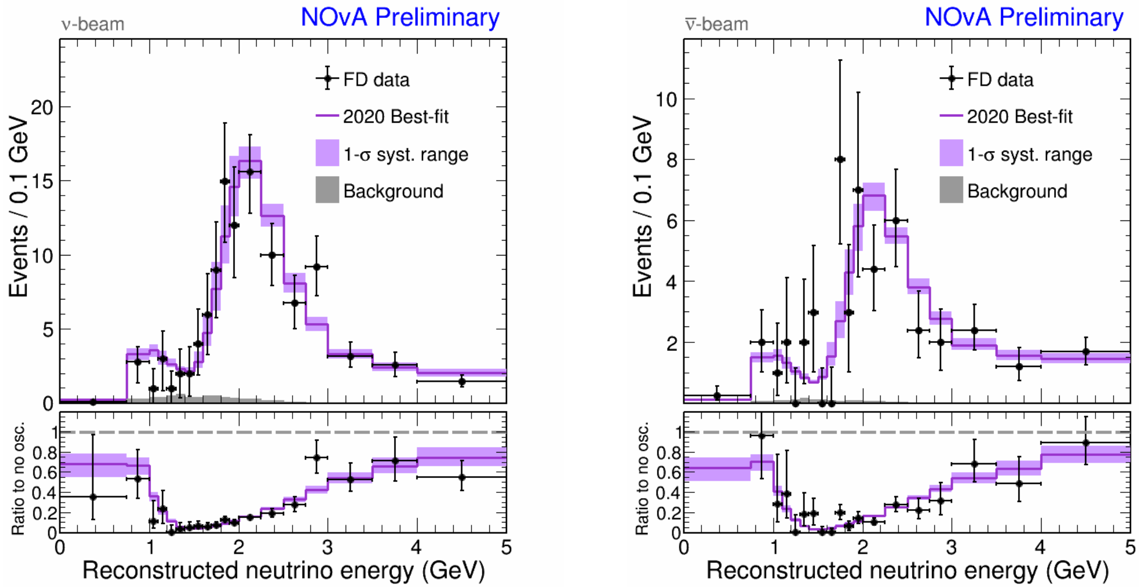


Figure 5.6: Oscillated spectra in the disappearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error

The results in the analysis region for FHC and RHC for both oscillation channels are shown in Figs. 5.5 - 5.8, along with a breakdown of the respective background components, shown in Tables 5.1 and 5.2. In total, 82 (33) ν_e ($\bar{\nu}_e$) candidates are seen for FHC (RHC), with an expected background of 26.83 (13.98) events. In the disappearance channel, a total of 211 (105) ν_μ ($\bar{\nu}_\mu$) candidates are seen summed over the 4 quantiles. The background expectation is 8.2 (2.1) events for FHC (RHC). The no oscillation expectation in the disappearance

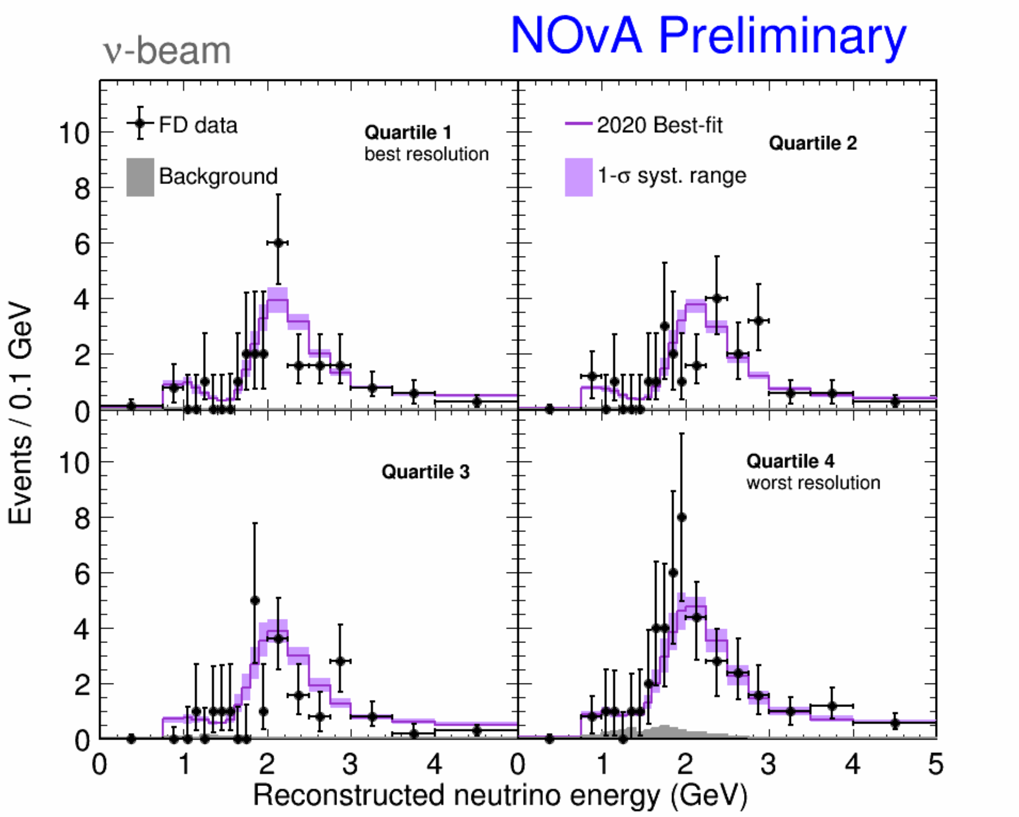


Figure 5.7: Oscillated spectra for each quantile in FHC for the disappearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error

channel is 1156.1 (488.1) events for FHC (RHC). Both channels show a clear sign of $\nu_\mu \rightarrow \nu_\mu$ disappearance and $\nu_\mu \rightarrow \nu_e$ appearance. The expected numbers are derived from a best fit set of oscillation parameters, done using a full joint fit of all samples in both FHC and RHC and are given by :

- $\Delta m^2_{32} = 2.41 \times 10^{-3}$ eV^2
- $\sin^2\theta_{23} = 0.57$
- $\delta_{CP} = 0.82\pi$

The value of θ_{13} is allowed to float in the fit but is constrained by precise measurements coming from reactor experiments such as Daya Bay, Double CHOOZ and RENO. The solar parameters are fixed in the fit. These values come from PDG and are given by :

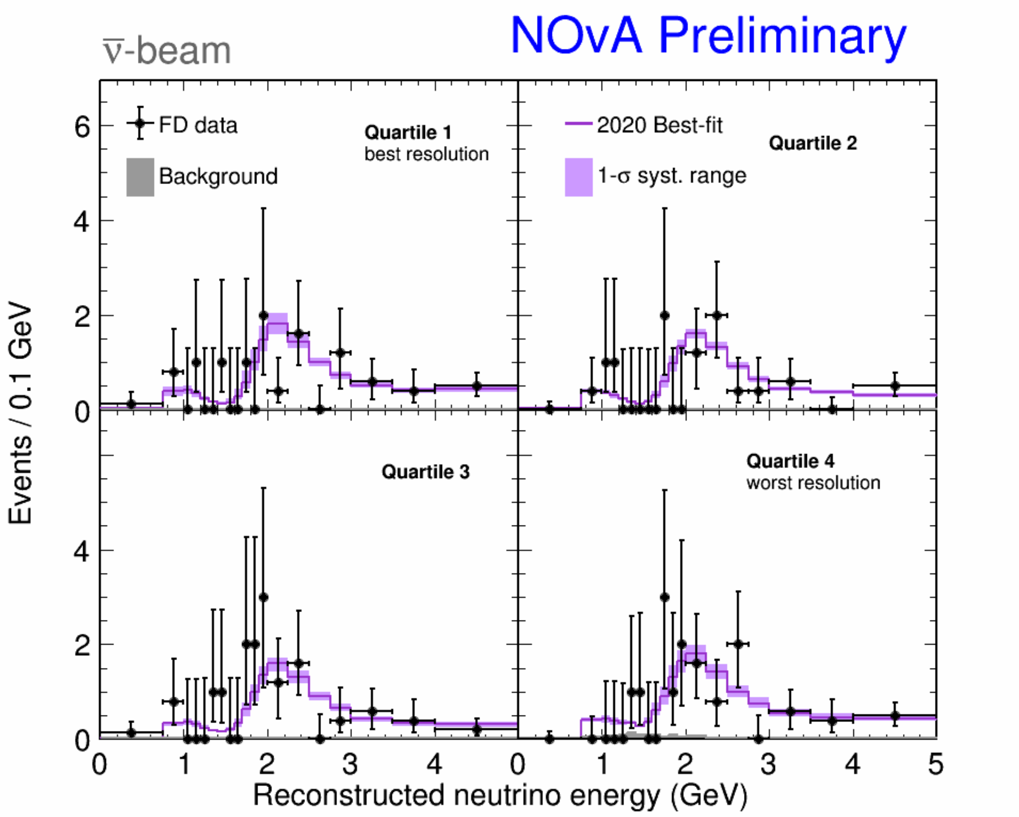


Figure 5.8: Oscillated spectra for each quantile in RHC for the disappearance channel plotted with the best-fit parameter expectation and the $1-\sigma$ systematic error

- $\sin^2 2\theta_{13} = 0.085 \pm 0.003$
- $\Delta m^2_{21} = 7.53 \times 10^{-5}$ eV^2
- $\sin^2 2\theta_{12} = 0.851$

The best fit favors the Normal Hierarchy and the Upper Octant of θ_{23} but not significantly. We will describe the constraints on the combinations of oscillation parameters more quantitatively in the following chapter but one can already start to make judgements about the measurement if one looks at the ν_e appearance data for FHC and RHC in a “bi-event” space or using a “S-curve” plot.

In a S-curve plot shown in Fig. 5.10, the number of expected events in the appearance channel are plotted as a function of δ_{CP} in the range $(0, 2\pi)$ for a range of $\sin^2\theta_{23}$ values $(0.45, 0.6)$

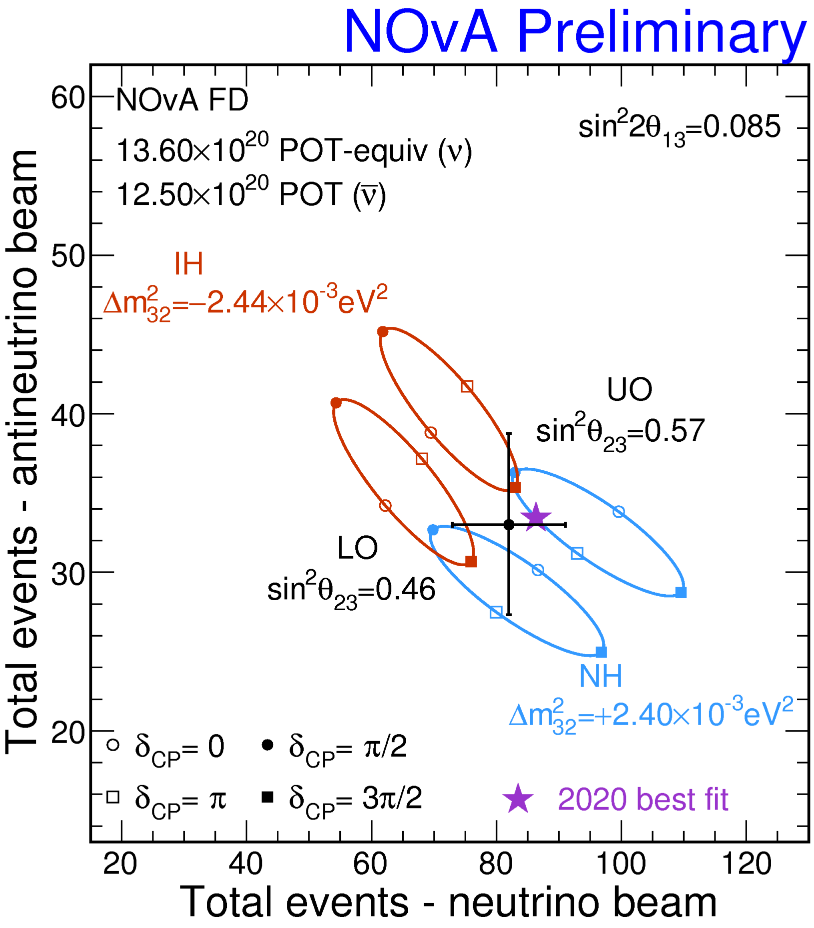


Figure 5.9: Total event counts in FHC and RHC in the appearance channel plotted with the best-fit expectations in the 4 quadrants of Hierarchy-Octant space

and Δm^2_{32} values in the Normal (2.4×10^{-3} eV^2) and Inverted Hierarchies (-2.44×10^{-3} eV^2) which shows that the data is close to the Inverted Hierarchy expectation for many θ_{23} values.

For the "bi-event" plot shown in Fig. 5.9, the observed number of data events for FHC and RHC are plotted for different best-fit oscillation parameters assuming different combinations of parameter space, (NH, IH) × (UO, LO). The variation of δ_{CP} traces out 4 ellipses for 4 such combinations. The range of expectations vary from 52 to 110 events for (IH, LO, $\delta_{CP} = \pi/2$) and (NH, UO, $\delta_{CP} = 3\pi/2$) respectively in FHC. For RHC the expectations vary from 25 to 45 for (NH, LO, $\delta_{CP} = 3\pi/2$) and (IH, UO, $\delta_{CP} = \pi/2$). The observed data is

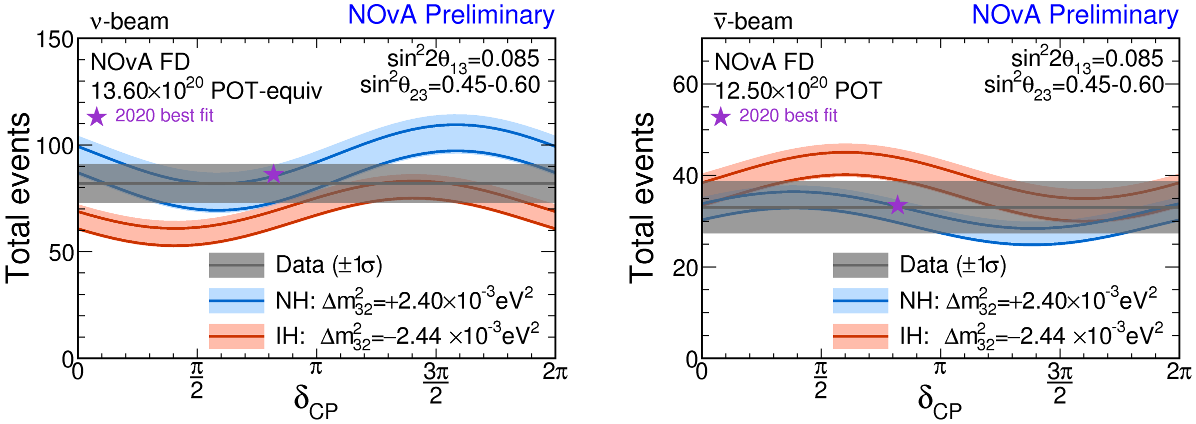


Figure 5.10: Total event counts in FHC (left) and RHC (right) in the appearance channel plotted with expectations at a range of oscillation parameters and as a function of δ_{CP}

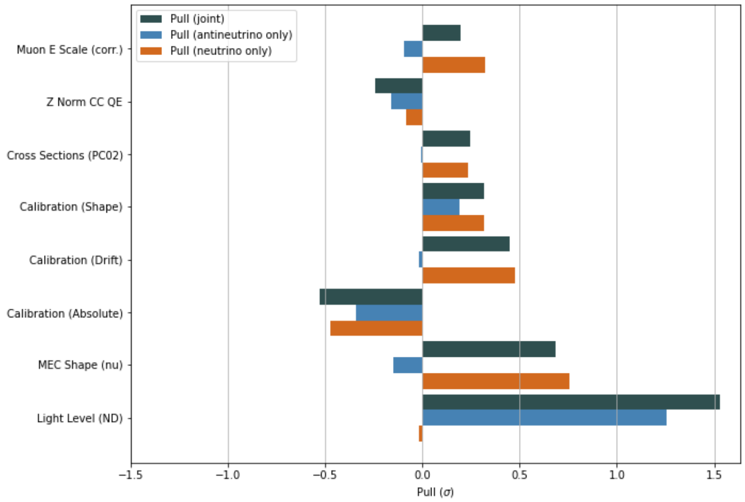


Figure 5.11: Systematic pulls in the best-fit expectation to the FD data

seen to lie in the middle of this range, thus being close to IH and LO as well as NH and UO and not showing a powerful preference for any δ_{CP}. The data, however, does seem to disfavor the regions of maximum asymmetry given by (IH, $\delta_{CP} = \pi/2$) and (NH, $\delta_{CP} = 3\pi/2$).

In the maximum likelihood fit, one can assess how hard the systematics are pulled to make the expectation compatible with data. As shown in Fig. 5.11, no systematics are seen to get pulled significantly however with the maximum pull coming from the ND light level systematic at 1.5σ. Everything else is $< 1\sigma$ with the next dominant pulls from the MEC $(q_0, |q|)$ shape uncertainty (0.6σ) and the energy scale uncertainty (-0.5σ).

5.2.1 Compatibility with 3-Flavor Mixing

The goodness of fit is assessed by a mock experiment procedure at the above best fit parameters in NH and UO for both FHC and RHC, combined and separately. A set of 1000 experiments are done by throwing Poisson fluctuations on top of the best-fit expectation and fit as before to get a distribution of the minimum negative log-likelihoods. The fit to the real data spectrum is very compatible with this distribution and the p-value measure of the goodness of fit, given by the number of mock experiments with a negative log-likelihood score worse (greater) than the real data score is 0.705 [142], as shown in Fig. 5.12.

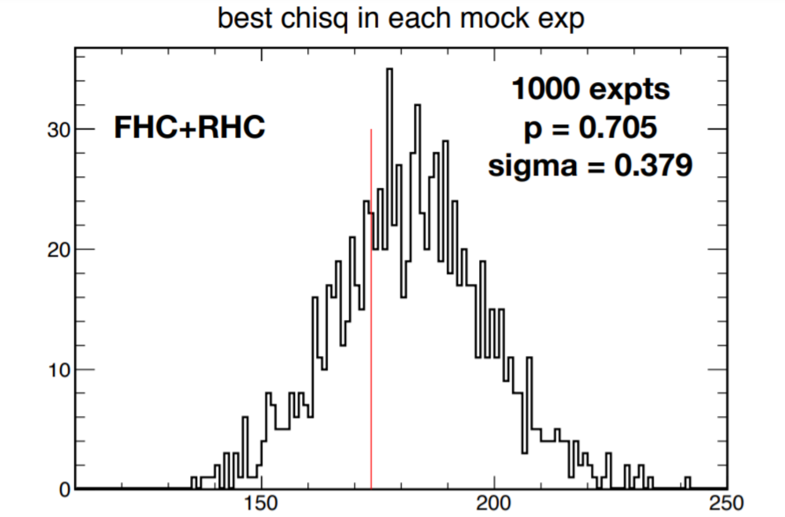


Figure 5.12: Goodness of fit $\Delta\chi^2$ for FD data (in red) shown along with the $\Delta\chi^2$ distribution for many pseudo-experiments

If one looks at only the FHC data or the RHC data, the p-values come out to 0.864 and 0.278 respectively, showing that the data in each of the beam modes are individually compatible with the best-fit expectation. This also suggests that the NOvA FHC and RHC data are very well explained by the 3-flavor mixing model. In addition, there is little indication that neutrinos and anti-neutrinos oscillate based on different mixing parameters which is allowed in various BSM scenarios that violate the CPT theorem. A quantitative discussion of this is however beyond the scope of this result.

5.2.2 Evidence of $\bar{\nu}_e$ Appearance

A key highlight of the observed FD oscillated spectra is a significant evidence for $\bar{\nu}_e$ appearance in the RHC beam. At the best fit, NOvA sees 33 candidate $\bar{\nu}_e$ events with a background expectation $\sim$ 14 events. The wrong-sign background expectation at the best fit, which is included in the above number is 2.25 events.

To establish the significance of $\bar{\nu}_e$ appearance quantitatively, a similar mock experiment procedure is used. However, given clear indications of appearance, throwing a large number of mock experiments to get a measure is not feasible. The procedure therefore utilizes a "fast"-MC methodology which generates a number of mock experiments for the RHC backgrounds quickly.

First, a probability distribution (pdf) is constructed for the oscillation parameters looking at only FHC data. This is done by varying the oscillation parameters from the FHC-only best fit measurement and evaluating the test statistic given by the log-likelihood ratio (or $\Delta\chi^2$) :

$$\Delta\chi^2(\vec{\theta}) = LL(\vec{\theta}) - LL(\vec{\theta}_{best}) \tag{5.1}$$

where LL denotes the negative log-likelihood at a given set of oscillation parameters, $\vec{\theta}$. The test statistic, $\Delta\chi^2$ is constructed in a 2D space $\delta_{CP} - \sin^2\theta_{23}$ by doing two fits, first by fixing the two oscillation parameters at each point in this grid and floating ("profiling") the rest, $(LL(\vec{\theta}))$ and then allowing all parameters to float in the fit, $(LL(\vec{\theta}_{best}))$. This is done under different assumptions of Normal and Inverted Hierarchy separately. Similarly, the test statistic distribution is constructed in 1D for Δm^2_{32}. The pdf is then constructed by transforming the test statistic distribution to $\exp(-\Delta\chi^2/2)$. The pdf for individual hierarchies are also accounted for by using the $\Delta\chi^2$ separating the two hierarchies. The pdf of $\sin^2 2\theta_{13}$ is generated from the reactor constraint assuming a gaussian distribution. The pdf for each systematic parameter is constructed assuming a normal distribution. A set of $10,000$ predictions are generated for the RHC wrong-sign and other backgrounds using the above pdfs for the oscillation parameters.

Then a distribution of mock RHC backgrounds is constructed by reweighting the nominal RHC background prediction (including wrong-sign) at the FHC best fit to the above predictions and adding Poisson fluctuations to each of them. In this way, one can generate millions of mock experiments with a background-only expectation quickly, as shown in Fig. 5.13. The p-value measure of compatibility of the observed 33 events with the RHC background-only expectation is calculated and translated to a significance measure (based on a Z-score) and is given by 4.2σ.

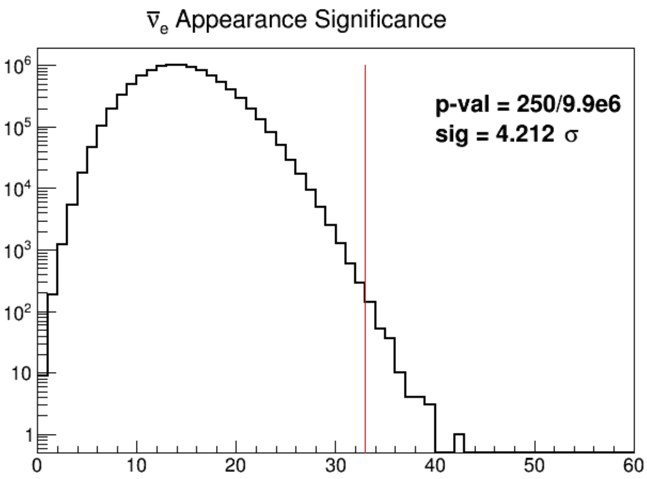


Figure 5.13: Significance of $\bar{\nu}_e$-appearance measured via many pseudo-experiments assuming no $\bar{\nu}_e$-appearance

While it falls below the conventional threshold of observation (5σ), this is nevertheless firm evidence of $\bar{\nu}_e$ appearance. This has not been seen by any other neutrino oscillation experiment at long baselines and serves to highlight the powerful NuMI beam utilized by NOvA, even in the traditionally weaker antineutrino mode of operation.

	FHC	RHC
Wrong-Sign	1.04	2.25
Beam $\nu_e + \bar{\nu}_e$ CC	14.08	7.29
NC	6.21	2.21
$\nu_\mu + \bar{\nu}_\mu$ CC	1.73	0.36
$\nu_\tau + \bar{\nu}_\tau$ CC	0.52	0.32
Cosmics	3.13	1.55
Total Bkg.	$26.83^{+1.46}_{-1.62}$	$13.98^{+0.76}_{-0.84}$
Total Signal	$58.96^{+2.50}_{-2.54}$	$19.20^{+0.64}_{-0.66}$
Total Expected	$85.79^{+2.96}_{-3.05}$	$33.18^{+1.08}_{-1.23}$
Total Observed	82	33

Table 5.1: Breakdown of individual components in the appearance spectrum at the best fit oscillation parameters. Errors shown are systematic uncertainties on the expectation

	FHC	RHC
Beam Bkg.	3.25	1.12
Cosmics	4.96	0.95
Total Bkg.	8.21	2.07
Wrong-Sign	12.64	26.10
Total Signal	214.1	103.4
Total Expected	$222.3^{+23.5}_{-22.4}$	$105.4^{+10.6}_{-10.1}$
Total Observed	211	105

Table 5.2: Breakdown of individual components in the disappearance spectrum at the best fit oscillation parameters. Errors shown are systematic uncertainties on the expectation

Chapter 6

Confidence Intervals

6.1 Feldman-Cousins

Beyond the best-fit oscillation parameter measurement, the statistical errors are determined using a frequentist approach. The confidence interval at a certain level denotes the range of parameters that are "compatible" with the observed data at the given level. To construct these intervals, the unified approach of Feldman and Cousins (FC) [143] is used which builds upon the equivalence of the principle of hypothesis testing to the Neyman confidence interval definition. An inverted likelihood ratio test at a particular point in oscillation parameter space, $\vec{\theta_0}$ with the null hypothesis defined by $\vec{\theta} = \vec{\theta_0}$ is used to assess compatibility with observed data. This is illustrated in Fig. 6.1. The test statistic used is given by the log-likelihood ratio as defined before :

$$\Delta\chi^2(\vec{\theta_0}) = LL(\vec{\theta_0}) - LL(\vec{\theta}_{best}) \tag{6.1}$$

where $\vec{\theta}_{best}$ denotes the minimum log-likelihood (negative) across the parameter space. $\vec{\theta_0}$ is then included in the $1 - \alpha$ confidence interval if we fail to reject the null hypothesis at

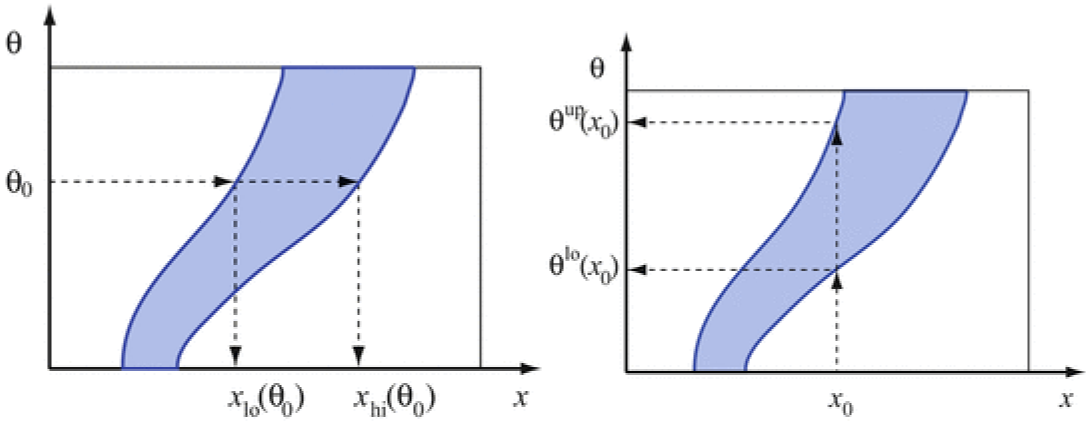


Figure 6.1: Illustration of the Neyman confidence interval construction. The confidence belts are created by a Likelihood ratio test for each θ_0 and then inverted for observed data to get the confidence interval. Taken from [144]

a level α, chosen apriori and based on the test statistic, $\Delta\chi^2(\vec{\theta}_0)$. The confidence level, α is a measure of the probability that one would observe a $\Delta\chi^2$ larger than or equal to the one actually observed if one would repeat the experiment many times and $\vec{\theta}_0$ was at the boundary of the confidence interval. This confidence level is therefore comparable to a p-value and is often translated to a signifcance measure in terms of an n-σ deviation on a normal distribution. For example, a 68% confidence interval is given at $\alpha = 0.32$ and is labelled the "1-σ" interval. Each point in parameter space can be assessed individually if it satisfies the inverted likelihood ratio test, i.e if the p-value of the test is > 0.32 and the construction of the confidence interval proceeds by including all such points. This results in an parameter interval that has the right "coverage" properties, i.e the probability that the true parameter lies within the $(1 - \alpha)\%$ confidence interval is not less than $1 - \alpha$, as required by the Neyman definition. Finally, the reason for the particular choice of the test statistic comes from the Neyman-Pearson Lemma which guarantees that the likelihood ratio maximises the statistical power of the hypothesis test. This ensures that the "over-coverage" is minimized, i.e the probability of the true parameter falling within the $(1-\alpha)\%$ confidence interval is not much larger than $1 - \alpha$.

Here we've described the Feldman-Cousins approach as would be applied for the purpose of oscillation parameter intervals. In general, the advantage of this approach is that the likelihood-ratio test statistic works equally well for null and non-null observations of a given Poisson process with an assumed background and can be therefore used for constructing both one-sided (upper and lower) and two-sided (central) confidence intervals. This is relevant for physicists since physical processes are often characterized by a Poisson distribution. This is the sense in which it is “unified” by the Feldman-Cousins procedure.

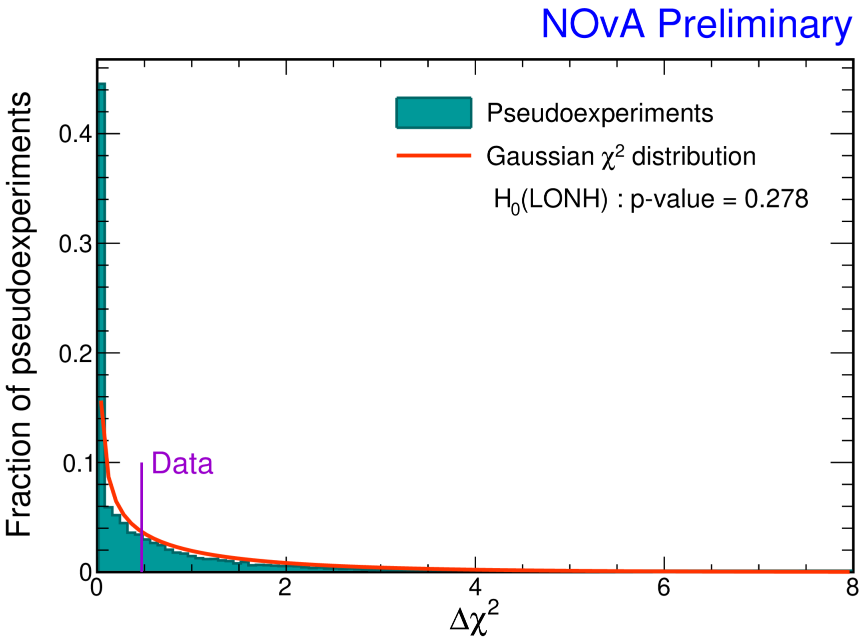


Figure 6.2: Illustration of the Feldman-Cousins method using a likelihood ratio ($\Delta\chi^2$) distribution generated using pseudo-experiments to find the p-value with respect to data. Also shown in red is the Wilks approximation

In many cases, the p-value of the likelihood ratio test at $\vec{\theta}_0$ is analytically related to the value of the test statistic itself, $\Delta\chi^2(\vec{\theta}_0)$. This is the case for example when the likelihood is given by a Gaussian distribution, as would happen if there are large enough statistics in the data [145]. Consequently, the $\Delta\chi^2(\vec{\theta}_0)$ p.d.f follows a χ^2_k distribution for some k degrees of freedom and every $\vec{\theta}_0$. In this case, one can readily construct the confidence interval by including parameter values that have a test-statistic itself below a certain threshold rather than relying on a threshold on the p-value. For example, $\vec{\theta}_0$ will belong to the 68% confidence interval if its $\Delta\chi^2 < 1$ (2.30) for the 1D (2D) case.

For our scenario, the statistics in the data aren't large enough to be able to use a Gaussian approximation for the likelihood and consequently a full Poisson form must be used. Moreover, particular features of the oscillation parameter space such as the physical boundary at $\sin^2\theta_{23} = 0.5$ where the amount of disappearance in the dip-region is $\sim$ maximal or the cyclic nature of δ_{CP} introduce further corrections from the Gaussian likelihood. This is easy to see for example close to $\sin^2\theta_{23} = 0.5$, seen in Fig. 6.2. Here the space of possibilities that the model can fluctuate to in order to make it compatible with the data is constrained along one direction since one can't have more disappearance than at $\sin^2\theta_{23} = 0.5$ (actually ~ 0.514 for NOvA, due to extra terms in the disappearance oscillation probability). Consequently, the likelihood distribution is skewed which induces a deviation from a χ^2_k approximation of the test-statistic p.d.f.

As a result, NOvA uses the FC procedure outlined above to produce all its results for the confidence intervals. Both in 1D and 2D, at each point in oscillation parameter space, many pseudo-experiments are thrown as Poisson fluctuations on top of the predicted spectrum at those parameters to generate the $\Delta\chi^2$ test-statistic distribution. The p-values are calculated at those points based on the distribution and the observed $\Delta\chi^2$ test-statistic for real data. These are then either mapped to intervals of 1σ, 2σ, 3σ in the 2D contour case or directly, the exact significance of rejection for that point in the 1D case. For the 2D contours, this is done for $\sin^2\theta_{23} - \delta_{CP}$ and $\Delta m^2_{32} - \sin^2\theta_{23}$, each under two different hypotheses of Normal Hierarchy (NH) and Inverted Hierarchy (IH). In the 1D case, curves denoting the significance of rejection of a particular parameter value are drawn for $\sin^2\theta_{23}$, δ_{CP}, Δm^2_{32} separately. These are done under 4 different hypotheses of (NH, IH) × (UO, LO) except for $\sin^2\theta_{23}$ where its drawn under separate hypotheses of NH and IH. Individual hypotheses (NH-LO, IH-LO, IH-UO, IH) are themselves evaluated via pseudo-experiments in terms of their overall significance of rejection compared to the best-fit expectation of NH and UO, as shown in Figs. 6.2 and 6.3.

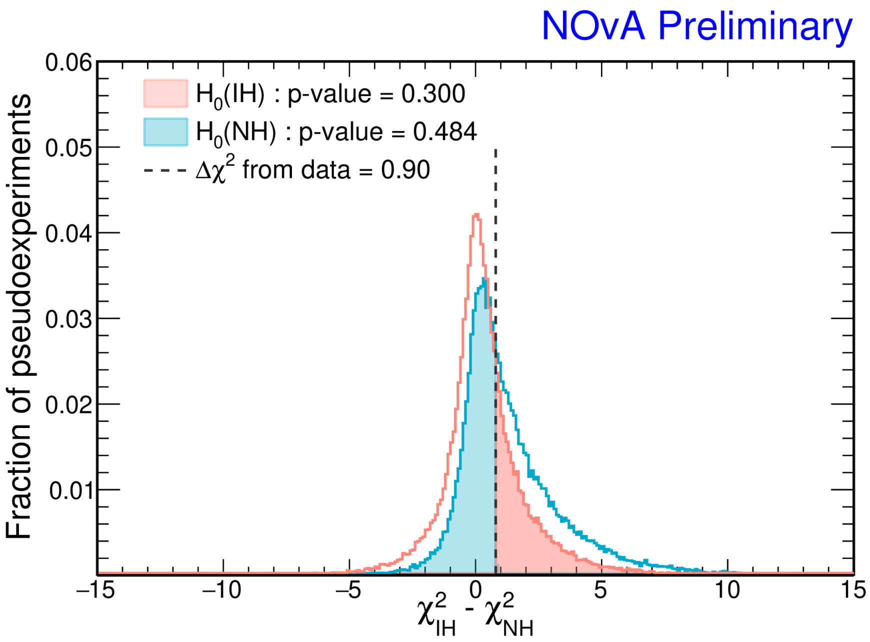


Figure 6.3: Significance of inverted hierarchy rejection using the FC method

Nuisance parameters are handled by "profiling" them out. This means that at each individual point in parameter space, either 1D or 2D, a fit to real data is first performed where the parameters in question are kept fixed and the nuisance parameters are allowed to float. The resulting best-fit nuisance parameters at that individual point is then used as an input for throwing pseudo-experiments at that point. Then similar to the real data fit, for each pseudo-experiment, two fits are performed, the first which keeps the parameters in question fixed and allows the nuisance parameters to float and the second which allows all parameters to float. The difference of the log-likelihoods is used to fill up the $\Delta\chi^2$ distribution. For example, at a given point (θ_0, δ_0) in $\sin^2\theta_{23} - \delta_{CP} - NH$ space, the pseudo-experiments are thrown at $\sin^2\theta_{23} = \theta_0$, $\delta_{CP} = \delta_0$ and nuisance parameters (Δm^2_{32}, $\sin^2\theta_{13}$, systematic parameters and so on), given by :

$$\hat{\vec{\nu}} \;\mid\; \chi^2_{data}(\hat{\vec{\nu}} \mid \theta_0, \delta_0) = min(\chi^2_{data}(\vec{\nu} \mid NH, \theta_0, \delta_0)) \tag{6.2}$$

Then $\Delta\chi^2_{expt}$ is calculated as :

$$\Delta\chi^2_{expt} = \chi^2_{true} - \chi^2_{best} \tag{6.3}$$

$$\chi^2_{true} = min(\chi^2(\vec{\nu} \mid NH, \theta_0, \delta_0)) \tag{6.4}$$

$$\chi^2_{best} = min(\chi^2(\theta, \delta, \vec{\nu})) \tag{6.5}$$

This choice of profiling nuisance parameters in the absence of a well defined prescription is done since it can be intuitively seen that if a point is accepted within a given confidence interval for the profiled values, then it ought to be accepted for other values too where the $\Delta\chi^2$ distribution would be larger [146].

6.2 Results

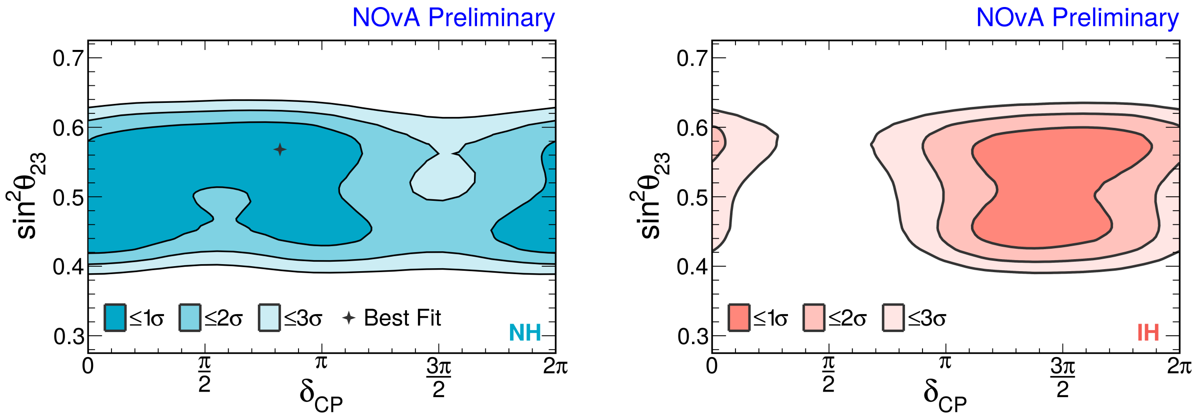


Figure 6.4: Confidence intervals in $\sin^2\theta_{23} - \delta_{CP}$ space for the Normal Hierarchy (left) and the Inverted Hierarchy (right)

The outcome of the FC procedure for the 1D "slices" and 2D contours of the oscillation parameters are shown in Figs. 6.4 - 6.7. As a consequence, the measured oscillation parameters are given by :

- $\Delta m^2_{32} = (2.41 \pm 0.07) \times 10^{-3}$ eV^2

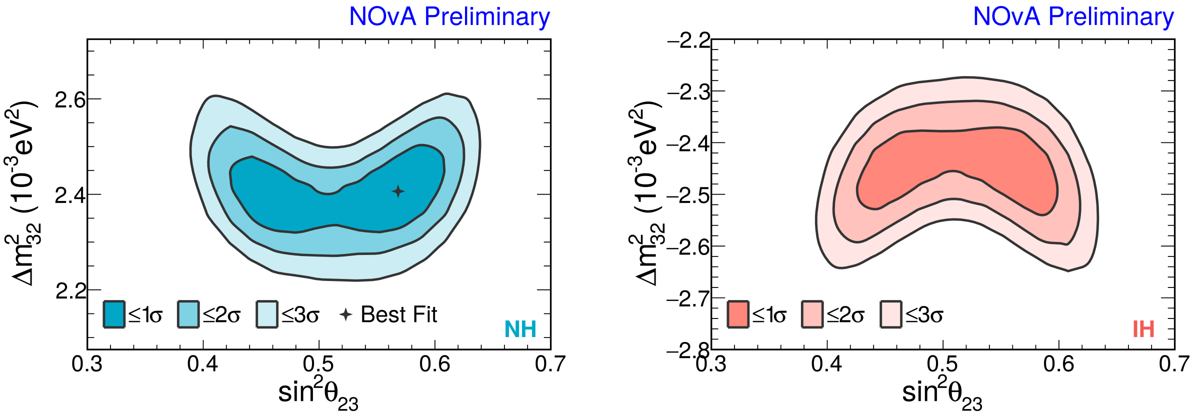


Figure 6.5: Confidence intervals in $\sin^2\theta_{23} - \Delta m^2_{32}$ space for the Normal Hierarchy (left) and the Inverted Hierarchy (right)

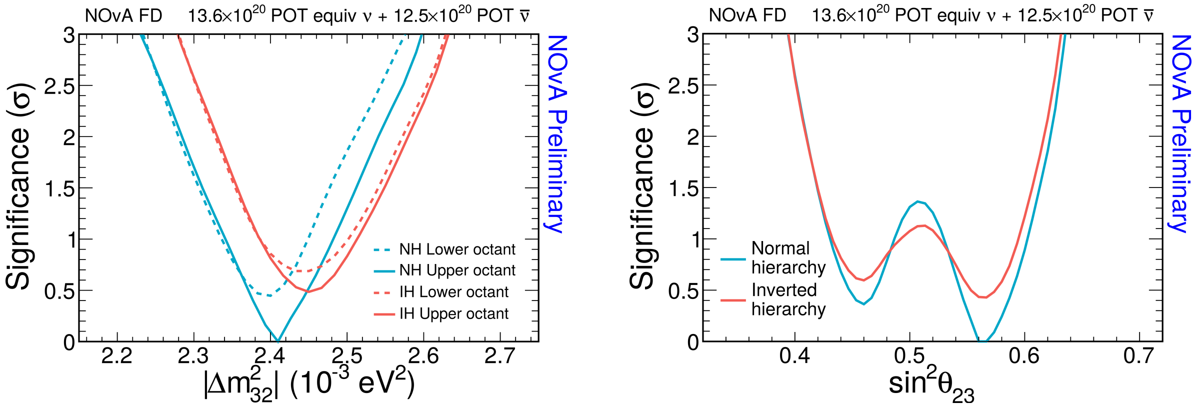


Figure 6.6: Significances of rejection for each value of the given oscillation parameter for various hypotheses of the profiled parameters.

- $\sin^2\theta_{23} = 0.57^{+0.04}_{-0.03}$ (UO)
- $\delta_{CP} = 0.82\pi^{+0.27\pi}_{-0.87\pi}$

which is a $\sim 3\%$ precision on Δm^2_{32} and $\sim 6\%$ precision on $\sin^2\theta_{23}$. The overall significance of rejection of the Normal Hierarchy is 1.0σ and the Lower Octant (LO) is 1.2σ. Regions of large asymmetry in $\nu_e - \bar{\nu}_e$ appearance are disfavored as mentioned before. (IH, $\pi/2$) is rejected at $> 3\sigma$ and (NH, $3\pi/2$) is rejected at $\sim 2\sigma$. There is only a weak constraint on δ_{CP} using this NOvA dataset. Table 6.1 shows the full confidence intervals for the measured oscillation parameters.

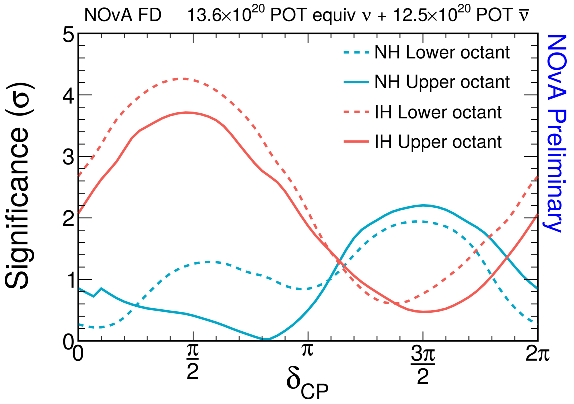


Figure 6.7: Significances of rejection for each value of δ_{CP} for various hypotheses of the profiled parameters.

	1-σ C.I
$\sin^2\theta_{23}$-NH	[0.431, 0.487] ∪ [0.53, 0.602]
$\sin^2\theta_{23}$-IH	[0.433, 0.492] ∪ [0.526, 0.595]
Δm^2_{32}-NH	[2.336, 2.482] ($\times 10^{-3}$ eV^2)
Δm^2_{32}-IH	[-2.513, -2.386] ($\times 10^{-3}$ eV^2)
δ_{CP}-NH	[0, 1.06] ∪ [1.82, 2] (π)
δ_{CP}-IH	[1.26, 1.73] (π)

Table 6.1: Confidence intervals for the individual oscillation parameters in the measurement for each mass hierarchy

6.2.1 Compatibility with Other Experiments

The NOvA measurement in the "atmospheric" space of $\Delta m^2_{32} - \delta_{CP}$ along with other experiments is shown in Fig. 6.8 and shows fairly good agreement. With this dataset, NOvA is able to provide a precise measurement of Δm^2_{32} which is competitive with measurements provided by Daya Bay [147] and T2K [148].

The $\sin^2\theta_{23} - \delta_{CP}$ contours in the Normal Hierarchy are shown along with the result from T2K in Fig. 6.8 . Notably, there is a tension at the level of $\sim 2\sigma$ in this space, owing to the fact that T2K heavily favors large ν_e appearance asymmetry over $\bar{\nu}_e$ appearance, which due to the shorter baseline results in a sensitivity for CP-violation. In contrast, NOvA is much more sensitive to the hierarchy than CP-violation. This complementarity in the experiment

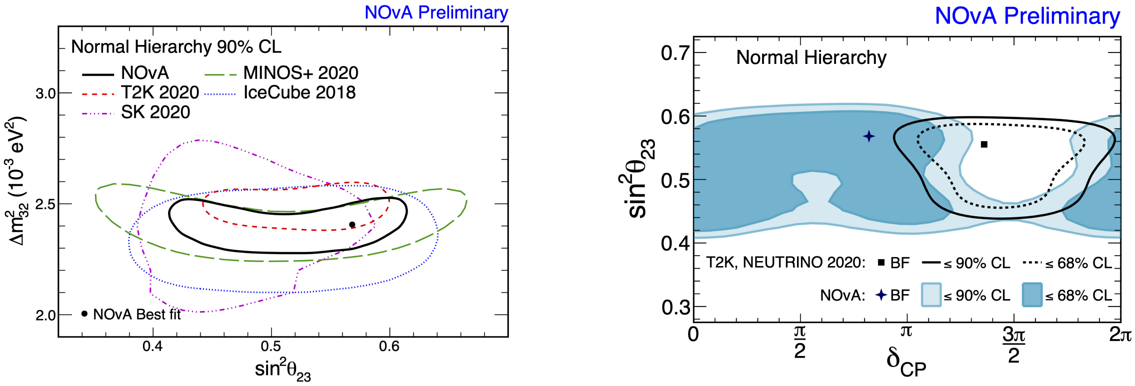


Figure 6.8: Compatibility of the NOvA measurement with various experiments in the atmospheric sector (left) and with T2K in $\sin^2\theta_{23} - \delta_{CP}$ space (right)

design has been noted and an agreement is in place to pursue a joint fit with T2K in the future. It is notable however that even based on this result, it might throw up a best fit in the Inverted Hierarchy even if the individual experiments favor the Normal Hierarchy [149].

One can also interpret this as a signal for Non-Standard Interactions (NSI) [150][151] and it will be interesting to see how this picture unfolds as both the experiments collect more data. However more discussion about this is beyond the scope of this document.

6.3 Accelerating Feldman-Cousins

It is easy to see that the Feldman-Cousins procedure is extremely computationally intensive since it relies on many pseudo-experiments to generate the $\Delta\chi^2(\vec{\theta})$ distribution for each point, $\vec{\theta}$ in parameter space. For NOvA, since all the confidence intervals are constructed using this procedure, it exploits the NERSC supercomputer system to make the task manageable. With an overall computational budget of more than $\sim$ 50 million CPU-hours, it is able to accomplish this within a week or two [152].

The FC procedure done for NOvA in practice tries to use as many pseudo-experiments as necessary to achieve a $\sim$ 5% precision on the confidence intervals themselves. This takes the form of a varying number of pseudo-experiments thrown across the parameter space and even with the HPC capabilities at NERSC, the resource usage is demanding. For contours especially, where the resource usage is the most, one can try to mitigate this burden with well motivated Bayesian techniques. Intuitively, one can see that one needn't expend computational resources at points in parameter space far from confidence interval boundaries and can demand a smaller precision in these regions, while trying to concentrate more on regions near the boundary. Since these boundaries are not known apriori, the Feldman-Cousins procedure is carried out in practice as essentially a grid search.

A variety of Bayesian optimization techniques, however, exist to solve just these class of problems [153]. Here, the model over the parameter space is itself assumed to come from a probability distribution. In our confidence interval construction, the model is the p-value profile/surface over the oscillation parameter space in question, 1D or 2D.

$$p = f(\vec{\theta}; \vec{\nu}_{prof}) \tag{6.6}$$

We are then interested in

$$\mathcal{C}_{1-\alpha} = \{\vec{\theta} \ | \ f(\vec{\theta}; \vec{\nu}_{prof}) > \alpha\} \tag{6.7}$$

for the $(1-\alpha)\%$ confidence interval. The Bayesian optimization techniques then assume $f(\cdot\ ,\ \cdot)$ as being generated from a probability distribution itself. Using observed data points, which in our case would refer to the p-values calculated at some point in oscillation parameter space, one can then refine that distribution iteratively to arrive at an accurate approximation of $f(\cdot\ ,\ \cdot)$. This is non-parametric to the extent that it depends only on general features of $f(\cdot\ ,\ \cdot)$ rather than relying on an analytic form of $f(\cdot\ ,\ \cdot)$ which isn't available. In usual problems, the objective is to find extremal points of $f(\cdot\ ,\ \cdot)$ from this approximation and iteratively get closer, but in our case, it can be suitably modified to find $\mathcal{C}_{1-\alpha}$. This has the advantage then of not needing to go through the entire parameter space in order to find $\mathcal{C}_{1-\alpha}$ and can therefore result in substantial computational savings. We will describe the application of just such a method to NOvA's Feldman-Cousins confidence intervals.

6.3.1 Gaussian Process

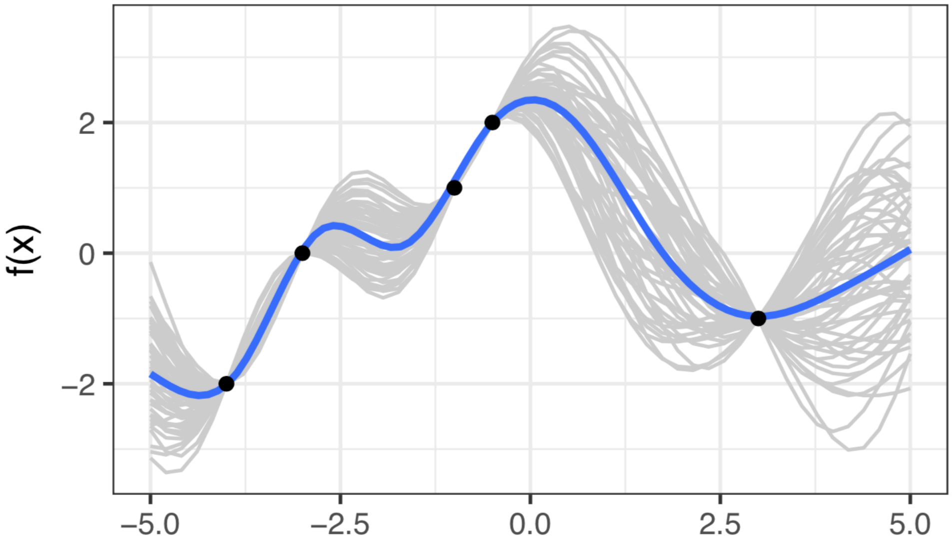


Figure 6.9: Illustration of the Bayesian interpolation procedure using Gaussian Processes for some example data. Blue curve represents the mean posterior prediction while the grey curves represent random draws from the $\mathcal{GP}$

A Gaussian Process ($\mathcal{GP}$) is an infinte extension of a multivariate Gaussian [154]. It is defined as a stochastic process over function space where any finite collection of points are jointly Gaussian distributed. Intuitively, one can interpret a $\mathcal{GP}$ as a random process that generates functions, $f(x)$ for each draw from it, as illustrated in Figs. 6.9 and 6.10. We say $f \sim \mathcal{GP}(\mu, \kappa(\cdot,\cdot))$ if :

$$\begin{pmatrix} f(x) \\ f(x') \end{pmatrix} \sim \mathcal{N}(\begin{pmatrix} \mu(x) \\ \mu(x') \end{pmatrix}, \begin{bmatrix} \kappa(x,x) & \kappa(x,x') \\ \kappa(x,x') & \kappa(x',x') \end{bmatrix}). \tag{6.8}$$

where $\mu(x)$ is the $\mathcal{GP}$ mean, defined as a prior function over x and $\kappa(\cdot,\cdot)$ parameterizes the covariance over any finite collection of x. In our case, the stochasticity comes from the finite grid resolution in the FC procedure as well as the p-value calculation of the $\Delta\chi^2$ distribution using a finite number of pseudo-experiments. In many problems, the prior function, μ is taken to be zero everywhere. One can then obtain a posterior predictive distribution given some observed data points, $\vec{x}_{obs}$ and their corresponding p-values, $f(\vec{x}_{obs})$ by :

$$f(x^*)|f(\vec{x}_{obs}) \sim N(\kappa(x^*, \vec{x}_{obs})(\kappa(\vec{x}_{obs}, \vec{x}_{obs}))^{-1} f(\vec{x}_{obs}), \tag{6.9}$$

$$\kappa(x^*, x^*) - \kappa(x^*, \vec{x}_{obs})(\kappa(\vec{x}_{obs}, \vec{x}_{obs}))^{-1} \kappa(\vec{x}_{obs}, x^*)) \tag{6.10}$$

Here x^* are new data points in parameter space and $f(x^*)$ denotes the distribution of predictions of the function values (p-values) at those points. As a result, the $\mathcal{GP}$ mean p-value prediction at x^* is given from above by :

$$p(x^*) = \kappa(x^*, \vec{x}_{obs})(\kappa(\vec{x}_{obs}, \vec{x}_{obs}))^{-1} f(\vec{x}_{obs}) \tag{6.11}$$

and the standard deviation in the $\mathcal{GP}$ prediction at x^* by :

$$\sigma_p(x^*) = \kappa(x^*, x^*) - \kappa(x^*, \vec{x}_{obs})(\kappa(\vec{x}_{obs}, \vec{x}_{obs}))^{-1} \kappa(\vec{x}_{obs}, x^*) \tag{6.12}$$

It is evident that the prediction and its errors depend only on the observed p-values at some points $\vec{x}_{obs}$ and the kernel function $\kappa(\cdot, \cdot)$. The kernel function is chosen apriori based on general features one expects from $f(x)$. For example, a widely used kernel function is the so-called Radial Basis Function (RBF) where

$$\kappa(x, x') = \exp\left(\frac{-(x - x')^2}{2l^2}\right) \tag{6.13}$$

The RBF form is infinitely smooth and is therefore chosen to model $f(x)$ where f is infinitely smooth itself. Moreover, it has one kernel parameter ("hyper-parameter"), l which

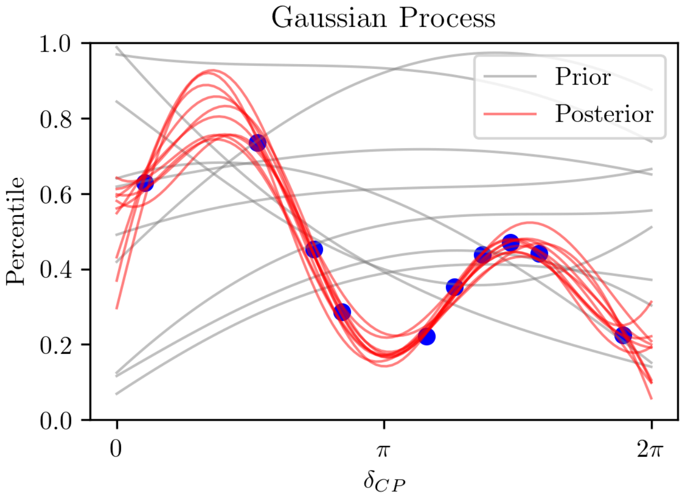


Figure 6.10: Illustration of the Bayesian interpolation procedure using Gaussian Processes for the percentile (1-pvalue) as a function of δ_{CP} obtained using FC. Blue dots represent points where the FC pseudo-experiments were thrown

is interpreted as a length scale over x, i.e for $|x - x'| \ll l$, one gets a large correlation between function values, $f(x)$ and $f(x')$ while for $|x - x'| \gg l$, one gets a small correlation. This means that the length-scale, l denotes the region of parameter space over which one is expected to see function values at one point influence the function values at other points. For distances much larger than l, we can then expect that the function values are mostly uncorrelated. This is similar to our scenario, where one expects the variation of p-values across oscillation parameter space to be smooth with the results of far-away regions being independent of nearby regions, as illustrated in Fig. 6.10.

Another example of an infinitely smooth kernel function is the Rational Quadratic Function (RQF) which bakes in a mixture of RBF kernels. This is useful when one doesn't have just a single characteristic length-scale to model the function over parameter space. The RQF function is given by :

$$\kappa(x, x') = \left(1 + \frac{(x - x')^2}{2\alpha l^2}\right)^{-\alpha} \tag{6.14}$$

There are two hyper-parameters associated with this kernel, α and l. α can be interpreted as a scale mixture over a given length-scale, l. It is easy to see that as $\alpha \to \infty$, we get back the RBF form.

Since the $\mathcal{GP}$ is uniquely characterized by $\kappa(\cdot, \cdot)$, one needs to choose the hyper-parameters embedded in the kernel. This is done by maximizing a marginal likelihood of the function values at observed data-points over the set of hyper-parameters. This is generally given by the form :

$$-2 \log \mathcal{L}(\vec{w}) = f(\vec{x}_{obs})^T \kappa(\vec{x}_{obs}, \vec{x}_{obs}; \vec{w})^{-1} f(\vec{x}_{obs}) + \log ||\kappa(\vec{x}_{obs}, \vec{x}_{obs}; \vec{w})|| + n \log 2\pi \tag{6.15}$$

The first term has the interpretation of a χ^2 over the observed function values and $||\kappa||$ denotes the determinant of the kernel matrix. The second term models the "complexity" of given $\vec{w}$ for describing the observed function values. For example, with the RBF kernel with a length-scale, l, the marginal likelihood penalizes smaller length scales than necessary to describe the observations. By minimizing the above equation, one has an appropriate choice for the hyper-parameters, $\vec{w}$ that can be used to obtain the prediction and its uncertainty.

6.3.2 Optimizing the FC Interval Construction

The idea of the Bayesian optmization procedure proceeds by calculating the p-values at some points in either 1D or 2D oscillation parameter space using a fraction of the pseudo-experiment budget at those points. One can then obtain a $\mathcal{GP}$ prediction for the p-values at all other points in this space, as outlined in the previous section. One then proceeds to use this prediction, via an "acquisition" function that prioritizes the allocation of pseudo-experiments at specific points [155]. We can thus refine the approximation based on the new pseudo-experiments at new or the original points iteratively and consequently, construct the confidence intervals, $\mathcal{C}_{1-\alpha}$. Iteratively, this will come closer and closer to the actual intervals constructed from a grid search application of the FC procedure, as can be seen from Fig. 6.11.

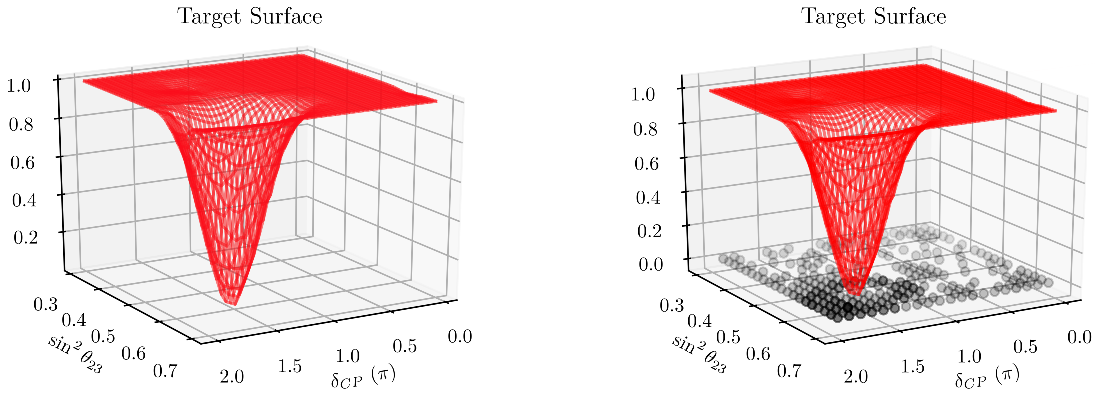


Figure 6.11: The percentile (1-p) surface for some toy oscillation data as a function of $\sin^2\theta_{23} - \delta_{CP}$. Also shown on the right is the points where the $\mathcal{GP}$-based algorithm chooses to throw pseudo-experiments to approximate the target surface

The acquisition function, $a(\vec{\theta})$ at points $\vec{\theta}$ aims to balance between "exploration", where new points are sought, and "exploitation", where the $\mathcal{GP}$ approximation error is reduced. A particular choice of the function can be given by :

$$a(\vec{\theta}) = \sum_{\alpha_i} |\frac{f(\vec{\theta}) - \alpha_i}{\sigma_{f(\vec{\theta})}}|^{-1} \tag{6.16}$$

where $f(\vec{\theta})$ is the $\mathcal{GP}$ mean prediction and $\sigma_{f(\vec{\theta})}$ is $\mathcal{GP}$ standard deviation as given above. α_i in our case will denote the 1σ, 2σ, 3σ intervals that we want to construct in oscillation parameter space. This form makes it evident that for points, $\vec{\theta}$ to have a large priority for pseudo-experiment allocation (large $a(\vec{\theta})$), one must need either a large error in the prediction or it must be close to the confidence interval boundary, as seen in Fig. 6.12. Therefore, iteratively this algorithm will seek out the necessary confidence interval boundaries and refine them by allocating pseudo-experiments in that region. In summary, the algorithm can be described by the pseudo-code :

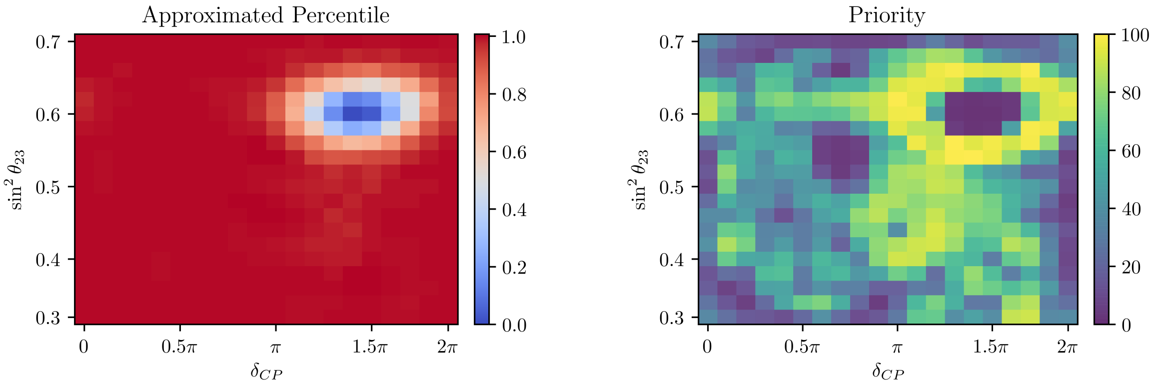


Figure 6.12: The approximate percentile surfaces in the Inverted Hierarchy (left) along with the acquistion function priority (right) for the toy study described below

Algorithm 1 $\mathcal{GP}$ iterative confidence interval construction

```
for each iteration $t = 1, 2, ...$ do
    Propose points in parameter space $\arg\max_\theta a(\theta)$
    for each point $\theta'$ do
        Simulate likelihood ratio statistic distribution
        for $k = 1, 2, ...$ do
            Perform a pseudo experiment
            Maximize the likelihood with respect to $(\theta, \delta)$
            Maximize the likelihood with constraint $\theta = \theta'$
            Calculate likelihood ratio statistic
        end for
        Calculate $p$-value based on the simulated distribution
    end for
    Train $\mathcal{GP}$ approximation $f(\theta)$ for the $p$-values
    Update confidence intervals
end for
```

6.3.3 Toy Studies

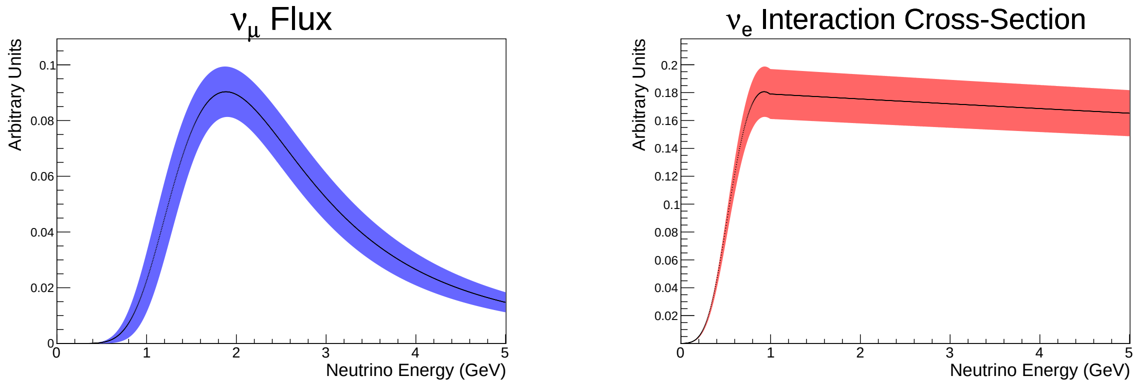


Figure 6.13: Lineshapes for the flux and cross-section used in the toy oscillation study to mimic NOvA shown along with its systematic error band

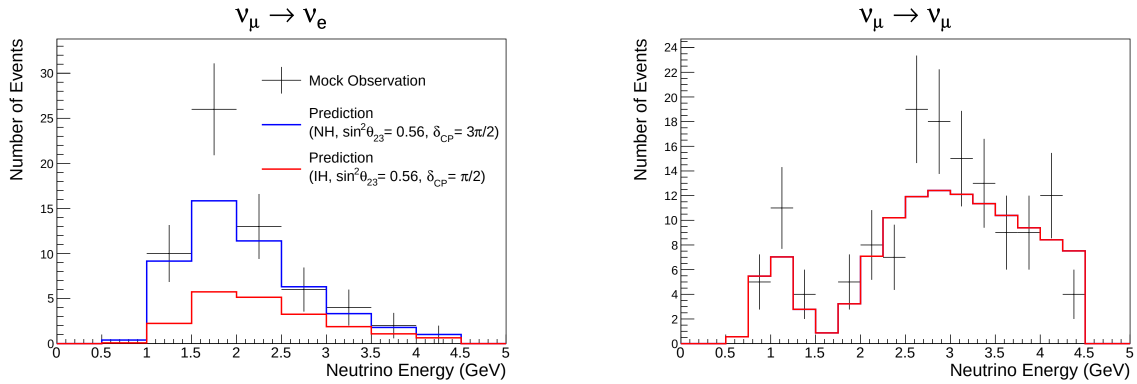


Figure 6.14: Mock datasets generated based on Poisson fluctuations of the expectation at a given set of oscillation parameters in the toy study. Shown for the appearance channel (left) and the disappearance channel (right)

To show that this algorithm works, we first implement this procedure on a toy NOvA-like long baseline oscillation experiment [155]. Here toy shapes approximating the NOvA flux and neutrino interaction cross-sections (shown in Fig. 6.13) are used to generate predictions for both oscillation channels at some given oscillation parameters and a baseline of 810 km. Matter effects are taken into account in the oscillation probability by assuming the Earth's crust density of 2.84 g/cm^3 as is the case for NOvA. A systematic uncertainty of 10% is taken on the flux and the cross-section as a normalization error on the prediction. This is then fit to a mock dataset generated by a Poisson fluctuation on top of the prediction (shown in Fig. 6.14) at the parameters :

- $\Delta m^2_{32} = 2.44 \times 10^{-3}$ eV^2
- $\sin^2\theta_{23} = 0.56$
- $\delta_{CP} = 1.5\pi$

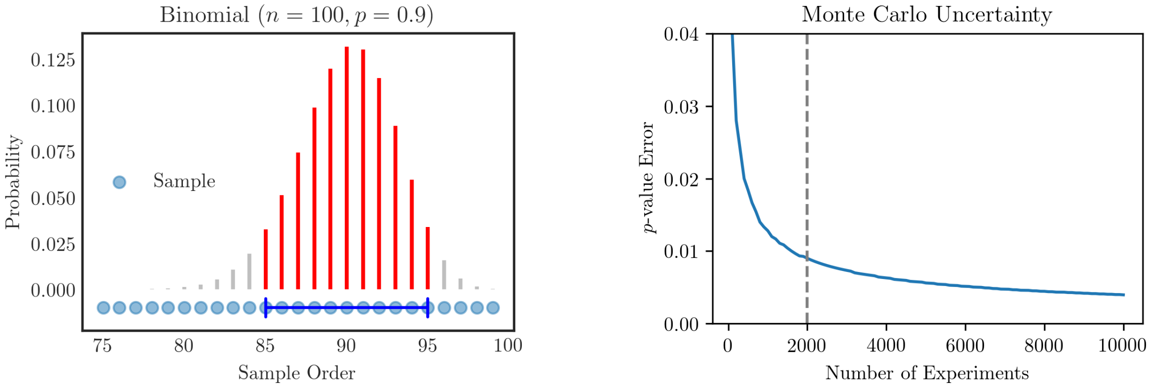


Figure 6.15: Illustration of the binomial proportion confidence interval for a given p-value calculated based on a certain number of pseudo-experiments. The right plot shows the p-value error as a function of the number of pseudo-experiments

The confidence intervals for these parameters are constructed, both in 1D and 2D, by a regular application of the FC procedure with 2000 pseudo-experiments at each point in parameter space. Only 68% and 90% intervals are constructed. More pseudo-experiments would have been needed to go further than this, which is avoided in the toy study for the sake of computational feasibility. The $\mathcal{GP}$ kernel chosen is a modification of the RBF kernel with :

$$\kappa(x, x') = \exp\left(\frac{-(x - x')^2}{2l^2}\right) + \sigma_p^2 \delta_{xx'} \tag{6.17}$$

The second term is just an extra error added to the diagonal of the kernel matrix in order to take into account the stochastic error from the p-value calculation using a fraction of pseudo-experiments. This error is therefore given by a binomial proportion confidence interval, as described in Fig. 6.15. For now, nuisance parameters aren't profiled but rather marginalised, i.e integrated over some prior distribution for those nuisance parameters. The priors chosen are flat across $(2, 3)$ $\times 10^{-3}$ eV^2 for $|\Delta m^2_{32}|$ and $(0.3, 0.7)$ for $\sin^2\theta_{23}$. The prior on the nuisance parameters for the systematic uncertainties in the toy model is assumed to be a standard normal distribution. This is done for computational feasibility and will not alter the conclusions. For 2D contours, a 20×20 grid is assumed for both the FC and the $\mathcal{GP}$ calculations.

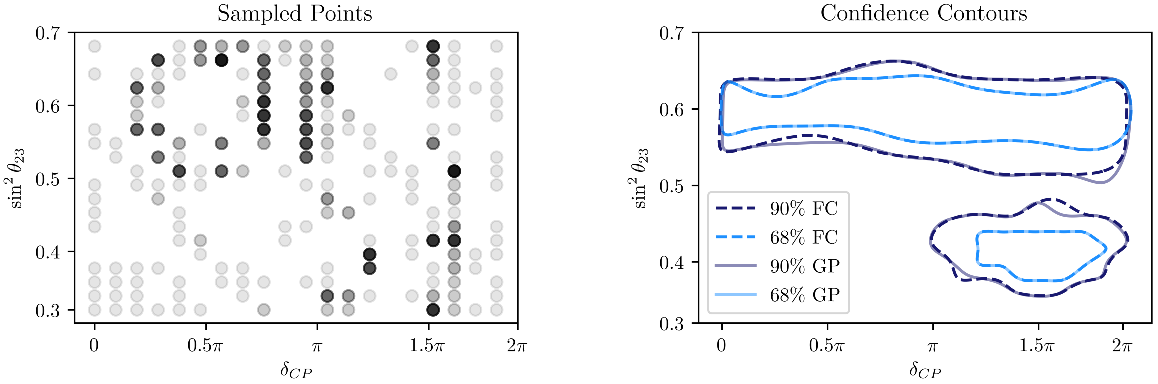


Figure 6.16: Results from the $\mathcal{GP}$ algorithm for approximate contours in $\sin^2\theta_{23} - \delta_{CP}$ space for the Normal Hierarchy. On the left are the points where the pseudo-experiments were thrown with the color scale denoting the fraction with respect to a traditional FC application. On the right are comparisons between the $\mathcal{GP}$-approximated contours and the FC contours

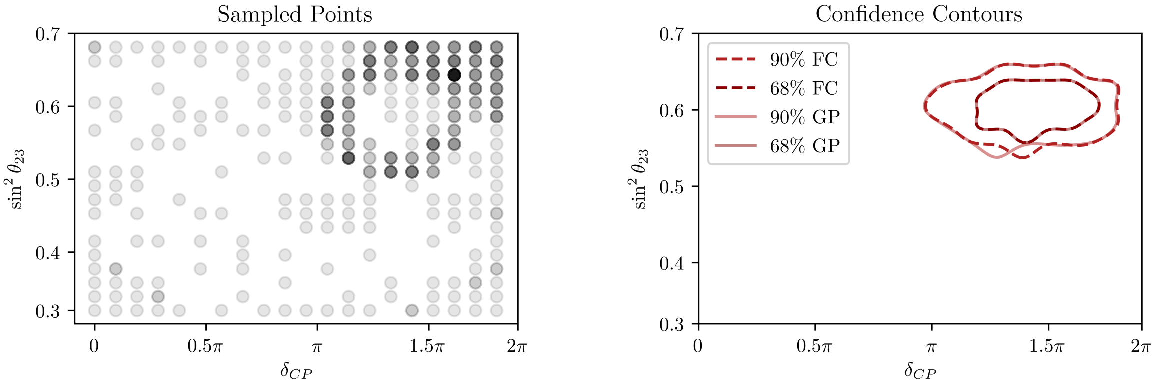


Figure 6.17: Results from the $\mathcal{GP}$ algorithm for approximate contours in $\sin^2\theta_{23} - \delta_{CP}$ space for the Inverted Hierarchy. On the left are the points where the pseudo-experiments were thrown with the color scale denoting the fraction with respect to a traditional FC application. On the right are comparisons between the $\mathcal{GP}$-approximated contours and the FC contours

In each iteration of the $\mathcal{GP}$ calculation, 40 points with the highest acquisition function score are chosen to throw 200 new pseudo-experiments. A point is exhausted after a few iterations

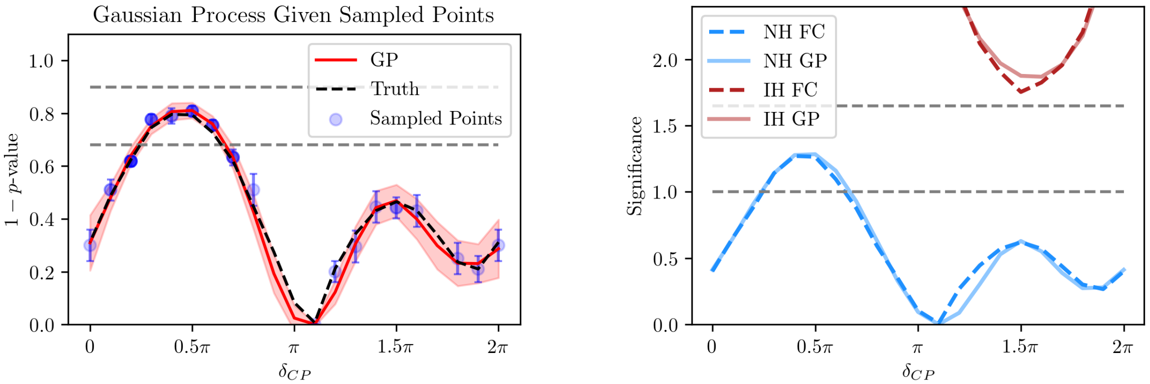


Figure 6.18: Comparison of the $\mathcal{GP}$-based significance curves as a function of δ_{CP} vs the FC result in the toy study. On the left is the comparison in terms of the p-value and on the right is the significance score. The shaded regions represent the error in the $\mathcal{GP}$-approximation and the blue points with the color scale represent where the pseudo-experiments were thrown and its fraction

if 2000 pseudo-experiments are thrown. For the 1D case, since there are only 20 points on the grid, 10 points are prioritized each iteration, again with 200 new pseudo-experiments per iteration per point chosen. In the first iteration, points are chosen at random.

As a result, the $\mathcal{GP}$ and regular grid-search FC results are compared and show great agreement for both 1D and 2D cases, as can be seen from Figs. 6.16 - 6.18. In the 2D case, it is notable that the $\mathcal{GP}$ algorithm is also able to capture disconnected regions in the confidence intervals. In the 1D case, the exact significance of rejection for each point of δ_{CP} are compared and also shows good agreement. The 68% confidence intervals for those parameters can then be constructed by collecting all points of the parameter which have a significance below 1σ, which matches the result from regular FC.

If the FC procedure is repeated for 200 different mock datasets and compared to the $\mathcal{GP}$ approximation, one can study the distribution of a metric that is essentially a “classification” accuracy, i.e the proportion of points in parameter space that are correctly identified by the $\mathcal{GP}$ algorithm to fall into its true confidence interval. One sees from Fig. 6.19 that within 5 iterations for the 2D contours, we get a median accuracy of $> 99.5\%$ for both NH and IH

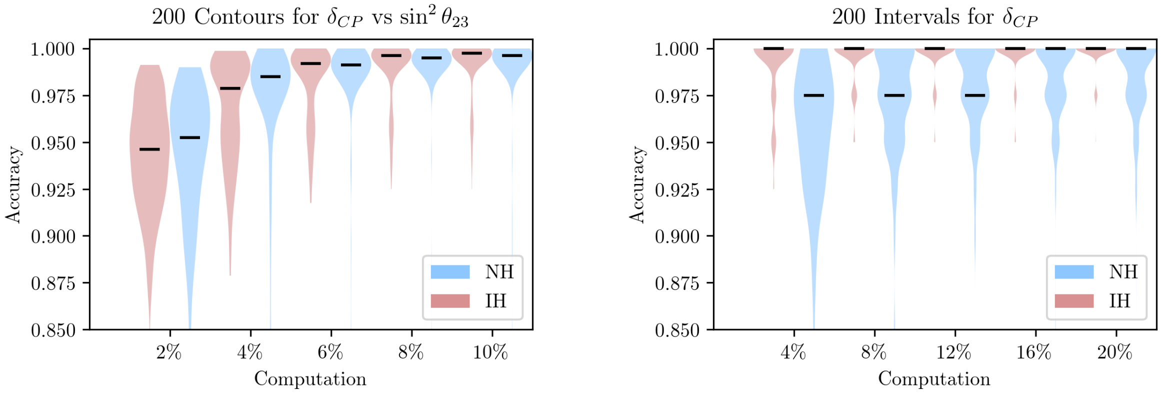


Figure 6.19: Distribution of accuracies of the $\mathcal{GP}$-based approximated confidence intervals for 200 different mock datasets as a function of fractional number of pseudo-experiments thrown. On the left is for the 2D contours and on the right is for the 1D δ_{CP} slice. The black bars represent medians of the accuracy distributions

in the $\sin^2\theta_{23} - \delta_{CP}$ contours after 5 iterations. For 1D, it essentially reaches 100% since all points on the 1D grid are correctly identified. Five iterations imply a speed-up of 10× for the 2D case and 5× for the 1D case, based on the fraction of pseudo-experiment budget used up by the $\mathcal{GP}$ algorithm. One sees therefore, that without sacrificing any accuracy in the confidence interval construction, one can have an order of magnitude gain in speed! Most of this speed comes from skipping a large number of points on the 2D grid where the approximate p-value is already known to be small and can be thus rejected at high confidence levels. Therefore, with this promising proof-of-concept, one can then study its application on the actual confidence intervals constructed using real NOvA data and simulation.

6.3.4 Application on NOvA Data

Some modifications are made from the previous section to study the $\mathcal{GP}$ algorithm on the actual NOvA 1D and 2D contours. The biggest one is the kernel choice, where now, the RQF kernel is chosen in order to capture variations in the contours across different length scales. The 2D NOvA contours are also produced using a 30×30 grid, while the 1D is produced using a 30-point grid. As a result, the $\mathcal{GP}$ algorithm proceeds by choosing 90 points having

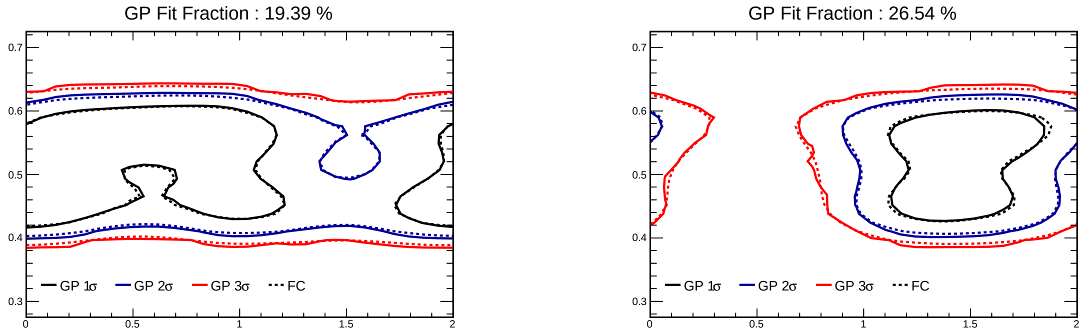


Figure 6.20: Comparison of the $\mathcal{GP}$ approximate contours after 4 iterations vs the actual NOvA contours in $\sin^2\theta_{23} - \delta_{CP}$ space. On the left is for Normal Hierarchy and on the right is Inverted Hierarchy

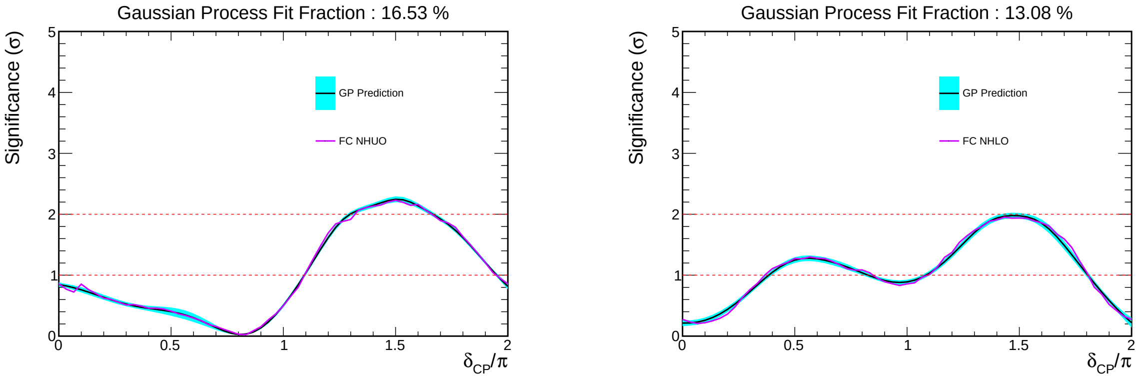


Figure 6.21: Comparison of $\mathcal{GP}$ approximated significance curves for δ_{CP} vs the actual NOvA result. Also shown in blue is the $\mathcal{GP}$ prediction error. Left is for NHUO and right is for NHLO

the maximum acquisition score in the 2D case. In the 1D case, 10 points are chosen. In the 2D case for each iteration per point, 25% of the actual FC pseudo-experiment budget is used in order to ensure atleast some pseudo-experiments with a $\Delta\chi^2$ larger than the one observed with real data. In the 1D case as well, a fraction of pseudo-experiments are chosen for each iteration and point in the grid, so the above condition is satisfied.

As can be seen from Figs. 6.20, for the $\sin^2\theta_{23} - \delta_{CP}$ contours assuming NH and IH separately, we have good agreement for the 1σ, 2σ and 3σ contours with enough freedom to capture all regions of interest, including those near $\delta_{CP} = 3\pi/2$ for NH. One can also draw the upper and lower bounds for the $\mathcal{GP}$ approximated contours based on the $\mathcal{GP}$ error (shown in Figs.

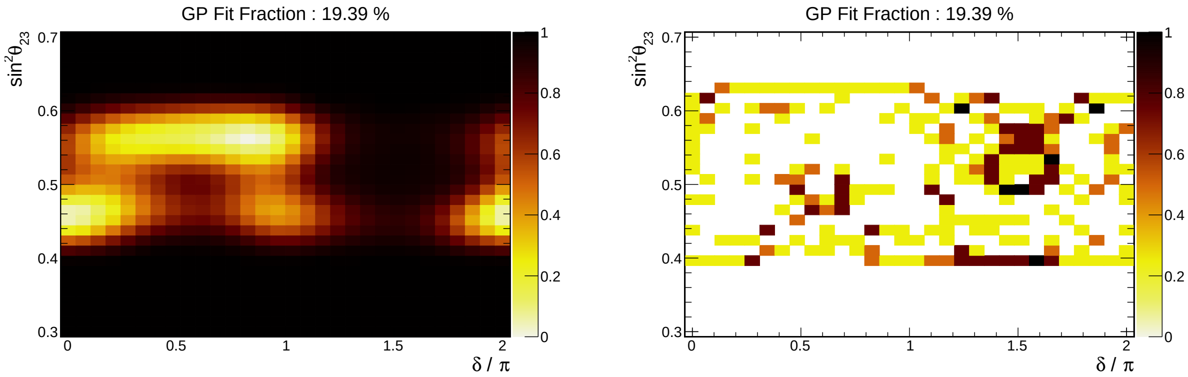


Figure 6.22: The approximate p-value surface in the $\mathcal{GP}$ procedure for NH after 4 iterations (left) and the points where the pseudo-experiments were thrown (right)

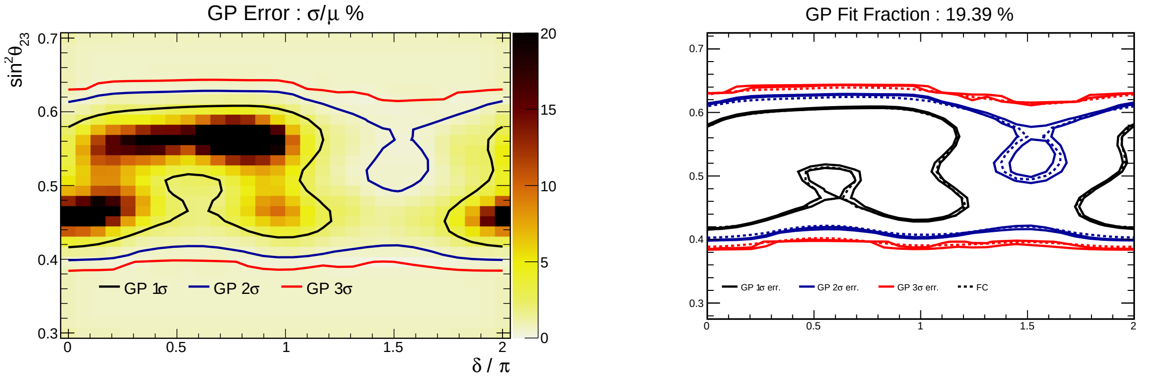


Figure 6.23: The $\mathcal{GP}$ approximate contours shown along with the error in the p-value prediction (left). On the right is the approximate contour bounds from the $\mathcal{GP}$ prediction for NH

6.23 and 6.25), which shows these bounds close iteratively to $< 5\%$ precision near the actual FC contour boundaries. As a result, the $\mathcal{GP}$ approximated contours fully enclose the original contours and are within a couple of linewidth of the original for most regions. This is seen within 4 iterations with $\sim 4 - 5\times$ the speed-up.

For the 1D case as well (shown in Fig. 6.21), one can see the significances of rejection for δ_{CP} for all 4 hypotheses of (NH, IH) × (LO, UO) within a similar number of iterations and fractional speed-up. This shows that the method is very well suited to actual confidence intervals as well.

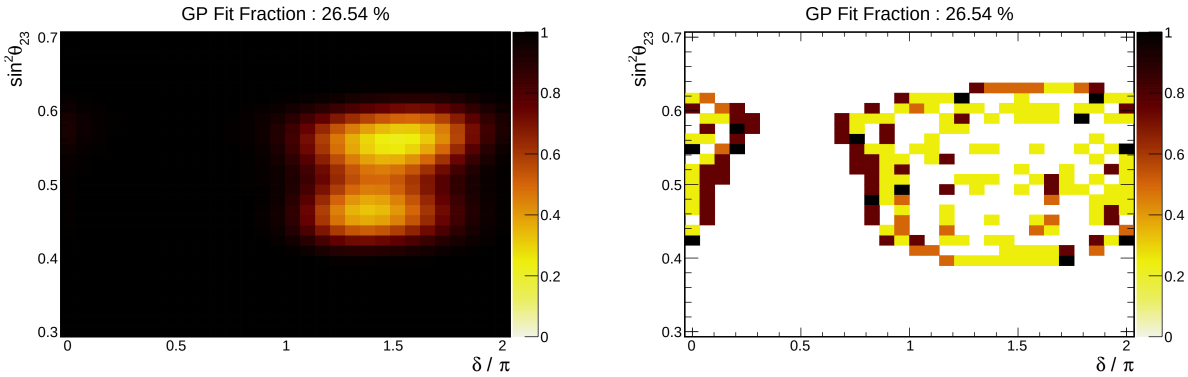


Figure 6.24: The approximate p-value surface in the $\mathcal{GP}$ procedure for IH after 4 iterations (left) and the points where the pseudo-experiments were thrown (right)

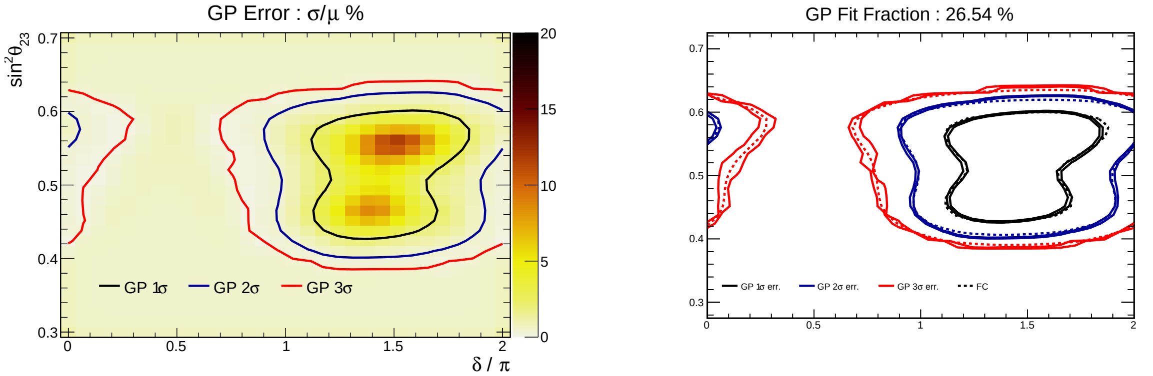


Figure 6.25: The $\mathcal{GP}$ approximate contours shown along with the error in the p-value prediction (left). On the right is the approximate contour bounds from the $\mathcal{GP}$ prediction for IH

6.3.5 Future Prospects for the $\mathcal{GP}$ Algorithm

We've seen that we we are able to derive real confidence intervals for the oscillation parameters at a fraction of the computational cost associated with a traditional application of the Feldman-Cousins procedure. A number of things can still be optimized further. For example, the fraction of pseudo-experiments chosen per iteration, the form of the acquistion function itself for 1D curves. In addition, one can conceivably incorporate prior information on the confidence interval boundaries based on the χ^2_k approximation itself. A large focus, however, will need to be on the implementation of this algorithm at NERSC. The algorithm is structurally serial in nature and optimization procedures will need to be suitably adapted for paralllelizing the algorithm on a distributed system like NERSC.

In terms of physics, this approach has a lot of potential in solving one of the key problems with the FC procedure which is widely used in a variety of contexts, especially in neutrino oscillation experiments. It provides a nice synthesis of Bayesian and frequentist ideas, albeit to construct confidence intervals that still have a frequentist interpretation. With a more intelligent and statistically well-motivated approach to confidence intervals compared to the normal brute force grid search and without sacrificing accuracy, one hopes that this technique will become common in the future.

6.4 Future Prospects at NOvA

NOvA is one of the flagship long-baseline neutrino oscillation experiments that will be able to make strong statements about atleast the neutrino mass hierarchy and the octant of θ_{23}. Along with T2K, the results would define the state of the field until next generation experiments like DUNE and HyperK come online.

NOvA is scheduled to run until 2026, with around two-thirds of the planned data still to be collected. NuMI beam upgrades are planned to boost the intensity of the neutrino beam upto 900 kW. It is expected that NOvA will run in 50% neutrino mode and 50% anti-neutrino mode, with a total planned exposure of 31.5×10^{20} for each mode. If the true oscillation parameters are favorable, NOvA expects to constrain the mass hierarchy at more than 3σ (shown in Fig. 6.26), possibly even $4 - 5\sigma$ if we are lucky.

The analysis technique described here is powerful enough to show that the dominant errors in the oscillation parameter measurement is statistical rather than systematic. However, more improvements are planned in order to reduce the systematic burden, especially in the detector response, which are some of the largest. A test-beam detector has been built and

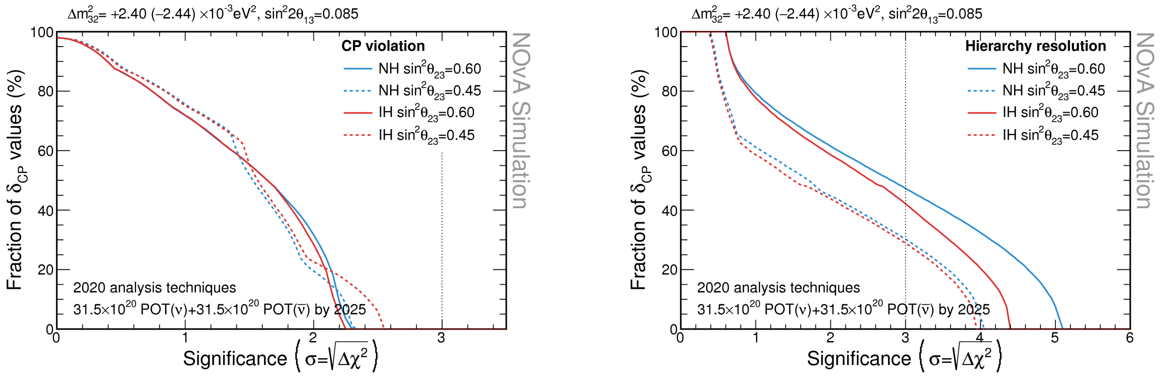


Figure 6.26: Sensitivities for the measurement of CP-violation (left) and Mass Hierarchy (right) at the end of the NOvA lifespan shown for different choices of true parameters

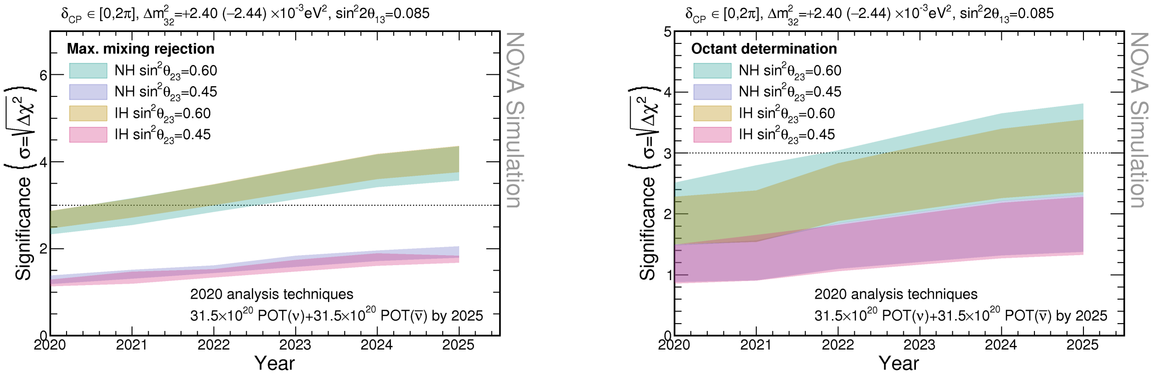


Figure 6.27: Potential for rejecting maximal mixing (left) and determination of the θ_{23} Octant (right) based on current analysis techniques at NOvA

has already started collecting data from a beam of hadrons and leptons. This will be vital in order to constrain the detector response uncertainties [156]. Further improvements related to the application of advanced machine learning techniques like for CVN will further improve the statistical error on the measurement [157]. We've discussed some of these improvements including a deep learning approach for a ν_e-CC energy estimator. Similar techniques are also being studied for ν_μ-CC as well.

Finally, the joint fit with T2K should also clarify the interesting tension that seems to be developing between the two experiments and their respective measurements. This will need a comprehensive study on the correlations between the various systematic models used by the individual experiments. However, the prospects for the joint NOvA-T2K measurements are big, with a potential for evidence of many interesting questions related to 3-flavor neutrino oscillations, including CP-violation, neutrino mass hierarchy and whether θ_{23} is maximal. All of these will be further refined by DUNE and HyperK, which should definitely rule out some of the different hypotheses. One can therefore be optimistic that the 3-flavor PMNS model for neutrino oscillations will be completed within the next decade. This is a huge achievement, being a culmination of decades of experimental work.

Chapter 7

Conclusions

This dissertation describes the analysis of NOvA data with an exposure of 13.60×10^{20} POT and 12.50×10^{20} POT in the neutrino beam mode and anti-neutrino beam mode of operation respectively, using the NuMI beam at Fermilab. This is done using a 14 kt Far Detector placed 810 km from the NuMI beam source along with a 300 t Near Detector placed 1 km from the beam source. The analysis is a measurement of the remaining unknowns in the 3-flavor PMNS model of neutrino oscillations, i.e for the parameters Δm^2_{32}, $\sin^2\theta_{23}$ and δ_{CP}. The measurement is done using a joint fit of $\nu_\mu \rightarrow \nu_\mu$ ($\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$) disappearance and $\nu_\mu \rightarrow \nu_e$ ($\bar{\nu}_\mu \rightarrow \bar{\nu}_e$) appearance.

As a result, the measurements are given by :

- $\Delta m^2_{32} = (2.41 \pm 0.07) \times 10^{-3}$ eV^2
- $\sin^2\theta_{23} = 0.57^{+0.04}_{-0.03}$ (UO)
- $\delta_{CP} = 0.82\pi^{+0.27\pi}_{-0.87\pi}$

A 4.2σ evidence is found for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations which hasn't been seen yet by any other experiment. The data weakly favors the normal neutrino mass hierarchy and the upper octant of θ_{23} by 1.0σ and 1.2σ respectively. The inverted mass hierarchy and $\delta_{CP} = \pi/2$ is rejected at $> 3\sigma$. The normal mass hierarchy for $\delta_{CP} = 3\pi/2$ is rejected at $\sim 2\sigma$ which is in tension with measurements done by the T2K experiment for those same parameters.

This author has contributed to various aspects of the analysis, including the detector attenuation calibration, the event reconstruction, the development of particle identification, cosmic rejection and the selection tuning for the Far Detector oscillation analysis samples, systematic studies, in particular for the flux model and sideband studies. Future improvements to the analysis also discussed here include a deep-learning based energy estimator for ν_e-CC events as well as algorithmic modifications to reduce the computational burden in producing confidence intervals.

Bibliography

[1] Enrico Fermi. Tentativo di una teoria dei raggi β. *Il Nuovo Cimento (1924-1942)*, 11(1):1–19, 1934.

[2] James Chadwick. Possible existence of a neutron. *Nature*, 129(3252):312–312, 1932.

[3] Wolfgang Pauli. Pauli Letter Collection: Letter to Lise Meitner. Technical report, 1930.

[4] DB Chitwood et al. Improved measurement of the positive-muon lifetime and determination of the Fermi constant. *Physical Review Letters*, 99(3):032001, 2007.

[5] Hans Bethe and Rudolph Peierls. The Neutrino. *Nature*, 133(3366):689–690, 1934.

[6] Clyde L Cowan, Frederick Reines, FB Harrison, HW Kruse, and AD McGuire. Detection of the free neutrino: a confirmation. *Science*, 124(3212):103–104, 1956.

[7] Tsung-Dao Lee and Chen-Ning Yang. Question of parity conservation in weak interactions. *Physical Review*, 104(1):254, 1956.

[8] Chien-Shiung Wu, Ernest Ambler, Raymond W Hayward, Dale D Hoppes, and Ralph Percy Hudson. Experimental test of parity conservation in beta decay. *Physical Review*, 105(4):1413, 1957.

[9] Maurice Goldhaber, Lee Grodzins, and Andrew W Sunyar. Helicity of neutrinos. *Physical Review*, 109(3):1015, 1958.

[10] Gaillard Danby, Jean Maurice Gaillard, Konstantin Goulianos, Leon M Lederman, Nariman Mistry, Melvin Schwartz, and Jack Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Physical Review Letters*, 9(1):36, 1962.

[11] K Kodama, N Saoulidou, G Tzanakos, B Baller, B Lundberg, R Rameika, JS Song, CS Yoon, SH Chung, S Aoki, et al. Detection and analysis of tau–neutrino interactions in DONUT emulsion target. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 493(1-2):45–66, 2002.

[12] Chen-Ning Yang and Robert L Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical Review*, 96(1):191, 1954.

[13] Sheldon L Glashow. Partial-symmetries of weak interactions. *Nuclear physics*, 22(4):579–588, 1961.

[14] Abdus Salam and John Clive Ward. Electromagnetic and weak interactions. Technical report, Imperial College of Science and Technology, London, 1964.

[15] J Goldstone. Field theories with "superconductor" solutions. *Nuovo Cimento*, 19:154–164, Aug 1960.

[16] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[17] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[19] Steven Weinberg. A model of leptons. *Physical review letters*, 19(21):1264, 1967.

[20] FJ Hasert et al. Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment. *Nuclear Physics B*, 73(1):1–22, 1974.

[21] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.

[22] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973.

[23] J. D. Bjorken. Asymptotic sum rules at infinite momentum. *Phys. Rev.*, 179:1547–1553, Mar 1969.

[24] Jerome I. Friedman and Henry W. Kendall. Deep inelastic electron scattering. *Ann. Rev. Nucl. Part. Sci.*, 22:203–254, 1972.

[25] Gerardus't Hooft. Renormalization of massless Yang-Mills fields. *Nuclear physics: B*, 33(1):173–199, 1971.

[26] G Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider. *Physics Letters B*, 126(5):398–410, 1983.

[27] J. Beringer et al. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012.

[28] Bruce T Cleveland, Timothy Daily, Raymond Davis Jr, et al. Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *The Astrophysical Journal*, 496(1):505, 1998.

[29] John N Bahcall, Walter F Huebner, Stephen H Lubow, Peter D Parker, and Roger K Ulrich. Standard solar models and the uncertainties in predicted capture rates of solar neutrinos. *Reviews of Modern Physics*, 54(3):767, 1982.

[30] W. Hampel et al. GALLEX solar neutrino observations: Results for GALLEX IV. *Phys. Lett. B*, 447:127–133, 1999.

[31] P. Hernandez. Neutrino Physics. *8th CERN–Latin-American School of High-Energy Physics*, 2016.

[32] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998.

[33] Q. R. Ahmad et al. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002.

[34] K. Eguchi et al. First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance. *Phys. Rev. Lett.*, 90:021802, Jan 2003.

[35] S. Abe et al. Precision Measurement of Neutrino Oscillation Parameters with KamLAND. *Phys. Rev. Lett.*, 100:221803, 2008.

[36] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[37] D. D'Angelo et al. Recent Borexino results and prospects for the near future. *EPJ Web Conf.*, 126:02008, 2016.

[38] L. Wolfenstein. Neutrino oscillations in matter. *Phys. Rev. D*, 17:2369–2374, May 1978.

[39] S. P. Mikheyev and A. Yu. Smirnov. Resonance Amplification of Oscillations in Matter and Spectroscopy of Solar Neutrinos. *Sov. J. Nucl. Phys.*, 42:913–917, 1985.

[40] Jose W. F. Valle and Jorge C. Romao. *Neutrinos in high energy and astroparticle physics*. Physics textbook. Wiley-VCH, Weinheim, 2015.

[41] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.

[42] Bruno Pontecorvo. Neutrino experiments and the problem of conservation of leptonic charge. *Sov. Phys. JETP*, 26(984-988):165, 1968.

[43] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Physical Review Letters*, 10(12):531, 1963.

[44] Makoto Kobayashi and Toshihide Maskawa. CP-violation in the renormalizable theory of weak interaction. *Progress of theoretical physics*, 49(2):652–657, 1973.

[45] Stefan Antusch, Carla Biggio, Enrique Fernandez-Martinez, M Belen Gavela, and Jacobo Lopez-Pavon. Unitarity of the leptonic mixing matrix. *Journal of High Energy Physics*, 2006(10):084, 2006.

[46] Stefan Antusch and Oliver Fischer. Non-unitarity of the leptonic mixing matrix: Present bounds and future sensitivities. *Journal of High Energy Physics*, 2014(10):94, 2014.

[47] CERN Courier. Neutrino production moves to an industrial scale. 2012.

[48] FP An et al. Observation of electron-antineutrino disappearance at Daya Bay. *Physical Review Letters*, 108(17):171803, 2012.

[49] Jung Keun Ahn et al. Observation of reactor electron antineutrinos disappearance in the RENO experiment. *Physical Review Letters*, 108(19):191802, 2012.

[50] Y Abe et al. Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector. *Journal of High Energy Physics*, 2014(10):86, 2014.

[51] B. Abi et al. Long-baseline neutrino oscillation physics potential of the DUNE experiment. *Eur. Phys. J. C*, 80(10):978, 2020.

[52] K. Abe et al. Physics potential of a long-baseline neutrino oscillation experiment using a J-PARC neutrino beam and Hyper-Kamiokande. *PTEP*, 2015:053C02, 2015.

[53] RB Patterson. The NOvA experiment: status and outlook. *Nuclear Physics B-Proceedings Supplements*, 235:151–157, 2013.

[54] K Abe et al. The T2K experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 659(1):106–135, 2011.

[55] R Acciarri et al. Long-baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report, Volume 4 The DUNE detectors at LBNF. *arXiv preprint arXiv:1601.02984*, 2016.

[56] K Abe et al. Letter of Intent: The Hyper-Kamiokande Experiment—Detector Design and Physics Potential—. *arXiv preprint arXiv:1109.3262*, 2011.

[57] DV Forero, S Morisi, M Tortola, and JWF Valle. Lepton flavor violation and non-unitary lepton mixing in low-scale type-i seesaw. *Journal of High Energy Physics*, 2011(9):142, 2011.

[58] P. Adamson, F. P. An, I. Anghel, A. Aurisano, et al. Limits on Active to Sterile Neutrino Oscillations from Disappearance Searches in the MINOS, Daya Bay, and Bugey-3 Experiments. *Phys. Rev. Lett.*, 117:151801, Oct 2016.

[59] Susanne Mertens. KATRIN : Recent Results and Future Perspectives. *The XXIX International Conference on Neutrino Physics and Astrophysics.*

[60] J. Wolf. The KATRIN Neutrino Mass Experiment. *Nucl. Instrum. Meth. A*, 623:442–444, 2010.

[61] A. Gando et al. Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with KamLAND-Zen. *Phys. Rev. Lett.*, 117:082503, Aug 2016.

[62] S. Pascoli, S. T. Petcov, and Antonio Riotto. Connecting low energy leptonic CP violation to leptogenesis. *Phys. Rev. D*, 75:083511, Apr 2007.

[63] S. Mihara, J. P. Miller, P. Paradisi, and G. Piredda. Charged Lepton Flavor-Violation Experiments. *Ann. Rev. Nucl. Part. Sci.*, 63:531–552, 2013.

[64] Todd Haines et al. Neutrinos From SN1987A in the IMB Detector. *Nucl. Instrum. Meth. A*, 264:28–31, 1988.

[65] K. Hirata et al. Observation of a neutrino burst from the supernova SN1987A. *Phys. Rev. Lett.*, 58:1490–1493, Apr 1987.

[66] E. N. Alekseev, L. N. Alekseeva, I. V. Krivosheina, and V. I. Volchenko. Detection of the Neutrino Signal From SN1987A in the LMC Using the INR Baksan Underground Scintillation Telescope. *Phys. Lett. B*, 205:209–214, 1988.

[67] Antonio Capone, Paolo Lipari, and Francesco Vissani. Neutrino Astronomy. *Multiple Messengers and Challenges in Astroparticle Physics*, 2018.

[68] C. Simpson et al. Sensitivity of Super-Kamiokande with Gadolinium to Low Energy Anti-neutrinos from Pre-supernova Emission. *Astrophys. J.*, 885:133, 2019.

[69] T. Kafka. MINOS Experiment at Fermilab. *Prog. Part. Nucl. Phys.*, 64:184–186, 2010.

[70] Fermilab Creative Services. Image Gallery. (https://vms.fnal.gov/asset).

[71] P. Adamson et al. The NuMI Neutrino Beam. *Nucl. Instrum. Meth. A*, 806:279–306, 2016.

[72] L. Aliaga et al. Neutrino Flux Predictions for the NuMI Beam. *Phys. Rev. D*, 94(9):092005, 2016. [Addendum: Phys.Rev.D 95, 039903 (2017)].

[73] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[74] L. Aliaga. PPFX tech-note for the 2017 analysis. *NOvA Internal Document*, (Docdb 23441).

[75] C. Alt et al. Inclusive production of charged pions in p+C collisions at 158-GeV/c beam momentum. *Eur. Phys. J. C*, 49:897–917, 2007.

[76] J. M. Paley et al. Measurement of charged pion production yields off the NuMI target. *Phys. Rev. D*, 90:032001, Aug 2014.

[77] Daniel Stephen Pershey. *A Measurement of ν_e Appearance and ν_μ Disappearance Neutrino Oscillations with the NOvA Experiment.* PhD thesis, Caltech, 2018.

[78] D. S. Ayres et al. The NOvA Technical Design Report. 10 2007.

[79] Micah C. Groh. Constraints on Neutrino Oscillation Parameters from Neutrinos and Antineutrinos with Machine Learning. 2 2021.

[80] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth. A*, 614:87–104, 2010.

[81] Grace Song. Model Tests of Inclusive CC Proton Production Over Full 5D Phase Space. (https://www.phys.ksu.edu/reu/2018/song.html).

[82] J. A. Formaggio and G. P. Zeller. From eV to EeV: Neutrino cross sections across energy scales. *Rev. Mod. Phys.*, 84:1307–1341, Sep 2012.

[83] J. Nieves, J. E. Amaro, and M. Valverde. Inclusive quasielastic charged-current neutrino-nucleus reactions. *Phys. Rev. C*, 70:055503, Nov 2004.

[84] Aaron S. Meyer, Minerba Betancourt, Richard Gran, and Richard J. Hill. Deuterium target data for precision neutrino-nucleus cross sections. *Phys. Rev. D*, 93(11):113015, 2016.

[85] R. Gran, J. Nieves, F. Sanchez, and M. J. Vicente Vacas. Neutrino-nucleus quasi-elastic and 2p2h interactions up to 10 GeV. *Phys. Rev. D*, 88(11):113007, 2013.

[86] Ch. Berger and L. M. Sehgal. Lepton mass effects in single pion production by neutrinos. *Phys. Rev. D*, 76:113004, 2007.

[87] A. Bodek and U. K. Yang. Modeling deep inelastic cross-sections in the few GeV region. *Nucl. Phys. B Proc. Suppl.*, 112:70–76, 2002.

[88] E. Oset, L. L. Salcedo, and D. Strottman. A Theoretical Approach to Pion Nuclear Reactions in the Resonance Region. *Phys. Lett. B*, 165:13–18, 1985.

[89] A. Aurisano, C. Backhouse, R. Hatcher, N. Mayer, J. Musser, R. Patterson, R. Schroeter, and A. Sousa. The NOvA simulation chain. *J. Phys. Conf. Ser.*, 664(7):072002, 2015.

[90] A. Aurisano et al. The 2017 NOvA Detector Simulation . *NOvA Internal Document*, (Docdb 23697).

[91] Chris Hagmann, David Lange, and Douglas Wright. Cosmic-ray shower generator (CRY) for Monte Carlo transport codes. 2:1143–1146, 2007.

[92] J. Blair. Fiber Brightness Technote. *NOvA Internal Document*, (Docdb 35674).

[93] T. Alion et al. Calibration Technotes. *NOvA Internal Document*, (Docdb 13579).

[94] M. Groh and F. Psihas. Technote Pi0 Mass Peak in RHC and FHC Data using Prong CVN. *NOvA Internal Document*, (Docdb 23817).

[95] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[96] M. Baird. Slicing Module Comparison Tech Note . *NOvA Internal Document*, (Docdb 9195).

[97] D. Pershey, J. Huang, and M. Judah. TDSlicer Technote. *NOvA Internal Document*, (Docdb 27689).

[98] R. C. Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957.

[99] N. Nayak and T. Alion. Technote: Comparing TDSlicer to Slicer4D using the ND Overlay sample. *NOvA Internal Document*, (Docdb 34597).

[100] M. Baird, J. Bian, M. Messier, E. Niner, D. Rocco, and K. Sachdev. Event Reconstruction Techniques in NOvA. *J. Phys. Conf. Ser.*, 664(7):072035, 2015.

[101] Richard O. Duda and Peter E. Hart. Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Commun. ACM*, 15(1):11–15, January 1972.

[102] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[103] A Aurisano et al. A convolutional neural network neutrino event classifier. *Journal of Instrumentation*, 11(09):P09001, 2016.

[104] M. Groh and T. Warburton. Technote: CVN Training for the 2020 Analysis. *NOvA Internal Document*, (Docdb 42897).

[105] N. Nayak and A. Back. Technote: event CVN validation for Production 5. *NOvA Internal Document*, (Docdb 43008).

[106] M. Wei. NuMu Energy Estimator Technote for Prod5 MC. *NOvA Internal Document*, (Docdb 43757).

[107] J. Porter. ReMId retraining 2019 technote. *NOvA Internal Document*, (Docdb 42277).

[108] S. Yu. Ana2020 Nue Energy Estimator Technote. *NOvA Internal Document*, (Docdb 43814).

[109] F. Psihas et al. Context-enriched identification of particles with a convolutional network for neutrino events. *Phys. Rev. D*, 100:073005, Oct 2019.

[110] Pierre Baldi, Jianming Bian, Lars Hertel, and Lingge Li. Improved energy reconstruction in NOvA with regression convolutional neural networks. *Physical Review D*, 99(1):012011, 2019.

[111] N. Nayak, L. Hertel, and J. Bian. RegCVN - Nue Energy Estimator Technote. *NOvA Internal Document*, (Docdb 42363).

[112] D. Torbunov. LSTM NuMu Energy Estimator Technote for Prod5. *NOvA Internal Document*, (Docdb 42304).

[113] N. Nayak, T. Warburton, E. Catano-Mur, and A. Sutton. Event selection for the 2020 3-flavor analysis. *NOvA Internal Document*, (Docdb 44040).

[114] A. Back et al. Data quality 2020 technote. *NOvA Internal Document*, (Docdb 44226).

[115] Moren, A. and Himmel, A. and Niner, E. *NOvA Internal Document*, (Docdb 42766).

[116] Luke Vinton. *Measurement of Muon Neutrino Disappearance with the NOvA Experiment.* PhD thesis, Sussex U., 2018.

[117] R. Gandrajula and L. Kolupaeva. Cosmic rejection BDT for ν_e appearance analysis. *NOvA Internal Document*, (Docdb 42473).

[118] N. Nayak and J. Bian. Technote: Cosmic Rejection For Detached Bremsstrahlung Showers. *NOvA Internal Document*, (Docdb 23141).

[119] K. Bays et al. Cross-section tuning 2020 tech note. *NOvA Internal Document*, (Docdb 43962).

[120] M. A. Acero et al. Adjusting neutrino interaction models and evaluating uncertainties using NOvA near detector data. *Eur. Phys. J. C*, 80(12):1119, 2020.

[121] G. Kafka. Far Over Near Extrapolation Tech Note. *NOvA Internal Document*, (Docdb 9579).

[122] A. Mislevic. Near-to-Far Extrapolation in Transverse Momentum for the NOvA 2020 3-Flavor Analysis. *NOvA Internal Document*, (Docdb 44401).

[123] M. Elkins. BEN+Michel Decomposition 2020 Tech Note. *NOvA Internal Document*, (Docdb 43712).

[124] R. Barlow. Extended Maximum Likelihood. *Nuclear Instruments and Methods in Physics, Volume 293, Issue 3*, 1990.

[125] R. Brun et al. TMinuit Class Reference. *ROOT Documentation.*

[126] L. Cremonesi. Beam systematic uncertainties for third analyses. *NOvA Internal Document*, (Docdb 17608).

[127] N. Nayak. PCA Flux Systematics for the 2018 Oscillation Analyses. *NOvA Internal Document*, (Docdb 27884).

[128] N. Nayak. PCA Flux Systematics for the 2020 Oscillation Analysis. *NOvA Internal Document*, (Docdb 44049).

[129] C. H. Llewellyn Smith. Neutrino Reactions at Accelerator Energies. *Phys. Rept.*, 3:261–379, 1972.

[130] Richard Gran. Model Uncertainties for Valencia RPA Effect for MINERvA. 5 2017.

[131] M. Dolce, J. Wolcott, and H. Gallagher. GENIE 3.0.6 hN CV and uncertainties for 2020 analysis. *NOvA Internal Document*, (Docdb 43724).

[132] T. Nosek. Genie Cross Section Systematics for the 2020 3Flavor Analysis Using PCA. *NOvA Internal Document*, (Docdb 44274).

[133] J. Hewes and O. Samoylov. NOvA Production 5 Detector Simulation Changes and Systematics. *NOvA Internal Document*, (Docdb 43935).

[134] R. Patterson and M. Wetstein. 2020 Neutron Systematic Tech Note. *NOvA Internal Document*, (Docdb 43811).

[135] M. Strait. Tech note: Muon Energy Scale Systematic. *NOvA Internal Document*, (Docdb 20816).

[136] Z. Vallari. Technote: Nue Acceptance Syst for Ana2020. *NOvA Internal Document*, (Docdb 43917).

[137] A. Booth. Evaluation of Systematic Uncertainty in Normalisation Due to Near Detector Data-MC Differences in Selection Efficiency. *NOvA Internal Document*, (Docdb 43838).

[138] J. Wolcott, L. Suter, and M. Baird. Executive summary for 2020 3-flavor analysis review. *NOvA Internal Document*, (Docdb 44422).

[139] A. Himmel. New Oscillation Results from the NOvA Experiment. *The XXIX International Conference on Neutrino Physics and Astrophysics*, 2020.

[140] Y. Zhang. FD numu sideband studies. *NOvA Internal Document*, (Docdb 44768).

[141] N. Nayak. 2020 FD Nue Sidebands - Final. *NOvA Internal Document*, (Docdb 45103).

[142] N. Nayak. Goodness of Fit 2020. *NOvA Internal Document*, (Docdb 45653).

[143] Gary J. Feldman and Robert D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, 1998.

[144] Luca Lista. *Statistical Methods for Data Analysis in Particle Physics*, volume 909. Springer, 2016.

[145] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals Math. Statist.*, 9(1):60–62, 1938.

[146] G. Feldman. Notes on the Inclusion of Nuisance Parameters in the Unified Approach. *NOvA Internal Document*, (Docdb 15884).

[147] Bedrich Roskovec. Latest Reactor Neutrino Oscillation Results from the Daya Bay Experiment. *PoS*, ICHEP2020:170, 2021.

[148] K. Abe et al. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. *Nature*, 580(7803):339–344, 2020. [Erratum: Nature 583, E16 (2020)].

[149] Kevin J. Kelly, Pedro A. N. Machado, Stephen J. Parke, Yuber F. Perez-Gonzalez, and Renata Zukanovich Funchal. Neutrino mass ordering in light of recent data. *Phys. Rev. D*, 103(1):013004, 2021.

[150] Peter B. Denton, Julia Gehrlein, and Rebekah Pestes. CP -Violating Neutrino Nonstandard Interactions in Long-Baseline-Accelerator Data. *Phys. Rev. Lett.*, 126(5):051801, 2021.

[151] Sabya Sachi Chatterjee and Antonio Palazzo. Nonstandard Neutrino Interactions as a Solution to the $NO\nu A$ and T2K Discrepancy. *Phys. Rev. Lett.*, 126(5):051802, 2021.

[152] Alexandre Sousa et al. Implementation of Feldman-Cousins Corrections and Oscillation Calculations in the HPC Environment for the NOvA Experiment. *EPJ Web Conf.*, 214:05012, 2019.

[153] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

[154] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[155] Lingge Li, Nitish Nayak, Jianming Bian, and Pierre Baldi. Efficient neutrino oscillation parameter inference using Gaussian processes. *Phys. Rev. D*, 101(1):012001, 2020.

[156] Andrew Sutton. The NOvA Test Beam Program. *PoS*, NuFACT2018:058, 2018.

[157] Alexander Radovic et al. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716):41–48, 2018.