Optimisation of the Search for CP-symmetry Violation at the Deep Underground Neutrino Experiment

Isobel Mawby

A thesis presented for the degree of
Doctor of Philosophy in Physics

University of Warwick
Department of Physics
April 2023

Contents

1 Introduction 22

2 Neutrino Theory 26

2.1 Evidence of Neutrino Oscillations 26
  2.1.1 The Solar Neutrino Problem 26
  2.1.2 The Atmospheric Neutrino Problem 28
2.2 Neutrino Oscillation Formalism 29
2.3 Two-Flavour Neutrino Oscillations 31
2.4 CP-Violation 32
2.5 Matter Effects 34
2.6 Current Experimental Status 36
  2.6.1 The Solar Parameters 36
  2.6.2 The Atmospheric Parameters 38
  2.6.3 The Final Mixing Angle, θ_{13} 40
  2.6.4 The CP-Phase 44
2.7 Sensitivity to the CP-Phase 44

3 The Deep Underground Neutrino Experiment 48

3.1 An Overview of the DUNE Experiment 48
3.2 The Far Detector 49
  3.2.1 LArTPC Technology 51
  3.2.2 Horizontal Drift 52
  3.2.3 Vertical Drift 53
3.3 Near Detector 55
  3.3.1 The Need for the Near Detector 55
  3.3.2 ND-LAr 56
  3.3.3 TMS/ND-GAr 57
  3.3.4 PRISM 57
  3.3.5 SAND 58
3.4 The Beam 58
3.5 The Prototypes 60

4 Event Reconstruction 63

4.1 Monte-Carlo Event Generation 63
  4.1.1 Event Generation 63
  4.1.2 Event Evolution 64
  4.1.3 Detector Response 64
4.2 Hit Finding 64
  4.2.1 Signal Processing 65

4.2.2 Gaussian Hit Finding 66
4.3 Pandora Pattern-Recognition Software 67
4.3.1 A Multi-Algorithm Approach 68
4.3.2 Consolidated Reconstruction 72
4.4 Particle Characterisation 75
4.4.1 Track Characterisation 75
4.4.2 Shower Characterisation 75
4.5 Neutrino Energy Estimation 77
4.5.1 Muon System 77
4.5.2 Electron System 77
4.5.3 Hadronic System 78
4.5.4 Performance 78

5 Calibration Motivated Reconstruction 80
5.1 2D → 3D Matching Studies 80
5.1.1 Metrics 81
5.1.2 Results 82
5.1.3 The Hit Width Cluster Merging Algorithm 83
5.2 Cosmic-Ray Reconstruction Improvements 86
5.2.1 Performance Metrics 87
5.2.2 Merging Across Detector Boundaries 88
5.2.3 Extending Through Showers 89
5.3 Delta-Ray Reconstruction Improvements 92
5.3.1 Performance Metrics 93
5.3.2 Delta-Ray 2D → 3D Matching 94

6 Probing the Sensitivity to CP Violation at DUNE 103
6.1 Fake Data Generation 104
6.2 CP-Violation Metrics 105
6.2.1 Sensitivity to CP Violation 105
6.2.2 Sensitivity Estimate without Uncertainties 106
6.2.3 Sensitivity Estimate with Oscillation Parameter Uncertainties 110
6.2.4 Sensitivity Estimate with Systematic Uncertainties 114
6.3 The $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ Selection 114
6.3.1 Overview 114
6.3.2 Tuning the Selection 116
6.3.3 Initial Modifications 116
6.4 The Muon-like BDT 118
6.4.1 BDT Variables 118
6.4.2 BDT Variable Correlations 124
6.4.3 BDT Performance 125
6.5 The Electron-like BDT 126
6.5.1 BDT Variables 126
6.5.2 BDT Variable Correlations 131
6.5.3 BDT Performance 132
6.6 $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ Selection Performance 133
6.6.1 ν_e Selection Performance 134
6.6.2 ν_μ Selection Performance 136

7 The Search for Improvements 140
7.1 Quantifying the Improvements 140
7.2 Cheating the Pandora Pattern-Recognition 141
  7.2.1 Motivation 141
  7.2.2 Cheating Details 141
  7.2.3 Results 142
7.3 Cheating the Neutrino Vertex 144
  7.3.1 Motivation 144
  7.3.2 Cheating Details 144
  7.3.3 Results 144
7.4 Cheating the Signal Electrons 145
  7.4.1 Motivation 145
  7.4.2 Cheating Details 146
  7.4.3 Results 146
7.5 Visible Vertexing 148
  7.5.1 Motivation 148
  7.5.2 Cheating Details 149
  7.5.3 Results 149
7.6 Shifted Vertexing 150
  7.6.1 Motivation 150
  7.6.2 Cheating Details 151
  7.6.3 Results 151
7.7 Cheating the Initial Shower Region 153
  7.7.1 Motivation 153
  7.7.2 Perfecting the Initial Electron Region Algorithm 154
  7.7.3 Perfecting the Initial Photon Region Algorithm 155
  7.7.4 Cheating Details 156
  7.7.5 Results 156
7.8 Outlook 158

8 Optimising the Sensitivity to CP-Violation at DUNE 160
8.1 The Shower Refinement Algorithm 160
8.2 The Hybrid Shower Refinement Algorithm 161
  8.2.1 Finding the 2D Connection Pathways 162
  8.2.2 Matching the Connection Pathways 165
  8.2.3 Identification of the Connection Pathway 167
  8.2.4 Electron Modification 168
  8.2.5 Photon Modification 169
  8.2.6 Hybrid Algorithm Results 169
8.3 Optimising the Hybrid Algorithms 172
  8.3.1 Smearing the Displacement 172
  8.3.2 Smearing the dE/dx 173
  8.3.3 Trustworthiness of the Smearing Configuration Results 174
  8.3.4 Outlook 175
8.4 Full Reconstruction Shower Refinement Algorithm 175
  8.4.1 The Connection Pathway BDT 175
  8.4.2 Shower Parameterisation 176
  8.4.3 BDT Variables 178

8.4.4 Connection Pathway BDT Correlations 184
8.4.5 Connection Pathway BDT Performance 184
8.5 Analysis Improvements 185
8.5.1 Hybrid Algorithm Cheating Method Removal 185
8.5.2 Merging the Electron-like and Connection Pathway BDTs 187
8.5.3 Neutrino Vertex Cheat Removal 191
8.5.4 Concluding Remarks 194

9 Validity of Results 195
9.1 Systematic Uncertainties 195
9.1.1 MC Model Uncertainties 195
9.1.2 Behaviour of the Systematics 197
9.1.3 Impact on the CP-Violation Analysis 202
9.1.4 Implementing the Systematics 202
9.2 MC/Data Comparison 206
9.2.1 ProtoDUNE Data 206
9.2.2 Data Sample Selection 207
9.2.3 MC Sample Selection 210
9.2.4 Muon-like BDT Variable Comparison 210
9.2.5 Electron-like BDT Variable Comparison 214
9.2.6 Conclusion 216

10 Conclusions and Outlook 217

List of Tables

2.1 Experimental specifics of the LBL experiments. 40

3.1 The details of the phased approach of DUNE. 49
3.2 The predicted time and exposure at which key physics milestones are reached by DUNE. 50

5.1 The topological distributions of the single muon sample. 81
5.2 The causes of specific angular regions inefficiencies. 83
5.3 The integrated correct event fraction with the improved stitching procedure included, and with the full volume approach utilised. 88
5.4 The integrated correct event fraction achieved by the initial Pandora reconstruction and after improvements have been made. 92
5.5 The tensor of the event of figure 5.15. 96
5.6 The delta-ray reconstruction performance metrics with the introduction of the *ThreeViewDeltaRayMatching* algorithm. 99
5.7 The delta-ray reconstruction performance metrics with the introduction of the *TwoViewDeltaRayMatching* algorithm. 101
5.8 The delta-ray reconstruction performance metrics with the introduction of the *OneViewDeltaRayMatching* algorithm. 102

6.1 The three file types used to obtain an oscillated fake data set. 104
6.2 The central values and uncertainties of the oscillation parameters. 107
6.3 The distributions that govern the throws of the oscillation parameters. . . 111
6.4 The fiducial volume definition. 115
6.5 The percentage of correctly identified candidate electrons/muons in signal CC ν_e events. 118
6.6 The integrated selection metrics of the ν_e selection procedure. 134
6.7 The integrated selection metrics of the ν_μ selection procedure. 137

7.1 The integrated efficiency, purity and background rejection of the ν_e selection, attained by fully cheating the Pandora pattern-recognition software. . 143
7.2 The integrated efficiency, purity and background rejection of the ν_e and ν_μ selection, attained when cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the ‘visible’ vertex. 149
7.3 The integrated efficiency, purity and background rejection of the ν_e selection, attained when cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the ‘shifted’ vertex. 154

7.4 The integrated efficiency, purity and background rejection of the ν_e selection, attained when cheating the placement of the neutrino vertex alongside the reconstruction of the electron and photon initial regions. 158

8.1 The integrated efficiency and purity of the ν_e selection, attained when cheating the placement of the neutrino vertex alongside the addition of the hybrid electron and photon refinement algorithms, where true information is used to inform the value of the displacement and dE/dx variables. . . . 173
8.2 The relative importance of the connection pathway BDT variables. 183
8.3 The integrated efficiency and purity of the ν_e selection, attained with the improved reconstruction (with cheated vertexing) and the two-pass ν_e selection. 189
8.4 The integrated efficiency and purity of the ν_e selection, attained with the improved reconstruction (without cheated vertexing) and the two-pass ν_e selection. 192

9.1 The relative uncertainties of the energy scale and particle response parameters. 196
9.2 The criteria used to identify different beam particle types at ProtoDUNE. . 208

List of Figures

2.1 The solar neutrino flux energy decomposition. 27
2.2 The selected ν_e and ν_μ as a function of zenith angle in the Super-Kamiokande detector. 29
2.3 Illustrations of the two-flavour neutrino oscillation probability. 33
2.4 The confidence limits and best-fit positions of the $(\sin^2\theta_{12}, \Delta m^2_{21})$ phase space obtained by KamLAND and the solar experiments. 37
2.5 The confidence limits and best-fit positions of the $(\sin^2\theta_{23}, |\Delta m^2_{31}|)$ phase space obtained by IceCube/DeepCore and Super-Kamiokande. 39
2.6 The confidence limits and best-fit positions of the $(\sin^2\theta_{23}, |\Delta m^2_{31}|)$ phase space obtained by T2K, MINOS and NOνA. 40
2.7 The normal and inverted ordering of the mass eigenstates. 41
2.8 The allowed regions of the $(\tan^2\theta_{12}, \sin^2\theta_{13})$ phase space obtained by KamLAND, the solar experiments, and from a combined fit. 41
2.9 The continuous curves of $(\theta_{13}, \delta_{\mathrm{CP}})$ solutions for a neutrino and antineutrino appearance probability. 42
2.10 The continuous curves of $(\theta_{13}, \delta_{\mathrm{CP}})$ solutions for a neutrino and antineutrino appearance probability, at different L/E values and of two different appearance channels. 43
2.11 The confidence limits and best-fit positions of the $(\sin^2\theta_{13}, \delta_{\mathrm{CP}})$ phase space obtained by T2K, MINOS and NOνA. 43
2.12 The confidence limits and best-fit positions of the $(\sin^2\theta_{13}, \delta_{\mathrm{CP}})$ phase space obtained by RENO and Daya Bay. 44
2.13 The confidence limits and best-fit positions of the $(\delta_{\mathrm{CP}}, \sin^2\theta_{23})$ phase space obtained by T2K and NOνA. 45
2.14 $P(\nu_\mu \to \nu_e)$ and $P(\bar{\nu}_\mu \to \bar{\nu}_e)$ as a function of neutrino energy at the 1300km baseline of DUNE. 46
2.15 The electron neutrino-antineutrino appearance asymmetry as a function of the baseline at the first and second oscillation peak. 46

3.1 A cartoon of the components of the DUNE experiment and the LBNF beamline. 49
3.2 The layout of the FD complex during phase I of the experiment. 50
3.3 The operation of a LArTPC. 51
3.4 The planned design of far detector modules I and II. 52
3.5 A schematic of a horizontal APA frame showing the wrapped U, V, W and G wires. 53
3.6 The bars of ARAPUCA devices and their placement on the APAs. 53
3.7 The planned design of far detector module III. 54

3.8 An illustration of the perforated PCB layers of a CRP. 55
3.9 A schematic of the phase II ND complex, which consists of ND-LAr, ND-GAr and SAND. 56
3.10 The energy composition of the predicted muon neutrino flux at the ND in the on-axis position and at several off-axis positions. 57
3.11 A diagram of the LBNF beamline. 59
3.12 The energy and flavour composition of the predicted flux at the FD in FHC mode and RHC mode. 59
3.13 The ProtoDUNE-SP detector module. 61
3.14 A schematic of the H4-VLE beamline. 62

4.1 The generated waveforms of a minimum ionising particle that travels parallel to the wire plane but perpendicular to the wire orientation in the DUNE FD. 65
4.2 The output of the signal processing of a neutrino candidate data event at MicroBooNE. 66
4.3 The hits created from a deconvolved waveform in ProtoDUNE-SP data. . . 67
4.4 The W view hits of a simulated event at ProtoDUNE-SP. 67
4.5 The 2D hits of the U, V and W view for a simulated CC quasi-elastic ν_e event in the DUNE FD. 68
4.6 The formed W view clusters of the CC ν_e example event. 69
4.7 The W view projection of the 3D candidate vertices, and identified 3D interaction vertex, of the CC ν_e example event. 70
4.8 The matched clusters of the created track-like pfos. 71
4.9 The 3D hits of the track-like and shower-like pfos of the CC ν_e event. . . . 72
4.10 The workflow of the Pandora consolidated reconstruction chain. 73
4.11 A simulated cosmic-ray which arrives at the ProtoDUNE-SP detector after the beam trigger time. 74
4.12 The formation of the initial track-like region of a shower using the crude cylinder search method, and the more sophisticated incremental growth method. 76
4.13 The fractional energy residuals of ν_e and ν_μ interactions. 78
4.14 The fractional energy residuals of contained and uncontained ν_μ interactions. 79

5.1 The reconstruction efficiency of single muon events as a function of θ_{xz} and θ_{yz}. 83
5.2 The reconstruction of a track that travels parallel to the orientation of the readout wires of one, or several, views. 84
5.3 An example of the treatment of a wide-hit cluster. 84
5.4 An example of the cluster associations of a cluster. 85
5.5 The result of the application of the *HitWidthClusterMerging* algorithm to the track presented in figure 5.2. 85
5.6 The efficiency improvements obtained with the application of the *HitWidthClusterMerging* algorithm. 86
5.7 The vertex and kinematic distributions of the cosmic-ray sample. 87
5.8 The U,V and W clusters of a reconstructed cosmic-ray event in which the cosmic-ray has been merged across one of the two detector gaps it passes through. 88

5.9 The correct event fraction as a function of energy with the improved stitching procedure included, and with the full volume approach utilised. 89
5.10 The U, V and W clusters of a cosmic-ray that has been split by regions of large electromagnetic behaviour. 90
5.11 The found extension points of the clusters of a broken cosmic-ray. 91
5.12 The collected hits that form the trajectory of the cosmic-ray track as it propagates through the electromagnetic shower. 91
5.13 The correct event fraction, as a function of energy, achieved by the initial Pandora reconstruction and after improvements have been made. 92
5.14 The true charge contributors to the hits of a delta-ray hierarchy within a ProtoDUNE-SP event. 94
5.15 A group of connected 2D cluster matches of a hypothetical event. 95
5.16 The outcome of the *DeltaRayRemoval* and *CosmicRayRemoval* tools. . . . 97
5.17 Examples of delta-rays that exhibit a 'broken by muon track' topology and a 'broken by gap' topology. 98
5.18 An example of the absorption (into the parent cosmic-ray) and fragmentation of delta-ray clusters. 100

6.1 The 'simple estimate' of the sensitivity achieved by the CP-violation analysis outlined in section 6.3. 108
6.2 The 'simple estimate' ν_e, ν_μ, $\bar{\nu}_e$ and $\bar{\nu}_\mu$ sensitivity contributions. 109
6.3 Poisson distributions defined by the expected number of selected ν_e events. 109
6.4 The importance of neutrino energy information in the achieved sensitivity to CP-violation. 111
6.5 The estimated median sensitivity and 68% contours achieved by the Pandora-based CP-violation analysis, with the inclusion of oscillation parameter uncertainties. 113
6.6 The workflow of the Pandora-based CP-violation analysis. 114
6.7 A flow chart outlining the Pandora-based selection procedure. 115
6.8 The muon-like BDT topology variable distributions of the signal and background. 119
6.9 Theoretical dE/dx distributions in liquid argon. 120
6.10 The muon-like BDT calorimetry variable distributions of the signal and background. 120
6.11 The muon-like BDT Michel electron variable distributions of the signal and background. 121
6.12 The muon-like BDT track/shower variable distributions of the signal and background. 123
6.13 The signal and background correlation matrices of the muon-like BDT. . . 124
6.14 The classification distributions and ROC curve of the muon-like BDT. . . . 125
6.15 The electron-like BDT displacement variable distributions of the signal and background. 127
6.16 Examples of electron topologies and reconstruction failures the smear the electron-like BDT variable distributions. 127
6.17 Examples of photon topologies and reconstruction failures the smear the electron-like BDT variable distributions. 128
6.18 The photon conversion distribution. 128

6.19 The electron-like BDT initial dE/dx variable distributions of the signal and background. 129
6.20 The electron-like BDT DCA variable distributions of the signal and background. 130
6.21 The electron-like BDT wideness and energy density variable distributions of the signal and background. 130
6.22 The electron-like BDT track/shower variable distributions of the signal and background. 131
6.23 The signal and background correlation matrices of the electron-like BDT. . 132
6.24 The classification distributions and ROC curve of the electron-like BDT. . 132
6.25 The ν_e selection metrics, as a function of true neutrino energy and the selected ν_e true and reconstructed energy distributions. 135
6.26 The bin-level sensitivity contributions of the ν_e selected spectrum. 136
6.27 The ν_μ selection metrics, as a function of true neutrino energy and the selected ν_μ true and reconstructed energy distributions. 137
6.28 The bin-level sensitivity contributions of the ν_μ selected spectrum. 139

7.1 The estimated sensitivity to CP-violation achieved by fully cheating the Pandora pattern-recognition software. 142
7.2 The ν_e and ν_μ selected reconstructed energy spectra obtained when the Pandora pattern-recognition is cheated. 143
7.3 The estimated sensitivity to CP-violation achieved by fully cheating the placement of the neutrino vertex. 145
7.4 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the shower reconstruction. 147
7.5 The electron-like BDT displacement variable distributions of the signal and background obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction. 147
7.6 An illustration of the 'interaction' and 'visible' shower vertex. 148
7.7 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the 'visible' vertex. 150
7.8 The electron-like BDT classification distributions obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the 'visible' vertex. 150
7.9 The selected ν_e reconstructed energy spectrum obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the 'visible' vertex. 151
7.10 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the 'shifted' vertex. 152
7.11 The electron-like BDT displacement variable distributions of the signal and background obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the 'shifted' vertex. 153
7.12 The electron-like BDT classification distributions obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction, with use of the 'shifted' vertex. 153

7.13 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the reconstruction of the electron and photon initial regions. 157
7.14 The electron-like BDT displacement variable distributions of the signal and background obtained when cheating the placement of the neutrino vertex alongside the electron and photon initial regions. 158
7.15 The electron-like BDT classification distributions obtained when cheating the placement of the neutrino vertex alongside the electron and photon initial regions. 159

8.1 An example CC ν_e event in which the primary electron shower is reconstructed as two separate particles. 162
8.2 The found initial directions out of the neutrino vertex in the example CC ν_e event. 163
8.3 The found shower spines in the example CC ν_e event. 164
8.4 The found shower start positions in the example CC ν_e event. 166
8.5 The matching of 2D connections pathways across the three views. 167
8.6 The 2D hits added to the electron shower in the example CC ν_e event. . . 169
8.7 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the addition of the hybrid photon refinement algorithm. 170
8.8 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the addition of the hybrid electron refinement algorithm. 171
8.9 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the addition of the hybrid electron and photon refinement algorithms. 171
8.10 The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the addition of the hybrid electron and photon refinement algorithms, where true information is used to inform the value of the displacement and dE/dx variables. 172
8.11 The electron-like BDT displacement variable distributions of the signal and background obtained when cheating the placement of the neutrino vertex alongside the addition of the hybrid electron and photon refinement algorithms, where true information is used to inform the value of the displacement and dE/dx variables. 173
8.12 The electron-like BDT dE/dx variable distributions of the signal and background obtained when cheating the placement of the neutrino vertex alongside the addition of the hybrid electron and photon refinement algorithms, where true information is used to inform the value of the displacement and dE/dx variables. 174
8.13 Illustrations of the initial region and connection pathway parameterisations. 177
8.14 Illustrations of the shower and ambiguous region parameterisations. 178
8.15 The connection pathway BDT initial region variable distributions of the signal and background. 179
8.16 The connection pathway BDT connection pathway variable distributions of the signal and background. 180

8.17 The connection pathway BDT shower region ‘sensible shower’ variable distributions of the signal and background. 181
8.18 The connection pathway BDT found hit ratio distributions of the signal and background. 181
8.19 The connection pathway BDT shower region electron/photon separation variable distributions of the signal and background. 182
8.20 The connection pathway BDT ambiguous region variable distributions of the signal and background. 183
8.21 The signal and background correlation matrices of the connection pathway BDT. 184
8.22 The classification distributions and ROC curve of the connection pathway BDT. 185
8.23 The estimated sensitivity to CP-violation achieved with the implementation of the *ShowerRefinement* algorithm (in electron extension mode) and with the neutrino vertex placement cheated. 186
8.24 The classification distributions and ROC curve of the enhanced electron-like BDT. 188
8.25 The classification distributions and ROC curve of the backup electron-like BDT. 189
8.26 The selected ν_e reconstructed energy spectrum attained with the improved reconstruction (with cheated vertexing) and the two-pass ν_e selection. . . . 190
8.27 The bin-level sensitivity contributions of the ν_e selected spectrum attained with the improved reconstruction (with cheated vertexing) and the two-pass ν_e selection. 190
8.28 The estimated sensitivity to CP-violation achieved with the improved reconstruction (with cheated vertexing) and the two-pass ν_e selection. 191
8.29 The classification distributions and ROC curve of the enhanced electron-like BDT when the neutrino vertexing is no longer cheated. 192
8.30 The selected ν_e reconstructed energy spectrum attained with the improved reconstruction (without cheated vertexing) and the two-pass ν_e selection. . 193
8.31 The estimated sensitivity to CP-violation achieved with the improved reconstruction (without cheated vertexing) and the two-pass ν_e selection. . . 193

9.1 The ν_e reconstructed energy spectrum of example throws of the flux systematics alongside that of the corresponding best-fit CP-phase value. . . . 198
9.2 The accuracy of the best fit CP-phase value across many flux systematic throws. 198
9.3 The ν_e reconstructed energy spectrum of example throws of the cross section systematics alongside that of the corresponding best-fit CP-phase value. 199
9.4 The accuracy of the best fit CP-phase value across many cross section systematic throws. 199
9.5 The ν_e reconstructed energy spectrum of example throws of the energy systematics alongside that of the corresponding best-fit CP-phase value. . . 200
9.6 The accuracy of the best fit CP-phase value across many energy systematic throws. 201
9.7 The selected ν_e spectra of two CP-phase values which are equidistant from a maximally violating CP-phase value. 201

9.8 The estimated sensitivities to CP-violation achieved by the improved Pandora-based CP-violation analysis, with the inclusion of oscillation parameter uncertainties and the individual systematic uncertainties in the fits. 203
9.9 The estimated median sensitivity and 68% contours achieved by the Pandora-based CP-violation analysis, with the inclusion of oscillation parameter and systematic uncertainties. 204
9.10 The estimated median sensitivity and 68% contours achieved by the improved Pandora-based CP-violation analysis, with the inclusion of oscillation parameter and systematic uncertainties. 205
9.11 The estimated median sensitivity and 68% contours achieved by the improved Pandora-based CP-violation analysis (in which the neutrino vertexing is cheated), with the inclusion of oscillation parameter and systematic uncertainties. 206
9.12 The muon, proton and pion dE/dx distributions obtained in this study compared to those obtained by the ProtoDUNE collaboration. 211
9.13 The muon-like BDT topology variable distributions of the signal and background in MC and data. 212
9.14 The muon-like BDT calorimetry variable distributions of the signal and background in MC and data. 213
9.15 The electron-like BDT calorimetry variable distributions of the signal and background in MC and data. 215
9.16 The electron initial dE/dx distribution obtained in this study compared to those obtained by the ProtoDUNE collaboration. 215
9.17 The electron-like BDT topology variable distributions of the signal and background in MC and data. 216

Acknowledgements

John, it's no understatement to say that I couldn't have asked for a better supervisor. Thank you for your academic (and emotional) support/advice, for always pushing me, our epic physics slack-chats, and for making my PhD so well rounded with all of your wildlife teachings. I'm extremely grateful for the time you've spent reading my drafts - your kindness and commitment to your students is an inspiration.

Alice, you helped me to become the person that I needed to be to do this, and for that I'm ever grateful. Thank you for taking me away from the world of academia with our 'spontaneously planned' adventures, for being the person that I want to celebrate with, and the one that that holds my hand through the darkest times. I can't believe you've put up with my physics rambles, but I'm glad that you do.

Alex, it's been great doing our PhD together and I'm so happy that we became friends. Thank you for keeping me going throughout the PhD and for listening to me complain (you deserve a medal). You bring so much fun to PhD life, and I look forwards to our future neutrino adventures!

Matthew, I feel very lucky to have met you. Thank you for your kindness and support, amazing sense of humour, deep conversations, starting my guitar journey, introducing me to the iconic Log Lady, and ultimately, for bringing so much joy to my PhD.

I'd also like to thank the Pandora group for their support and friendship, especially: Maria for being an awesome housemate, and one of the kindest people I've ever known; Andy Chappell for his humour, guitar chats and infinite wisdom; Dom for developing the original CP-violation analysis, and his advice, sarcasm and kindness; and Andy Blake for inspiring me to enter the realms of neutrino physics.

Finally, Greggs. Thank you for your vegan sausage rolls and mince pies - they have fuelled this PhD.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- The Pandora Modular Shower module introduced in chapter 4, and used in chapter 8, was developed by Dominic Barker and Ed Tyley.
- The initial $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection procedure, muon-like BDT, and electron-like BDT introduced in chapter 6 were inherited from, and created by, Dominic Brailsford.
- The parameters used to obtain the sensitivity to CP-violation estimates, discussed in chapters 6 and 9, are motivated by the DUNE TDR analysis. These choices are made clear in the text, but, to give an example, includes the oscillation parameter uncertainties.
- The e, p, π and μ selections of chapter 9, and the analysis methods used, are those of the ProtoDUNE collaboration.
- All event samples were generated using the DUNE collaboration's event generation and detector simulation software. This includes samples for the DUNE Far Detector (used in chapters 6, 7, 8 and 9), and the single-phase ProtoDUNE module (used in chapters 5 and 9). Real data (used in chapter 9) was provided by the single-phase ProtoDUNE module.
- The cosmic-ray samples investigated in chapter 5 were created by Viktor Pěč and collaborators at the University of Sheffield.

This is for you mum, you’ve given everything you could to get me here.

Abstract

The Deep Underground Neutrino Experiment (DUNE) is a next-generation long baseline experiment, which will be situated in South Dakota. Its detectors will utilise liquid-argon time projection chamber technology, which is able to capture neutrino interactions with an incredible spatial and calorimetric resolution. With what will become the world's most intense neutrino beam, a highly capable near detector, and four (10kt fiducial mass) far detector modules, DUNE will be able to achieve an ambitious physics programme. Most notably, DUNE will determine whether charge-parity symmetry is broken in neutrino oscillations - a finding that would have significant implications for the understanding of the matter-antimatter asymmetry in our Universe.

This thesis presents the optimisation of a CP-violation analysis at DUNE using the Pandora pattern-recognition software. The analysis assumes a 3.5 year exposure (1.36×10^{23} protons on target) to a neutrino and an antineutrino beam (7 year total). Only the predicted data of the far detector modules is used; near detector samples are not included. The initial sensitivity to CP-violation is found to be $3.8\sigma^{+0.9\sigma}_{-1.1\sigma}$ in an estimate that includes oscillation parameter uncertainties, systematic uncertainties and statistical fluctuations, and assumes a normal-ordering of the neutrino mass hierarchy. The performance of the Pandora event reconstruction is linked to that of the analysis, which is found to be limited by the reconstruction of the initial track-like region of electrons and photons. The *ShowerRefinement* algorithm is developed in response to this and its implementation into the analysis workflow results in an improved sensitivity to CP-violation of $4.6\sigma^{+0.9\sigma}_{-1.0\sigma}$. With a perfected neutrino interaction vertex placement, this is further increased to $5.1\sigma^{+1.0\sigma}_{-1.1\sigma}$.

Acronyms

ADC Analogue-to-Digital Converter

APAs Anode Plane Assemblies

ArgoNeuT Argon Neutrino Test Project

BDT Boosted Decision Tree

BNL Brookhaven National Laboratory

CC Charge-Current

CDR Conceptual Design Report

CNO Carbon-Nitrogen-Oxygen Cycle

CP Charge-Parity

CPC CP-Conservation

CPV CP-Violation

CRPs Charge Readout Planes

CSDA Continuous-Slowing-Down-Approximation

DCA Distance of Closest Approach

DONUT Direct Observation of the Nu Tau

DP Dual-Phase

DUNE Deep Underground Neutrino Experiment

FD Far Detector

Fermilab Fermi National Accelerator Laboratory

FHC Forward Horn Current

FV Fiducial Volume

GALLEX Gallium Experiment

GAr Gaseous-Argon

GENIE Generates Events for Neutrino Interaction Experiments

HD Horizontal Drift

HK Hyper-Kamiokande

IBD Inverse Beta Decay

ICARUS Imaging Cosmic and Rare Underground Signals

IMB Irvine-Michigan-Brookhaven

IO Inverted Ordering

Kamiokande Kamioka Nucleon Decay Experiment

KamLAND Kamioka Liquid Scintillator Anti-Neutrino Detector

LAr Liquid-Argon

LArIAT LArTPC in a Testbeam

LArTPC Liquid-Argon Time Projection Chamber

LBL Long-Baseline Accelerator

LBNF Long-Baseline Neutrino Facility

LEP Large Electron-Positron

MACRO Monopole Astrophysics and Cosmic Ray Observatory

MC Monte-Carlo

MCS Multiple Coulomb Scattering

MicroBooNE Micro Booster Neutrino Experiment

MINOS Main Injector Neutrino Oscillation Search

MUSIC Muon Simulation Code

MUSUN Muon Simulations Underground

NC Neutral-Current

ND Near Detector

NO Normal Ordering

NOνA NuMI Off-Axis Electron Neutrino Appearance

PCA Principle Component Analysis

PCB Printed Circuit Board

pfo Particle Flow Object

PID Particle Identification

POT Protons on Target

PRISM Precision Reaction-Independent Spectrum Measurement

RENO Reactor Experiment for Neutrino Oscillation

RHC Reverse Horn Current

ROC Receiver Operating Characteristic

SAGE Soviet–American Gallium Experiment

SAND System for on-Axis Neutrino Detection

SCE Space Charge Effect

SK Super-Kamiokande

SLAC Stanford Linear Accelerator Center

SNO Sudbury Neutrino Observatory

SP Single-Phase

SPS Super Proton Synchroton

SURF Sanford Underground Research Facility

T2K Tokai to Kamioka

TDR Technical Design Report

TMS Temporary Muon Spectrometer

TOF Time of Flight

VD Vertical Drift

Chapter 1

Introduction

Picture it, Tübingen, 1930: Wolfgang Pauli has written his, now famous, open letter [1] to the radioactive attendees of the Gauverein meeting. The letter is a response to Charles Ellis and William Wooster's results [2], which have found a continuous energy spectrum of electrons to be emitted in β-decays, seemingly violating the conservation of energy. Pauli offers a 'desperate remedy' and proposes that a neutral, spin 1/2 particle (which he called the 'neutron') is created alongside the electron in β-decays. He ends by imploring the audience to deliberate on how this 'neutron' can be found.

In 1932, James Chadwick discovered a neutral particle [3], but it was far too heavy to be Pauli's 'neutron'. It was in fact the modern day neutron, and its discovery advanced the understanding of atomic nuclei and led, in 1934, to Enrico Fermi's theory of β-decay [4]. Fermi's theory incorporated Pauli's 'neutron' (now known as the neutrino) and was able to correctly predict the observed properties of β-decays, such as the shape of the electron energy spectrum. This provided strong evidence that Pauli was correct. With Fermi's theory in hand, Hans Bethe and Rudolph Peierls calculated the cross section (σ) of the Inverse Beta Decay (IBD) process

$$\bar{\nu}_e + p \rightarrow n + e^+ \tag{1.1}$$

and, found $\sigma < 10^{-44}\text{cm}^2$ [5]. They concluded that there is no 'practically possible way of observing the neutrino', but they were about to be proved wrong.

Antineutrinos are produced in copious amounts in nuclear explosions, a fact known by Los Almos scientists Frederick Reines and Clyde Cowan. In the early 1950s they proposed perhaps one of the wildest experiments in particle physics history [6]. They would create a nuclear explosion and detect the released antineutrinos via IBD in an underground, free-falling, liquid-scintillator based detector. It was envisaged that the detector would land on a bed of feathers and rubber, where it would stay until the site was safe enough to allow its collection. Their idea was actually approved but, perhaps fortunately, the pair followed Jerome Kellogg's suggestion and decided to use reactor antineutrinos instead [6].

They detected neutrinos by demanding a pair of back-to-back, 0.51MeV photons (from positron annihilation) and a delayed light signal (from neutron capture). By 1956, the neutrino had been discovered [7].

In 1962, an experiment at Brookhaven National Laboratory (BNL) investigated whether the neutrinos produced in β-decays and pion decays - denoted as ν_e and ν_μ respectively - were the same [8]. This would imply that in the processes

$$\begin{aligned}\nu_e + n &\rightarrow e^- + p\\ \bar{\nu}_e + p &\rightarrow e^+ + n\end{aligned}$$

and,

$$\begin{aligned}\nu_\mu + n &\rightarrow \mu^- + p\\ \bar{\nu}_\mu + p &\rightarrow \mu^+ + n\end{aligned}$$

$\nu_e = \nu_\mu$. In their experiment, the decay products of a pion beam, created by the collision of high-energy protons with a beryllium target, were intercepted by a 13.5m thick steel wall, which only neutrinos could penetrate. The neutrinos would enter a following spark chamber, a detector of alternating aluminium and gaseous layers, where their interactions with the aluminium nuclei were detected. Only muons were found to be produced, which proved that electron and muon neutrinos were different.

In 1975, physicists at the Stanford Linear Accelerator Center (SLAC) found energy to be missing in e^+e^- annihilation processes. This could be explained by the existence of a new particle, the tau lepton [9, 10], which decayed to produce 'invisible' neutrinos alongside the observed final state. The τ implied the existence of the ν_τ which was experimentally observed by the Direct Observation of the Nu Tau (DONUT) experiment at the turn of the 21st century [11]. With this discovery, the results of the Large Electron-Positron (LEP) experiment [12], which measured the decay width of the Z^0, suggest that all active, light neutrinos had been found. These neutrinos are now known to 'oscillate' between flavours as they propagate [13, 14, 15], which is explained by their non-zero mass and 'mixing' between their mass and flavour eigenstates.

Since the 1960s, neutrino oscillations have been studied and our understanding of the governing parameters has grown, yet some questions remain, such as whether Charge-Parity (CP) symmetry is broken in neutrino oscillations. The CP operation is the conjunction of the application of

- the **charge conjugation operator**, which replaces left(right)-handed (anti)neutrinos with left(right)-handed antineutrinos(neutrinos) and,
- the **parity operator**, which replaces left(right)-handed (anti)neutrinos with right(left)-handed (anti)neutrinos.

Since only left-handed neutrinos and right-handed antineutrinos participate in the weak interaction, the individual charge conjugation and parity symmetries are violated in the weak interaction [16, 17]. However, CP-symmetry may be respected. Current long baseline accelerator experiments hint at this not being the case [18, 19], but do not have the sufficient sensitivities to say for sure. The discovery of CP-violation in the lepton sector would be a significant advance in our understanding of the observed matter-antimatter asymmetry in the Universe.

The Deep Underground Neutrino Experiment (DUNE) is a next-generation long baseline experiment, which aims to determine whether CP is violated as part of its wide and ambitious physics programme. This thesis presents a body of work which investigates and optimises the sensitivity of the DUNE experiment to CP-violation in neutrino oscillations.

In chapter 2, the theory of neutrino oscillations is presented, and a brief account of the current status of neutrino oscillation parameter measurements is given. The need for DUNE will be highlighted and the chapter will end with a summary of how the experiment's design will be tailored to allow a determination of whether CP is violated in neutrino oscillations.

In chapter 3, the design and operation of the beam and detectors of DUNE will be discussed. Special focus will be given to the far detector modules and their prototypes, with which the CP-violation analysis of this thesis is performed.

Chapter 4 details the reconstruction algorithms of the DUNE far detector, i.e. how the detector response is translated into a fully characterised particle hierarchy. Much of this chapter will focus on the workings of the Pandora pattern-recognition software, to which modifications will be made in the later chapters.

Chapter 5 illustrates the optimisation process of Pandora algorithms to aid DUNE's physics goals. The chapter initially discusses the optimisation of the cosmic-ray reconstruction and, later, the delta-ray reconstruction. The reader should interpret this project as a 'service task', i.e. a piece of work completed to benefit the wider experiment.

In chapter 6, the CP-violation analysis is introduced. The 'sensitivity to CP-violation' metric and its estimates are first discussed, and the rest of the chapter is dedicated to the description and characterisation of the Pandora-based $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection procedure.

In chapter 7, the reconstruction-to-analysis continuum is established. Well-motivated reconstruction improvements are assessed in terms of the resulting gains to the CP-violation analysis metric, and the leading limitation of the reconstruction with respect to the CP-violation analysis is identified.

Chapter 8 begins with a presentation of the reconstruction algorithm that was created in response to the leading reconstruction limitation identified in chapter 7. The remainder of the chapter discusses the implementation of the algorithm into the analysis and presents the subsequent improvements.

In chapter 9, the validity of the CP-violation analysis improvements of chapter 8 are examined. The first half of the chapter investigates the impact of the inclusion of oscillation parameter and systematic uncertainties in the CP-violation sensitivity estimate. The latter half presents a simulation/data study in which ProtoDUNE-SP data is used to compare the performance of the electron-like Boosted Decision Tree (BDT) and muon-like BDT, which drive the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection.

The conclusions of the work of this thesis will be discussed in chapter 10, where the next steps for the CP-violation analysis will also be presented.

Chapter 2

Neutrino Theory

2.1 Evidence of Neutrino Oscillations

2.1.1 The Solar Neutrino Problem

The sun is powered by two main fusion chains: the *pp*-chain and the Carbon-Nitrogen-Oxygen Cycle (CNO) cycle which respectively contribute ~99% and ~1% to the solar output [20]. Electron neutrinos are the only neutrinos produced and are typically labelled by their creation processes, e.g. ^{8}B neutrinos are those created in ^{8}B decays. The composition of the solar electron neutrino flux as a function of energy is shown in figure 2.1.

In the 1960s, Ray Davis set out to measure the solar neutrino flux [22] after a prediction of its magnitude by John Bahcall [23]. Davis' experiment consisted of a 3.9×10^5 litre tank of tetrachloroethylene (C_2Cl_4) which lay ~1.5km underground in the Homestake gold mine, in Lead, South Dakota. Incoming electron neutrinos would interact via the process

$$\nu_e + {}^{37}\mathrm{Cl} \rightarrow {}^{37}\mathrm{Ar} + e^- \tag{2.1}$$

and the resulting ^{37}Ar atoms would be collected and used to infer the incident electron neutrino flux. The experiment ran from 1968 until 2002, and reported an electron neutrino flux that was approximately a factor of three lower than that predicted [24]. The reason for this discrepancy was debated. Was Bahcall's solar model incorrect? Was the experiment, which had no pointing or energy information, at fault? Or was there something else at play?

Between 1987-1995, the solar neutrino deficit was once again observed, this time by the water Cherenkov Kamioka Nucleon Decay Experiment (Kamiokande) during its phase II and III periods [25]. Kamiokande detected electron neutrinos via the elastic scattering process

$$\nu_e + e^- \rightarrow \nu_e + e^-, \tag{2.2}$$

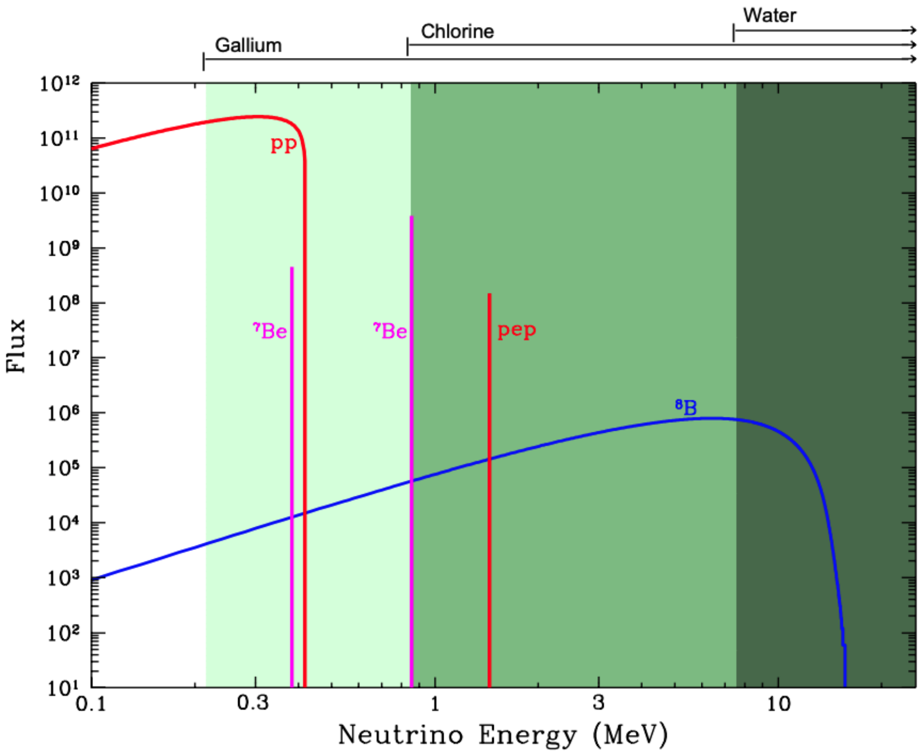


Figure 2.1: The decomposition of the solar neutrino flux in terms of the processes from which the solar neutrinos originate. The energy threshold for detection of the Gallium-, Chlorine- and water-based experiments is illustrated. Figure reproduced from reference [21].

which had an energy threshold of ~7MeV (~9MeV) in Kamikande II(III). Their measurement was more robust, since they were able to infer the direction of the incoming neutrinos and confirm their solar origins. However, the Kamiokande experiment was only sensitive to the high-energy ^{8}B neutrinos, a small component of the solar flux.

In the early 1990s, the Soviet–American Gallium Experiment (SAGE) [26] and the Gallium Experiment (GALLEX) [27] began operation, their detection principle analogous to the Homestake experiment (but using ^{71}Ga instead of C_2Cl_4). SAGE and GALLEX reported a solar flux that was half the theoretical prediction. The energy threshold for neutrino absorption in ^{71}Ga is 233 KeV and so the experiments were sensitive to the dominant *pp* neutrino flux and thus demonstrated the energy dependence of the deficit.

These tensions were ultimately resolved by the Sudbury Neutrino Observatory (SNO) in the early 2000s [13, 14]. SNO's detection principle was based on the measurement of Cherenkov light much like Kamiokande, however it had a key advantage: its target medium was heavy water (D_2O). This lowered the energy threshold of the neutral current process

$$\nu_x + d \to n + p + \nu_x, \tag{2.3}$$

where $x = e, \mu, \tau$, such that the **total** neutrino flux could be measured. This agreed well with Bahcall's solar model prediction and, by measuring the process

$$\nu_e + d \rightarrow p + p + e^-, \tag{2.4}$$

the combined ν_μ and ν_τ flux was inferred and found to be inconsistent with zero. These results resolved the solar neutrino problem: the solar model was correct, but it seemed that electron neutrinos were 'turning into' muon and tau neutrinos.

2.1.2 The Atmospheric Neutrino Problem

The Earth's atmosphere is subject to a significant flux of cosmic-rays, which interact with atmospheric nuclei to create hadronic cascades. The hadronic showers, are dominated by the decay chain

$$\pi^+(\pi^-) \rightarrow \nu_\mu(\bar{\nu}_\mu) + \mu^+(\mu^-) \tag{2.5}$$

$$\mu^+(\mu^-) \rightarrow e^+(e^-) + \bar{\nu}_\mu(\nu_\mu) + \nu_e(\bar{\nu}_e) \tag{2.6}$$

which leads to the prediction that the ratio

$$\frac{N(\nu_\mu + \bar{\nu}_\mu)}{N(\nu_e + \bar{\nu}_e)} \tag{2.7}$$

should be close to two, below 1GeV [28].

Measurements of the atmospheric flux ratio came in the 1980s, when several experiments were built to search for proton decay. It was imperative that these experiments made precise measurements of the atmospheric neutrino flux, as this provided the main background to their results. There was overwhelming evidence from Kamiokande [29], the Irvine-Michigan-Brookhaven (IMB) experiment [30], the Monopole Astrophysics and Cosmic Ray Observatory (MACRO) [31] and Soudan-2 [32] that the atmospheric flux ratio fell below its predicted value. This suggested that the muon flux was underestimated and(or) the electron flux was overestimated.

Perhaps the most revealing results [15] came in 1998, from Super-Kamiokande (SK), a 50kt water Cherenkov detector situated in the Mozumi Mine of Kamioka, Japan. Super-Kamiokande measured the electron and muon neutrino fluxes as a function of the zenith angle, where a positive(negative) cosine of the zenith angle corresponds to down-going(up-going) neutrinos which enter the top(bottom) of the detector. Since up-going muons travel through the atmosphere and Earth to reach the detector, SK probed baselines in the $15 - 1.3 \times 10^4$km range. The results are shown in figure 2.2, which are a sharp contrast to the isotropic prediction.

The results show a deficit of the muon neutrino flux, which suggests that muon neutrinos are disappearing in a clear analogy to the solar electron neutrinos. This phenomena is

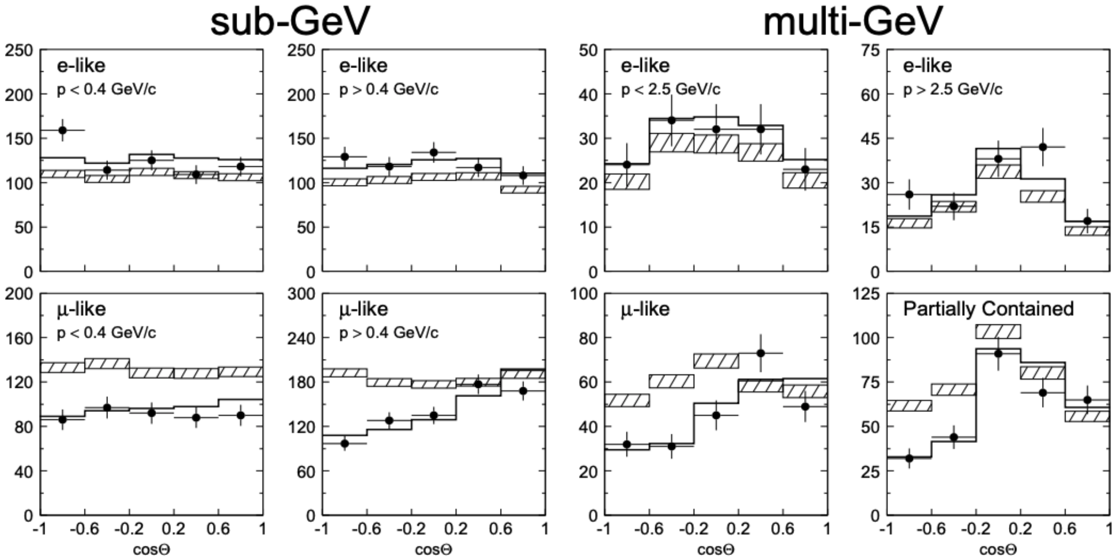


Figure 2.2: The selected ν_e (top) and ν_μ as a function of zenith angle and separated by energy where a positive(negative) cosine of the zenith angle corresponds to down-going(up-going) neutrinos which enter the top(bottom) of the SK detector. Data points are shown by black dots, the expectation in the case of no oscillations is illustrated by the hatched regions and the black line is the best-fit expectation for $\nu_\mu \to \nu_\tau$ oscillations. Figure reproduced from reference [15].

once again found to be energy dependent, but is now seen to also depend on the neutrino path length. Additionally, we see that the electron neutrino flux agrees with prediction. If we take the conclusions of subsection 2.1.1, we can thus infer that the muon neutrinos must be changing into tau neutrinos and that this ‘neutrino conversion’ depends on the neutrino flavour. So what’s happening?

2.2 Neutrino Oscillation Formalism

Experimentally we can only probe the neutrino eigenstates associated with the weak interaction: these are termed ‘flavour eigenstates’. If one considers that neutrinos are massive particles, there also exist the ‘mass eigenstates’, which are those of the Hamiltonian operator and that have a defined energy. One can express the flavour eigenstates in terms of the mass eigenstates as

$$|\nu_\alpha\rangle = \sum_{i=1,2,3} U_{\alpha i} |\nu_i\rangle\,, \tag{2.8}$$

where $|\nu_\alpha\rangle$ are the flavour eigenstates with indices e, μ, τ; $|\nu_i\rangle$ are the mass eigenstates with indices 1, 2, 3; and U is a 3×3 matrix, which is referred to as the ‘mixing matrix’ as it describes the mixing between the mass and flavour eigenstates [33]. The mixing matrix must be unitary to conserve probability, i.e. to ensure that an initially observed neutrino is later found to have one of the three possible flavours. It therefore satisfies the condition $UU^\dagger = U^\dagger U = I$ and so has n^2 degrees of freedom: three of which are real and, 6 complex. But how many of these are **physical**?

In the Lagrangian of the weak interaction, the neutrino flavour eigenstates are coupled to the lepton eigenstates ($|l_\alpha\rangle$) which can be redefined as

$$|l'_\alpha\rangle = e^{-i\phi_\alpha}\,|l_\alpha\rangle \tag{2.9}$$

to absorb three of the complex phases without any physical consequences. If neutrinos are Dirac, the neutrino flavour eigenstates can be similarly redefined such that we are left with a single **physical** phase. This cannot be done if neutrinos are Majorana, as the mass eigenvalues would become complex. Whether neutrinos are Dirac or Majorana is not known, but has an impact on the conservation of lepton number: in the Dirac(Majorana) case, lepton number is conserved(violated) and, neutrinos and antineutrinos have different(the same) lepton numbers and are (in)distinguishable [34].

The unitary matrix is termed the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix, after those who contributed to its formation [35, 36]. It is typically expressed as

$$U_{\text{PMNS}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\text{CP}}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{\text{CP}}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{\text{CP}}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{\text{CP}}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{\text{CP}}} & c_{13}c_{23} \end{pmatrix} \mathbf{P}, \tag{2.10}$$

where c_{ij} and s_{ij} denote $\cos\theta_{ij}$ and $\sin\theta_{ij}$ respectively, θ_{ij} are the 'mixing angles', δ_{CP} is the CP-phase and $\mathbf{P}$ is a 3×3 matrix which is equal to the identity matrix in the Dirac case and contains two complex phases in the Majorana case [33]. The Majorana phases do not appear in the neutrino oscillation probabilities and thus will not be discussed further.

So what is the probability of measuring a neutrino with flavour α at one position and flavour β at a position distance L away? We will first consider that the neutrino propagates through a vacuum, then this simplification will be removed in section 2.5. The neutrino is initially described by the flavour eigenstate $|\nu_\alpha\rangle$ and as the neutrino propagates, the time-dependent Schrodinger equation tells us that the mass eigenstates evolve such that

$$|\nu_i(t,\vec{L})\rangle = e^{-i(E_i t - \vec{p}_i\cdot\vec{L})}\,|\nu_i(0,\vec{0})\rangle\,. \tag{2.11}$$

The probability of measuring a ν_β neutrino at the downstream position is then given by

$$P(\nu_\alpha \to \nu_\beta) = |\,\langle\nu_\beta|\nu_\alpha(t,\vec{L})\rangle\,|^2 = \Big|\sum_{i=1,2,3} e^{-i(E_i t - \vec{p}_i\cdot\vec{L})}\,\langle\nu_\beta|\,U_{\alpha i}\,|\nu_i(0,\vec{0})\rangle\Big|^2, \tag{2.12}$$

which can be simplified by using equation 2.8, and the relationship $\langle\nu_\beta|\nu_\alpha\rangle = \delta_{\alpha\beta}$ (where

$\delta_{\alpha\beta}$ is the Kronecker delta) to give

$$P(\nu_\alpha \to \nu_\beta) = \Big| \sum_{i=1,2,3} e^{-i(E_i t - \vec{p}_i \cdot \vec{L})} U^*_{\beta i} U_{\alpha i} \Big|^2 . \tag{2.13}$$

If one assumes that the mass eigenstates have the same energy and that the mass eigenvalues are negligible relative to their momentum, the relationship

$$Et - \vec{p}_i \cdot \vec{L} = Et - \sqrt{E^2 - m_i^2} L \approx \frac{m_i^2 L}{2E} \tag{2.14}$$

is obtained. The 'constant energy' approximation is necessary because the mass eigenstates have been modelled as plane waves instead of wavepackets. However, it can be shown that both approaches yield the same result [37]. The oscillation probability of the $\nu_\alpha \to \nu_\beta$ transition [38] therefore becomes

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} &- 4 \sum_{i>j} \mathrm{Re}\big[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}\big] \sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right) \\ &+ 2 \sum_{i>j} \mathrm{Im}\big[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}\big] \sin\left(\frac{\Delta m^2_{ij} L}{4E}\right), \end{aligned} \tag{2.15}$$

where Δm^2_{ij} are the 'mass splitting terms' which are equal to $m_i^2 - m_j^2$. By convention the three mixing angles are denoted by θ_{12}, θ_{23} and θ_{13} and, usually, only the mass splitting terms Δm^2_{31} and Δm^2_{21} feature in the oscillation probabilities. The final mass splitting is given by $\Delta m^2_{32} = \Delta m^2_{31} - \Delta m^2_{21}$.

2.3 Two-Flavour Neutrino Oscillations

To ease the interpretation of equation 2.15 and to build our intuition of the physical impact of the mixing angles and mass splitting terms it is useful to consider a simplified scenario in which only two neutrino flavours exist [39]. In the two-flavour case, the mixing matrix and oscillation probabilities are:

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{2.16}$$

$$P(\nu_\alpha \to \nu_\beta) = \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 L}{2E}\right) \tag{2.17}$$

$$P(\nu_\alpha \to \nu_\alpha) = 1 - P(\nu_\alpha \to \nu_\beta) \tag{2.18}$$

where $\alpha \neq \beta$.

It is seen in equation 2.17 that the mixing angle controls the amplitude of the disappearance probability. This is understood with reference to the mixing matrix (equation 2.16) where we see that the mixing angle tells us how similar the flavour and mass eigenstates are. If they are very similar, each flavour eigenstate will have a large contribution from one mass eigenstate and a small contribution from the other. As time progresses these contributions change and, in this case, will have to change significantly to resemble the decomposition of the other flavour eigenstate. Therefore the more similar the flavour and mass eigenstates are, the more oscillations are suppressed. Namely, if $\theta = 0°$ the oscillation probability is zero and if $\theta = 45°$ the oscillation probability is maximised in what is referred to as 'maximal mixing'.

What about the mass splitting? If the 'constant energy' approximation is made then the momentum of each mass eigenstate is

$$p_i = \sqrt{E^2 - m_i^2}, \tag{2.19}$$

which depends on the value of m_i. If the mass splitting is small, the mass eigenstates will have a similar momentum and the phase difference between them will be negligible. The mass eigenstate decomposition of a propagating neutrino will therefore remain almost constant and the probability of oscillation will be small. This tells us that in order to have neutrino oscillations, the mass splitting must be non-zero. It is noted that the mass splitting enters the oscillation probability as $\sin^2$ and so its sign cannot be determined from neutrino oscillations in a vacuum.

Finally, we turn to the L/E term. This term defines the moment at which the mass eigenstate decomposition of the propagating neutrinos are probed. As will be seen, neutrino beam experiments are often designed such that their L/E value maximises the investigated oscillation probability. Figure 2.3 illustrates the oscillation probability as a function of L for a given energy E.

2.4 CP-Violation

Our current models tell us that the energy of the early universe would have been converted into equal amounts of matter and antimatter, which would have subsequently undergone complete annihilation. As we can observed our Universe (and everything in it), these models must be wrong.

In 1967, Andrei Sakharov [40] proposed three conditions that, if met, would result in the matter-antimatter asymmetry:

1. **Violation of baryon number:** to allow a process to exist that creates an excess of baryons over anti-baryons.

2. **Violation of C and CP symmetries:** to allow processes that create an excess of

baryons to be favoured.

3. **Deviation from thermal equilibrium:** to prevent an inverse process from removing the created excess.

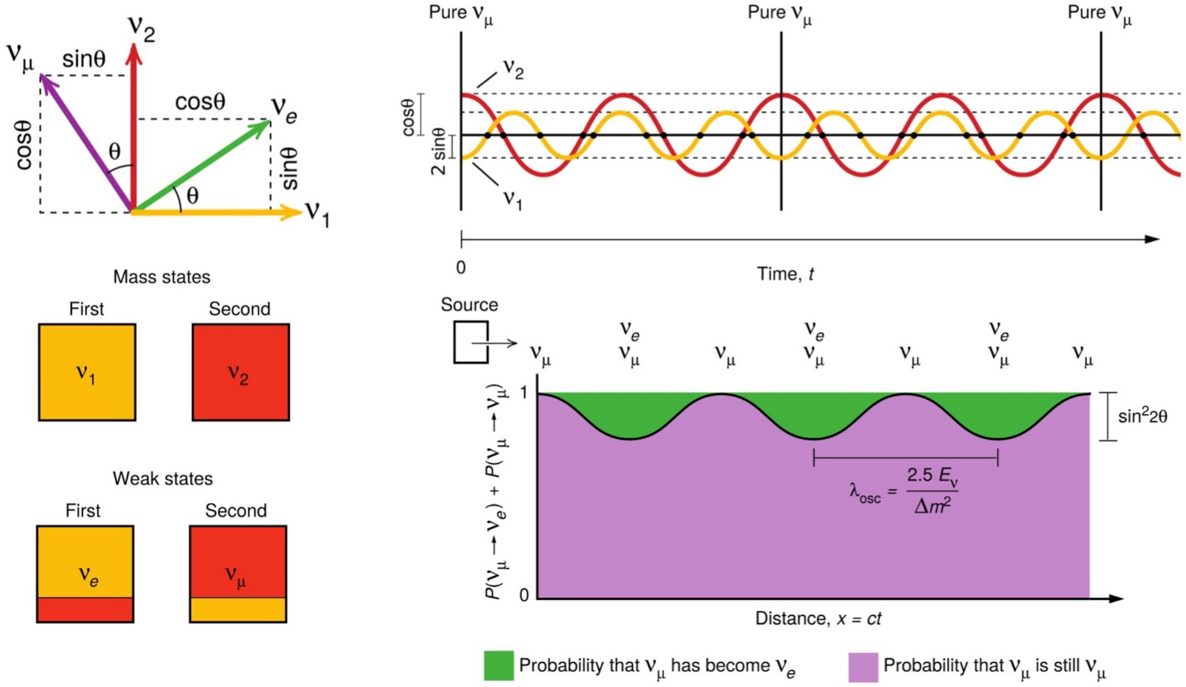


Figure 2.3: Top-left: the relationship between the mass basis (ν_1, ν_2) and the flavour/weak basis (ν_e, ν_μ). Bottom-left: the decomposition of the flavour eigenstates in terms of the mass eigenstates, where yellow and red correspond to ν_1 and ν_2 respectively. Top right: the time evolution of the amplitude of an initial ν_μ beam in terms of the mass eigenstates. Bottom-right: $P(\nu_\mu \to \nu_\mu)$ and $P(\nu_\mu \to \nu_e)$ as a function of propagation distance. Figure adapted from reference [41].

CP-violation has been observed in the quark sector but not to the extent necessary to meet Sakharov's second condition. But perhaps the leptonic sector holds the answer? CP symmetry violation in neutrino oscillations implies that

$$P(\nu_\alpha \to \nu_\beta) \neq P(\bar{\nu}_\alpha \to \bar{\nu}_\beta), \tag{2.20}$$

and so an experiment that measures the oscillation probability of a given channel with neutrinos and antineutrinos is sensitive to CP-violation. However, if one assumes CPT symmetry then

$$P(\nu_\alpha \to \nu_\beta) = P(\bar{\nu}_\beta \to \bar{\nu}_\alpha) \tag{2.21}$$

and,

$$P(\nu_\alpha \to \nu_\alpha) = P(\bar{\nu}_\alpha \to \bar{\nu}_\alpha), \tag{2.22}$$

which tells us that CP-violation can only be directly observed in 'appearance' experiments, where one investigates how many neutrinos (of a different flavour to the original flavour) appear after some distance. As an example, electron appearance experiments examine the probability

$$\begin{aligned}P\left(\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e\right) \approx{}& \sin^2\theta_{23}\sin^2 2\theta_{13}\sin^2\left(\frac{\Delta m^2_{31}L}{4E}\right) + 2\left[\sin\theta_{23}\sin 2\theta_{13}\sin\left(\frac{\Delta m^2_{31}L}{4E}\right)\right.\\ &\left.\times\cos\theta_{23}\cos\theta_{13}\sin 2\theta_{12}\sin\left(\frac{\Delta m^2_{21}L}{4E}\right)\times\cos\left(\frac{\Delta m^2_{32}L}{4E}\pm\delta_{\mathrm{CP}}\right)\right]\\ &+\cos^2\theta_{23}\cos^2\theta_{13}\sin^2 2\theta_{12}\sin^2\left(\frac{\Delta m^2_{21}L}{4E}\right),\end{aligned} \tag{2.23}$$

where the sign of δ_{CP} is positive (negative) in the case of neutrinos (antineutrinos) [38]. The asymmetry in the electron neutrino and antineutrino oscillation probability is reported as

$$A = \frac{P(\nu_\mu \to \nu_e) - P(\bar{\nu}_\mu \to \bar{\nu}_e)}{P(\nu_\mu \to \nu_e) + P(\bar{\nu}_\mu \to \bar{\nu}_e)}, \tag{2.24}$$

which is non-zero if CP is violated. The Jarlskog invariant (J) can be used to quantify the amount of CP-violation present in neutrino oscillations. In the representation used to express the PMNS matrix in equation 2.10, it is written as

$$J = \frac{1}{8}\cos\theta_{13}\sin 2\theta_{12}\sin 2\theta_{13}\sin 2\theta_{23}\sin\delta_{\mathrm{CP}}. \tag{2.25}$$

Here we see that the extent of CP-violation is quantified in terms of the CP-phase, where CP is only conserved if $\delta_{\mathrm{CP}} = 0, \pm\pi$ and is maximally violated if $\delta_{\mathrm{CP}} = \pm\pi/2$. The Jarlskog invariant is not unique to neutrino oscillations and allows for a comparison between the amount of CP-violation in the quark sector, where $J = 3.08^{+0.15}_{-0.13}\times 10^{-5}$ [42].

2.5 Matter Effects

In reality we study neutrinos that have, at some point, travelled through matter. Neutrinos interact with the electrons of the matter: ν_e through charged and neutral current weak interactions, and $\nu_{\mu/\tau}$ through neutral current interactions only. The difference in the potential that ν_e and $\nu_{\mu/\tau}$ experience can be written as:

$$\Delta V = 2\sqrt{2}G_F E N_e, \tag{2.26}$$

where G_F is the Fermi constant, E is the neutrino energy and N_e is the electron number density of the matter [39]. With this potential, the time evolution of the flavour eigenstates is now governed by

$$i\frac{d}{dt}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} = \left[H_V^F + \begin{pmatrix}\Delta V & 0\\ 0 & 0\end{pmatrix}\right]\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix}, \tag{2.27}$$

where H_V^F is the vacuum Hamiltonian in the flavour basis. Using the relationship $H_V^F = UH_V^M U^\dagger$, where H_V^M is the vacuum Hamiltonian in the mass basis, one can show that

$$H_V^F = H_V^M + \frac{\Delta m^2}{2E}\begin{pmatrix}-\cos 2\theta & \sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{pmatrix}, \tag{2.28}$$

where

$$H_V^M = \frac{1}{2E}\begin{pmatrix}m_1^2 & 0\\ 0 & m_2^2\end{pmatrix}. \tag{2.29}$$

So how does this additional term affect the mass eigenstates, the mass eignenvalues and, ultimately, the oscillation probability? To answer this, we use the relationship $H_M^M = U^\dagger H_M^F U$, where H_M^M and H_M^F are respectively the matter Hamiltonian in the mass and flavour bases, to obtain

$$H_M^M = \frac{1}{2E}\begin{pmatrix}m_1^2 + \Delta V\cos^2\theta & \Delta V\cos\theta\sin\theta\\ \Delta V\cos\theta\sin\theta & m_2^2 + \Delta V\sin^2\theta\end{pmatrix}. \tag{2.30}$$

The modified mass eigenstates and eigenvalues are retrieved from the diagonalisation of equation 2.30 and can be interpreted as those of an ‘effective neutrino mass’. They satisfy

$$P(\nu_e \to \nu_\mu) = \sin^2 2\theta_m \sin^2\left(\frac{\Delta m_m^2 L}{2E}\right), \tag{2.31}$$

where θ_m is the effective mixing angle and Δm_m^2 is the effective mass splitting. Their expressions, in terms of the vacuum mixing angle and mass splitting terms, are:

$$\Delta m_m^2 = \Delta m^2\lambda \tag{2.32}$$

$$\sin 2\theta_m = \frac{\sin 2\theta}{\lambda}, \tag{2.33}$$

where

$$\lambda = \sqrt{\left(\frac{\Delta V}{\Delta m^2} - \cos 2\theta\right)^2 + \sin^2 2\theta}. \tag{2.34}$$

The effect of matter on neutrino oscillations is referred to as the ‘matter effect’, or as the Mikheyev-Smirnov-Wolfenstein (MSW) effect, after those who contributed to its formulation [43, 44]. From these equations we see that:

- If the vacuum mixing angle is zero, then the effective mixing angle is zero i.e.

neutrino oscillations will not occur in matter unless they occur in a vacuum.

- If the matter is very dense then $\sin 2\theta_m$ and, consequently the neutrino oscillation probability, will be suppressed.
- If $\Delta V/\Delta m^2 = \cos 2\theta$ then a maximal effective mixing angle is achieved for any non-zero value of the vacuum mixing angle. This is referred to as the MSW resonance.
- Linear terms of the mass splitting enter the oscillation probability allowing their sign to be determined.

This discussion of the matter effect has focused on neutrinos however, the picture is almost identical for antineutrinos apart from the sign of ΔV. This sign change means that the oscillation probability of neutrinos and antineutrinos differ which mimics the signature of CP-violation making it harder to measure. As we will see in section 2.7, experiments which aim to measure δ_{CP} can be designed to handle this.

2.6 Current Experimental Status

By the end of this section, we will have identified that:

- $|\Delta m^2_{21}| \ll |\Delta m^2_{32}|$,
- θ_{23} is near maximal,
- θ_{12} is relatively large,
- θ_{13} is relatively very small.

The dominant and subdominant terms of the oscillation probabilities are therefore different and depend on the oscillation channel. This motivates the need for several experimental configurations, which probe different oscillation channels and can be tuned to enhance the sensitivity to a given set of oscillation parameters. This idea will be illustrated in this section, where the determination of the oscillation parameters will be discussed in order of the ease with which they are measured and, therefore, the advancement of experimental progress over time.

2.6.1 The Solar Parameters

The solar neutrino experiments of subsection 2.1.1 probed the disappearance probability of the solar ν_e. These types of experiments have long baselines approximately equal to the distance between the Sun and Earth (1 AU $\approx 1.5 \times 10^8$km) and measure neutrinos with an energy < 10MeV (see figure 2.1). They satisfy

$$\frac{\Delta m^2_{31} L}{E} \gg 1 \qquad \text{and} \qquad \frac{\Delta m^2_{21} L}{E} \geq 1, \tag{2.35}$$

such that, to leading order, the oscillation survival probability in a vacuum [33] is given by

$$P(\nu_e \to \nu_e) = P(\bar{\nu}_e \to \bar{\nu}_e) \approx 1 - \sin^2\theta_{12} \sin^2\left(\frac{\Delta m^2_{21} L}{4E}\right). \tag{2.36}$$

The solar experiments therefore measure θ_{12} and Δm^2_{21}, where the sign of the latter can be determined by the MSW contribution arising from the propagation of the neutrinos through the sun [45]. These parameters are referred to as the 'solar parameters'.

The solar parameters can also be determined by long baseline reactor experiments, which measure the disappearance of low energy antineutrinos. One such example is Kamioka Liquid Scintillator Anti-Neutrino Detector (KamLAND) [46], a liquid scintillator detector which sits in the Kamioka Mine, Japan. With an average baseline of 180km and a peak antineutrino energy of ~4MeV, the oscillation probability probed by KamLAND is also given by equation 2.36.

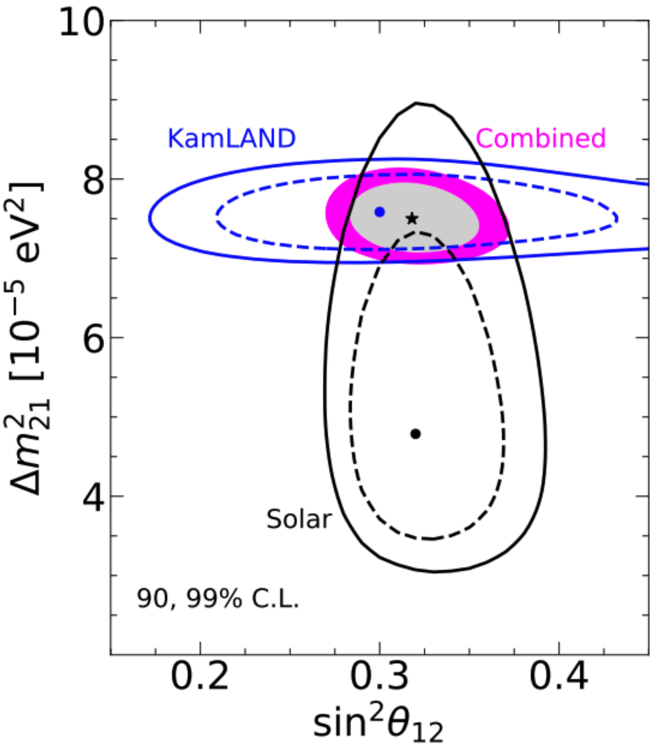


Figure 2.4: The 90% (dashed lines) and 99% (solid lines) confidence limits and best-fit positions (dots) of the ($\sin^2\theta_{12}, \Delta m^2_{21}$) phase space obtained by the solar experiments (black) and KamLAND (blue). The results of a combined fit are shown, where the 90% and 99% confidence regions are shown in grey and pink respectively and the best-fit position is denoted with a star. Figure reproduced from reference [47].

Figure 2.4 presents the allowed region of the solar parameters as determined by the solar and long baseline reactor experiments. It is seen that the solar neutrino experiments have a higher sensitivity to θ_{12} (a consequence of the higher statistics and therefore greater precision to the oscillation amplitude), whilst long-baseline reactor experiments have a higher sensitivity to Δm^2_{21} (a consequence of the higher energy resolution and hence greater precision to the oscillation frequency).

2.6.2 The Atmospheric Parameters

Both the IceCube [48] experiment, which has a similar detection principle to SK but uses the Antarctic ice as its detector medium, and SK [49] probe the disappearance probability of atmospheric muon neutrinos. The baseline of these experiments lies between $15 - 1.3 \times 10^4$km whilst their energies are high: of the order of tens of GeV. The condition

$$\frac{\Delta m^2_{21} L}{E} \ll 1 \tag{2.37}$$

is therefore met. With this approximation, and assuming that $\Delta m^2_{32} \approxeq m^2_{31}$, the dominant terms of the muon neutrino survival probability [33] can be shown to be

$$P(\nu_\mu \to \nu_\mu) \approx 1 - \sin^2 2\theta_{23} \sin^2 \left(\frac{\Delta m^2_{31} L}{4E} \right) \tag{2.38}$$

These experiments have been able to determine θ_{23} and $|\Delta m^2_{31}|$, which are consequently known as the 'atmospheric parameters'. Their results are shown in figure 2.5 where, it is clear that:

- The sign of Δm^2_{31} cannot be determined. This is because there is no significant MSW contribution to the oscillation probability i.e. Δm^2_{31} only enters equation 2.38 via squared sinusoidal terms and so $P(\Delta m^2_{31}) = P(-\Delta m^2_{31})$.
- The octant of θ_{23} (whether $\theta_{23} > \pi/4$ or $\theta_{23} < \pi/4$) cannot be determined. This is because θ_{23} appears in the oscillation probability as $\sin^2 2\theta_{23}$ and so $P(\theta_{23}) = P(\pi/2 - \theta_{23})$.
- θ_{23} is measured with less precision than the solar mixing angle. This is because its true value is close to (or equal to) the maximal mixing value and here the change in the disappearance rate as a function θ_{23} is minimised.

So called Long-Baseline Accelerator (LBL) experiments also satisfy the condition of equation 2.37. These experiments - illustrated in figure 3.1 - consist of a $\overset{(-)}{\nu}_\mu$ beam, a near detector and a far detector, the latter of which often lies underground to reduce backgrounds. In a greatly simplified overview, the near detector measures the unoscillated

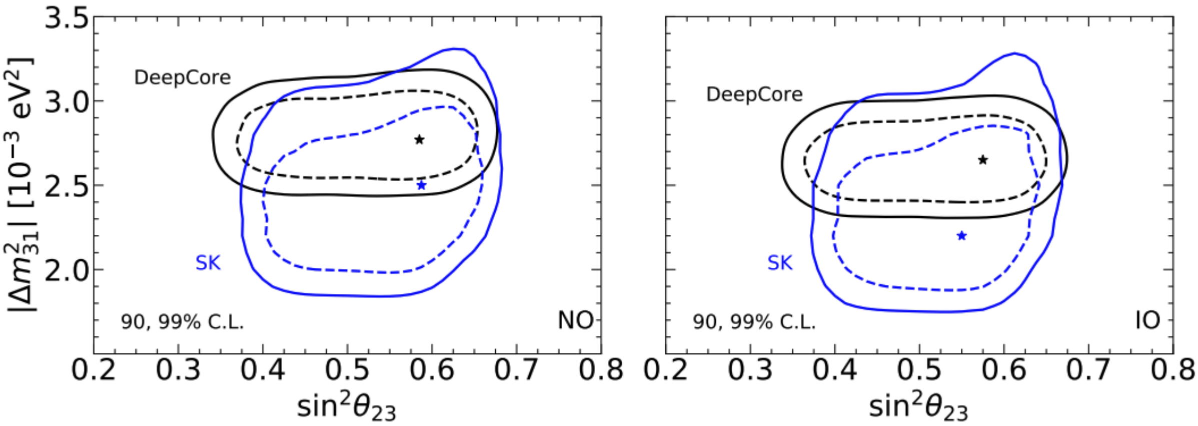


Figure 2.5: The 90% (dashed lines) and 99% (solid lines) confidence limits and best-fit positions (stars) of the $(\sin^2\theta_{23}, |\Delta m^2_{31}|)$ phase space obtained by IceCube/DeepCore (black) and Super-Kamiokande (blue) assuming the Normal Ordering (NO) and Inverted Ordering (IO) of the mass-hierarchy. Figure reproduced from reference [47].

$\overset{(-)}{\nu}_\mu$ beam flux whilst the far detector measures the oscillated flux, and comparison of the two enables one to infer the oscillation parameters. The placement of the near and far detectors can be tuned such that the L/E value maximises the oscillation probability and the statistics of the experiment:

- **The baseline (L)** is determined by the separation of the near and far detectors. The beam widens with distance from the source, its intensity following a $1/L^2$ dependence, which reduces the statistics attained at the far detector and thus places an upper limit on the baseline length.

- **The neutrino energy (E)** is determined by the beam design, and has both an upper and lower limit. Experiments also make use of the dependence of the flux energy composition on the transverse distance from the beam axis (see figure 3.10). One can choose to situate the detectors along the beamline, where the energy spectrum is broad, allowing a wide range of L/E values to be probed. Alternatively, one can place the detectors 'off-axis', where the beam becomes almost monoenergetic. As one moves further off-axis the peak energy falls which allows particular energies to be sampled. However, the off-axis extent is limited as the beam intensity decreases with off-axis distance.

The LBL experiments that have contributed most to the determination of the atmospheric parameters are: the Main Injector Neutrino Oscillation Search (MINOS) [50], Tokai to Kamioka (T2K) [51] and the NuMI Off-Axis Electron Neutrino Appearance (NOνA) [52]. They are summarised in table 2.1 and their oscillation results are presented in figure 2.6.

It is seen that, despite an MSW contribution, that arises in the case of an underground far detector, the LBL experiments have not determined the sign of Δm^2_{31}. This is because

Experiment	Detection Principle	Off/On-Axis	Baseline [km]	Peak Energy [GeV]
MINOS	Steel-scintillator calorimeter	On-axis	735	3
T2K	Water Cherenkov	2.5° off-axis	295	1
NOνA	Liquid scintillator	0.8° off-axis	810	2

Table 2.1: Details of the LBL experiments, where ‘detection principle’ refers to that used by the experiment’s far detector.

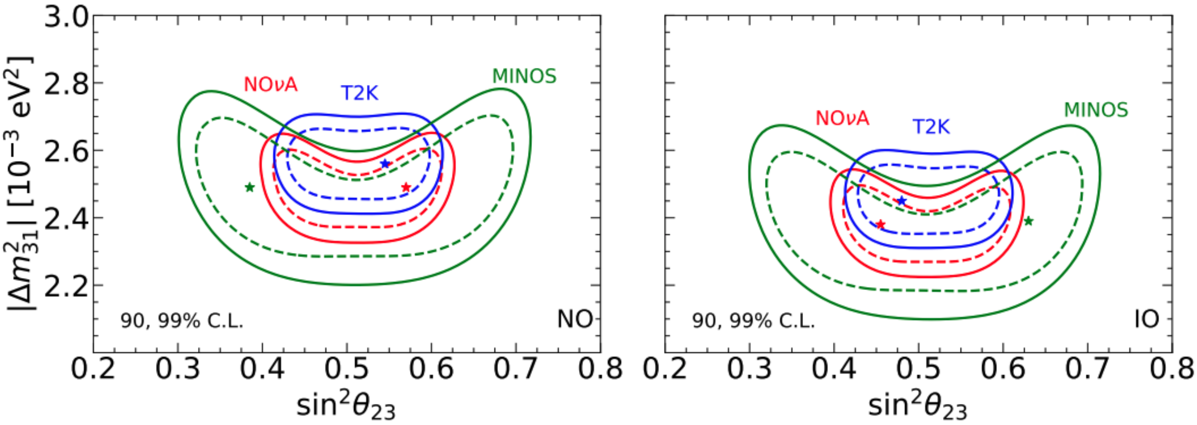


Figure 2.6: The 90% (dashed lines) and 99% (solid lines) confidence limits and best-fit positions (stars) of the $(\sin^2\theta_{23}, |\Delta m^2_{31}|)$ phase space obtained by T2K (blue), MINOS (green) and NOνA (red) assuming a normal mass-ordering (NO) and inverted mass-ordering (IO). Figure reproduced from reference [47].

the MSW contribution is too small to be disentangled from the effects of CP-violation. Consequently, there are two potential orderings of the mass eigenstates, or the mass hierarchy: normal-ordering/hierarchy ($m_1^2 < m_2^2 < m_3^2$) and inverted-ordering/hierarchy ($m_3^2 < m_1^2 < m_2^2$). These are illustrated in figure 2.7.

2.6.3 The Final Mixing Angle, θ_{13}

The vacuum oscillation probability [54] probed by the solar parameter experiments of subsection 2.6.1 is more accurately given by

$$P(\nu_e \to \nu_e) = P(\bar{\nu}_e \to \bar{\nu}_e) \approx \sin^4\theta_{13} + \cos^4\theta_{13}\left[1 - \sin^2 2\theta_{12}\sin\left(\frac{\Delta m^2_{21}L}{4E}\right)\right], \tag{2.39}$$

which can be approximated to

$$P(\nu_e \to \nu_e) = P(\bar{\nu}_e \to \bar{\nu}_e) \approx (1 - 2\sin^2\theta_{13})\left[1 - \sin^2 2\theta_{12}\sin\left(\frac{\Delta m^2_{21}L}{4E}\right)\right], \tag{2.40}$$

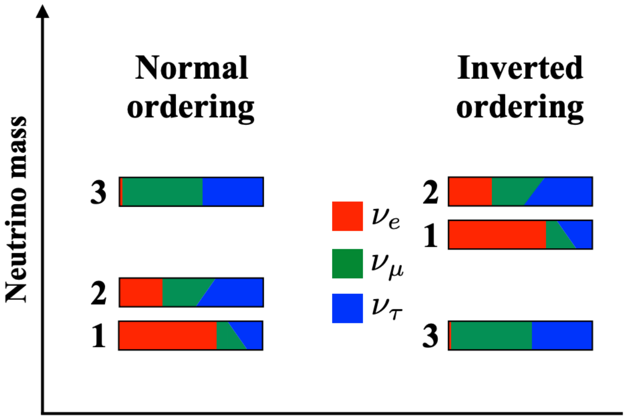


Figure 2.7: The mass eigenstates, ordered by their mass in the normal (left) and inverted (right) hierarchies. Their composition, in terms of the ν_e (red), ν_μ (green) and ν_τ (blue) flavour eigenstates is shown. Within each coloured band, δ_{CP} is varied from 0 (bottom) to π (top). Figure reproduced from reference [53].

where the affect of a non-zero θ_{13} is to suppress the oscillation probability. The allowed region of θ_{13} values determined by KamLAND and the solar experiments are shown in figure 2.8. The experiments prefer a non-zero value of θ_{13} as a result of an observed suppression of the oscillation probability but were limited by low statistics.

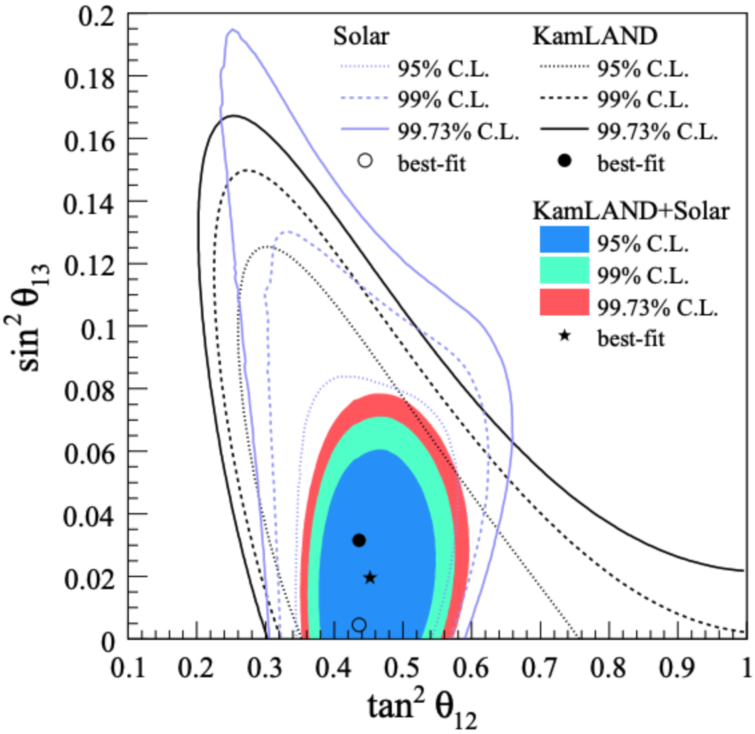


Figure 2.8: The allowed regions of the $(\tan^2\theta_{12}, \sin^2\theta_{13})$ phase space obtained by the solar experiments (blue), KamLAND (black) and from a combined fit (various shaded colours). Figure reproduced from reference [54].

The LBL experiments of subsection 2.6.2 also have access to θ_{13} via the electron appearance channel, which to leading-order, can be shown to be

$$P(\nu_\mu \to \nu_e) \approx \sin^2\theta_{23}\sin^2 2\theta_{13}\sin\left(\frac{\Delta m^2_{31}L}{4E}\right) \tag{2.41}$$

in a vacuum [33]. It is noted that $P(\nu_\mu \to \nu_e)$ has sub-dominant contributions from δ_{CP} (see equation 2.23). Determination of θ_{13} via electron appearance is not trivial. Firstly, despite the L/E optimisation of the LBL experiments, statistics for this channel are low, which places high demands on one's ability to constrain systematics and select events. But, the bigger issue here is that a degeneracy exists between δ_{CP} and θ_{13} such that, for a fixed baseline and neutrino energy, $P(\nu_\mu \to \nu_e)$ has a continuous curve of solutions (figure 2.9) [55]. It is seen how this 'infinite degeneracy' can be reduced to a 'two-fold degeneracy' by incorporating a measurement of $P(\bar{\nu}_\mu \to \bar{\nu}_e)$ and figure 2.10 demonstrates how the degeneracy can be lifted with a multi-valued L/E measurement, or with the data of another oscillation channel. Figure 2.11 presents the allowed $(\theta_{13}, \delta_{\mathrm{CP}})$ region obtained by the most-sensitive LBL experiments (T2K [56] and NOνA [57]), where it is seen that a non-zero value of θ_{13} is preferred.

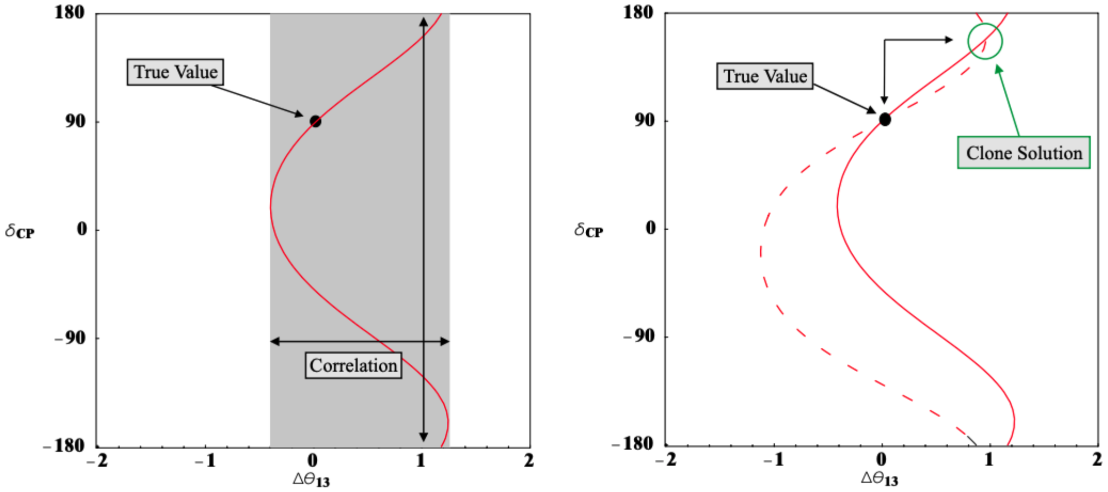


Figure 2.9: The continuous curves of $(\theta_{13}, \delta_{\mathrm{CP}})$ solutions for a neutrino (solid line) and antineutrino (dashed line) appearance probability. With just neutrino data, an 'infinite degeneracy' is present (left) but with the addition of antineutrino data this is reduced to a 'two-fold' degeneracy (right). Figure reproduced from reference [55].

The most precise measurements of θ_{13} come from the short baseline reactor experiments (Double Chooz [58], the Reactor Experiment for Neutrino Oscillation (RENO) [59] and Daya Bay [60]), which probe the disappearance of reactor electron antineutrinos. These experiments are liquid-scintillator based and measure the low-energy (of the order of a few MeV) antineutrinos via IBD which has a threshold of 1.8MeV [33]. With knowledge that the oscillations are driven by Δm^2_{31}, a baseline in the 1-2km region is needed to optimise the oscillation probability and maximise statistics. These experiments

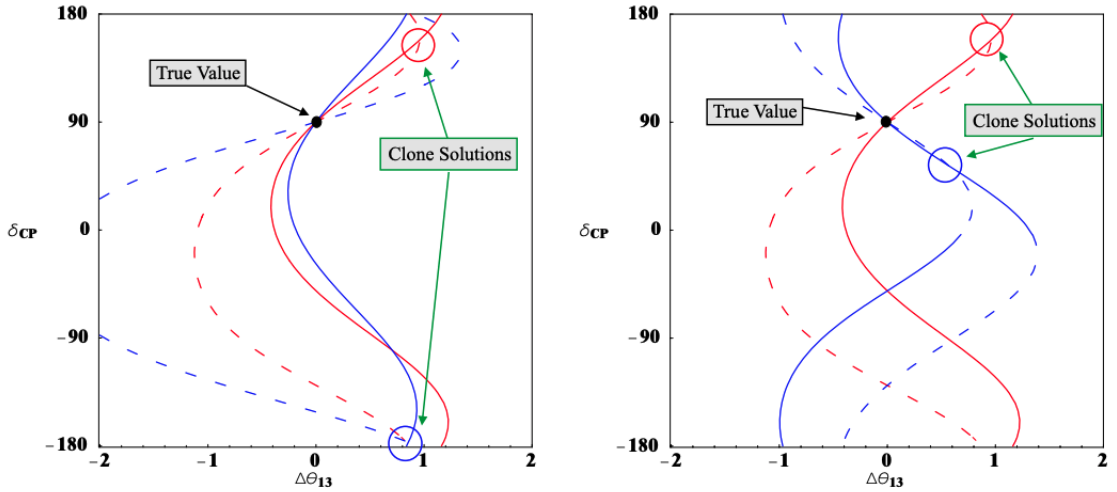


Figure 2.10: The continuous curves of $(\theta_{13}, \delta_{\text{CP}})$ solutions for a neutrino (solid line) and antineutrino (dashed line) appearance probability. The different colours denote the measurement at two different L/E values (left) and two different appearance channels (right). Figure reproduced from reference [55].

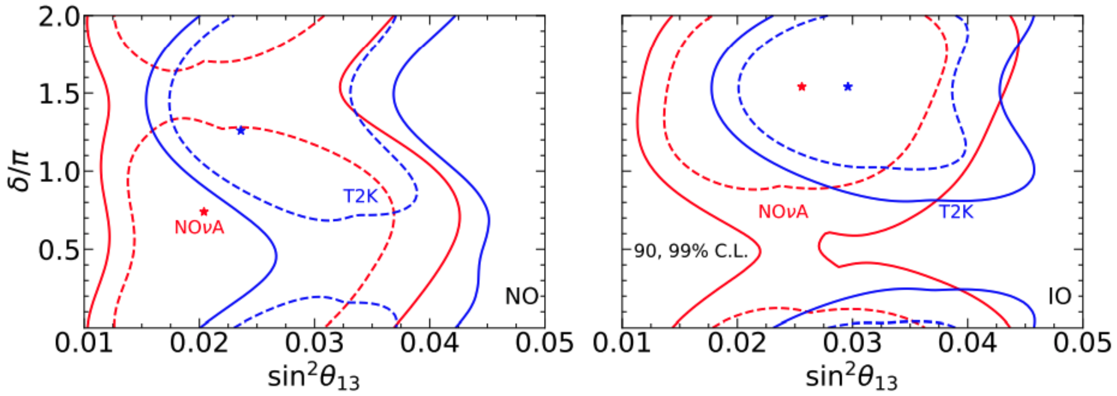


Figure 2.11: The 90% (dashed lines) and 99% (solid lines) confidence limits and best-fit positions (stars) of the $(\sin^2\theta_{13}, \delta_{\text{CP}})$ phase space obtained by the LBL experiments, T2K (blue) and NOνA (red), assuming a normal mass-ordering (NO) and inverted mass-ordering (IO). Figure reproduced from reference [47].

therefore meet the condition of equation 2.37 such that, to leading order, the relevant vacuum oscillation survival probability is given by

$$P(\bar{\nu}_e \rightarrow \bar{\nu}_e) = 1 - \sin^2 2\theta_{13} \sin^2\left(\frac{\Delta m^2_{31} L}{4E}\right). \tag{2.42}$$

Comparison of equations 2.41 and 2.42 reveals that short baseline reactor experiments will have much higher statistics than LBL experiments. Furthermore, the electron antineutrino disappearance channel does not suffer from the same degeneracy issue that affected the electron neutrino appearance channel. This allows the reactor experiments to determine θ_{13} to a high precision, as shown in figure 2.12.

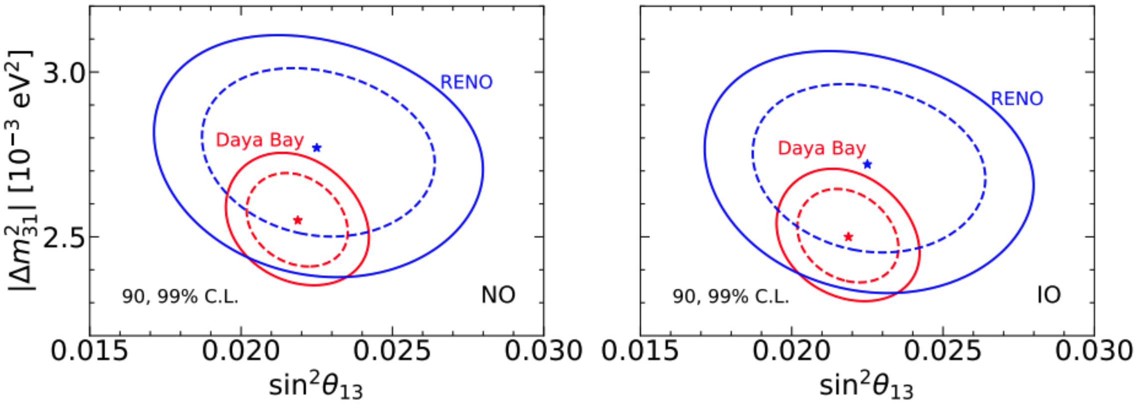


Figure 2.12: The 90% (dashed lines) and 99% (solid lines) confidence limits and best-fit positions (stars) of the ($\sin^2\theta_{13}, \delta_{CP}$) phase space, as obtained by the short baseline reactor experiments, RENO (blue) and Daya Bay (red), assuming a normal mass-ordering (NO) and inverted mass-ordering (IO). Figure reproduced from reference [47].

2.6.4 The CP-Phase

The ν_e appearance channel of the LBL experiments is most sensitive to the value of δ_{CP}: that it has any sensitivity at all is ensured by the non-zero value of θ_{13} (see equation 2.10). T2K and NOνA can use the reactor measurements of θ_{13}, which is referred to as the 'reactor constraint', to gain an insight into the value of δ_{CP}. They are unable to measure its precise value as a result of their statistical and systematic uncertainties, and because they cannot resolve the degeneracies that arise from the mass hierarchy and θ_{23} octant. NOνA disfavours the maximally violating values of the CP-phase and rejects $\delta_{CP} = \pi/2(3\pi/2)$ at $3\sigma(2\sigma)$ in the case inverted(normal) ordering [18]. T2K however has a strong preference for maximally violating values of the CP-phase and rejects both CP-conserving values ($\delta_{CP} = 0, \pi$) at 2σ [19]. Figure 2.13 illustrates the slight tension between the LBL experiments. It is the aim of the next generation of LBL experiments, DUNE [61] and Hyper-Kamiokande [62], to precisely measure the value of the CP-phase.

2.7 Sensitivity to the CP-Phase

So, how should DUNE be designed to determine whether CP is violated, and the value of the CP-phase? It has been shown in this chapter how matter effects mimic the CP-violation signal, and that the current LBL experiments are unable to disentangle the two contributions. However, these effects can be resolved if data is taken at several L/E points, such that multiple oscillation peaks are probed [63]. To do this, one either focuses on L, and builds multiple detectors with different baselines, or on E, and uses a wide-band beam and detectors with a good energy resolution. Considering the size of the required detectors, it is cheaper to do the latter. Constraints on the beam energy and baseline (see subsection 2.6.2) mean that, if the baseline is long enough, one can be sensitive to

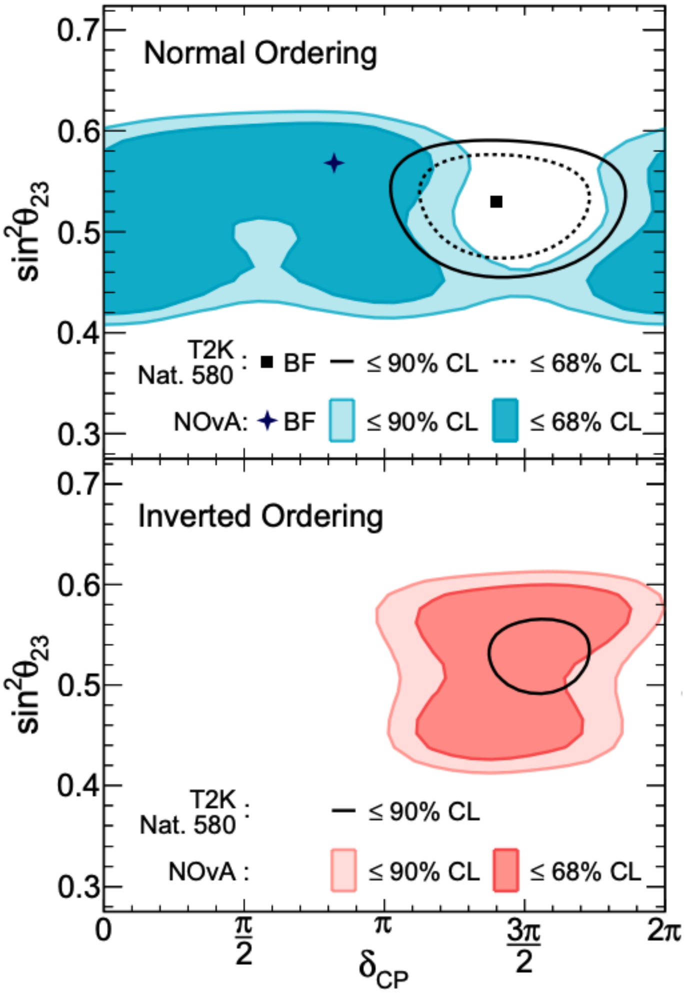


Figure 2.13: The confidence limits (CL) and best-fit (BF) positions of the $(\delta_{\mathrm{CP}}, \sin^2\theta_{23})$ phase space obtained by the T2K and NOνA experiments, assuming a normal mass-ordering (top) and inverted mass-ordering (bottom). Figure reproduced from reference [18].

two oscillation peaks. This is demonstrated in figure 2.14 which shows $P(\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e)$ as a function of neutrino energy at a baseline of 1300km (that chosen by DUNE). Detector and reconstruction thresholds, and energy resolution capabilities, mean that DUNE is only able to access the two oscillation peaks $\gtrsim$ 500MeV. But why has DUNE picked a baseline of 1300km?

Figure 2.15 demonstrates the CP-violation and matter-effect contributions to the electron neutrino-antineutrino appearance asymmetry, at the first and second oscillation peaks, as a function of the baseline. It is seen that for NO (IO):

- The first peak is dominated by the CP-violation (matter effect) contribution at small baselines and the matter effect (CP-violation) at long baselines.
- The CP-violation (matter effect) contribution dominates the second peak at all baselines.

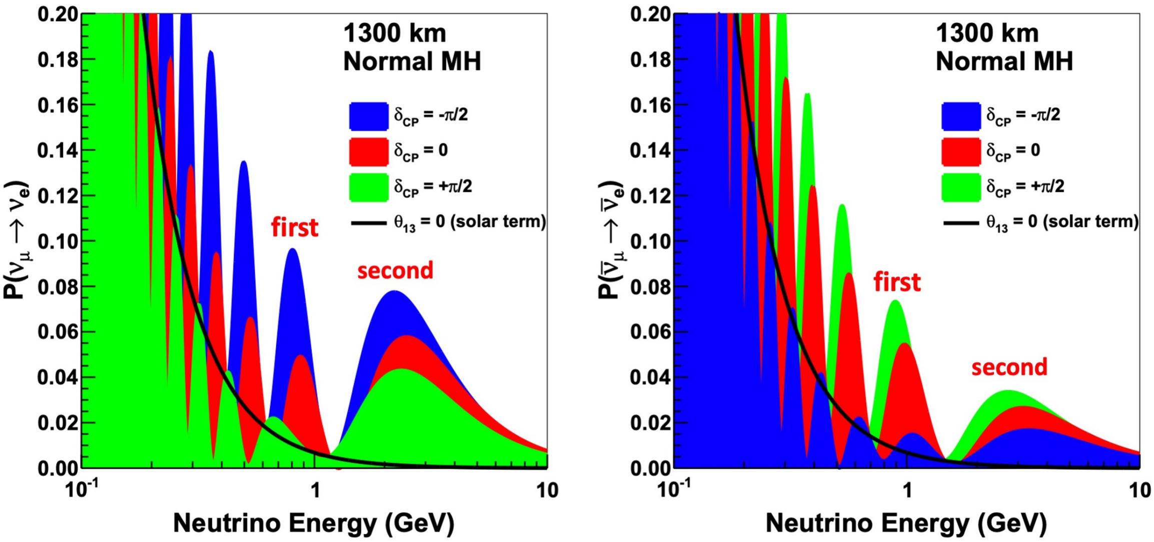


Figure 2.14: $P(\nu_\mu \to \nu_e)$ (left) and $P(\bar{\nu}_\mu \to \bar{\nu}_e)$ (right) as a function of neutrino energy at a baseline of 1300km (that chosen by DUNE). The first and second oscillation maxima that are experimentally accessible are highlighted. Figure reproduced from reference [53].

Therefore, assuming NO (IO), at 1300km one is able to use the first(second) oscillation peak to constrain the matter contribution (and measure the sign of Δm^2_{31}) such that one can determine whether CP is violated (and the value of δ_{CP}).

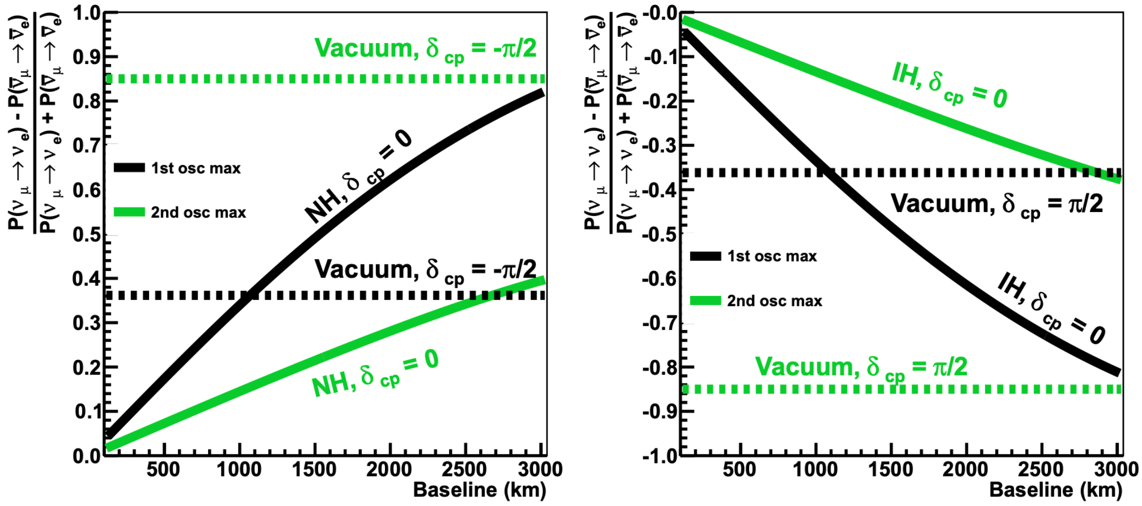


Figure 2.15: The electron neutrino-antineutrino appearance asymmetry as a function of the baseline at the first (black) and second (green) oscillation peak, assuming a normal hierarchy (left) and inverted hierarchy (right), where the matter-effect and CP-violation contributions are shown by solid and dashed lines respectively. The CP-violation contribution corresponds to the maximally violating phase. Figure adapted from reference [63].

Equation 2.23 demonstrates that the CP-violation contribution is a subdominant effect. So to even achieve any sensitivity, DUNE will need high statistics and well constrained systematics. This translates into the need for a high-intensity neutrino beam, a large far detector, an efficient and pure electron neutrino selection procedure, and a sophisticated near detector suite. The next chapter will discuss the detectors of DUNE and how they meet these requirements. The selection procedure will then be the focus of chapter 6 and subsequent chapters.

Chapter 3

The Deep Underground Neutrino Experiment

DUNE is a next generation long-baseline neutrino experiment. The collaboration consists of more than 1000 members, from over 200 institutions in upwards of 30 countries, each striving to answer the biggest questions in modern day neutrino physics. DUNE will benefit from the success of its predecessors, using the most intense (anti-)neutrino beam ever made, operating liquid-argon time projection chambers at an unprecedented scale, and using innovative ideas to constrain systematics to a world-leading precision.

This chapter will present an overview of the physical elements of the DUNE experiment and will motivate their need and design in context of the flagship neutrino oscillation analysis of DUNE, which is the analysis of this PhD. The following studies are performed with the horizontal drift far detector module and its prototype, ProtoDUNE-SP, and so particular attention is paid to these detectors.

3.1 An Overview of the DUNE Experiment

The DUNE experiment primarily aims to:

- precisely measure the neutrino oscillation parameters determining the value of the CP-phase, the mass hierarchy ordering and the octant of θ_{23} [64, 65],
- search for beyond the standard model physics e.g. proton decay [66], and
- detect low energy neutrinos, such as those from a supernova burst [67].

The design of the DUNE experiment has been tailored to meet these goals. In particular, to achieve precision measurements of the oscillation parameters, DUNE - shown in figure 3.1 - will consist of a neutrino beam, Near Detector (ND) and Far Detector (FD). The neutrino beam (section 3.4) will be provided by the Long-Baseline Neutrino Facility (LBNF) and begins its journey at the Fermi National Accelerator Laboratory (Fermilab), from where it travels 574m downstream to the ND site (section 3.3) and 1300km further to the FD complex (section 3.2) at the Sanford Underground Research Facility (SURF) which lies 1.5km underground. By comparing the observed neutrino interactions at the far and near detectors, the oscillation parameters can be determined.

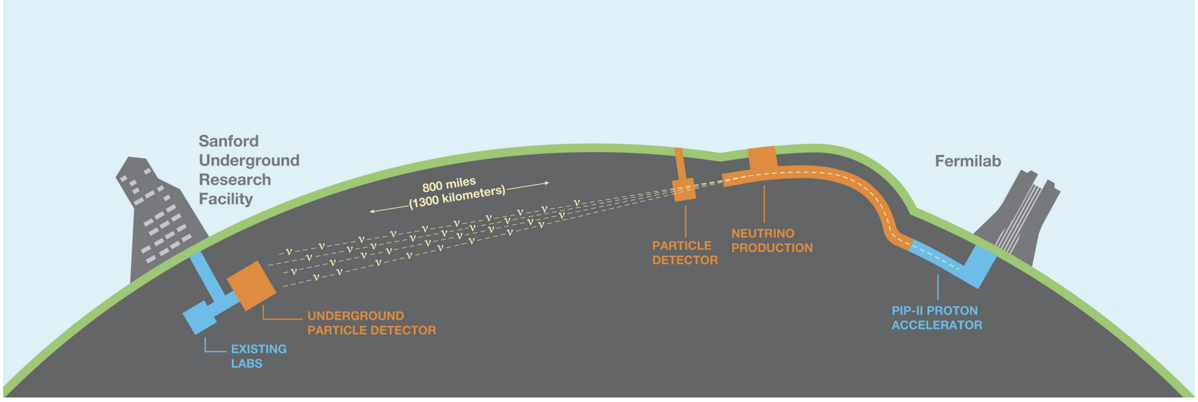


Figure 3.1: A cartoon of the components of the DUNE experiment and the LBNF beamline. Figure reproduced from reference [61].

Reference [68] details the timeline of DUNE. At the time of writing, the beamline, ND and FD complex designs are complete, and the FD site is under preparation. The upgrade of the proton accelerator at Fermilab is ongoing and is due to be completed by 2029. It is predicted that DUNE will begin taking data towards the end of this decade and that the neutrino beam will be operational by 2032.

The evolution of the experiment's construction is benchmarked by two phases, which are detailed in table 3.1. Phase I corresponds to a reduced far detector (with 50% of the planned fiducial mass), a limited near detector and a 1.2MW proton beam, whilst phase II sees a complete FD complex, upgraded near detector and a 2.4MW proton beam. Details of the key physics milestones achieved in each stage are given in table 3.2.

Parameter	Phase I	Phase II	Benefit
FD mass	20kt fiducial	40kt fiducial	FD statistics
Beam power	up to 1.2MW	2.4MW	FD statistics
ND configuration	ND-LAr, TMS, SAND	ND-LAr, ND-GAr, SAND	Systematic constraints

Table 3.1: The details of the phased approach of DUNE and the benefit of each upgrade with respect to DUNE's oscillation programme. Table adapted from reference [68].

3.2 The Far Detector

Figure 3.2 illustrates the layout of the FD complex, which consists of three underground caverns: two housing the detectors, and one the cryogenic systems [61]. A detector cavern

Experiment Stage	Physics Milestone	Exposure (kt·MW·yrs)	Years (Staged)
Phase I	5σ MO ($\delta_{\mathrm{CP}} = -\pi/2$)	16	1-2
	5σ MO (100% of δ_{CP} values)	66	3-5
	3σ CPV ($\delta_{\mathrm{CP}} = -\pi/2$)	100	4-6
Phase II	5σ CPV ($\delta_{\mathrm{CP}} = -\pi/2$)	334	7-8
	δ_{CP} resolution of 10° ($\delta_{\mathrm{CP}} = 0$)	400	8-9
	5σ CPV (50% of δ_{CP} values)	646	11
	3σ CPV (75% of δ_{CP} values)	936	14
	$\sin^2(2\theta_{13})$ resolution of 0.004	1079	16

Table 3.2: The predicted time and exposure at which key physics milestones are reached where, CPV and MO denote a measurement of CP-violation and the determination of the mass hierarchy ordering. Table adapted from reference [68].

will contain two 65.8m (along beam) × 18.9m (horizontal) × 17.8m (height) cryostats, each with a 10kt fiducial and 17.5kt total mass. To put this into perspective, a cryostat is approximately the length of two and a half tennis courts and as tall as three adult giraffes stacked on top of one another.

Three of the four detector modules will use Liquid-Argon Time Projection Chamber (LArTPC) technology: modules I and II will have a Horizontal Drift (HD) direction whilst module III will have a Vertical Drift (VD) direction. The technology of module IV is, as yet, undecided and is an exciting opportunity for detector innovation.

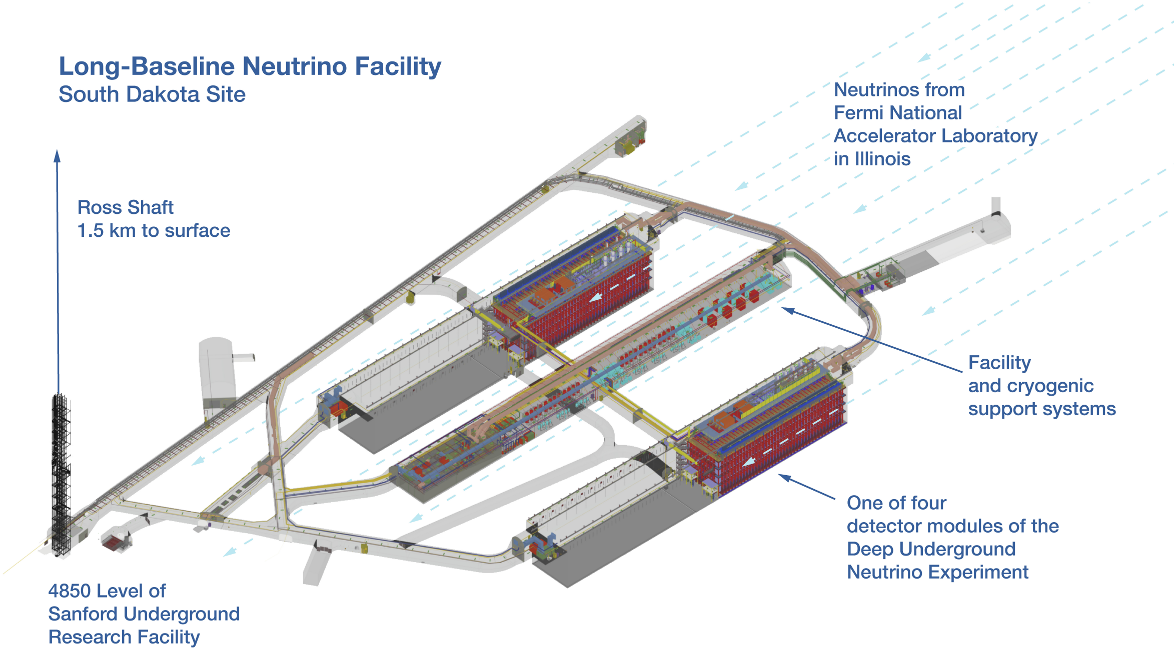


Figure 3.2: The layout of the FD complex during phase I of the experiment, where only modules I and II have been implemented. Figure reproduced from reference [69].

3.2.1 LArTPC Technology

In 1974, David R. Nygren proposed the ‘axially-focused time-projector detector’, a gas filled cylinder with a 2D spatial readout at one end [70]. Charged particles entering the detector would ionise the gas, and the resulting ionisation electrons would be drifted to the readout plane by an applied axial electric field, their transverse diffusion limited by a parallel magnetic field. The combination of the drift times and the 2D coordinates of the incident electrons would allow one to reconstruct the 3D trajectory of the particle.

Carlo Rubbia developed Nygren’s idea, losing the impractical - on large scales - magnetic field, and replacing the gaseous medium with Liquid-Argon (LAr) to form a LArTPC [71]. A liquid medium is preferable for neutrino experiments as a denser medium increases the neutrino interaction probability. Furthermore, LAr has a high electron mobility, low ionisation thresholds, high scintillation yields and, compared to other noble liquids, is relatively abundant and cheap [72]. Rubbia suggested that the 2D coordinate could be inferred by combining the readout of strips with different orientations or by using grids.

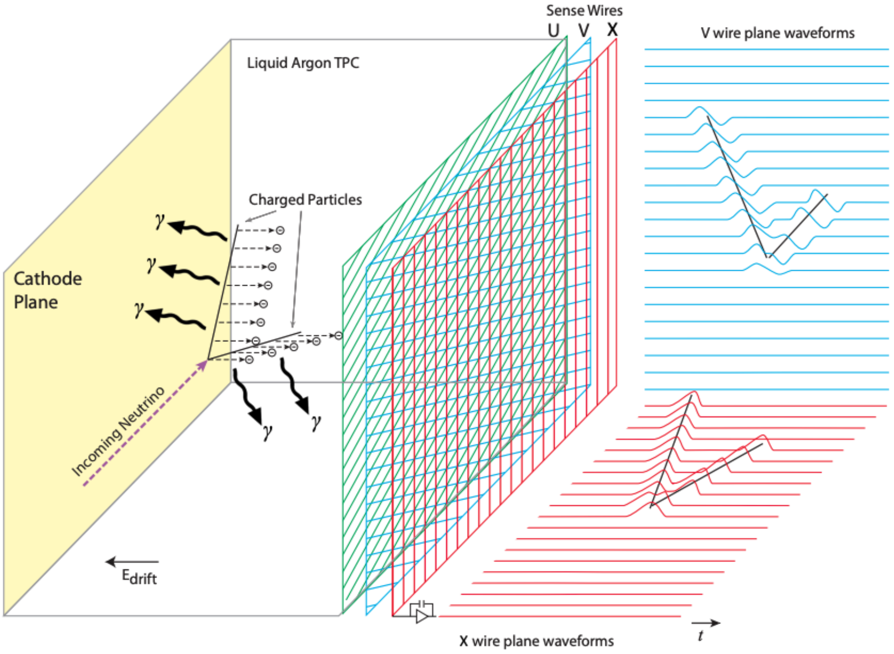


Figure 3.3: The operation of a LArTPC: a charged particle traverses the detector ionising the argon nuclei to create: ionisation electrons, which drift under an applied electric to the readout plane, and scintillation light, which is collected by the photon detection system (not shown). Figure reproduced from reference [73].

Figure 3.3 illustrates the LArTPC detection principle using the example of a HD module with a wire readout. Often overlooked is the scintillation photon signal, which can be collected and used to identify the start time of the ionisation electron drift with a high resolution. It is seen how LArTPCs capture neutrino interactions in a fine spatial and calorimetric resolution but this is can only be realised if a high LAr purity and low electronic noise is achieved.

The success of LArTPC technology in neutrino experiments has been demonstrated by the Micro Booster Neutrino Experiment (MicroBooNE) [72], and the Imaging Cosmic and Rare Underground Signals (ICARUS) [74], Argon Neutrino Test Project (ArgoNeuT) [75], LArTPC in a Testbeam (LArIAT) [76], and ProtoDUNE-SP [77] experiments.

3.2.2 Horizontal Drift

Modules I and II of the DUNE FD will be Single-Phase (SP) HD LArTPCs, where SP refers to the fact that the detector medium is solely liquid. A Technical Design Report (TDR) [73] details the design and operation of the HD modules, which are illustrated in figure 3.4.

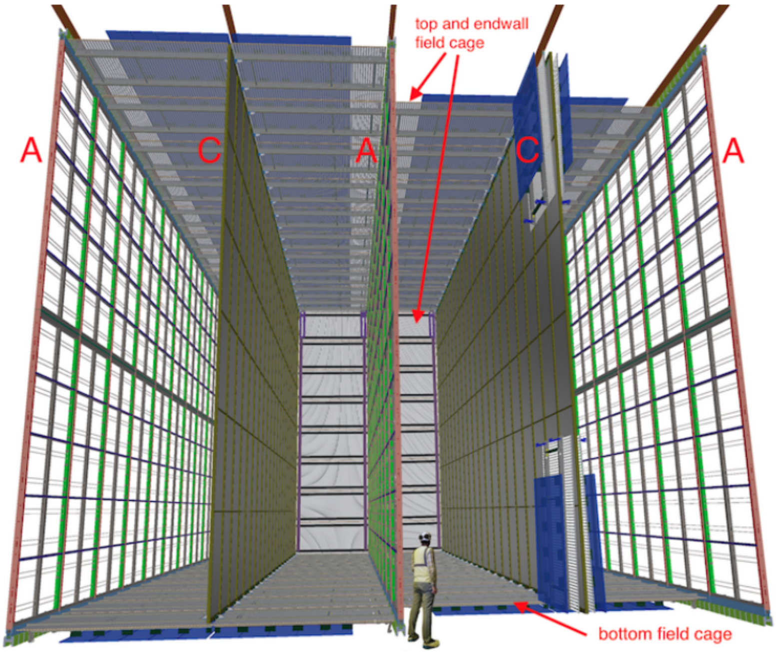


Figure 3.4: The planned design of modules I and II, where 'A' and 'C' denote the anode and cathode planes respectively. Figure reproduced from reference [73].

The modules will be segmented in the drift direction, by alternating anode and cathode walls, into four chambers, each with a maximum drift length of 3.5m. The cathodes are kept at a high voltage of -180kV, such that a horizontal 500V/cm electric field is present across the drift volumes; its uniformity is ensured by a surrounding field cage. The anode wall, which provides the detector readout, is a 25 (length) × 2 (height) grid of Anode Plane Assemblies (APAs).

An APA - shown in figure 3.5 - is a metal frame with four layers of wires wrapped around it, such that on each side there is a group of U and V wires, orientated at $\pm 35.7°$ to the vertical, and W and G wires orientated parallel to the vertical. The voltage of the wires is such that the ionisation electrons drift past the U and V wires of the 'induction planes' and are collected by the W wires of the 'collection plane'. The G wires form the 'shielding plane' which protects the U, V and W wires from long-range induction effects. The separation of the wires is ~5mm.

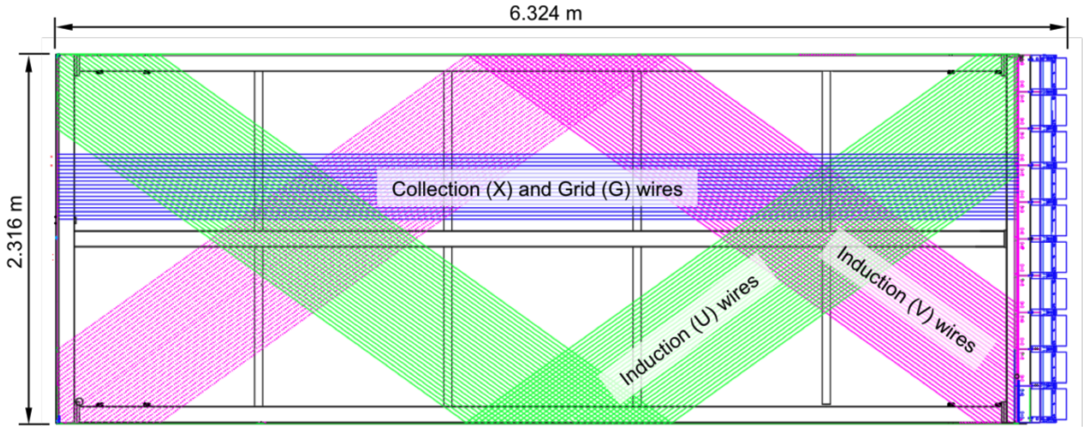


Figure 3.5: A schematic of a horizontal APA frame showing the wrapped U, V, W (denoted as X) and G wires. APA frames will hang vertically in the detector. Figure reproduced from reference [73].

The photon detection system uses the novel ARAPUCA devices which modify the wavelength of the incident scintillation light, trapping the photons until they are collected. They are housed on 209cm (length) × 12cm (height) × 2cm (depth) bars which are implemented on the APAs between the wire layers as demonstrated in figure 3.6.

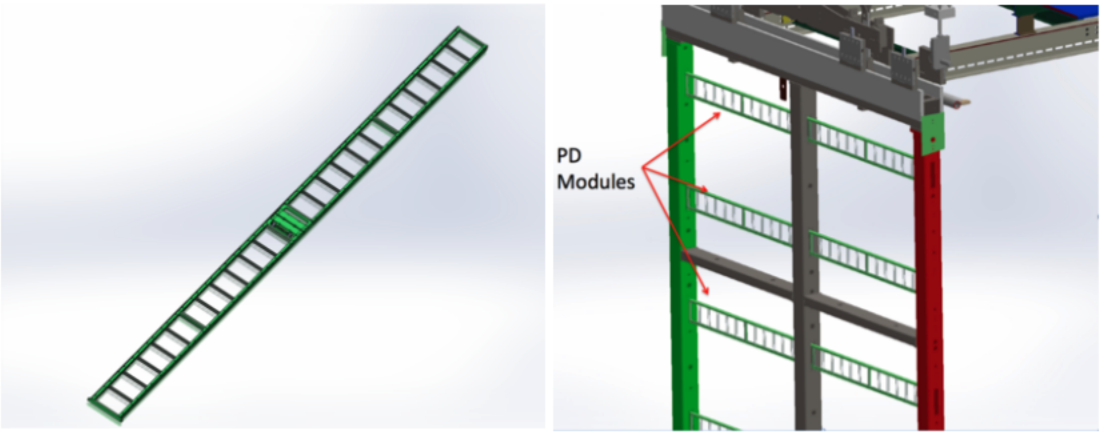


Figure 3.6: The bars of ARAPUCA devices (left) and their placement on the APAs (right). Figure reproduced from reference [73].

3.2.3 Vertical Drift

Module III is planned to be a SP VD LArTPC. Its design is an iteration on the proposed Dual-Phase (DP) module, which sandwiched a Gaseous-Argon (GAr) layer between the LAr and readout plane to amplify the ionisation electron signal [78]. The VD detector is advantageous as it meets the expectations of the DP design without the complications that arise from maintaining the LAr-GAr boundary. A Conceptual Design Report (CDR) and a TDR for the VD module is currently in progress, and its design is detailed in references [79, 80].

In the current design - shown in figure 3.7 - anode planes will be connected to the top and bottom walls of the detector and a cathode plane will split the volume, along the drift direction, into two chambers, each with a maximum drift distance of 6.5m. With this positioning, the readout electronics are accessible during the experiment's run time and can be fixed, allowing the dead regions that may otherwise occur to be addressed. This is not the case for the HD modules.

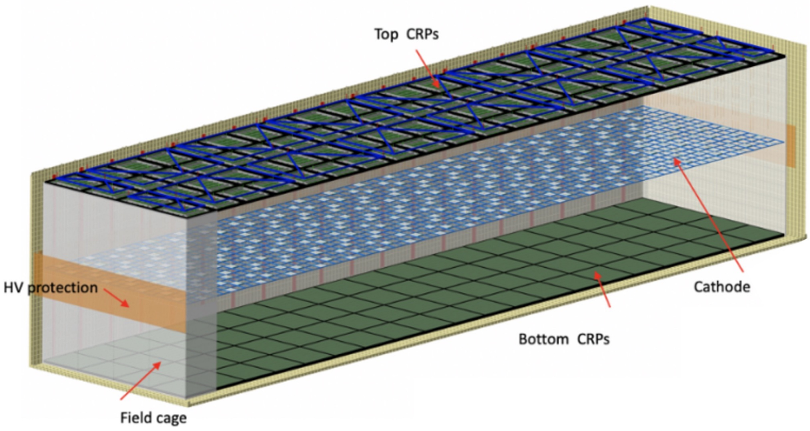


Figure 3.7: The planned design of module III. Figure reproduced from reference [79].

The cathode will be kept at a high voltage of −300kV, resulting in a 500V/cm electric field whose uniformity is maintained by a surrounding electric field cage. The anode planes are built from 3.4m × 3m Charge Readout Planes (CRPs), which are composed of perforated Printed Circuit Board (PCB) layers, as seen in figure 3.8. Holes are needed between the layers to allow the electrons to pass between them. Etched onto the PCBs are readout strips and their orientation and voltage are such that the shielding, induction, induction, collection plane structure is created.

The photon detection system will use the ARAPUCA devices that were developed for the HD detector. These are planned to be placed on the vertical walls of the cryostat, as well as on the cathode, to maximise the photon yield, which will exceed that obtained by the HD modules.

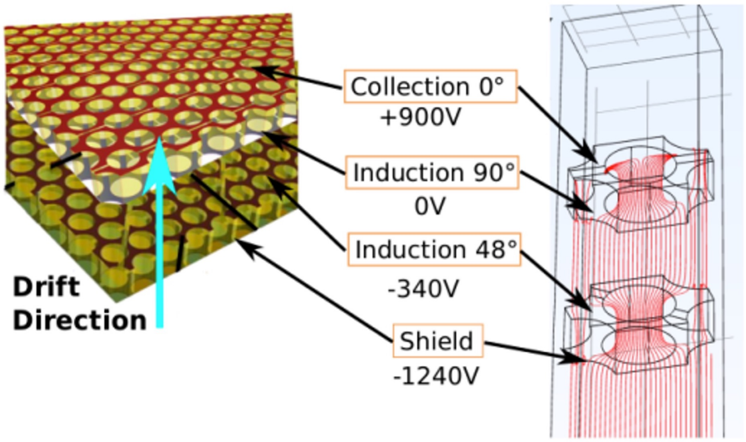


Figure 3.8: An illustration of the perforated PCB layers of a CRP, where the orientation and voltage of their readout strips is given. Figure reproduced from reference [80].

3.3 Near Detector

3.3.1 The Need for the Near Detector

The FD observables - the selected reconstructed neutrino energy spectra - are each a combination of the neutrino beam flux, neutrino cross section, detector acceptance, neutrino energy resolution and the sought oscillation probability. The oscillation probability can be extracted by using models to predict the contributions of the other terms. These models are based on theoretical and experimental knowledge, are by no means exact, and so place uncertainties on the reported oscillation parameters. In the era of 'precision measurements' these uncertainties are required to be highly constrained, thus motivating the ND.

The ND is placed in the same beamline, and as close to the beam origin as is possible, such that one can assume an unoscillated flux. One could imagine that a ratio of the FD and ND observation could be used to cancel the model uncertainties but, as ever, it isn't that simple. Complete cancellation would require the near and far detectors to be **identical** and for the flavour and energy composition of the probed flux to be **exactly** the same. This isn't the case:

- The position of the ND in the beamline places limitations on the extent to which it can match the FD design, which results in different detector acceptance and energy resolution abilities.
- The smaller size of the ND means that the energy of the contained neutrino events is lower and, therefore, the probed flux differs.
- The 1300km separation of the near and far detectors allow for neutrino oscillations and spreading of the beam, which alters the neutrino energy and flavour composition of the incident neutrino flux.

One is therefore left with residual uncertainties, which the ND can constrain via the following workflow:

1. The beam, neutrino cross section etc. models are combined such that the observation at the ND can be predicted.
2. The model parameters are tuned by comparing the prediction to the ND data.
3. The tuned model is used to predict the FD unoscillated spectrum (with uncertainties), which is compared to the observed spectrum to obtain the oscillation parameters.

The DUNE ND has three components, shown in figure 3.9: ND-LAr, the Temporary Muon Spectrometer (TMS)/ND-GAr, and the System for on-Axis Neutrino Detection (SAND). ND-LAr and TMS/ND-GAr will have the ability to move transverse to the beam direction in what is referred to as the Precision Reaction-Independent Spectrum Measurement (PRISM) concept. The ND components will now be discussed in brief, the reader is referred to the DUNE ND CDR [81] for more information.

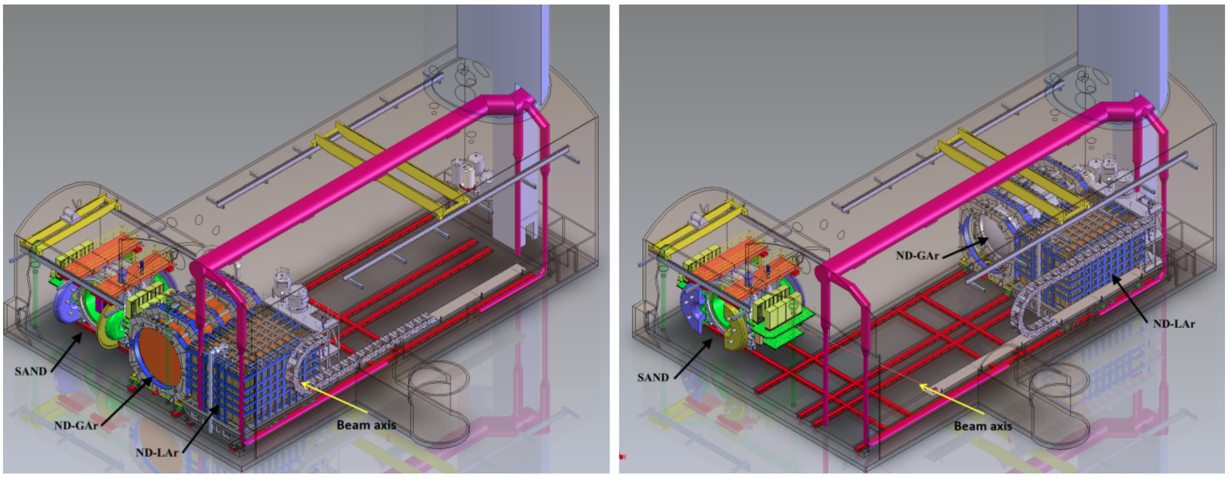


Figure 3.9: A schematic of the phase II ND complex, which consists of ND-LAr, ND-GAr and SAND. The PRISM concept, whereby ND-LAr and ND-GAr have the ability to move off-axis is demonstrated. Figure reproduced from reference [81].

3.3.2 ND-LAr

ND-LAr, sometimes referred to as ArgonCube, is a LArTPC, to ensure commonality of the near and far detector technology. However, the large drift volumes and wire readout of the FD are unsuitable for the ND, where the higher multiplicity of neutrino interactions (~55 per 10μs beam spill) would make reconstruction impossible. ND-LAr is therefore composed of many small LArTPCs, each optically isolated and with a pixel readout to ease reconstruction. The 5m (along beam) × 7m (horizontal) × 3m (height) dimensions and 67t fiducial mass of ND-LAr is sufficient to contain the hadronic system but muons with a momentum exceeding 0.7GeV/c will leave the detector.

3.3.3 TMS/ND-GAr

The momenta of the exiting muons must be measured to ensure a high neutrino energy resolution. This is done by TMS in phase I of the experiment and by ND-GAr in phase II. TMS is a 7m (along beam) × 7m (horizontal) × 3.2m (height) magnetised sampling calorimeter with steel and plastic scintillator layers. The magnetisation of the detector allows one to infer whether the interaction originated from a neutrino or antineutrino. ND-GAr is a central high-pressure gaseous argon TPC with a surrounding electromagnetic calorimeter layer and a magnetic field throughout.

The ND upgrade will help constrain the residual uncertainties. To give an example, GAr has a lower energy threshold for particle reconstruction which results in a more complete 'picture' of the products of neutrino interactions and a better understanding of their mechanics.

3.3.4 PRISM

In the PRISM concept, ND-LAr and TMS/ND-GAr have the ability to move up to 3.2° off-axis. As one does so, the flux becomes closer to monoenergetic, and the peak energy and beam intensity decrease, as shown in figure 3.10. This provides additional data sets to which our beam and neutrino interaction (etc.) models can be fitted, which helps to constrain their uncertainties and biases.

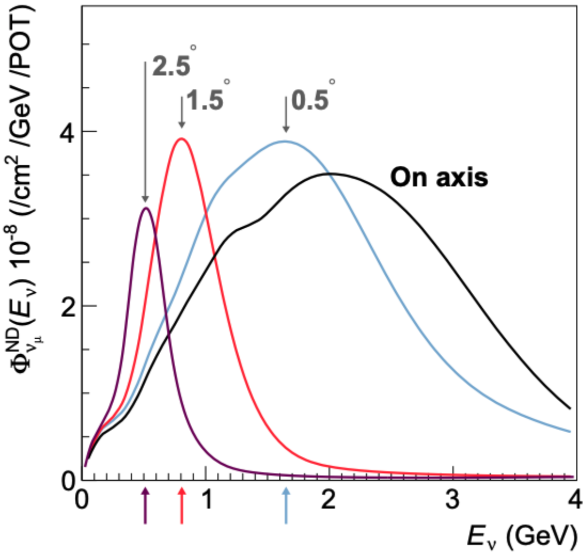


Figure 3.10: The energy composition of the predicted muon neutrino flux at the ND in the on-axis position and at several off-axis positions, where the peak energies are identified by arrows on the x-axis. Figure reproduced from reference [81].

Each data set corresponds to a different energy composition of the flux, and they can be linearly combined (with positive and negative coefficients) to create a target energy composition. This principle can be used to make neutrino energy estimations based on data rather than simulation, and therefore reduces the dependence on the neutrino interaction model. In the data-driven method, one creates a target energy composition with a narrow Gaussian centred on a given true energy, and uses the corresponding linear coefficients to build a sample of reconstructed neutrino interactions with which to determine the mapping between true and reconstructed energies.

A complementary PRISM oscillation analysis determines the oscillation parameters by using the linear combination approach to create the oscillated flux probed by the FD. This method still relies on the beam model but the dependence on the lesser known interaction model is reduced.

3.3.5 SAND

The success of the PRISM concept relies heavily on one's faith in their ability to accurately predict the flux energy composition. This motivates SAND, which monitors the stability of the beam, allowing any deviations to be identified and the causes, if possible, fixed. The SAND detector has an inner tracker core, a surrounding electromagnetic calorimeter layer and a magnetic field throughout. The detector is therefore able to reconstruct neutrino interactions such that the impact of the beam change on the neutrino energy spectrum can be modelled and the models applied to data if the root cause cannot be fixed.

3.4 The Beam

The LBNF beamline is detailed in figure 3.11 and described in detail in the LBNF/DUNE CDR [82]. It will operate as follows:

1. A primary beam of $60 - 120$GeV protons is obtained from the Main Injector accelerator at the Fermilab site, and aimed towards the target hall.

2. The protons collide with a 2.2m (length) × 16mm (diameter) cylindrical, graphite target to produce a secondary beam of pions and kaons.

3. The secondary beam is focused by magnetic horns, which if in Forward Horn Current (FHC) mode, will focus(lose) the positively(negatively) charged particles. The opposite is true if the current of the horns is reversed such that they run in Reverse Horn Current (RHC).

4. The FHC(RHC) secondary beam enters a 194m long decay pipe. The final state leptonic systems of the pion and kaon dominant decay modes consist of $\mu^+\nu_\mu(\mu^-\bar{\nu}_\mu)$ pairs.

5. If not removed, the $\mu^+(\mu^-)$ will decay and contaminate the beam with intrinsic $\nu_e(\bar{\nu}_e)$ and 'wrong sign' $\bar{\nu}_\mu(\nu_\mu)$ components. A hadron absorber is placed at the end of the decay pipe and removes any undecayed hadrons and $\mu^\pm$ from the beam.

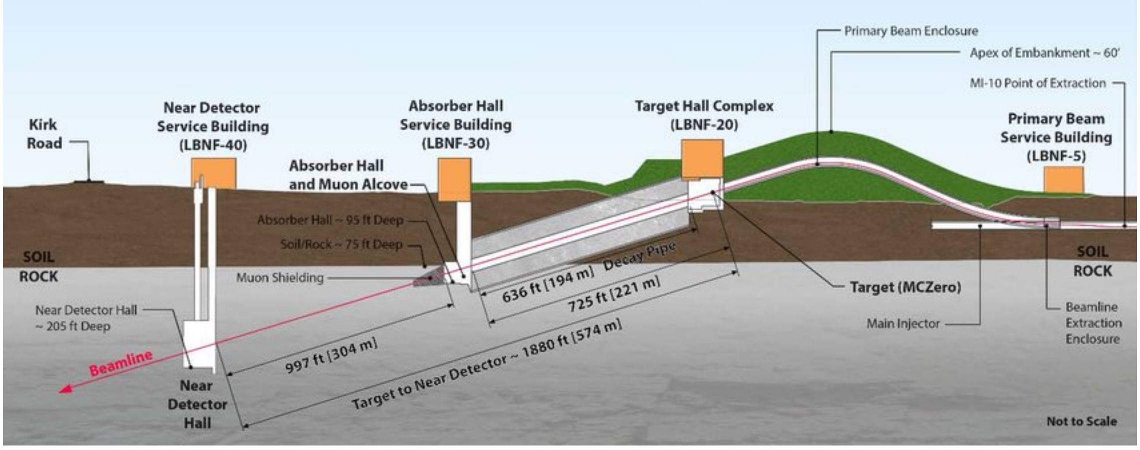


Figure 3.11: A diagram of the LBNF beamline. Figure reproduced from reference [82].

The predicted far detector FHC and RHC flux is shown in figure 3.12, where the intrinsic $\overset{(-)}{\nu}_e$ and 'wrong sign' components are seen to be fairly significant. These arise from the decay products of wrong sign mesons, neutral kaons and $\mu^\pm$ which decay before they reach the hadron absorber.

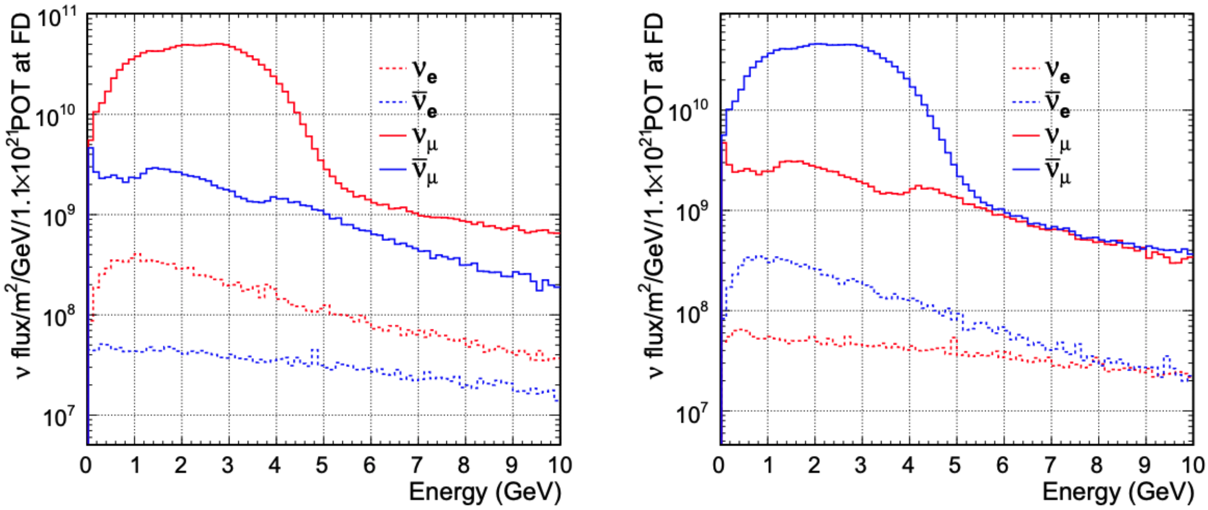


Figure 3.12: The energy and flavour composition of the predicted flux at the FD in FHC mode (left) and RHC mode(right). Figure reproduced from reference [53].

3.5 The Prototypes

Before the DUNE programme, the largest LArTPCs ever built were those of ICARUS T600. The ICARUS T600 detector had a total mass of 600t and was composed of two half-modules. Each was vertically split by a central cathode and had a maximum drift distance of 1.5m [74]. Now we can compare this to the proposed 17kt mass and 3.5m maximum drift distance of the FD modules, or the sophisticated modular structure of ND-LAr, and we see just how uncharted the waters that DUNE is entering are.

An intense prototype programme is therefore underway allowing the proposed detector designs, assembly procedures, operation and reconstruction software to be tested and developed to ensure DUNE's physics goals are met. At the time of writing, the following prototypes have been realised:

- **ProtoDUNE-SP:** A prototype of the HD DUNE FD modules housed at the CERN Neutrino Platform. It had an active volume of 7.2m × 6.1m × 7.0m and took testbeam and cosmic-ray data between 2018 and 2020 [83].
- **ProtoDUNE-DP:** A prototype of the proposed DP module III also housed at the CERN Neutrino Platform. It had an active volume of 6m × 6m × 6m and took cosmic-ray data between 2019 and 2020 [84].
- **ProtoDUNE-SP Vertical Drift:** ProtoDUNE-DP will be re-instrumented to create a prototype of the VD SP FD module. Cosmic-ray and testbeam data will be taken between 2023 and 2024 [79].
- **ArgonCube 2×2:** A prototype of the ND-LAr detector. In 2021, a modular (2 × 2) LArTPC was piloted with cosmic-rays at the University of Bern [85]. The 0.6m × 0.6m × 1.2m module has been moved to Fermilab where it will be joined by three more modules and tested in the NuMI neutrino beam [86].

As its operation is of relevance to the studies of this thesis, the ProtoDUNE-SP detector will now be discussed further.

The ProtoDUNE-SP detector module - shown in figure 3.13 - is vertically split by a central cathode into two chambers, each with a maximum drift distance of 3.6m [83]. Anodes bookend the detector, each containing three APAs that match the scale and design proposed for the DUNE FD HD modules (see subsection 3.2.2). Electron diverters sit in the gaps between APAs highlighted in figure 3.13. Their purpose is to modify the surrounding electric field so as to direct incoming ionisation electrons away from the inactive gap and onto the readout wires. Unfortunately the electron diverters malfunctioned as they were switched on, this is relevant for the study of section 9.2.

Practicalities result in some key differences between ProtoDUNE-SP and the DUNE FD modules. Firstly, ProtoDUNE-SP is a surface based detector and so is exposed to a

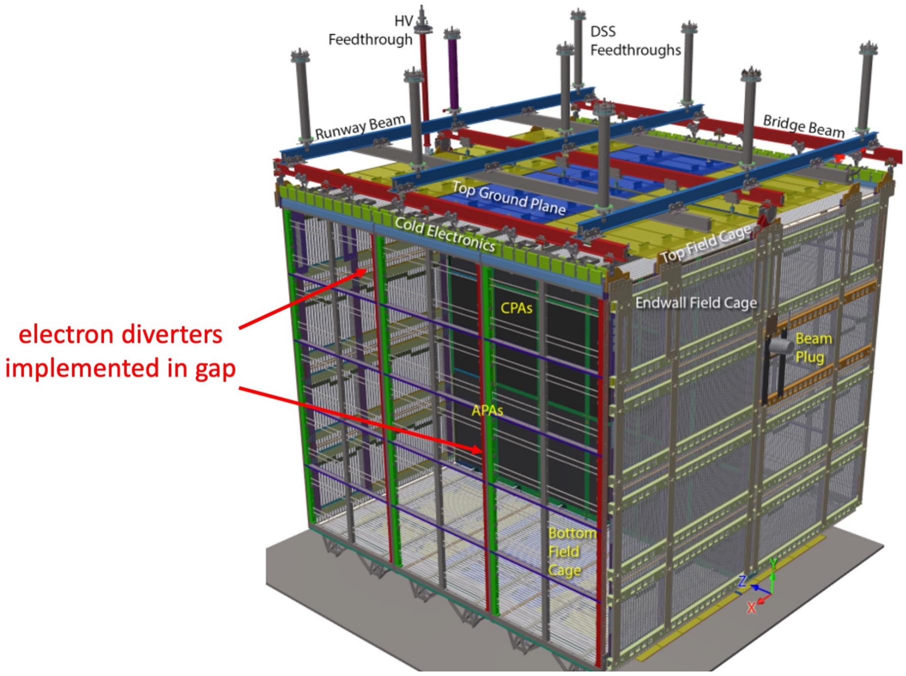


Figure 3.13: The ProtoDUNE-SP detector module, where the beam plug demonstrates the entry point of the testbeam and the APA gaps (which are instrumented with electron diverters) are highlighted. Figure adapted from reference [83].

significant cosmic-ray flux. The large number of ionised argon ions results in the build up of positive charge within the detector, which distorts the electric field and, consequently, the measured charge and inferred trajectory of reconstructed particles. This effect is referred to as the Space Charge Effect (SCE) and must be corrected in ProtoDUNE-SP studies. Furthermore, ProtoDUNE-SP is serviced by the H4-VLE (very low energy) testbeam, which is not a neutrino beam.

The H4-VLE beam is created by impinging the high-energy protons of the CERN Super Proton Synchroton (SPS) on a beryllium target, and the resulting products on a secondary copper, or tungsten, target. The beam has $0.3-7$GeV/c momentum range and is composed of $\pi^{\pm}, \mu^{\pm}, e^{\pm}, K, p$ where, electrons/positrons dominate at low momentum values and pions at high momenta [87]. During the ProtoDUNE-SP runtime, the beam operation was such that the beam only contained positively charged particles. The beam was transported to ProtoDUNE-SP via the H4-VLE beamline - shown in figure 3.14 - which is instrumented with:

- **Profile monitors (XBPMs):** track the position of the beam particles allowing the momentum to be measured.

- **Trigger counters (XBTFs):** provide the beam trigger and measure the time of flight of the beam particles.

- **Cherenkov detectors (XCETs):** low and high pressure XCETs report Cherenkov signals.

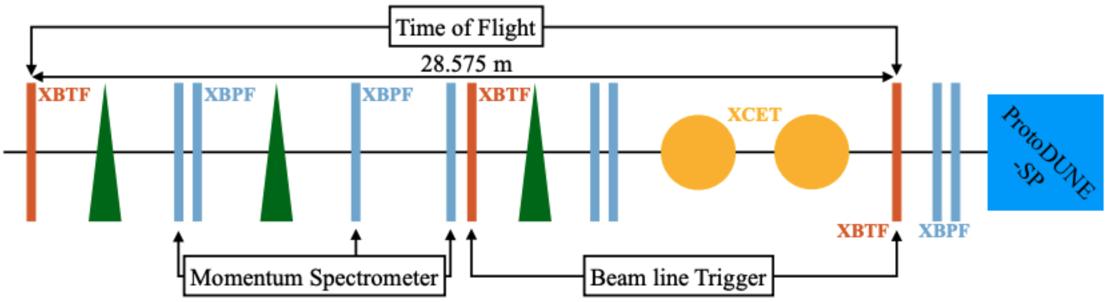


Figure 3.14: A schematic of the H4-VLE beamline showing the positions of the trigger counters (XBTFs), bending magnets (triangles), profile monitors (XBPMs) and Cherenkov detectors (XCETs). Figure reproduced from reference [83].

The information obtained from the beamline instrumentation gives ProtoDUNE-SP excellent Particle Identification (PID) capabilities, allowing one to examine the response of the detector to e.g. a high purity positron sample.

ProtoDUNE-SP has been a demonstrable success, achieving an electron lifetime, photon time resolution, photon yield and signal-to-noise ratio that exceed the requirements of the DUNE FD HD modules. The H4-VLE beam has allowed the detector response to the charged particles seen in the neutrino interactions of DUNE to be understood and the reconstruction software to be developed with respect to higher multiplicity events. Furthermore, several cross sections involving argon nuclei have been determined, which will be useful to DUNE by e.g. improving our understanding of the subsequent interactions of the particles emerging from neutrino interactions.

Chapter 4

Event Reconstruction

The wealth of information captured by the high spatial and calorimetric resolution of LArTPCs is very exciting but requires a highly sophisticated event reconstruction chain to exploit. This chapter details the reconstruction of events in the DUNE FD and the ProtoDUNE-SP detector to the extent required for an understanding of the studies introduced in chapters 6 and 9 respectively.

The reconstruction chain is implemented within LArSoft [88, 89]: a framework in which event generation, reconstruction and analysis can be performed. LArSoft uses an event data model to parameterise events and, this forms the core of the LArSoft workflow in which each stage uses the stored information to further characterise the event and add new information. Most LArTPC experiments adopt LArSoft and construct a reconstruction chain tailored to their detectors. As the ProtoDUNE-SP detector is a prototype of the DUNE FD, their reconstruction chains are very similar but there are several notable differences that will be discussed.

4.1 Monte-Carlo Event Generation

Before the DUNE experiment is operational, we rely on Monte-Carlo (MC) simulations to generate the interactions that will occur in the DUNE FD and to simulate the electronic readout. MC studies are very important as they enable the capabilities of DUNE to be predicted such that its detectors can be designed, the reconstruction chain optimised and the analyses developed to the quality needed to reach the physics goals of DUNE. The key stages of the MC simulation will now be discussed.

4.1.1 Event Generation

First, the primary particles of the event are generated. The main type of generated events seen in this thesis are:

- **Particle Gun Events (section 5.1):** The user designs an event by specifying the particle types, momenta and vertices. Particle gun events are hypothetical events used for specific reconstruction investigations.

- **Beam Neutrino Interactions (section 6.1):** The Generates Events for Neutrino Interaction Experiments (GENIE) event generator [90] takes the G4LBNF simulated beam flux input and applies nuclear interaction and cross section models to simulate neutrino-nucleus interactions in the detector.

- **Cosmic-Ray Muon Events (section 5.2):** The Muon Simulation Code (MUSIC) simulation is used to propagate cosmic-ray muons through the matter of the Earth to the DUNE FD, where the Muon Simulations Underground (MUSUN) simulation is used to generate the muons seen in the detector [91].

4.1.2 Event Evolution

GEANT4 [92, 93, 94] is used to simulate the evolution of the event. It transports the primary particles through the detector medium and calculates the energy deposition at each step, whilst simulating any decays or interactions that may occur. Following this, the detector simulation converts the deposited energy into scintillation photons and ionisation electrons and the latter are propagated to the readout planes. The effects of recombination, transverse and longitudinal diffusion, impurities and electric field distortions during their journey are simulated.

4.1.3 Detector Response

The Wire-Cell Toolkit [95] is used to simulate the Analogue-to-Digital Converter (ADC) waveform output from the wire readout in response to transported ionisation electrons. This process is described in detail in reference [96]. An example of a wire waveform is shown in figure 4.1, where it is seen that the U and V wires measure bipolar waveforms whilst the W wires measure unipolar waveforms. This is because opposite signed currents are induced when the ionisation electrons move towards, and then away from a wire plane, as is the case for the U and V wires but not for the W wire. In addition to the ionisation electron signal, electronic noise is also simulated.

The response of the photon detection system will not be discussed as it is not relevant to the work of this thesis.

4.2 Hit Finding

With the detector response in hand, whether obtained from simulation or reality, the reconstruction chain begins. The first stage is to convert the ADC waveforms to 2D

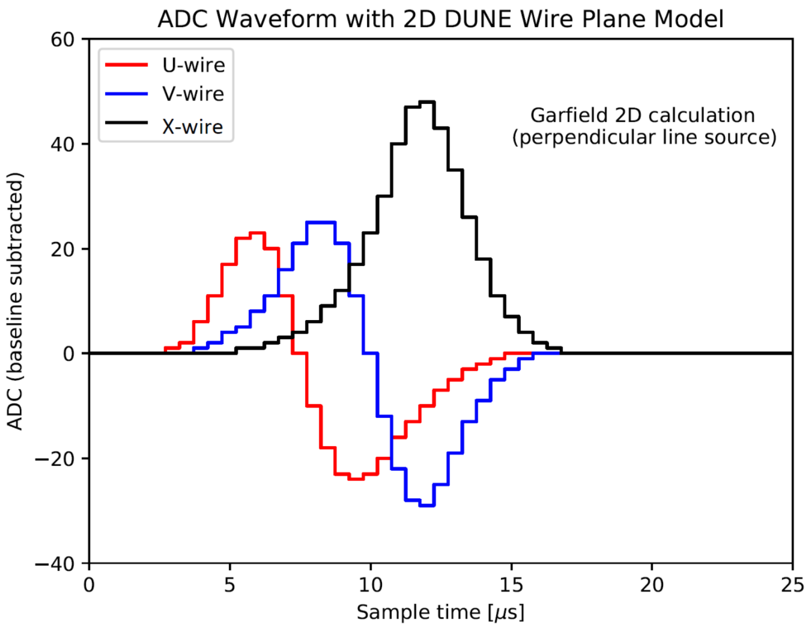


Figure 4.1: Shown for each wire plane is the generated waveform of a minimum ionising particle that travels parallel to the wire plane but perpendicular to the wire orientation in the DUNE FD, where plane X is equivalent to plane W. Figure reproduced from reference [53].

‘hits’, the building blocks used to construct the particles of the event. The key stages of this process will now be discussed.

4.2.1 Signal Processing

The waveform of a group of arriving ionisation electrons is a convolution of the:

1. Induced current from the signal ionisation electrons.
2. Electronics response i.e. signal amplification.
3. Induced current from any other ionisation charge in the wire’s vicinity.
4. Electronic noise.

The aim of the deconvolution stage is to extract the contribution of the ‘signal’ electron distribution and to convert this to the number of arrival ionisation electrons. In a 1D deconvolution algorithm, the waveform $M(t')$ measured on a wire i at time t' is given by

$$M_i(t') = \int_{-\infty}^{+\infty} R(t,t')S(t)dt, \tag{4.1}$$

where $R(t,t')$ is the response function, which encapsulates the field and electronic response to the signal $S(t)$ at time t'. The response function is often coupled with a filter function (e.g. the Wiener filter) which suppresses high-frequency noise. The DUNE and ProtoDUNE reconstruction chains use a 2D deconvolution algorithm which was developed by the MicroBooNE collaboration [96]. A 2D deconvolution algorithm considers the contribution from ionisation electrons that are collected by neighbouring wires. In this regime, equation 4.1 becomes

$$M_i(t') = \int_{-\infty}^{+\infty} (... + R_{i-1}(t'-t)S_{i-1}(t) + R_i(t'-t)S_i(t) + R_{i+1}(t'-t)S_{i+1}(t) + ...)dt, \quad (4.2)$$

where the contribution from the ionisation electrons of wire i and its neighbours are considered (one can consider as many neighbours as they wish). An example of the result of a 1D and 2D convolution algorithm is shown in figure 4.2, where the superior performance of a 2D, over a 1D, convolution algorithm is demonstrated.

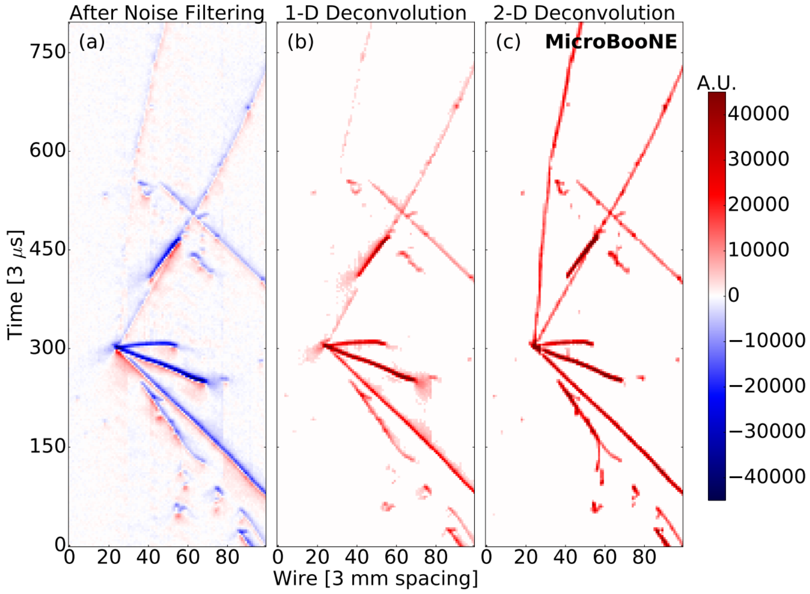


Figure 4.2: The output of the signal processing of a neutrino candidate data event at MicroBooNE with a) only noise filtering b) noise filtering and a 1D convolution algorithm c) noise filtering and a 2D convolution algorithm. Figure reproduced from reference [96].

4.2.2 Gaussian Hit Finding

The Gaussian hit finder module is next applied to the deconvolved waveforms. It identifies the pulses within an input waveform and fits to each a single, or multiple, Gaussian hypothesis obtaining a hit from each fitted Gaussian curve [97] as demonstrated in figure 4.3.

A hit is a 2D object defined by a central position and an extent in the drift direction, where the drift-coordinate is derived from the centre of the fitted Gaussian curve, the 'wire coordinate' is given by the position of the wire to which the waveform belongs, and

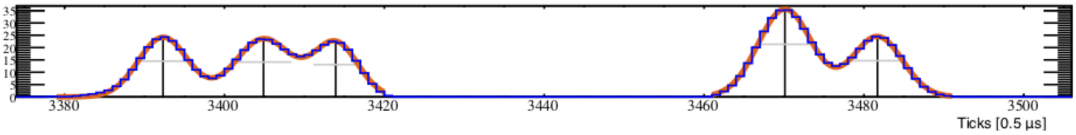


Figure 4.3: The hits created from a deconvolved waveform (blue) in ProtoDUNE-SP data where, the fitted Gaussian curves are shown in red, the hit extent is shown by horizontal grey lines and the drift-coordinate of the central position is indicated by black vertical lines. Figure reproduced from reference [53].

the hit width corresponds to the width of the fitted Gaussian curve. An example of the hits obtained in a single readout window at ProtoDUNE-SP is shown in figure 4.4.

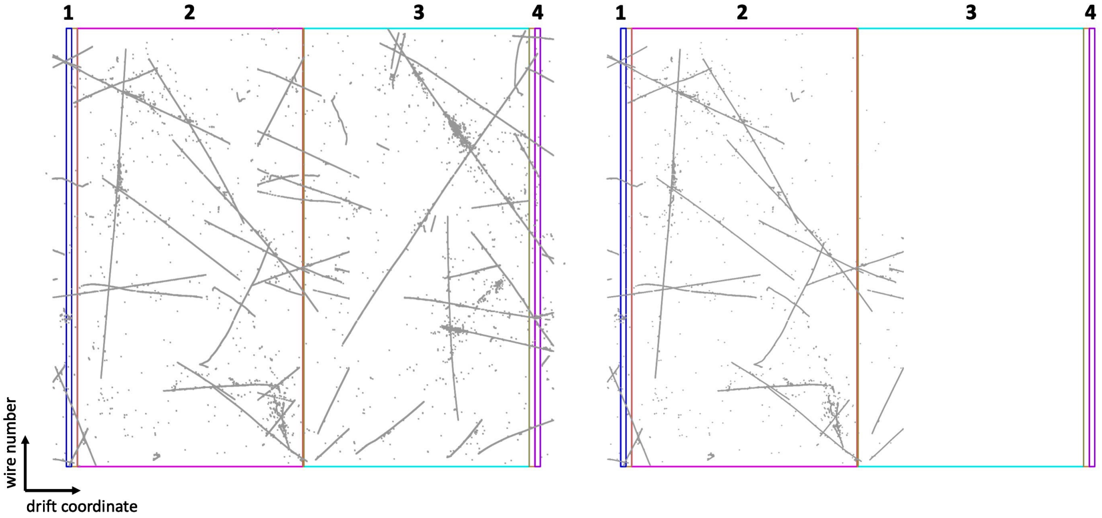


Figure 4.4: Shown in grey are the W view hits of all four drift volumes (left) and of the second drift volume (right) for a simulated event at ProtoDUNE-SP. Some hits in this image appear to lie outside the detector volume, the reason for this is explained in subsection 4.3.2.

4.3 Pandora Pattern-Recognition Software

The event reconstruction software is now applied to construct the particle hierarchies of the event i.e. so that we can say 'a track-like and shower-like particle originate from *this* interaction vertex, are responsible for *these* energy deposits, followed *these* trajectories and, at *this* interaction point the track-like particle decays to *these* children etc.'. Several event reconstruction software have been developed for the detectors of DUNE, but here only the Pandora pattern-recognition software will be discussed.

Pandora takes a multi-algorithm approach to event reconstruction, in which a chain of algorithms gradually builds and connects the particles in the event. Each algorithm performs a single, well-defined task e.g. the *CrossGapsAssociation* algorithm, which connects particles that traverse dead wire regions in the detector. Algorithms sometimes exploit detector features, the commonality of the drift-coordinate between the readout views of LArTPCs for example, but most algorithms are generally designed to be detector independent. This results in a flexible pattern-recognition software which is currently utilised across a wealth of LArTPC experiments and will be, and has been, applied to other detector technologies.

4.3.1 A Multi-Algorithm Approach

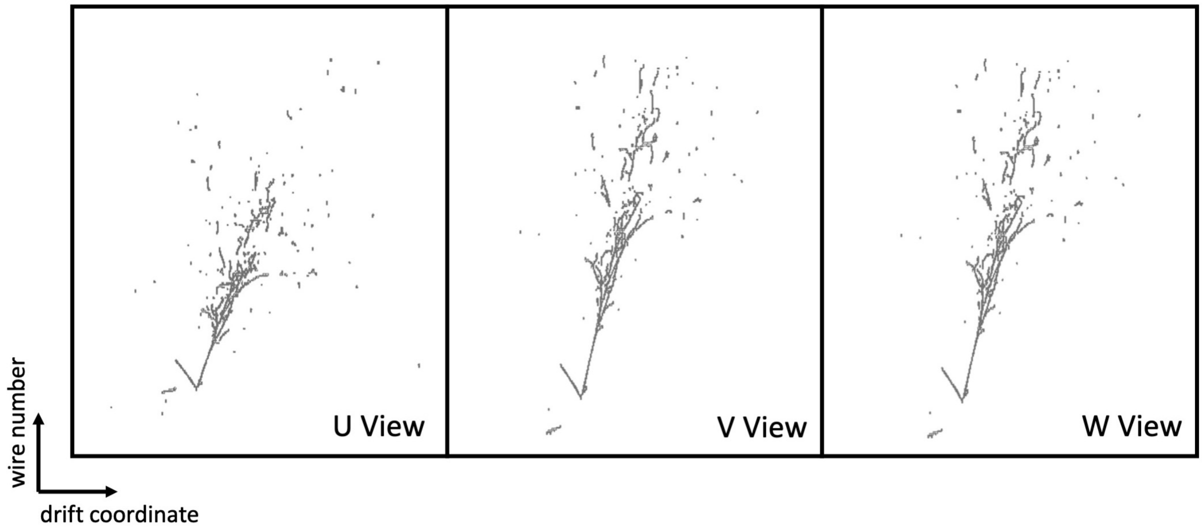


Figure 4.5: The 2D hits of the U, V and W view for a simulated CC quasi-elastic ν_e event in the DUNE FD. The primary track corresponds to a proton, and the shower an electron. To give the reader a sense of scale, the electron shower has a length of approximately 1.8m.

To understand the key stages of the Pandora reconstruction, let's consider figure 4.5, which shows the input 2D hits of a simple Charge-Current (CC) ν_e quasi-elastic interaction in the DUNE FD. In this interaction, a primary electron shower and proton track are seen to be produced. To reconstruct this event, the neutrino-interaction optimised reconstruction chain, *PandoraNu* [98], is applied:

1. **2D Clustering:**
    - In each view, a cautious clustering algorithm uses proximity based logic to group hits into seed clusters, which are then grown by subsequent algorithms based on their topological associations. This is demonstrated in figure 4.6.
    - The 2D clustering algorithms are track-orientated and so we expect track-like clusters to be well reconstructed but shower-like clusters to be split into 'shower spines' and 'shower branches'.

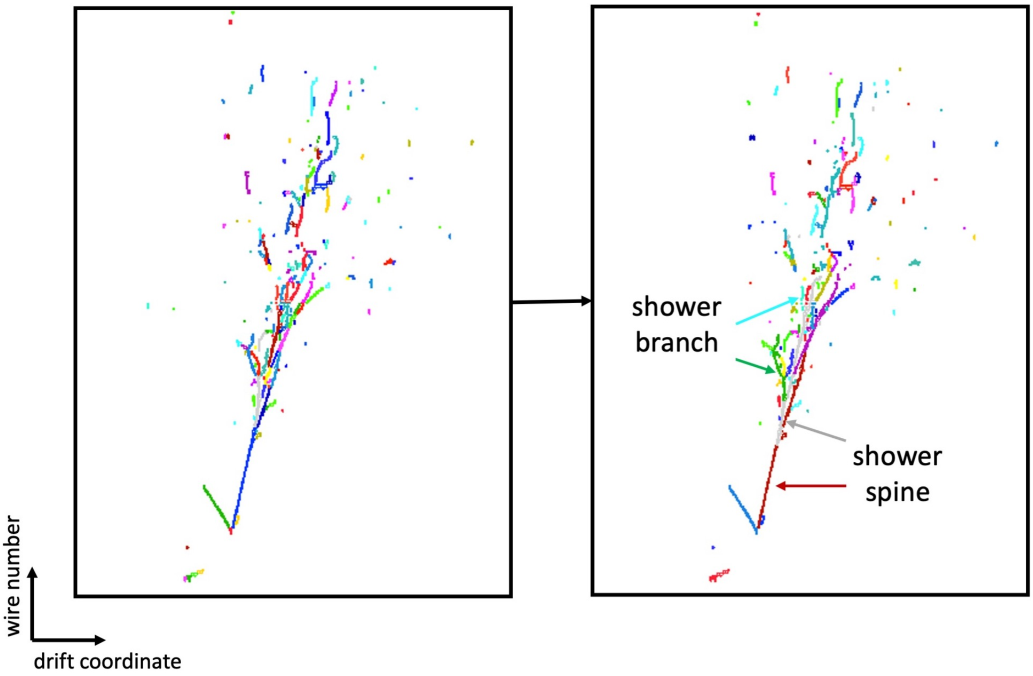


Figure 4.6: The formed clusters of the CC ν_e example event after the initial clustering algorithm (left) and 2D clustering stage (right) in the W view. Each colour denotes an individual cluster, some colours may be repeated.

2. **3D Interaction Vertex Reconstruction:**

    - The three views of 2D clusters are considered simultaneously and the commonality of the drift-coordinate is used to create candidate 3D vertices from places where a particle looks to originate in at least one view.
    - A combination of BDTs is used to identify the most likely initial vertex candidate, demonstrated in figure 4.7.

    The interaction presented in figure 4.7 is a CC quasi-elastic ν_e event, and consequently the vertexing algorithm is able to utilise the combined pointing information of the electron shower and proton track. In the antineutrino version of this interaction, a positron is produced alongside a neutron. As neutrons cannot be directly seen in LArTPCs, less pointing information is accesible to the vertexing algorithm, and this can result in the neutrino vertex being incorrectly placed at the position at which the shower cascade looks to begin, thus splitting the electron shower.

3. **Track Reconstruction:**

    - The commonality of the drift-coordinate is used to identify, across the three views, the track-like clusters that belong to the same particle. This is illustrated in figure 4.8.
    - The combined 3D information of the matched clusters is used to refine the

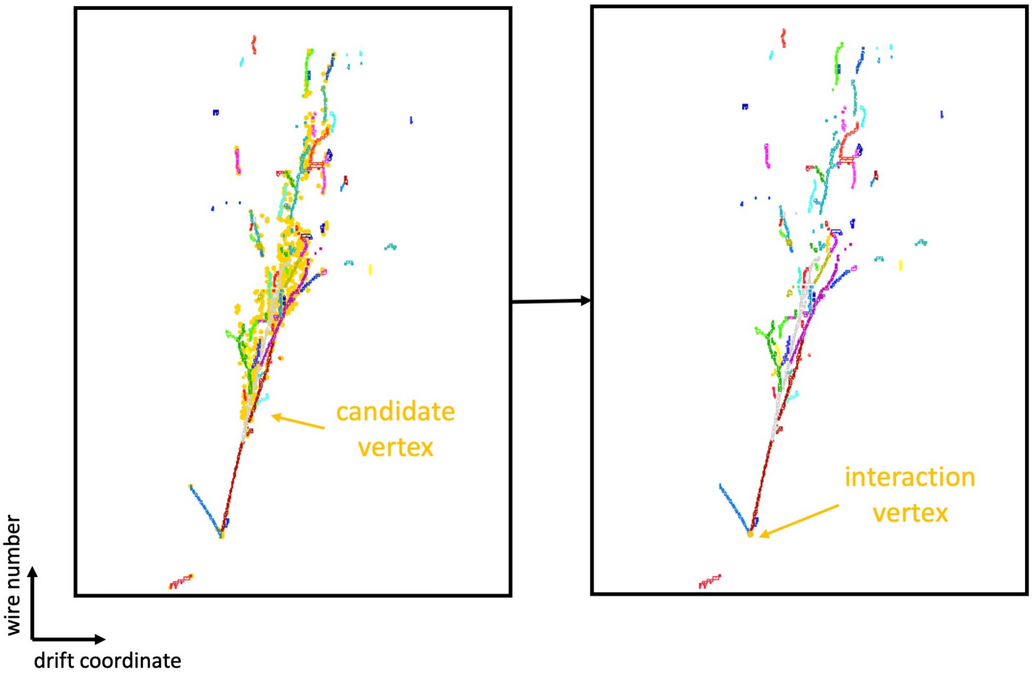


Figure 4.7: The W view projection of the 3D candidate vertices (left) and identified 3D interaction vertex (right) of the CC ν_e example event. Each colour denotes an individual cluster, some colours may be repeated.

clusters, merging and/or splitting them if necessary. This idea is revisited in more detail in subsection 5.3.2.

- Each triplet of matched clusters is used to create - what is referred to as - a Particle Flow Object (pfo), which represents the reconstructed particle.
- Any pfos created at this stage are assumed to be track-like, it is seen how this can sometimes include shower spines.

4. **Shower Reconstruction:**
   - A BDT is used to classify created pfos as track or shower-like and any shower-like pfos, created from shower spines, are disbanded so that their clusters can be properly grown.
   - Track-like pfos are protected and the *ShowerGrowing* algorithm is used to grow shower clusters in 2D by merging in shower branches to the shower spines in an iterative approach.
   - The grown shower clusters are matched across the three views and, from each triplet, a shower-like pfo is created.

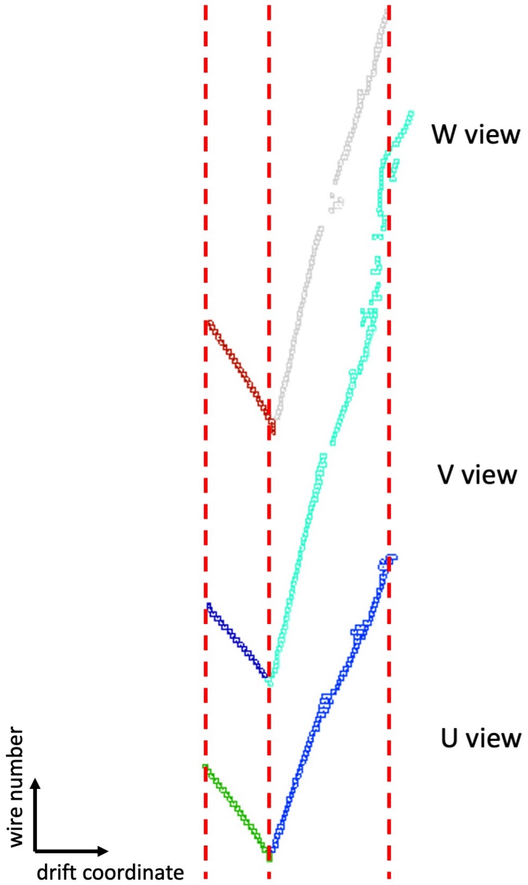


Figure 4.8: The matched clusters of the created track-like pfos, where the dashed red lines demonstrate the overlap of the clusters in the drift-coordinate. The right-most pfo has been created from shower spine clusters.

5. **3D Hit Creation:**

    - 3D hits are created by several algorithm tools, each designed to target a specific topology e.g. longitudinal tracks.
    - The tools have the same workflow: each pfo is examined and the commonality of the drift-coordinate is used to find, for each 2D hit, the corresponding position(s) on the cluster in each 'other view'. An analytic χ^2 minimisation is used to identify the 'best' 3D position from all possible combinations, creating a 3D hit.
    - The output is demonstrated in figure 4.9.

6. **Particle Hierarchy Creation:**

    - A neutrino pfo is created and is assigned the interaction vertex.
    - 3D topological and pointing information is used to find the children, grandchildren, great-grandchildren etc. of the neutrino hierarchy.

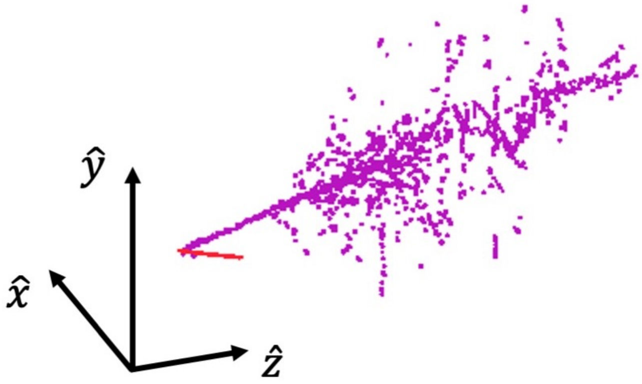


Figure 4.9: The 3D hits of the track-like (red) and shower-like (violet) pfos of the CC ν_e event.

  - Each pfo (except for the neutrino) is classified as track or shower-like and its 3D vertex is set equal to the closest position of the pfo to its parent.

The test-beam interactions of ProtoDUNE are reconstructed using the test-beam optimised reconstruction chain, *PandoraTestBeam*. *PandoraTestBeam* strongly resembles *PandoraNu* but has several modifications [99]:

- The vertex finding algorithms target the interaction vertex of the test-beam particle i.e. the position at which it interacts with the liquid-argon.
- The particle hierarchy creation is modified such that the primary pfo which is most consistent with the beamline is promoted to the tier occupied by the neutrino.

4.3.2 Consolidated Reconstruction

The input 2D hit images of the DUNE FD (figure 4.5) are a sharp contrast to those of ProtoDUNE-SP (figure 4.4), as the former detector is situated underground whilst the latter sits on the surface where it is subject to significant cosmic-ray flux. The consolidated reconstruction chain, shown in figure 4.10, has been developed to handle this, alongside complications which arise from having multiple drift volumes. It will now be described.

1. *PandoraCosmic* is a cosmic-ray optimised reconstruction chain, which is more strongly orientated towards finding long, straight tracks, rather than multiple particles that emerge from a vertex. It follows the structure outlined in subsection 4.3.1 but has several differences:

    - The interaction vertexing is performed in the final stages of Pandora, where the cosmic-ray vertex is placed at the 3D position which lies closest to the top of the detector.

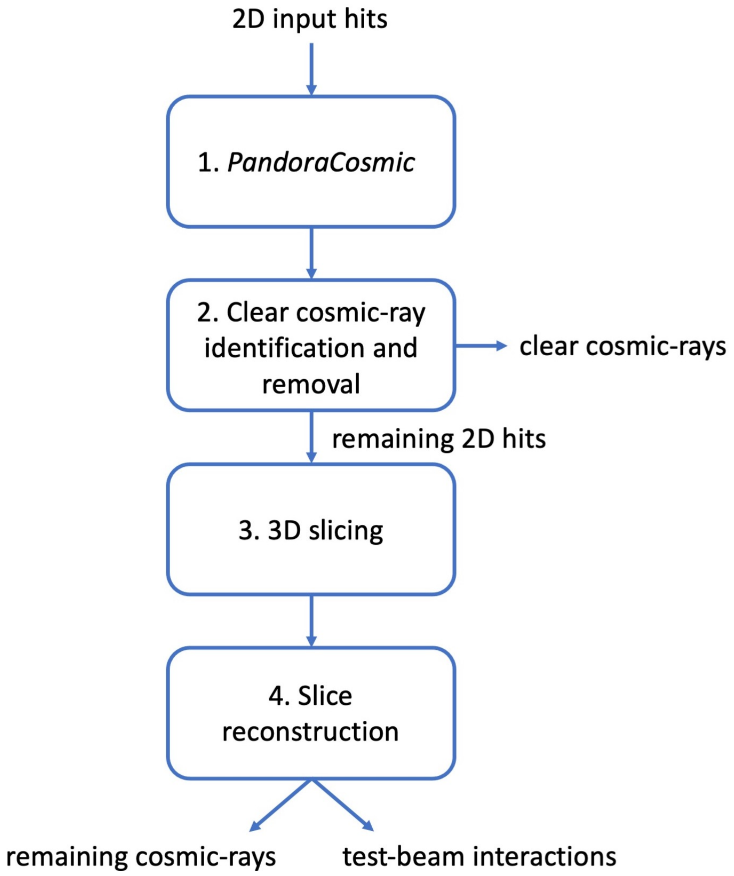


Figure 4.10: The workflow of the Pandora consolidated reconstruction chain.

- The shower reconstruction is tailored to the lower energy, and more ‘blob-like’ structures, of delta-ray and Michel electron showers.
- Each cosmic-ray hierarchy is two tiers deep, where the first(second) tier is occupied by a cosmic-ray(the child delta-rays and Michel electrons).

In the DUNE FD and ProtoDUNE-SP, the beam particle trigger time defines the time from which the ionisation electron drift time is measured. Since cosmic-rays arrive at the detector before or after the beam trigger, their reconstructed hits are displaced in the drift-coordinate from their true trajectory in the resulting spatial representation/image. Some cosmic-ray hits will therefore appear to lie outside the detector volume and, in the case of multi-chamber detectors, can appear to be shifted into different chambers. The hits of each detector volume are therefore reconstructed in isolation to prevent clustering errors.

2. A pfo is identified as a cosmic-ray muon if it satisfies any of the following:
   - The pfo enters and leaves the detector.
   - Part of the pfo has been shifted outside of the detector.
   - The pfo corresponds to a cosmic-ray that is believed to span multiple detector

chambers. An example is given in figure 4.11, in which the cosmic-ray has arrived after the beam trigger. The reported drift time is therefore too long, shifting the hits away from the anode planes. The *StitchingCosmicRayMerging* algorithm considers the combined reconstruction output and searches for pairs of 3D clusters that, if shifted by equal amounts and in opposite directions, align. Found cluster pairs are 'stitched' together and the resulting pfos are tagged as cosmic-rays.

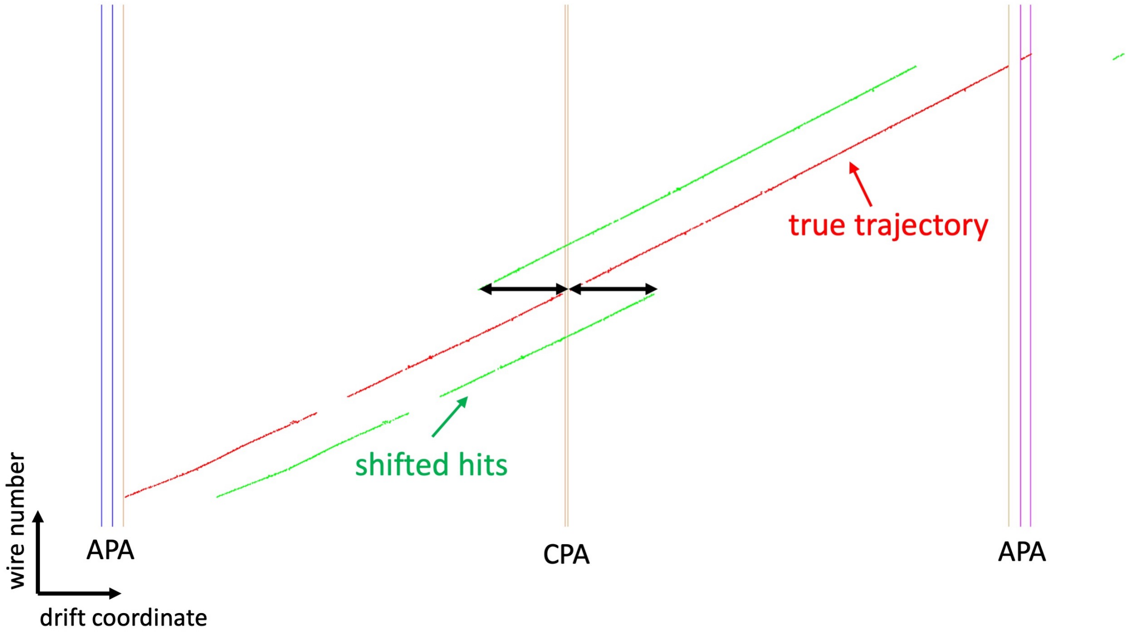


Figure 4.11: A simulated cosmic-ray which arrives at the ProtoDUNE-SP detector after the beam trigger time. The true trajectory is shown in red, the shifted trajectory in green and the acronyms 'CPA' and 'APA' denote the cathode and anode plane assemblies respectively. The arrival time (and thus shifted distance) of the stitched cosmic-rays can be calculated by considering the [-250, 2750]μs readout window of ProtoDUNE and its 2250μs CPA to APA drift time. The CPA(APA)-crossing cosmic-rays correspond to arrival times in the region [-2500, 500]μs ([-250, 2750]μs) [99].

3. The hits of the unidentified cosmic-rays and test-beam interaction remain: cosmic-rays should be reconstructed with *PandoraCosmic* and the test-beam interaction, *PandoraTestBeam*.

   A 'slicing' procedure is applied, which aims to slice up the reconstructed 3D hits such that each slice contains either a cosmic-ray or a test-beam interaction. The slicing is performed on the output of a streamlined reconstruction chain which considers the hits of all detector chambers in order to prevent the splitting of the test-beam interaction into different slices.

4. Each slice is reconstructed under a test-beam hypothesis and a cosmic-ray hypothesis and a BDT is used to chose the appropriate outcome for each.

4.4 Particle Characterisation

At this stage in the reconstruction chain, each reconstructed pfo is converted to either a LArSoft track or LArSoft shower object depending on the Pandora track/shower ID. These objects parameterise the particles in terms of the characteristics that are useful to analyses, e.g. those that aid particle identification and/or event selection. An analysis whose event selection demands a shower can be limited by incorrect Pandora track/shower IDs and so some analysis groups choose to process all pfos under both a track and shower hypothesis.

4.4.1 Track Characterisation

The *PandoraTrackCreation_module* [100] is used to parameterise track-like pfos in terms of trajectory points: each with a 3D position, 3D direction and an associated 2D hit (and therefore charge information).

To do this, a Principle Component Analysis (PCA) is first performed to determine the best fitting coordinate basis in which to fit the 3D hits. This is denoted by $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$, where $\hat{e}_1$ points in the direction that has the largest extent, $\hat{e}_2$ points in the perpendicular direction with the largest extent and $\hat{e}_3$ completes the basis. A combination of two 2D sliding fits in $\{\hat{e}_1, \hat{e}_2\}$ and $\{\hat{e}_1, \hat{e}_3\}$ is used to fit the 3D hits of the pfo, creating from each, a trajectory point.

At each trajectory point a dE/dx value is calculated by the *CalorimetryAlg* [101], which applies electron lifetime and recombination corrections to translate the associated hit charge into an energy and, using the wire pitch, a dE/dx.

4.4.2 Shower Characterisation

In the DUNE FD reconstruction chain, LArSoft shower objects are created by the *PandoraModularShowerCreation_module* [102], previously known as TRACS [103, 104]. In this module, a series of tools build up the parameterisation of shower-like pfos in terms of their:

- **Start position:** set equal to that determined by Pandora in the particle hierarchy creation stage.
- **Start direction:** obtained from a PCA of the shower 3D hits and points in the direction of the largest spatial extent.
- **Length:** the length along the shower axis which contains 90% of the 3D shower hits, with the shower axis defined to originate from the shower start position and to point in the start direction.

- **Opening angle:** the opening angle of a cone defined by the shower axis, length and width, where the latter is the radius of a cylinder that is centred on the shower axis and contains 90% of the 3D shower hits.

- **Total energy:** calculated for each readout plane, with use of the *CalorimetryAlg* to convert the total charge of the corresponding 2D hits into an energy by applying lifetime and recombination corrections.

- **Initial track stub:** a LArSoft track object created from the 3D hits of the track-like region of the shower. These are found by incrementally growing a 3D hit seed until the residual of a PCA fit significantly deviates (shown in the right panel of figure 4.12).

- **Initial dE/dx:** calculated for each readout plane, it is the median value of the dE/dx values of the initial track stub trajectory points. If the initial track stub could not be found, it is calculated from the hits that fall within an initial cylinder. The two methods are illustrated in figure 4.12.

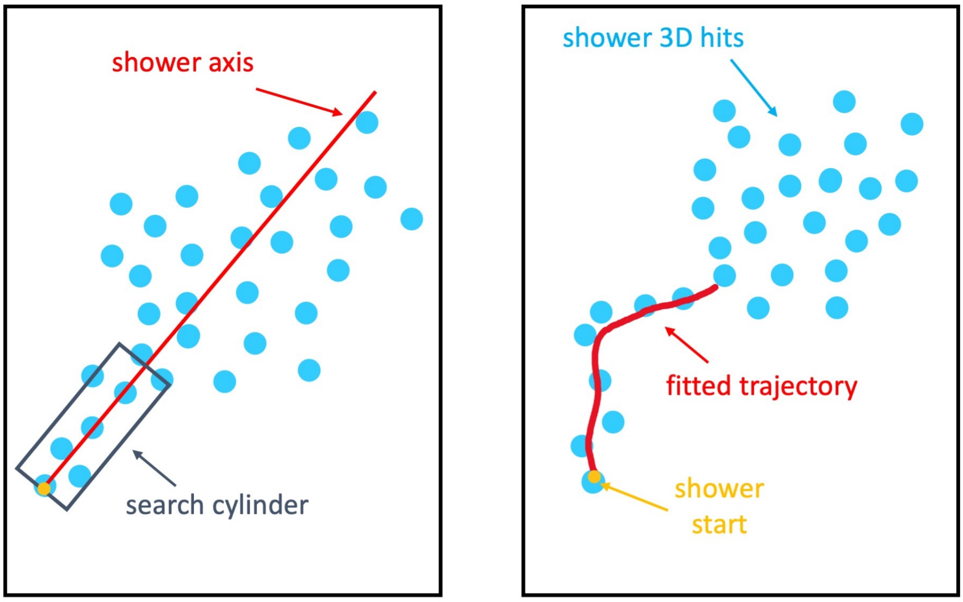


Figure 4.12: A schematic to demonstrate the formation of the initial track-like region of a shower using the crude cylinder search method (left) and the more sophisticated incremental growth method (right). Figure adapted from reference [105].

At ProtoDUNE-SP, the *PandoraCreationShower_module* [106] is used to create the LArSoft shower objects, which are parameterised in terms of only topological variables: the shower start position is taken from the Pandora reconstruction output and the start direction, length and opening angle are obtained from a PCA.

4.5 Neutrino Energy Estimation

Finally, the energy of ν_e and ν_μ interactions in the DUNE FD is estimated. This is achieved by the *NeutrinoEnergyAlgorithm* [107, 108] which calculates the energy of the leptonic and hadronic systems in isolation before combing their values. The *NeutrinoEnergyAlgorithm* allows the user to specify the leptonic system.

4.5.1 Muon System

If the muon track terminates within the detector, the muon momentum is inferred from its linear dependence on the reconstructed track length, with the linear parameters obtained from the MC simulation. If the track leaves the detector, the momentum is estimated by Multiple Coulomb Scattering (MCS) following the method outlined by the ICARUS collaboration in reference [109]. The method works as follows:

1. The track is split into windows of length l.
2. A straight-line fit is performed for the hits in each window.
3. The scattering angle between each pair of connected windows is calculated.
4. A Gaussian is fitted to the obtained scattering angle distribution and the width of the distribution ($\theta^{rms}_{measured}$) is identified.
5. Steps 1-4 are repeated with different window lengths and the equation

$$\theta^{rms}_{measured} = \sqrt{\left(\frac{13.6MeV}{\beta cp} z \sqrt{\frac{l}{X_0}} \left[1 + 0.038 \ln\left(\frac{l}{X_0}\right)\right]\right)^2 + (Cl^{-3/2})^2} \tag{4.3}$$

   is fitted to the results, with the noise proportionality constant (C) and the particle momentum (p) free to vary. βc and z denote the particle's velocity and charge respectively, whilst X_0 represents the radiation length of the material.

Finally, the MCS estimated momentum is calibrated using the linear relationship between reconstructed and true muon momentum obtained from the MC simulation.

4.5.2 Electron System

The energy estimation of electron showers is, in principle, simpler: the charge of the 2D hits of the collection plane is summed and converted to an energy by applying lifetime and recombination corrections. The electron shower energy is calibrated using the linear relationship between true and reconstructed electron energy obtained from the MC simulation.

4.5.3 Hadronic System

The energy estimation of the hadronic system mirrors that of the electron system. The energy resolution of the hadronic system is much poorer than that of the muon or electron system as a result of missing energy i.e. energy lost by methods not 'seen' in the detector such as that carried away by neutrons. The calibration correction somewhat accounts for this but is inaccurate as the correction is not specific to the event.

4.5.4 Performance

The fractional energy residuals of the estimated energy of the ν_μ and ν_e interactions of the FHC sample described in section 6.1 are shown in figure 4.13, where only neutrino interactions that have a reconstructed energy of less than 8GeV are considered. The *NeutrinoEnergyAlgorithm* is seen to estimate the energy of ν_e interactions most accurately, achieving an energy resolution of $\sim$11%, as electron showers are most often contained within the detector. In each case, a negative bias is seen, suggesting that, on average, the neutrino energy is underestimated. The fractional energy residuals are shown for contained and uncontained events in figure 4.14, where the former achieves an energy resolution of $\sim$14% and the latter an energy resolution of $\sim$22%, which demonstrates the poorer accuracy of the MCS approach. Here, the reconstructed energy of (un)contained events is seen to be, on average, underestimated (overestimated).

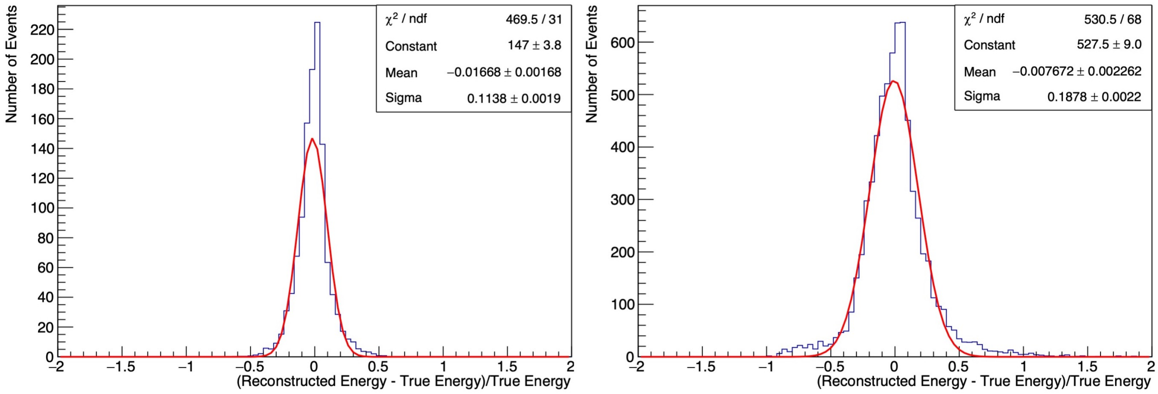


Figure 4.13: The fractional energy residuals (blue) of the ν_e (left) and ν_μ (right) interactions to which Gaussian curves have been fitted and are shown in red.

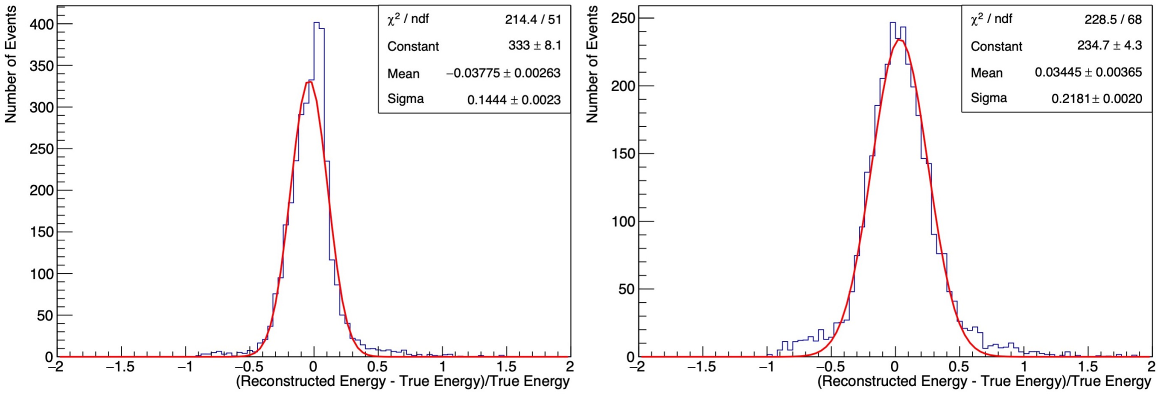


Figure 4.14: The fractional energy residuals (blue) of contained (left) and uncontained (right) ν_μ interactions to which Gaussian curves have been fitted and are shown in red.

Chapter 5

Calibration Motivated Reconstruction

This PhD began with a task to characterise the 2D → 3D matching of tracks in Pandora, which motivated improvements to the reconstruction of those that travel parallel to the wires of the readout planes. This work was of interest to the calibration working group and initiated a project to develop the cosmic and delta ray reconstruction for their calibration needs. The reader should interpret this project as a ‘service task’, a piece of work completed to benefit the wider experiment.

Section 5.1 will detail the initial 2D → 3D study whilst sections 5.2 and 5.3 will present the improvements made to cosmic and delta ray reconstruction respectively.

5.1 2D → 3D Matching Studies

The 2D → 3D track matching procedure is the application of the following algorithms in the noted order:

1. ***ThreeDTransverseTracks:*** matches the clusters of particles that have a significant extent in the drift direction, where it is possible to correlate features in multiple readout planes at a common drift-coordinate.
2. ***ThreeDLongitudinalTracks:*** matches the clusters of particles that are almost parallel to the beam direction i.e. the so-called isochronous case.
3. ***ThreeDTrackFragments:*** matches the clusters of particles in cases where the track has been fragmented in one view.

The study aimed to determine whether there were regions of the parameter space in which the majority of the particle creation was performed by the *ThreeDTrackFragments* algorithm. This is undesired as the *ThreeDTrackFragments* algorithm lacks sophistication, and so can result in the creation of particles that have missing hits and/or high levels of contamination. We began with a sample of $\sim 58,000$ single muon events that were generated within the DUNE FD test volume, a smaller detector volume used to reduce computational stress. The coordinate system of the DUNE FD is defined by: an origin, which lies in the centre of the beam-facing end of the detector; a y-axis, which points vertically upwards; an orthogonal z-axis, which points in the beam direction; and an x-axis, which completes the right-handed basis. The generation specifics are given in table 5.1 where particle directions are parameterised in terms of:

- θ_{xz} : the angle from the z-axis to the projection of the momentum vector onto the $x - z$ plane.
- θ_{yz} : the angle from the z-axis to the momentum vector after it has been rotated to lie above the z-axis.

Quantity	Distribution	Range
Momentum Magnitude	Uniform	[0, 5] GeV
θ_{xz}	Uniform	$[-90°, 90°]$
θ_{yz}	Uniform	$[-90°, 90°]$
Vertex x position	Uniform	[-360, 360] cm
Vertex y position	Uniform	[-600, 600] cm
Vertex z position	Uniform	[0, 1200] cm

Table 5.1: The distributions used to generate the single muon sample for the 2D → 3D track matching study.

5.1.1 Metrics

The metrics used to quantify performance should parameterise one's goal, whether that be a pattern-recognition goal e.g. to reconstruct all particles to a given standard or, a physics goal e.g. to improve the accuracy of a π^0 mass estimate.

We will use pattern-recognition metrics. These first require us to define a realistic estimate of the perfect reconstruction output, which for this study is chosen to be parameterised by the hits we expect the reconstructed muon to own. In general these are identified by:

1. Collecting the hits that belong to the MC particle.
2. Removing the hits that are deemed to be unattainable; for instance that lie too far away to have been reasonably associated to the MC particle by a reconstruction algorithm.

Here, we inherit the procedure used in the Pandora analysis for the MicroBooNE collaboration [98]: the hits for which the dominant contributor belongs to the muon hierarchy are collected, and those that lie downstream of a far-travelling photon or neutron are removed. This allows us to define the metric used for this study:

$$\text{efficiency} = \frac{\text{number of reconstructed MC particles}}{\text{number of reconstructable MC particles}}, \tag{5.1}$$

where a 'reconstructable' MC particle is one that we expect to have been reconstructed: it is required to have at least 5 hits in at least 2 views and at least 15 hits overall. A 'reconstructed' MC particle is one with a matched reconstructed particle, where a match is made if:

- there are at least 5 hits shared between the reconstructed and MC particle,
- the purity of the match exceeds 70%, and
- the completeness of the match exceeds 70%,

where

- the purity is the fraction of hits in the matched reconstructed particle that are shared with the MC particle, and
- the completeness is the fraction of MC hits that are shared with the matched reconstructed particle.

The completeness and purity thresholds are chosen to be higher than that used for MicroBooNE as a response to the simplicity of single-particle events.

5.1.2 Results

Much was learned when the efficiency was observed as a function of θ_{xz} and θ_{yz}, shown in figure 5.1. The *ThreeDTransverseTracks* algorithm is seen to be quite successful and is responsible for 74.1% of the particle creation, where uncertainties are negligible as a result of the number of events in the MC sample. The *ThreeDLongitudinalTracks* algorithm is shown to recover the particles that travel along the beam direction ($\theta_{xz} = 0°$) as intended. Large inefficiencies are however seen in particular angular regions, which align with the beam direction and/or the orientation of the readout wires of one, or several, views. For example, the efficiency after the *ThreeDLongitudinalTracks* algorithm falls to ~20% when $|\theta_{xz}| \approx 90°$ and to ~65% when $|\theta_{yz}| \approx 54°$. The regions with large inefficiencies are detailed in table 5.2.

Particles that travel along the beam direction have little/no extent in the drift direction. This makes it extremely challenging to identify the U, V and W hits that originate

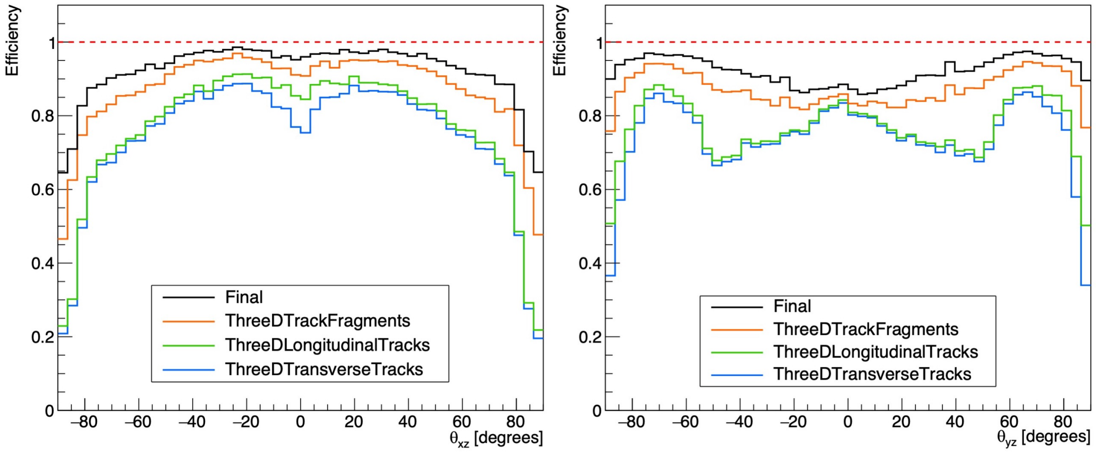


Figure 5.1: The reconstruction efficiency of the single muon events as a function of θ_{xz} (left) and θ_{yz} (right) after the successive application of the *ThreeDTransverseTracks* (blue), *ThreeDLongitudinalTracks* (green), *ThreeDTrackFragments* (orange) algorithms, and at the end of the full Pandora reconstruction chain (black).

Angular Region	Cause of Inefficiency
$\theta_{xz} = 0°$	Parallel to beam direction
$\theta_{xz} = \pm 90°$	Parallel to U, V and W wires
$\theta_{yz} =\sim \pm 54°$	Parallel to U, V wires
$\theta_{yz} = \pm 90°$	Parallel to beam direction and W wires

Table 5.2: The cause of the inefficiencies in the angular regions seen in figure 5.1.

from the same ionisation charge, and therefore to reconstruct the particle. This is the isochronous case that was mentioned earlier, and is a well-known feature of the LArTPCs modelled here.

When tracks travel parallel to the readout wires, their ionisation charge is collected by the same readout wires, resulting in hits with large widths. Examples of such hits are shown in figures 5.2 and 5.3. The Pandora 2D clustering algorithms only use a hit's central position in the hit merging/cluster formation logic and, since the hit centres lie far away from each other, parallel tracks are often poorly clustered (as seen in figure 5.2). Figure 5.1 demonstrates that the *ThreeDTrackFragments* algorithm will create a particle in these situations but it will likely be of a low completeness. These findings motivate the creation of a new 2D clustering algorithm, the *HitWidthClusterMerging* algorithm.

5.1.3 The Hit Width Cluster Merging Algorithm

The aim of the *HitWidthClusterMerging* algorithm is to identify pairs of wide-hit clusters that look like they should be merged together. These are referred to as cluster associations. The longest pathway of associated clusters is then found, using functionality that is inherited from the *ClusterAssociation* Pandora algorithm, and the chain of associated

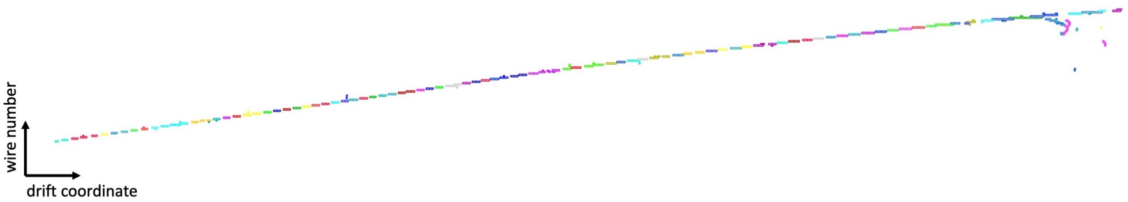


Figure 5.2: The reconstruction of a track that travels parallel to the orientation of the readout wires of one, or several, views. Each colour denotes an individual cluster, some colours may be repeated.

clusters are ultimately merged together. A simplified workflow of the *HitWidthClusterMerging* algorithm is:

1. Identify the clusters between which to search for associations; these are clusters that have a high proportion of wide hits.
2. Break up the hits of the wide-hit clusters, allowing the cluster to be parameterised in terms of its:
    - **Extremal points:** these are the central positions of the constituent hits with the highest and lowest drift or beam coordinate (depending on which has the largest extent).
    - **Cluster direction:** this is the result of a linear fit to the constituent hits (using the least-squares method).

   This is illustrated in figure 5.3.

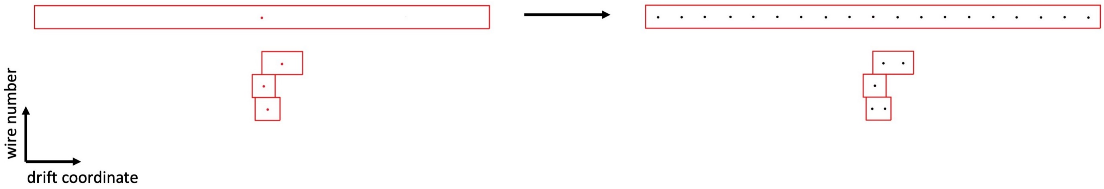


Figure 5.3: An example of the treatment of a wide-hit cluster. The cluster is shown before (left) and after (right) it is broken into its constituent hits, where the dots indicate the hit positions considered by the clustering algorithm.

3. Walk along the wide-hit clusters in order of the drift-coordinate of their extremal points, starting from the lowest. Cluster associations between the current cluster and those in the forwards direction are made if:
    - The distance between the highest drift-coordinate extremal point of the current cluster and the lowest drift-coordinate extremal point of the forwards cluster is:
        - less than 5cm in the drift-coordinate, and

    - less than 2cm in the beam-coordinate

  If the clusters overlap in the drift-coordinate, the assessment is instead made using the position of the constituent hit of the forwards cluster that lies closest to the current cluster's highest drift-coordinate extremal point.

- The cosine of the opening angle between the current and forwards cluster directions exceeds 0.97.
- The cosine of the opening angle between the cluster direction of the clusters, if merged, and each of the original clusters exceeds 0.9.

An example of a current cluster's associations is shown in figure 5.4.

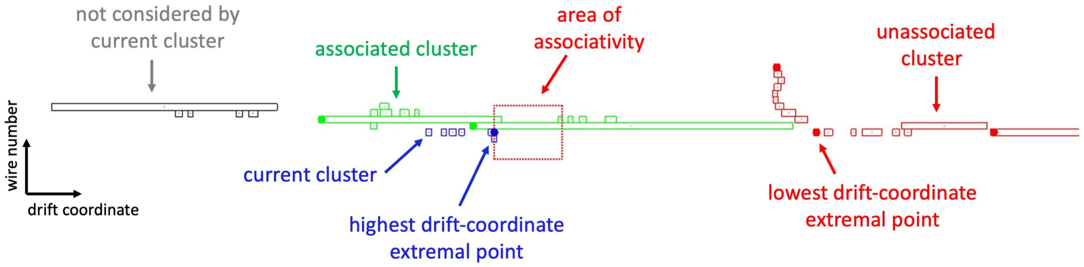


Figure 5.4: An example of the cluster associations of a current cluster (blue), where the (un)associated forwards clusters are shown in green(red).

4. Use the *ClusterAssociation* algorithm functionality to find and merge the longest pathway of associated clusters. An example of the output is shown in figure 5.5 which is in stark contrast to that presented in figure 5.2.

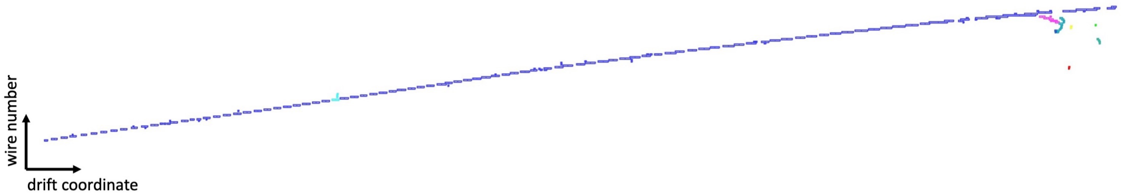


Figure 5.5: The result of the application of the *HitWidthClusterMerging* algorithm to the track presented in figure 5.2. Each colour denotes an individual cluster, some colours may be repeated.

The parameters used in the *HitWidthClusterMerging* algorithm were tuned by eye rather than by an automated minimisation procedure. This is a consequence of the complexity of the latter method.

The *HitWidthClusterMerging* algorithm was implemented into the Pandora reconstruction chain. The efficiency after the running of the *ThreeDTransverseTracks* algorithm and before the *ThreeDLongitudinalTracks* algorithm is shown in figure 5.6. It is seen that tracks that travel along the $\theta_{yz} = \sim\pm 54°$ direction are now created by the *ThreeDTransverseTracks* algorithm rather than the *ThreeDTrackFragments* algorithm, evidencing the success of the *HitWidthClusterMerging* algorithm. The integrated efficiency, at the end of the Pandora reconstruction chain, increases from 91.3% to 96.0%.

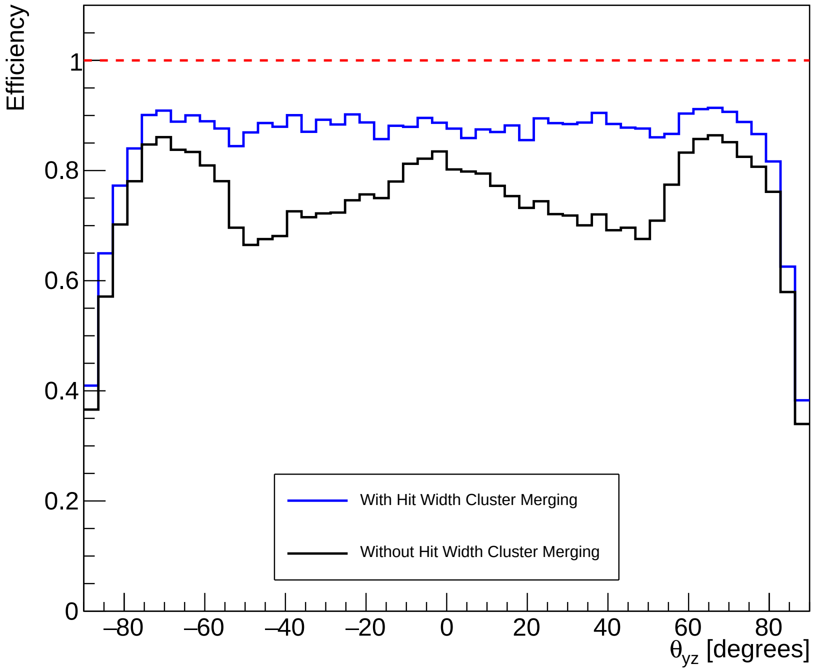


Figure 5.6: The efficiency observed after running the *ThreeDTransverseTracks* algorithm, but before the *ThreeDLongitudinalTracks* algorithm, with (and without) the *HitWidthClusterMerging* algorithm enabled.

5.2 Cosmic-Ray Reconstruction Improvements

The improvements achieved by the *HitWidthClusterMerging* algorithm caught the interest of the calibration working group who develop procedures to translate measured ionisation charge to energy. Their work is important, as accurate energy estimation is vital to the success of most DUNE analyses. For example, the analysis topic of this thesis, which examines DUNE's sensitivity to CP-violation.

One approach to calibration is the use of 'standard candles', which are particles whose energy is related to a measurable kinematic quantity. Stopping and through-going cosmic-ray muons are commonly used standard candles [110]. The calibration working group had found that cosmic-rays were often split in the Pandora reconstruction which hindered their ability to separate stopping and through-going cosmic-rays. Their cosmic-ray samples were shared with the Pandora team in order to fix this issue.

The cosmic-ray samples investigated were generated using the MUSUN generator [91] and version 8_32_00 of LArSoft. They are characterised by the distributions of figure 5.7, where it is seen that the cosmic-rays are generated just outside the boundaries of the detector and propagate in the downwards direction. The important thing to note is that the energies of the MUSUN sample are significantly larger than those of the single particle events of section 5.1.

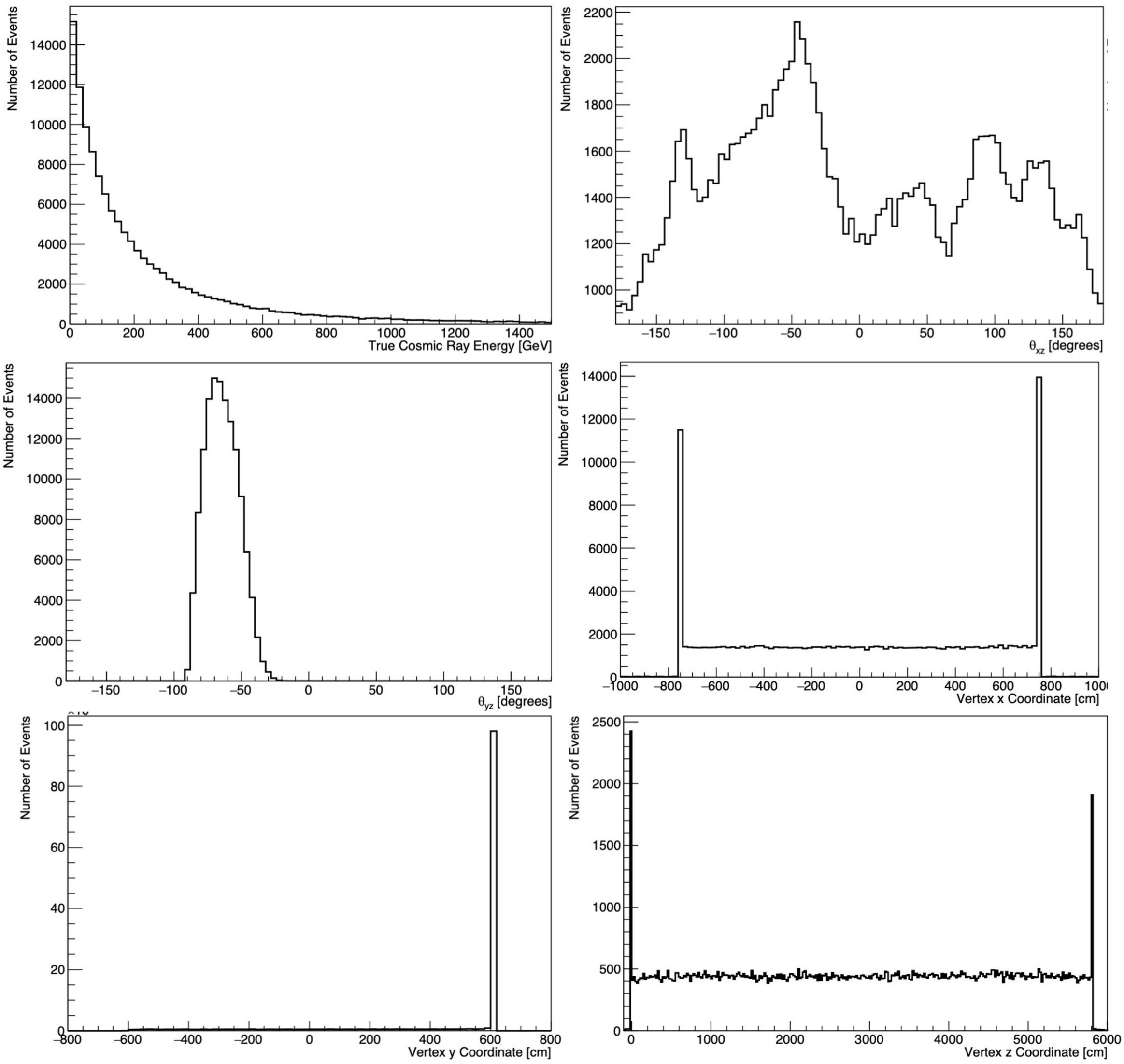


Figure 5.7: The true energy (top-left), θ_{xz} (top-right), θ_{yz} (middle-left), vertex x (middle-right), vertex y (bottom-left) and vertex z (bottom-right) distributions of the cosmic-ray sample.

5.2.1 Performance Metrics

The performance metrics chosen for this study are:

- **Completeness:** as defined in subsection 5.1.1.
- **Correct event fraction:** the fraction of events deemed correctly reconstructed, where a 'correct reconstruction' is one in which a reconstructable MC cosmic-ray muon has one, and only one, matched reconstructed particle.

The MC ↔ reconstructed particle matching procedure used in these metrics is that described in section 5.1 with the purity and completeness thresholds lowered to 50% and 10% respectively. This ensures that the correct event fraction is sensitive to splitting. If a MC particle is matched to more than one reconstructed particle, the reported completeness is that of the match with the highest number of shared hits.

5.2.2 Merging Across Detector Boundaries

Hand scanning events that were deemed ‘incorrect’ revealed that cosmic-rays were being broken across detector gaps, as illustrated in figure 5.8.

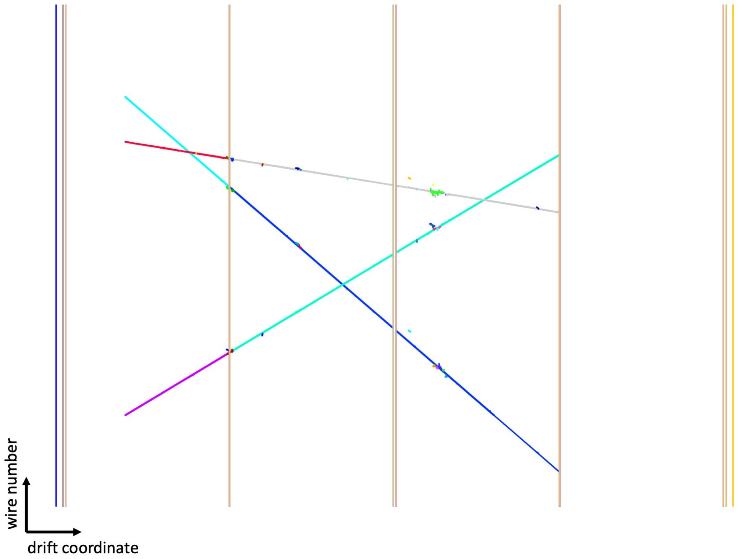


Figure 5.8: The U,V and W clusters of a reconstructed cosmic-ray event in which the cosmic-ray has been merged across one of the two detector gaps it passes through. Each colour denotes an individual cluster, some colours may be repeated.

After investigation, it was discovered that this was caused by a lack of functionality to merge tracks across multiple detector gaps in the stitching procedure described in subsection 4.3.2. This was therefore implemented and the subsequent improvements are presented in table 5.3 and figure 5.9.

Configuration	Correct Event Fraction
Initial	84.5%
Initial + Stitching	89.9%
Initial + FV	90.2%

Table 5.3: The integrated correct event fraction achieved by the initial Pandora reconstruction and after improvements have been made, where FV denotes the full volume approach. Uncertainties are negligible as a result of the number of events in the MC sample.

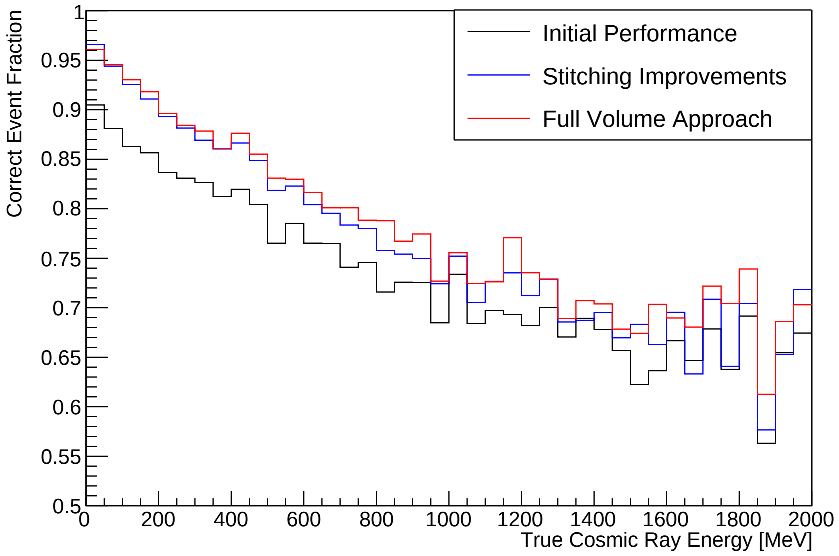


Figure 5.9: The correct event fraction, as a function of energy, achieved by the initial Pandora reconstruction (black), with the improved stitching procedure included (blue) and with the full volume approach utilised (red).

However, since the DUNE FD is underground, it can be assumed that each readout window will only ever feature an isolated cosmic-ray muon or neutrino beam event and consequently one can trigger on the event. The stitching procedure is therefore redundant and one can instead reconstruct the energy deposits of each drift volume in conjunction, this will be referred to as the 'full volume' approach. In the full volume approach, the clusters either side of a detector gap are merged together by the clustering and 2D → 3D matching algorithms, rather than by the explicit stitching procedure. This is more likely to be successful and thus results in the improvements shown in table 5.3 and figure 5.9.

5.2.3 Extending Through Showers

Further hand scanning of the incorrect events revealed that cosmic-rays were being broken at points of large electromagnetic behaviour as exampled in figure 5.10.

A 2D clustering algorithm called the *TrackInEMShower* algorithm was developed in response to this failure mechanism. Its approach is to trace the cosmic-ray trajectory through the shower region, so that a continuous cosmic-ray track can be created. In more detail, the workflow is:

1. Identify 'significant' clusters to consider in the algorithm; these are clusters that exceed 25cm in length.
2. Find the associated cluster pair with the longest combined length, where an associated cluster pair is one in which each cluster can be extended from its **extension point** to reach the other cluster. To check for association we:

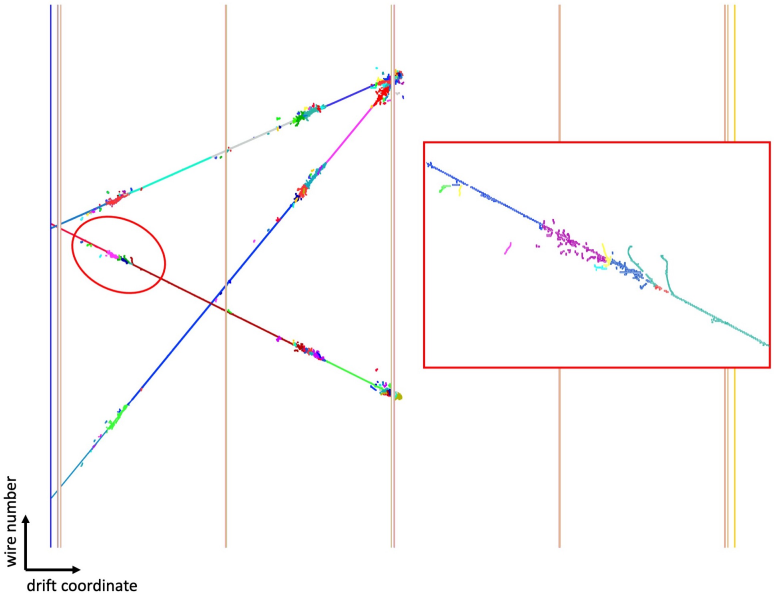


Figure 5.10: The U, V and W clusters of a cosmic-ray that has been split by regions of large electromagnetic behaviour, where each colour denotes an individual cluster (some colours may be repeated). The image on the right shows the region in the red circle with a larger scale, and will be used to illustrate the workflow of the *TrackInEMShower* algorithm.

(a) Find the **extension point** of each cluster:

   i. Identify the cluster endpoints.

   ii. Move along the cluster from its endpoint until the local cluster direction is consistent with the global direction of the other cluster.

   iii. Place the **extension point** at the start of this region.

   This procedure allows associations to be found in cases such as that illustrated in figure 5.11, where clustering errors have displaced the cluster endpoint from the true cosmic-ray trajectory.

(b) Apply the association criteria:

   i. The combined length of the clusters exceeds 75cm.

   ii. The cosine of the opening angle between the average cluster directions exceeds 0.99.

   iii. The extrapolated connection point of each cluster lies within 5cm of the other cluster's **extension point**. The extrapolated connection point is calculated by extending the cluster from its **extension point**, along the cluster direction, with a length equal to the distance between the extension points. This is shown in figure 5.11.

3. If an associated cluster pair is found, determine whether the **extension points** are separated by an electromagnetic shower region:

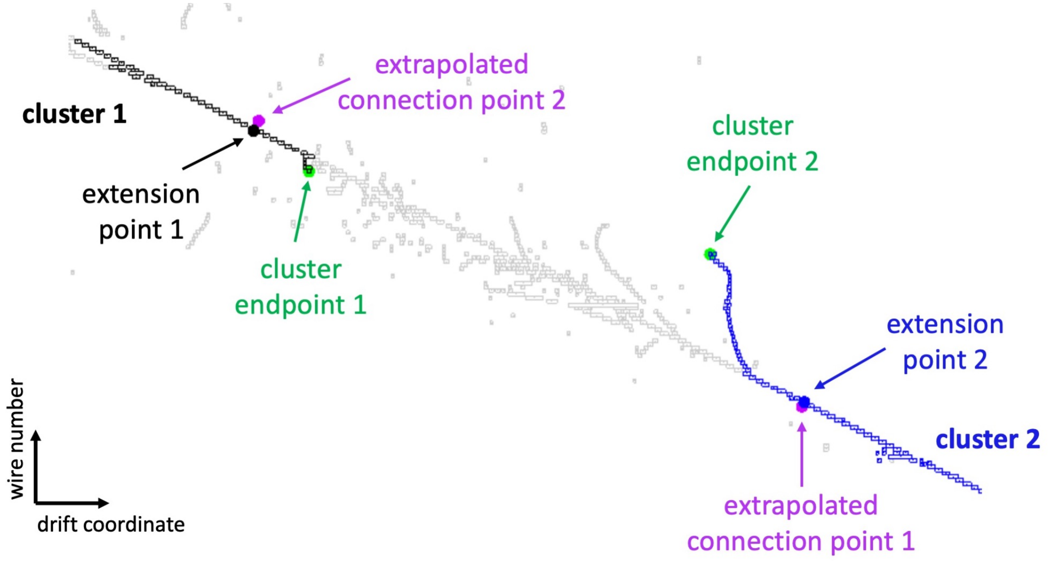


Figure 5.11: The found extension points of the clusters of a broken cosmic-ray, where clustering errors have displaced the cluster endpoint of the lower cluster.

(a) Collect any hits that lie between the **extension points** and within 0.35cm of the straight line that separates them, as shown in figure 5.12.

(b) Determine whether the collected hits are continuous, accounting for any detector gaps.

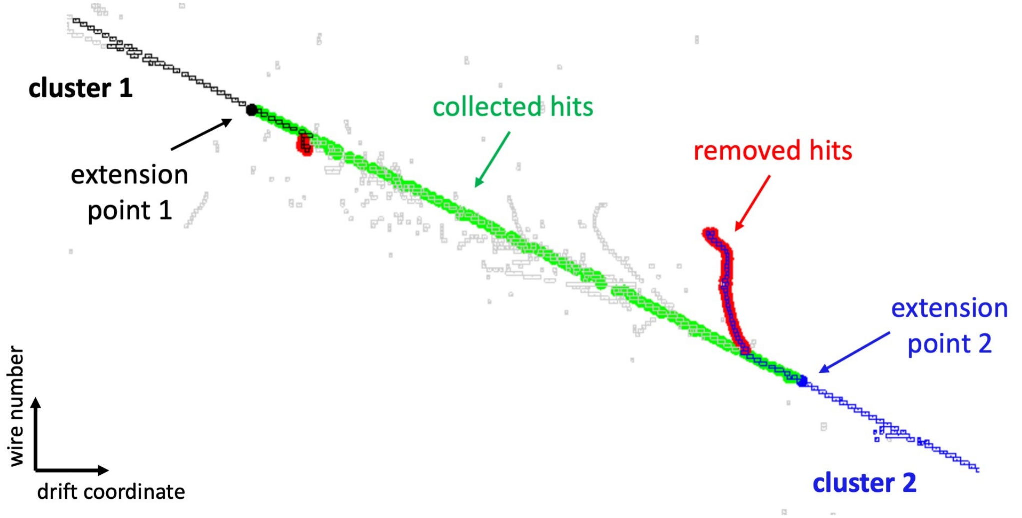


Figure 5.12: The collected hits (green) that form the trajectory of the cosmic-ray track as it propagates through the electromagnetic shower.

4. If a continuous pathway of hits is found, any clustering errors in the cluster endpoint region are fixed. This is demonstrated in figure 5.12 where substantial offshoots are removed and reclustered.

5. The collected hits and refined clusters are merged together.

With the implementation of *TrackInEMShower* algorithm, the correct event fraction is seen - in table 5.4 - to increase from 90.3% to 95.9%, demonstrating the algorithm's success. Figure 5.13 shows the correct energy fraction as a function of energy, where the algorithm is seen to be particularly useful in the high-energy region, as here regions of electromagnetic behaviour are larger and more frequent. Now that the cosmic-ray reconstruction performance was satisfactory, this side of the project was concluded.

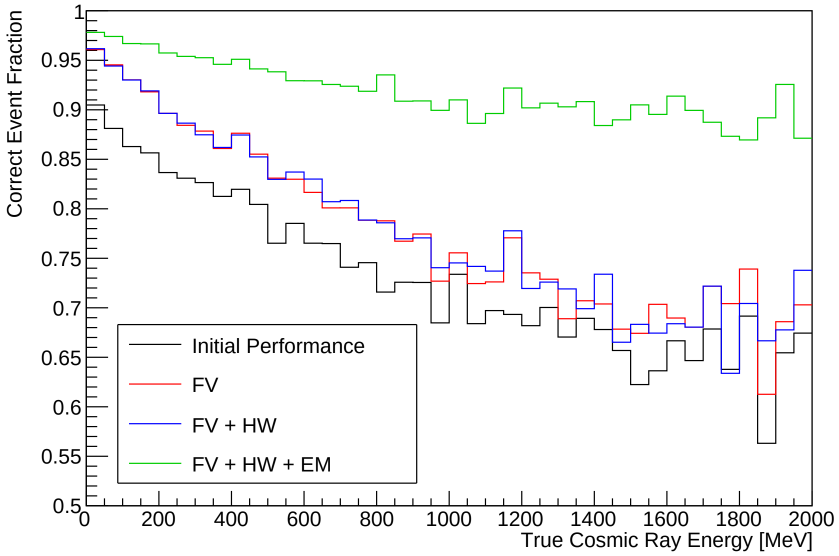


Figure 5.13: The correct event fraction, as a function of energy, achieved by the initial Pandora reconstruction (black) and after improvements have been made. FV, HW and EM denote the full volume approach (red), the addition of the *HitWidthClusterMerging* algorithm (blue), and the addition of the *TrackInEMShower* algorithm (green).

Configuration	Correct Event Fraction
Initial	84.5%
Initial + Stitching	89.9%
Initial + FV	90.2%
Initial + FV + HW	90.3%
Initial + FV + HW + EM	95.9%

Table 5.4: The integrated correct event fraction achieved by the initial Pandora reconstruction, and after improvements have been made where FV, HW and EM denote the full volume approach, the addition of the *HitWidthClusterMerging* algorithm, and the *TrackInEMShower* algorithm respectively.

5.3 Delta-Ray Reconstruction Improvements

An analysis began within the calibration working group which aimed to study the possibility of using delta-rays for calibration, as demonstrated by ICARUS [111]. A delta-ray is a shower created by the hard scattering of an ionisation electron from the parent cosmic-ray muon, it is seen in the LArTPC readout as an offshoot from the parent muon track.

Delta-rays are standard candles: the opening angle between their direction and that of the parent muon track is directly related to their energy [112]. However, for delta-rays to be a viable source of calibration, an accurate estimate of their direction is vital. This then depends on the ability of Pandora to reconstruct a sufficient proportion of the delta-ray.

This motivated the development of new reconstruction metrics, detailed in subsection 5.3.1, and led to the complete overhaul of the delta-ray reconstruction, which is described in section 5.3.2. These developments were made in the context of ProtoDUNE-SP, where the analysis was initially performed, and used the RITM1086087 MC data set, which:

- simulates a 1 GeV beam and cosmic-ray overlay, and
- was generated with LArSoft version 09_16_00.

5.3.1 Performance Metrics

Since we aim to understand the reconstruction quality of the delta-rays, we will use pattern-recognition metrics. But what does the perfect reconstruction goal look like for a delta-ray?

Following the procedure outlined in subsection 5.1.1 we first collect the hits for which the dominant charge contributor belongs to the delta-ray hierarchy; these are shown in the first panel of figure 5.14. Here, we see that:

1. The delta-ray initially travelled along the muon track, so both particles contribute to the charge of the hits in this region. These hits should be chosen to belong to the cosmic-ray as its reconstruction takes priority.
2. Bremsstrahlung radiation results in large gaps between hits, and so it is deemed impossible for an algorithm to reasonably identify their owner.

These are common occurrences, so, to obtain the perfect reconstruction goal shown in the second panel of figure 5.14, we:

1. Remove any shared hits to which the MC delta-ray hierarchy contributes less than 90% of the total hit charge.
2. Remove unconnected Bremsstrahlung radiation hits:
    (a) create a cluster seed out of the delta-ray hits that lie upstream of all Bremsstrahlung radiation in the delta-ray hierarchy,
    (b) use an iterative approach to grow the seed by collecting the hits that lie within 2.5cm of the seed.

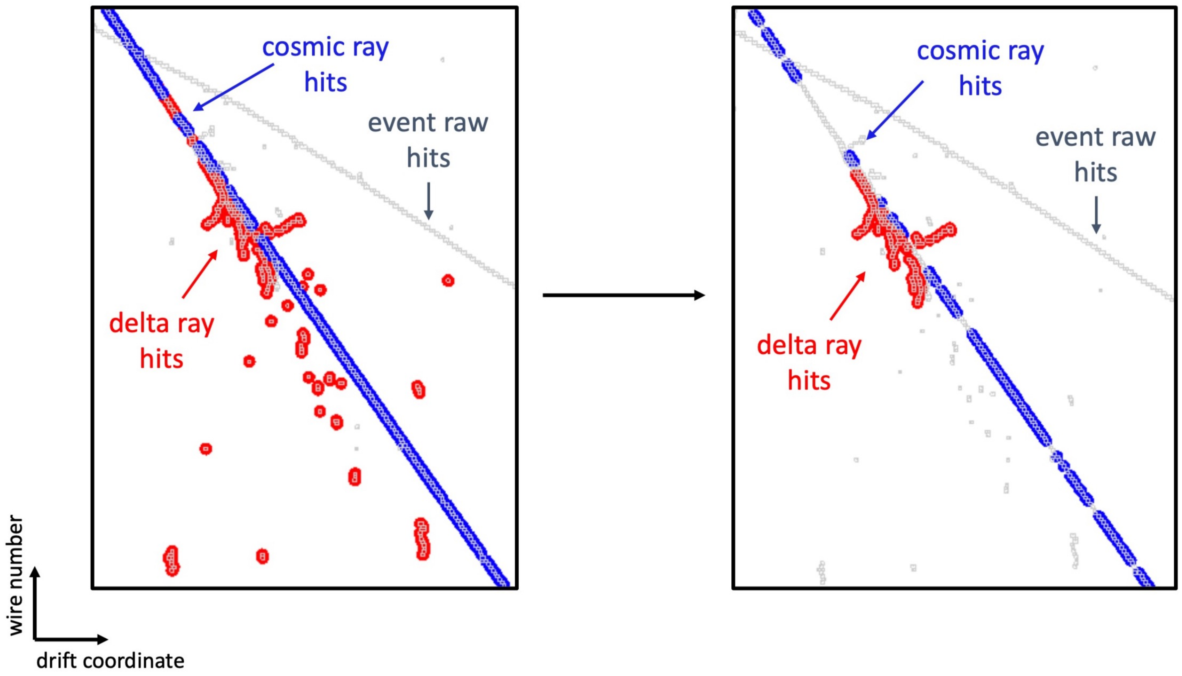


Figure 5.14: The 2D hits of a ProtoDUNE-SP event, where the grey squares are the input raw hits. A delta-ray hierarchy is shown, where the blue(red) dots denote the hits for which a cosmic(delta)-ray dominates the charge contribution.

The above procedure defines the hits that belong to a MC delta-ray, which we can now identify as reconstructable if it, and its parent cosmic-ray muon, have at least 5 hits in at least 2 views (and at least 10 hits overall). A reconstructed particle is matched to the MC particle if they share at least 5 hits and have a completeness and purity of at least 10% and 50% respectively.

Finally, we can define the delta-ray reconstruction performance metrics:

- **Correct Delta-Ray Fraction:** the fraction of reconstructable delta-rays that are correctly reconstructed, where the reconstruction is deemed correct if a single reconstructed particle has been matched to the MC delta-ray, and if the correct cosmic-ray has been identified as the delta-ray parent.
- **Completeness:** as defined in subsection 5.1.1.
- **Purity:** as defined in subsection 5.1.1.
- **Breakdown of Matches:** the percentage of reconstructable MC delta-rays with 0, 1, 2 or 3+ matches.

5.3.2 Delta-Ray 2D → 3D Matching

By hand scanning the event displays of incorrectly reconstructed delta-rays, several issues were seen to frequently occur in one, or multiple views:

- The delta-ray cluster had been split.
- Delta-ray hits had been merged into the parent cosmic-ray.
- Cosmic-ray hits had been merged into the child delta-ray.

These issues were found to be a consequence of the simplicity of the 2D → 3D delta-ray matching procedure. This finds the U, V and W clusters of the delta-ray, using the knowledge that they have a common span in the drift-coordinate, and creates a particle from the cluster triplet, doublet or singlet with the highest number of hits. It was thought that improvements could be made by replacing this procedure with the Pandora 'tensor' approach in which clusters can be merged together and/or refined before particle creation, increasing their completeness and/or purity.

The Tensor Approach

The loosely-named tensor approach revolves around a 'tensor' object. This object does not fit the mathematical description of a tensor - it is instead a 3D array which holds the cluster matches in the event. An example tensor, for the demonstrative event of figure 5.15, is shown in table 5.5. One typically asks the tensor for groups of connected matches i.e. in the given example we would be given the group u_1, v_1, v_2, w_1 and the group u_2, v_3, w_2 which correspond to 1:2:1 and 1:1:1 connected matches. In the former case, we can infer that the V clusters should be merged together, whilst in the latter case, we have an unambiguous match from which a particle can be created.

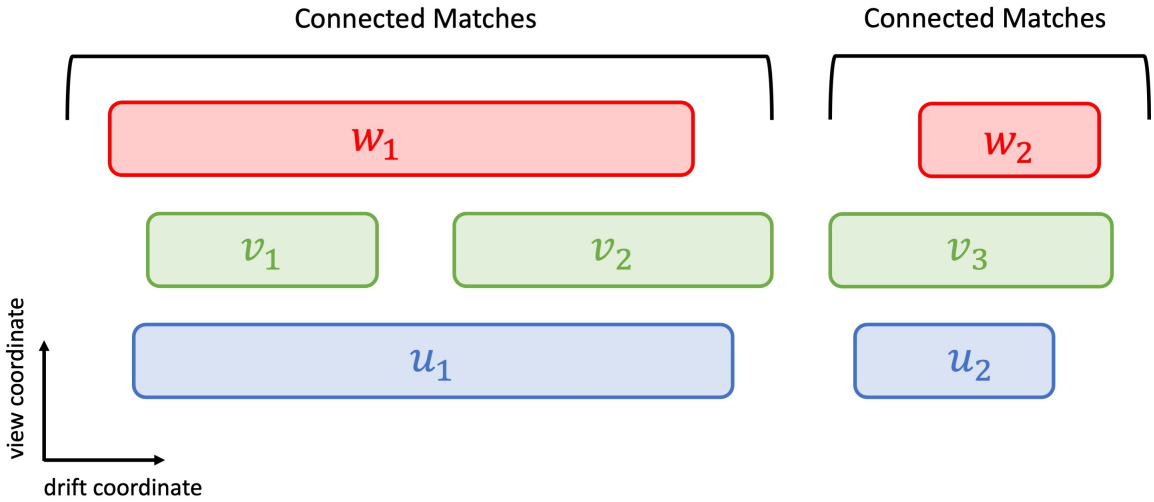


Figure 5.15: The U, V and W clusters of a hypothetical event, where the clusters that form a connected group of matches, and so likely belong to the same particle, are identified.

The tensor approach follows this demonstrated workflow: an algorithm builds and manages the tensor and applies a number of tools that modify the clusters and create particles. Each tool has a specific job and run in a chain such that:

- A tool is only activated if that before can make no changes.
- The tool chain begins again if a tool has made a change.

If a tool modifies a cluster, the tensor is updated.

U View Cluster	V View Cluster	W View Cluster
u_1	v_1	w_1
u_1	v_2	w_1
u_2	v_3	w_2

Table 5.5: The tensor of the event of figure 5.15, where each row is an element of the tensor and details a cluster match.

The ThreeViewDeltaRayMatching Algorithm

The *ThreeViewDeltaRayMatching* algorithm was developed to create delta-ray particles from three-view matches. The algorithm considers all U, V and W cluster combinations and inserts the triplet into the tensor if:

1. each of the clusters are available i.e. do not already belong to a particle,
2. each of the clusters has more than 5 hits,
3. the postulated delta-ray should have a muon parent, so also require that each cluster lies within 3cm of a cluster of the same muon particle,
4. there is a common overlap in the drift-coordinate,
5. the clusters correspond to a consistent set of 3D positions; for each combination (U:V→W, V:W→U, U:W→V) the projection of the two clusters agrees well with the cluster in the third view.

With the tensor filled, the following tool chain is applied:

1. ***DeltaRayRemoval:***
    - A refinement tool, which reassigns hits that have been clustered into a cosmic-ray, but which are thought belong to a child delta-ray.
    - It considers each delta-ray cluster in the context of its match, so that the parent muon is known. The cosmic-ray trajectory will likely deviate near the delta-ray connection point if a clustering mistake has been made (as in the left panel of figure 5.16). If this is the case, the muon clusters of the other two views are projected into the investigated view, to create a muon trajectory, and a proximity-based logic is used to reassign hits.

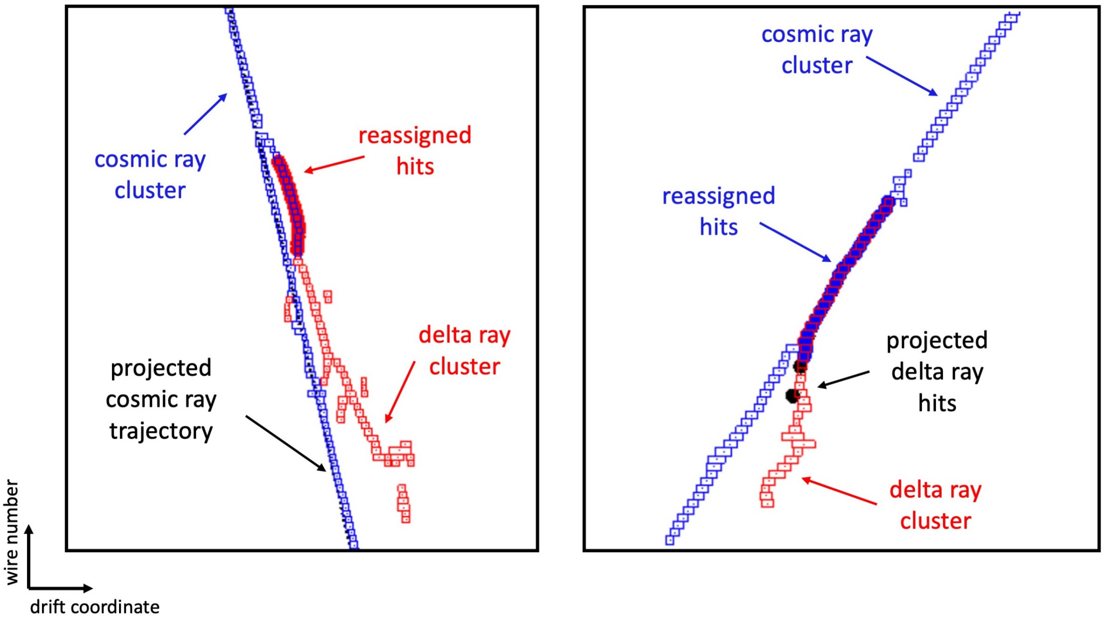


Figure 5.16: Left: The outcome of the *DeltaRayRemoval.* Cosmic-ray (blue) hits that are reassigned to the delta-ray (red) are shown by red dots. Right: The outcome of the *CosmicRayRemoval* tool. Delta-ray (red) hits that are reassigned to the cosmic-ray (blue) are shown by blue dots.

2. ***CosmicRayRemoval:***
    - A refinement tool, which reassigns hits that have been clustered into a delta-ray, but are thought belong to its parent cosmic-ray.
    - It considers each delta-ray cluster in context of its match, so that the parent muon is known. A clustering mistake has likely been made if a significant section of the delta-ray lies within the cosmic-ray trajectory (as in the right panel of figure 5.16). In this case, the remaining delta-ray clusters of the match are projected into the investigated view, and proximity-based logic is used to identify and reassign the cosmic-ray hits.
3. ***UnambiguousDeltaRay:***
    - A particle creation tool that creates delta-ray particles from a group of 1:1:1 connected matches if at least one cluster is connected (lies within 2cm) of the parent cosmic-ray muon.
4. ***DeltaRayMerge:***
    - A cluster merging tool, which merges clusters that have been split in one or two views.
    - Pairs of matches with two common clusters are considered and merges are made if the broken clusters are associated, i.e. if:

  - The matches have a common muon parent.
  - The broken clusters exhibit a ‘broken by muon track’ topology (left panel of figure 5.17), where the clusters connect (within 1.5cm) to the same cosmic-ray and their connection points lie within 10cm of each other.
  - The broken clusters exhibit a ‘broken by gap’ topology (right panel of figure 5.17), where the clusters are close together (separated by less than 3cm) and at least one cluster connects (within 1.5cm) to the cosmic-ray.
- Pairs of matches with one common cluster are considered and merges are made if:
  - Both broken cluster pairs are associated.
  - The postulated cluster triplet corresponds to a consistent set of 3D positions.

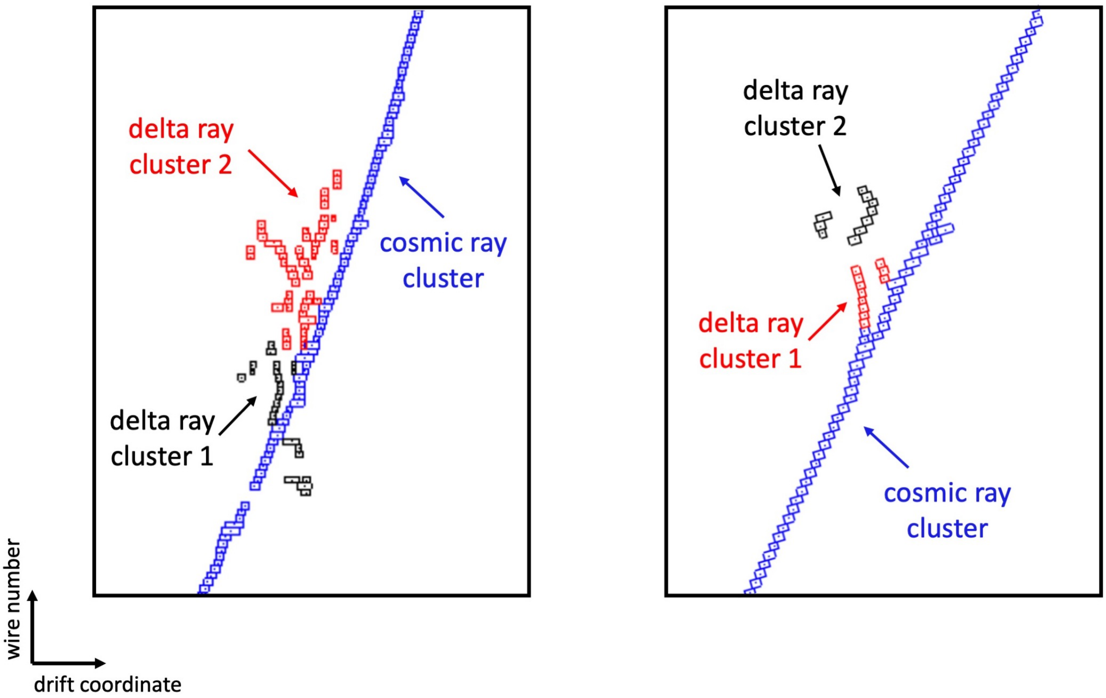


Figure 5.17: Examples of delta-rays that exhibit a ‘broken by muon track’ topology (left) and a ‘broken by gap’ topology (right). In each case, the *DeltaRayMerge* tool merges the clusters together.

5. ***AmbiguousDeltaRay:***

- A particle creation tool that creates particles from $l{:}n{:}m$ connected matches where $l \neq m \neq n \neq 1$.
- The connected matches are considered in order of their total hit number, and a delta-ray particle is created if, for each view, the consistency of the 3D positions of the cluster triplet exceeds a threshold.

Table 5.6 details the delta-ray reconstruction metrics before and after the implementation of the tensor approach. The correct event fraction is seen to fall by ∼46.4% with the alternative use of the tensor approach. This is explained by the increase of the percentage of reconstructable delta-rays that have no matched reconstructed particle, which suggests that three-view matches cannot be made. This can occur if the delta-ray has been poorly reconstructed in at least one-view: for example the cluster may have been merged into the parent cosmic-ray track or has been fragmented.

The average completeness and purity is calculated from reconstructed delta-rays i.e. those that have at least one matched reconstructed particle. If there are multiple matches, the completeness and purity is that of the match with the highest number of shared hits. The average completeness is found to increase by ∼7.9%, which occurs because now only particles with a good quality three-view match are created. However, there is also evidence that the cluster merging and refinement tools contribute to this, as there is a ∼1.4% purity degradation and the percentage of split particles decreases by ∼8.9%. Their success is therefore demonstrated, and the need for a two-view approach is motivated.

Configuration	Correct Event Fraction	Average Completeness	Average Purity	0	1	2	3+
Initial	73.9%	73.0%	93.4%	9.4%	76.5%	12.0%	2.1%
Three-View Matching	27.5%	80.9%	92.0%	26.2%	68.6%	4.4%	0.8%

Table 5.6: The delta-ray reconstruction performance metrics before and after the introduction of the tensor approach where, the average completeness and purity is calculated from the delta-rays that have at least one matched reconstructed particle and the breakdown of the multiplicity of the matches is detailed in the final four columns.

TwoViewDeltaRayMatching **Algorithm**

The *TwoViewDeltaRayMatching* algorithm was developed to create delta-ray particles from two-view matches. The algorithm runs three times with each two-view combination and, in each case, finds doublet matches, which are made if:

1. there is evidence that the third view cluster has been merged into the parent cosmic-ray track (left panel of figure 5.18) or has been fragmented (right panel of figure 5.18), and
2. the postulated triplet match satisfies the three-view consistency requirements (of the *ThreeViewDeltaRayMatching* algorithm).

The *TwoViewAmbiguousDeltaRay* tool creates particles from the doublet matches, in order of their total hit counts, and attempts to recover the third-view cluster with the following workflow:

1. Identify the 'best matched' third-view cluster.
2. If it belongs to a parent muon then the functionality of the *DeltaRayRemoval* tool is used to remove the merged delta-ray hits from the cosmic-ray (left panel of figure 5.18). The created cluster is grown from any remaining matched clusters such that the consistency of the 3D positions of the triplet is maintained.
3. If it does not belong to a muon, or if no hits could be removed, the 'best matched' available cluster is identified and grown (right panel of figure 5.18).

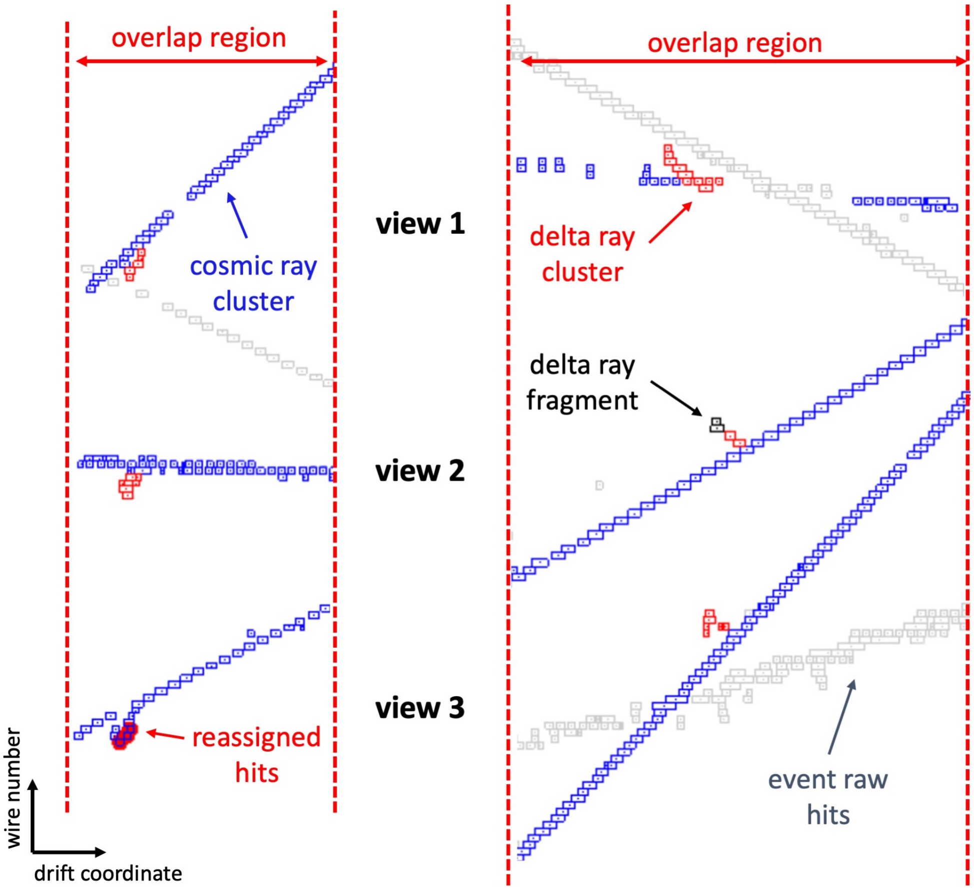


Figure 5.18: Left: An example of a delta-ray cluster, which has been absorbed into the parent cosmic-ray track. The *TwoViewAmbiguousDeltaRay* tool removes these hits, creating the third-view cluster, and subsequently a three-view delta-ray particle. Right: An example of a delta-ray cluster, which has been fragmented in one-view. The *TwoViewAmbiguousDeltaRay* tool merges these clusters together, and subsequently creates a three-view delta-ray particle.

The correct event fraction increases from 27.5% to 71.4% with the addition of the *TwoViewDeltaRayMatching* algorithm, however this still falls short of the 73.9% achieved by the initial Pandora reconstruction. The performance metrics of table 5.7 show that the quality of the reconstructed delta-rays achieved by the tensor approach is superior and that the inferior correct event fraction arises because some delta-rays are still not reconstructed at all.

In the initial Pandora reconstruction, a delta-ray particle is created even if no match can be found, resulting in 'single view' reconstructed delta-rays. This implementation handles cases in which the delta-ray has been fragmented or, merged into the cosmic-ray track, in multiple views and is thought to be responsible for the difference in the correct event fraction. It is, however, important to note that such single view reconstructed delta-rays are not useful to the calibration group since we cannot infer their 3D direction. This finding motivates the *OneViewDeltaRayMatching* algorithm.

Configuration	Correct Event Fraction	Average Completeness	Average Purity	0	1	2	3+
Initial	73.9%	73.0%	93.4%	9.4%	76.5%	12.0%	2.1%
Three-View Matching	27.5%	80.9%	92.0%	26.2%	68.6%	4.4%	0.8%
+ Two-View Matching	71.4%	79.4%	93.8%	14.3%	77.4%	7.1%	1.2%

Table 5.7: The delta-ray reconstruction performance metrics before and after the introduction of the tensor approach, where the average completeness and purity is calculated from the delta-rays that have at least one matched reconstructed particle, and the breakdown of the multiplicity of the matches is detailed in the final four columns.

***OneViewDeltaRayMatching* Algorithm**

The *OneViewDeltaRayMatching* algorithm was developed to create delta-ray particles in cases where a multiple-view match cannot be found. At the time of running, most delta-ray particles have been made and this enables us to take a much more aggressive approach to match and merge clusters such that we avoid creating single-view delta-rays.

The algorithm runs in each view and first identifies the available clusters that are likely to belong to delta-rays: these are significant clusters (more than 3 hits) which have an identifiable parent cosmic-ray (those which lie within 3cm of a cosmic-ray track). The algorithm considers each cluster, and first attempts to merge it into an existing delta-ray particle:

1. Identify the most likely parent cosmic-ray: this is that to which the cluster lies closest to.

2. For each 'other' view, search amongst the unavailable clusters that lie close to the parent cosmic-ray (within 3cm) and identify any which overlap in the drift-coordinate.

3. Take the best matched cluster in each view to be that with the highest number of hits.

4. If the best matched clusters (from the other two views) belong to the same created delta-ray, merge in the available cluster.

If it was not possible to merge the available cluster into an existing delta-ray, it is assumed that the delta-ray particle has not yet been created. In this case, the algorithm attempts to create the delta-ray:

1. For each potential parent cosmic-ray, find the best matched available cluster in each 'other view' by following steps 1-3 in the above, but instead searching amongst available clusters.
2. Grow the available cluster, and each best matched cluster, by merging in close available clusters (those within 3cm) using an iterative approach
3. Create the delta-ray particle.

Configuration	Correct Event Fraction	Average Completeness	Average Purity	0	1	2	3+
Initial	73.9%	73.0%	93.4%	9.4%	76.5%	12.0%	2.1%
Three-View Matching	27.5%	80.9%	92.0%	26.2%	68.6%	4.4%	0.8%
+ Two-View Matching	71.4%	79.4%	93.8%	14.3%	77.4%	7.1%	1.2%
+ One-View Matching	76.8%	77.9%	92.8%	8.5%	78.9%	10.9%	1.7%

Table 5.8: The delta-ray reconstruction performance metrics before and after the introduction of the tensor approach, where the average completeness and purity is calculated from the delta-rays that have at least one matched reconstructed particle, and the breakdown of the multiplicity of the matches is detailed in the final four columns.

The delta-ray reconstruction metrics are shown in table 5.8, where it is seen that, with the introduction of the *OneViewDeltaRayMatching* algorithm, the correct event fraction exceeds that obtained by the initial Pandora reconstruction by ∼2.9%. A small ∼ 1% loss in the purity is seen, which arises as a result of the aggressive nature of the cluster merging/matching approach. It is concluded that in the tensor approach, delta-rays are reconstructed more frequently and to a higher quality. These results are expected to benefit the calibration working group as, with the aforementioned improvements, the accuracy of the delta-ray initial direction estimator will increase. This will result in an improved mapping from reconstructed to true energy and make the use of delta-rays as standard candles more achievable.

Chapter 6

Probing the Sensitivity to CP Violation at DUNE

The oscillation channel of most interest to DUNE is the $\overset{(-)}{\nu}_\mu$ to $\overset{(-)}{\nu}_e$ appearance channel for which the oscillation probability [38] is given by

$$\begin{aligned} P\left(\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e\right) \approx{}& \sin^2\theta_{23}\sin^2 2\theta_{13}\frac{\sin^2(\Delta_{31}-aL)}{(\Delta_{31}-aL)^2}\Delta_{31}^2 \\ &+ \sin 2\theta_{23}\sin 2\theta_{13}\sin 2\theta_{12}\frac{\sin(\Delta_{31}-aL)}{(\Delta_{31}-aL)}\Delta_{31} \\ &\times \frac{\sin(aL)}{aL}\Delta_{21}\cos(\Delta_{32}\pm\delta_{\mathrm{CP}}) \\ &+ \cos^2\theta_{23}\sin^2 2\theta_{12}\frac{\sin^2(aL)}{(aL)^2}\Delta_{21}^2, \end{aligned} \tag{6.1}$$

where $a = \pm G_F N_e/\sqrt{2}$ and $\Delta_{ij} = \Delta m_{ij}^2/4E$. The CP-phase is seen to enter equation 6.1 at leading order and thus observably impacts the composition of the flux at the DUNE FD providing a handle through which the extent of CP-violation can be determined. However the ability of DUNE to detect CP-violation is limited by: our ability to determine the identity and energy of the interacting neutrinos, the performance of the detector and beam, the statistics of the experiment, the magnitude of our uncertainties and our understanding of them, and by the true value of the oscillation parameters themselves.

The impact of these limitations can be estimated by performing fake data studies. This is where the universes allowed by the aforementioned uncertainties are probed and for each, one:

1. reconstructs the generated neutrino interactions at the DUNE FD, corresponding to a given Protons on Target (POT) exposure;

2. selects the CC $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ neutrino interactions, creating the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ reconstructed energy spectra;

3. fits the spectra to determine whether CP is violated.

These estimates can be used to guide hardware, reconstruction and analysis improvements such that DUNE's ability to discover CP-violation is maximised. This approach is illustrated in this thesis, in which the ability to determine the neutrino energy spectra is linked to the performance of the CP-violation analysis, allowing the development of reconstruction and analysis improvements to be motivated and implemented.

The aim of this chapter is to detail the above workflow in the context of a Pandora-based CP-violation analysis. Section 6.1 will detail the generation of the MC samples used in the Pandora-based CP-violation analysis, section 6.2 will discuss how one can estimate whether DUNE is sensitive to CP-violation, and section 6.3 will present the CC $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection procedure. This ordering ensures that the reader gains an understanding of how DUNE's ability to determine CP-violation depends on the accuracy of the selected $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ reconstructed energy spectra, allowing them to appreciate the chapter's ending discussions of the limitations of the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection procedure.

6.1 Fake Data Generation

To simulate neutrino oscillations, DUNE provides three types of files: the first contains the CC and Neutral-Current (NC) interactions of the beam $\overset{(-)}{\nu}_\mu$ and intrinsic $\overset{(-)}{\nu}_e$ components, the other two files correspond to a scenario where each beam component fully oscillates via an oscillation channel specified in table 6.1.

File Name	Oscillation Channels
Beam	$\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_\mu$, $\overset{(-)}{\nu}_e \to \overset{(-)}{\nu}_e$
Flavour Swap 1	$\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e$, $\overset{(-)}{\nu}_e \to \overset{(-)}{\nu}_\tau$
Flavour Swap 2	$\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_\tau$, $\overset{(-)}{\nu}_e \to \overset{(-)}{\nu}_\mu$

Table 6.1: The three file types used to obtain an oscillated fake data set. Their obtainment from the beam file is detailed in the final column.

To demonstrate how these files can be used to determine what might be seen in the DUNE FD during a specified run time and projected POT, start by considering a beam file with 10 events. The flavour swap files are created from those same 10 events in the sense that the neutrino energy is the same but their flavours are changed as prescribed above.

A weight is assigned to each of the 30 events that is equal to the oscillation probability of the appropriate oscillation channel in the final column in table 6.1. The oscillation weight is calculated with the true neutrino energy of the event. If the neutrino energies are common between the files, 10 beam neutrinos will correspond to 10 oscillated neutrinos i.e. no neutrinos are gained or lost. Of course in practise the files used in an analysis do not share the same base events, but they will share the underlying neutrino energy spectrum which, with large statistics, is equivalent.

Finally, to get an oscillated FD spectrum prediction which corresponds to the projected POT, the oscillation weight is multiplied by a POT weight given by

$$\text{POT weight} = \frac{\text{projected POT}}{\text{POT of generated file}}. \tag{6.2}$$

In practice a large number of input events is used to avoid statistical fluctuations. There are consequently many files of each file type used and within each file type, each file is generated with a different POT. Each event is therefore given a POT weight of

$$\text{POT weight} = \frac{\text{projected POT}}{\sum_{i=0}^{n} \text{POT of file } i}, \tag{6.3}$$

where i is the enumeration of a file within the file type and n is the total number of files of the file type.

6.2 CP-Violation Metrics

The performance of the Pandora-based CP-violation analysis, presented in this thesis, will be quantified by the **sensitivity to CP-violation**, i.e. whether one is able to say that CP is violated. In this section, the fits through which the sensitivity is obtained will be detailed, the behaviour of the sensitivity metric with respect to the CP-phase understood, and several different estimates of the sensitivity, that vary in their sophistication, will be discussed.

6.2.1 Sensitivity to CP Violation

The sensitivity to CP-violation is equal to the level of confidence with which one can reject CP-Conservation (CPC) in preference to CP-Violation (CPV), and is given by the equation

$$\text{sensitivity} = \sqrt{\Delta\chi^2_{\text{CPV}}} = \Bigg(\underbrace{\min\{\chi^2_{\delta_{\text{CP}}=0}, \chi^2_{\delta_{\text{CP}}=\pm\pi}\}}_{\text{CPC fit}} - \underbrace{\chi^2_{\text{CPV}}}_{\text{CPV fit}} \Bigg)^{1/2}, \tag{6.4}$$

where each chi-squared value is the sum of the chi-squared values obtained from a simultaneous fit to the ν_e, $\bar{\nu}_e$, ν_μ, and $\bar{\nu}_\mu$ selected neutrino reconstructed energy spectra. Each chi-squared value represents the likelihood of a deviation from a hypothesised expected spectrum, to quantify how well an observed spectrum agrees with a hypothesis or to what extent the hypothesis can be rejected. The CPC fit assumes a CPC hypothesis, with δ_{CP} fixed to 0 or $\pm\pi$, and, the CPV fit assumes a hypothesis in which δ_{CP} is allowed to vary.

Motivated by the DUNE TDR analysis [64], this analysis considers the [0,8]GeV region of the spectra, using energy bins of width 0.25GeV. It is hypothesised that each bin of an observed spectrum will follow a Poisson distribution defined by a mean equal to the corresponding bin height of the expected spectrum. For each bin, the probability of obtaining the observed value [113] is given by

$$P(n_i) = \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \tag{6.5}$$

where μ_i and n_i are the expected and observed bin content of the i'th bin. However, it is often easier to work with the natural logarithm of this the, so called, log-likelihood. The test statistic,

$$\lambda = \frac{\ln P(n_i)}{\ln P(\mu_i)} \tag{6.6}$$

can then be used to quantify the difference between the likelihood of obtaining the observation and expectation, where a larger value indicates a bigger disagreement between the observation and hypothesis. Systematic and oscillation parameter uncertainties allow for multiple predictions of the expectation and, when fitting the hypothesis to an observation, one seeks the prediction that best fits the observation, i.e. that with the smallest test statistic. This is equivalent to minimising the chi-squared, which can be expressed as:

$$\begin{align} \chi^2 &= -2\ln\lambda \tag{6.7} \\ &= 2\sum_{i}^{N_{\text{bins}}} \Big(E_i - O_i + O_i \ln\frac{O_i}{E_i} \Big) \tag{6.8} \end{align}$$

where the sum is over the bins of the spectrum, E_i is the 'expected' bin count, that predicted by the hypothesis, and O_i is the 'observed' bin count.

To gain the most informed and realistic estimate of the sensitivity, one would explore the entire phase space defined by the uncertainties of the MC model. This is extremely time consuming and computationally expensive, and a much simpler estimate will often suffice. The sensitivity estimates used in this thesis will now be detailed and their interpretations highlighted.

6.2.2 Sensitivity Estimate without Uncertainties

In the simplest estimate of the sensitivity to CP-violation, one only considers the uncertainty of the CP-phase and neglects the consideration of statistical fluctuations and all other uncertainties. This means that:

1. Observations are generated across the δ_{CP} phase space with no systematic shifts applied, and with the assumption that the oscillation parameters are known exactly

and are equal to their central values, defined in table 6.2.

2. The CPC evaluation of equation 6.4 has no degrees of freedom, but the CPV evaluation can explore the δ_{CP} phase space.

Oscillation Parameter	Central Value	Uncertainty
θ_{12}	0.5903	2.3%
θ_{23} (Normal Ordering)	0.866	4.1%
θ_{23} (Inverted Ordering)	0.869	4.0%
θ_{13} (Normal Ordering)	0.150	1.5%
θ_{13} (Inverted Ordering)	0.151	1.5%
Δm^2_{21}	$7.39 \times 10^{-5}\text{eV}^2$	2.8%
Δm^2_{32} (Normal Ordering)	$2.451 \times 10^{-3}\text{eV}^2$	1.3%
Δm^2_{32} (Inverted Ordering)	$-2.512 \times 10^{-3}\text{eV}^2$	1.3%
ρ	2.848 g cm^{-3}	2.0%

Table 6.2: The central values and uncertainties of the oscillation parameters. These values match those used in the DUNE TDR CP-violation analysis detailed in reference [64].

The simplistic estimate is very quick to obtain, and is easily interpreted, but has several caveats:

- The sensitivity, as a function of δ_{CP}, is only understood at a single position in the phase space defined by the full oscillation and systematic uncertainties.

- With no other degrees of freedom in the fit, the fit will always be able to assign the CPV signal to the CP-phase and thus the maximal sensitivity is reported.

The 'simple estimate' of the sensitivity achieved using the Pandora-based CP-violation analysis, presented in this chapter, is shown in figure 6.1. Its individual ν_e, $\bar{\nu}_e$, ν_μ and $\bar{\nu}_\mu$ contributions are illustrated in figure 6.2. The characteristics of the sensitivity are best understood by thinking in terms of the underlying Poisson distributions that govern the chi-squared values of the CPC and CPV evaluations of equation 6.4. In the following explanations, for simplicity, we will pretend that DUNE has no energy reconstruction capabilities - such that its observations and expectations are numbers of events, i.e. counts rather than a spectra.

In the 'simple estimate', a CPV prediction can always be found that exactly matches the observation and hence the corresponding chi-squared evaluation will equal zero. We therefore only consider the Poisson distributions of the CPC fit, for which there is a single prediction for each CPC prediction ($\delta_{\text{CP}} = 0$ and $\delta_{\text{CP}} = \pm\pi$). The corresponding Poisson distributions of the expected ν_e counts are shown in figure 6.3. There are three key points to take from figure 6.3:

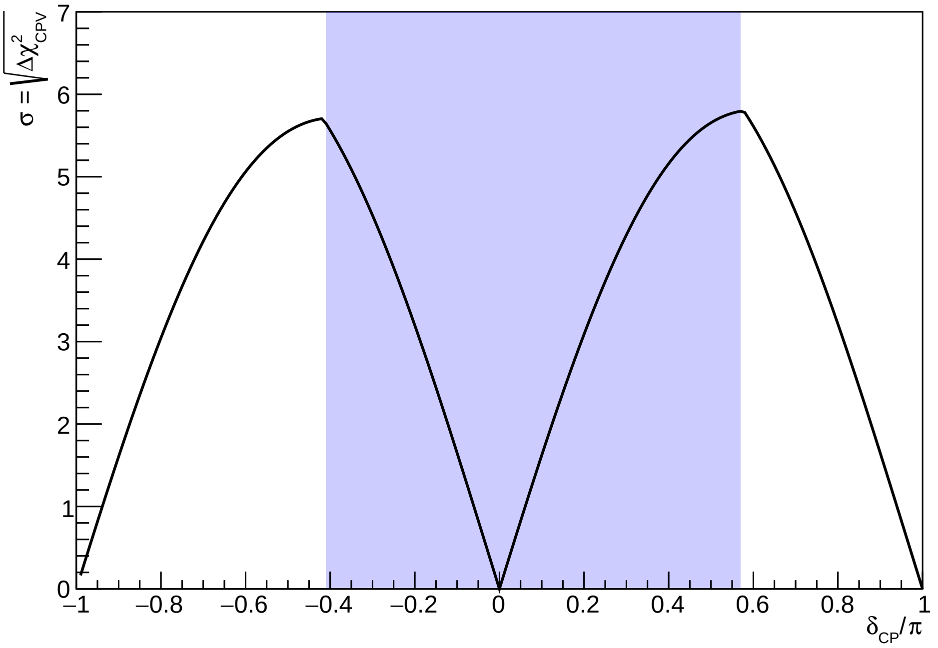


Figure 6.1: The 'simple estimate' of the sensitivity achieved by the CP-violation analysis outlined in section 6.3 with a 3.5 year exposure (1.36 $\times 10^{23}$ POT) in each beam mode (FHC and RHC) and an assumed normal ordering of the mass hierarchy. In the blue(white) region, the CPC fit of equation 6.4 is minimised by the CPC $\delta_{\mathrm{CP}} = 0(\delta_{\mathrm{CP}} = \pi)$ expectation.

1. At the count rates expected at DUNE, the Poisson distribution loses its asymmetry and can be approximated by a Gaussian distribution.

2. A larger difference between expectation and observation results in a larger test statistic and thus chi-squared value. This reflects the better confidence one has to distinguish two spectra if the difference between them is large.

3. A larger expected mean results in a broader shape of the distribution, and so an equivalent difference gives a larger chi-squared value if the count rate is smaller. This reflects the reduced confidence one has to discriminate the spectra if the difference is insignificant with respect to the scale of the expectation.

With these three statements, we can now explain the structure and features of the sensitivity and its contributions:

- **Shape:** Equation 6.1 demonstrates that the deviation of the observed count from the CPC expectation follows a sinusoidal dependence, where the maximally-violating phase corresponds to maximal deviation. By statement 2, the sensitivity will follow the same dependence.

- $\overset{(-)}{\nu}_e$ **Peak Asymmetry:** The peak of the positive(negative) maximally violating

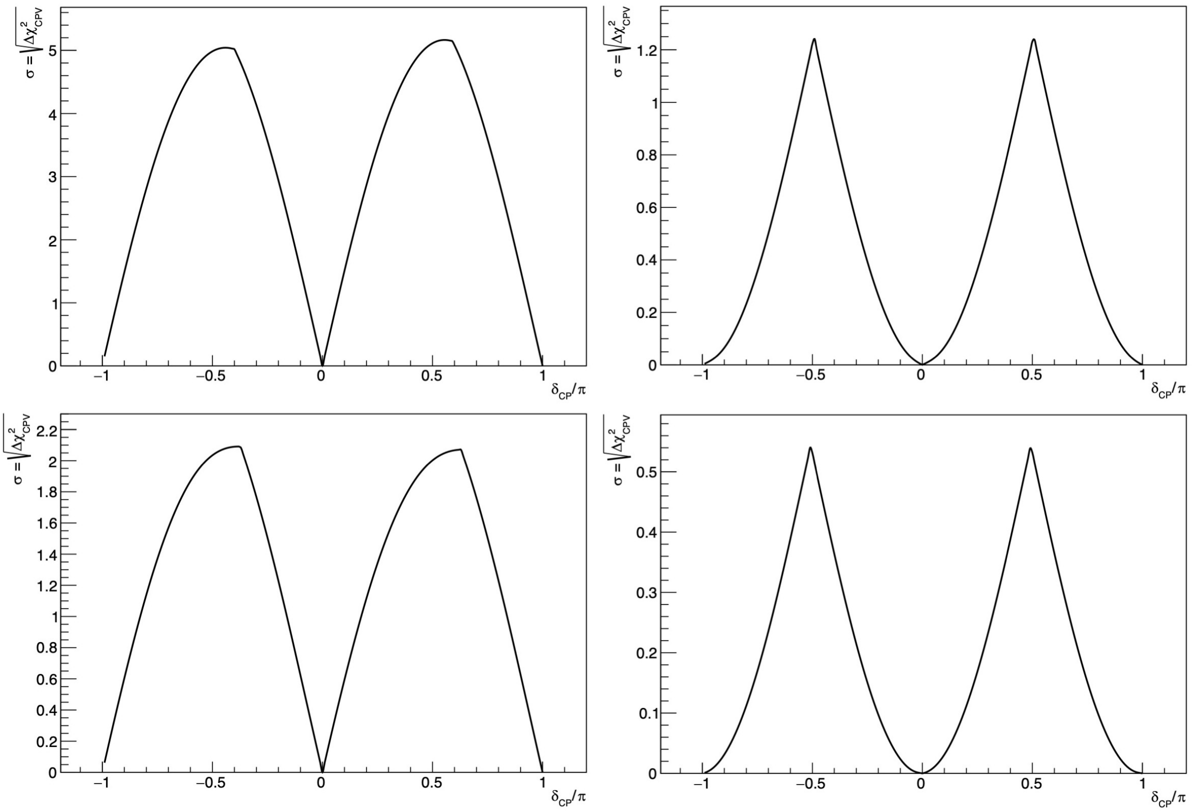


Figure 6.2: The 'simple estimate' ν_e (top-left), ν_μ (top-right), $\bar{\nu}_e$ (bottom-left) and $\bar{\nu}_\mu$ sensitivity contributions to figure 6.1.

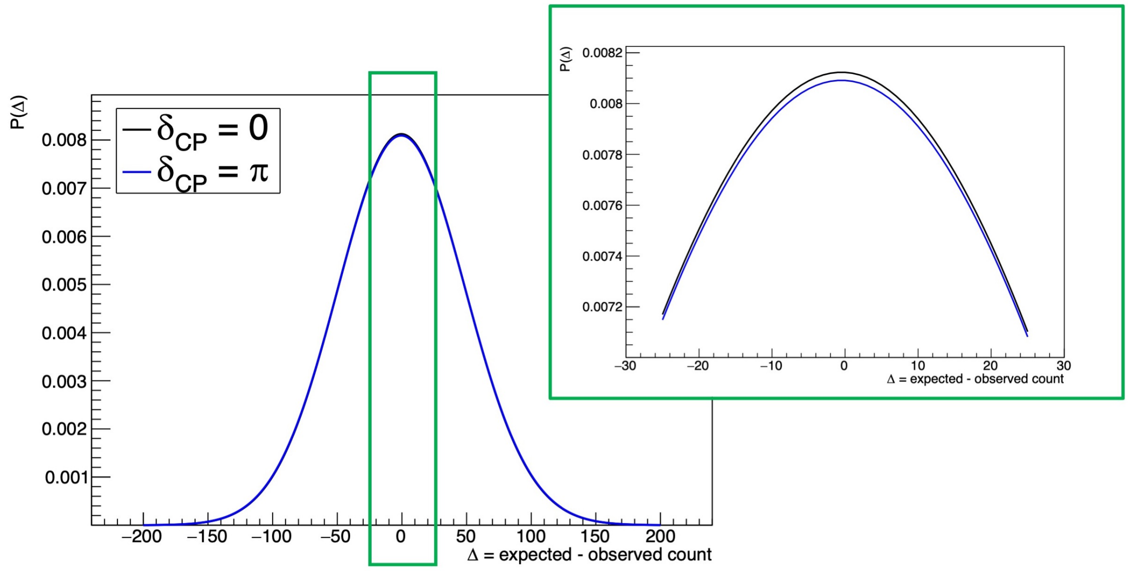


Figure 6.3: Poisson distributions defined by the expected number of selected ν_e events, where count numbers are obtained from the CP-violation analysis outlined in section 6.3 in which a 3.5 year exposure (1.36×10^{23} POT) in each beam mode (FHC and RHC) and an assumed normal ordering of the mass hierarchy are assumed. The expected $\delta_{\text{CP}=0}(\delta_{\text{CP}=\pi})$ count is 2412(2431), which includes backgrounds.

phase is largest in the FHC(RHC) contribution. This is because a different CPC expectation minimises the CPC fit (of equation 6.4) at each peak **and**, because the corresponding expected count numbers differ. For the neutrino case, the RHS peak is minimised by $\delta_{\text{CP}} = 0$ (see figure 6.1), which has a lower number of counts than $\delta_{\text{CP}} = \pi$ (which minimises the LHS peak) as illustrated in figure 6.3). By statement 3, this corresponds to an asymmetry.

- $\overset{(-)}{\nu}_e > \overset{(-)}{\nu}_\mu$: The CP contribution to the $\overset{(-)}{\nu}_\mu$ disappearance probability is less than that to the $\overset{(-)}{\nu}_e$ appearance probability. Consequently the difference in the CPC and CPV counts is smaller, especially with respect to the large scale of the $\overset{(-)}{\nu}_\mu$ counts, which, by statements 2 and 3, results in a reduced sensitivity contribution.
- **FHC > RHC:** The cross section of FHC neutrino interactions is larger than that for RHC, which has conflicting effects on the sensitivity. By statement 2, a larger difference increases the chi-squared but, by statement 3, a larger predicted count reduces the chi-squared. Equation 6.8 demonstrates that, if the expected and observed counts are multiplied by a common factor, the chi-squared is scaled by the same factor.

In the above explanations, we have considered a single count for each observation/-expectation. When moving to a binned energy spectrum, one examines the differences between observation and expectation at the bin level, i.e. performs a counting experiment for each energy bin. Figure 6.4 demonstrates that the sum of chi-squared values is higher than that of a single count. The CP contribution is energy dependent, and so some energy bins contribute will more to the chi-squared than others. This is discussed more in subsections 6.6.1 and 6.6.2.

6.2.3 Sensitivity Estimate with Oscillation Parameter Uncertainties

One can make the previous estimate more realistic by considering the oscillation parameter uncertainties. This means that:

1. Observations are generated across the phase space defined by the uncertainties of the oscillation parameters, such that each observation corresponds to a ‘thrown’ oscillation parameter set, which represents a simultaneous variation of the oscillation parameters, where each variable is sampled according to the relevant distribution given in table 6.3.
2. The fits of equation 6.4 explore the predictions allowed by the oscillation parameter uncertainties. As in the DUNE TDR analysis [64], $\sin^2\theta_{23}$, Δm^2_{32} and δ_{CP} are allowed to vary freely, whilst θ_{12}, Δm^2_{21} and ρ are implemented as ‘nuisance’ parameters.

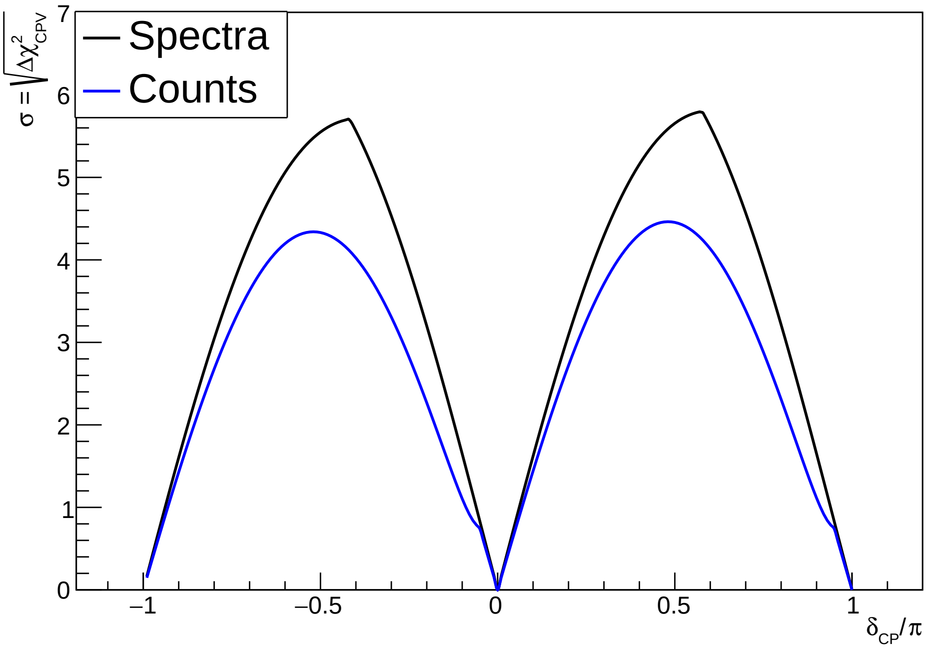


Figure 6.4: The 'simple estimate' of the sensitivity achieved by the CP-violation analysis outlined in section 6.3 with a 3.5 year exposure (1.36 $\times 10^{23}$ POT) in each beam mode (FHC and RHC) and an assumed normal ordering of the mass hierarchy. To obtain the black(blue) curve, observations take the form of reconstructed energy spectra (counts).

Parameter	Distribution	Range
θ_{12}	Fixed	N/A
$\sin^2\theta_{23}$	Uniform	$[0.4, 0.6]$
θ_{13} (Normal Ordering) θ_{13} (Inverted Ordering)	Gaussian Gaussian	0.1503 ± 0.0023 0.1510 ± 0.0023
Δm^2_{21}	Fixed	N/A
Δm^2_{32} (Normal Ordering) Δm^2_{32} (Inverted Ordering)	Uniform Uniform	[2.3, 2.7] $\times 10^{-3}$eV^2 [-2.7, -2.3] $\times 10^{-3}$eV^2
ρ	Fixed	N/A
δ_{CP}	Uniform	[-1, 1]

Table 6.3: The distributions that govern the throws of the oscillation parameters, where a fixed parameter is set equal to its central value defined in table 6.2. These values match those used in the DUNE TDR CP-violation analysis detailed in reference [64].

The fits may struggle to navigate the complex, multi-dimensional chi-squared surface and can fall into false minima. To avoid this, the fits are seeded in both mass orderings, in both octants of θ_{23} and, in the case of the CPV fit, at several values of δ_{CP}.

With the above procedure, approximately 1% of throws were found to return a negative sensitivity. This should never happen as, with an extra degree of freedom, the CPV fit should always find a best-fit position that is as good as, or better than, that found by the CPC fit. This suggests that even with the above seeds, the fits are struggling to find the correct minima. The CPV fit was further seeded at the best-fit position of the CPC fit to prevent this.

Following this, a large asymmetry between the sensitivity of the maximally violating CP-phase throws was seen. To confirm that this was still a fitting problem, a 'cheated' seed was added such that the CPV fit was additionally seeded at the true position of the throw and the CPC fit seeded at the same position but with a CPC value of δ_{CP}. The results of the throws now looked correct. To demonstrate the effect of the cheated seed, a rough estimate of the sensitivity was obtained at $\delta_{\text{CP}} = \pi/2$. This value was chosen as here the asymmetry was seen to be the largest. With the cheated seed, the sensitivity estimate at this point changed from $6\sigma^{+4\sigma}_{-1\sigma}$ to $5\sigma^{+1\sigma}_{-1\sigma}$. Such issues were also reported during the DUNE TDR analysis, and were addressed by the use of many seeds.

The median sensitivity and 68% contours calculated from 600 throws at each of 200 δ_{CP} values (20 were used in the analysis of reference [64]) are shown in figure 6.5. Time constraints and grid system changes limited the number of throws that could be made, which results in a lack of smoothness to the sensitivity curves. However, this is not believed to hinder the conclusions that are to be drawn. In order to estimate what one would see with a higher number of throws, a smoothing algorithm was applied, which averaged each sensitivity value by considering its two nearest neighbours. The smoothing algorithm is seen - in figure 6.5 - to be behaving sensibly.

But can we trust these results? In section 9.1, the systematic uncertainties will be included in our sensitivity estimate, and we will compare the attained sensitivity to that in literature. The similarities give us faith that, with the cheated seed, the fits correctly converge. It is understood that the cheated seed cannot be replicated for fits to real data, but it is not within the scope of this PhD to improve the fitting procedure and this is left as future work.

The peak of the smoothed median sensitivity (obtained from the $\delta_{\text{CP}} = \pi/2$ peak) is $5.1\sigma^{+0.4\sigma}_{-0.5\sigma}$ which, when compared to 5.8σ (achieved in the 'simple estimate'), illustrates that the sensitivity is harmed by including the oscillation parameter uncertainties. Why is this? Oscillation parameter uncertainties result in a range of CPC predictions and it is possible that one of these predictions can better match a CPV observation, lessening our ability to reject CP-conservation; this is known as having a 'degeneracy'.

Parameters that are implemented as nuisance parameters in the fit are limited in their ability to influence the prediction. Each nuisance parameter adds a penalty term to equation 6.8, such that it becomes

$$\chi^2 = 2\sum_{i}^{N_{\text{bins}}} \left(E_i - O_i + O_i \ln\frac{O_i}{E_i} \right) + \sum_{j}^{N_{\text{nuisance}}} \left(\frac{\Delta x_j}{\sigma_j} \right)^2, \tag{6.9}$$

where the sum is over the nuisance parameters, Δx_j is the deviation of the j'th nuisance

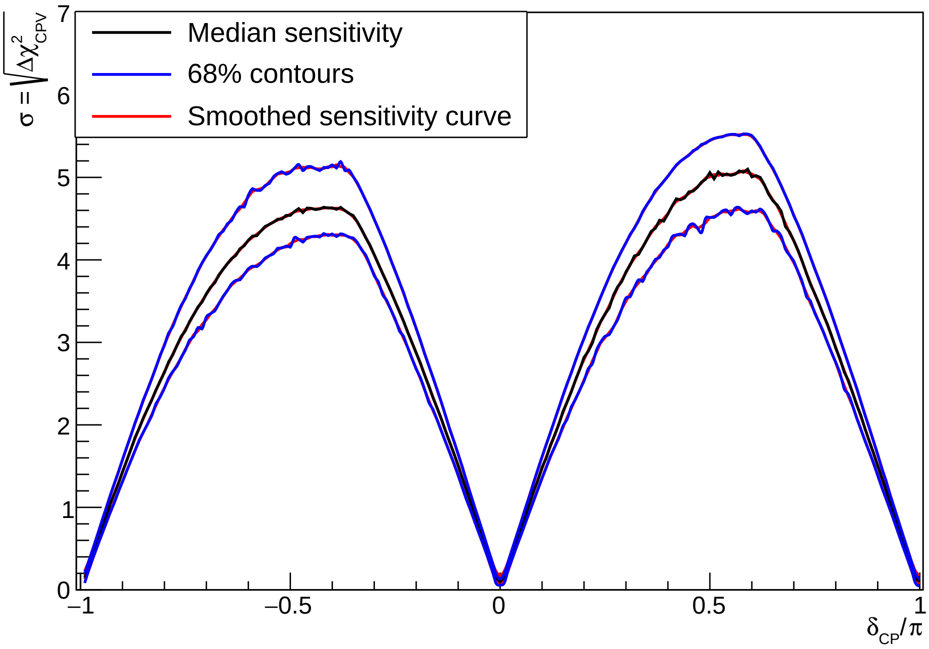


Figure 6.5: The median sensitivity (black line) and 68% contours (blue) achieved - before smoothing - by the Pandora-based CP-violation analysis outlined in this chapter, with a 3.5 year exposure (1.36 $\times 10^{23}$ POT) in each beam mode (FHC and RHC). Oscillation parameter uncertainties are included in the estimate and normal-ordering of the neutrino mass hierarchy is assumed. The smoothed sensitivity curves are shown in red.

parameter from its central value, and σ_j is the corresponding uncertainty given in table 6.2.

Penalty terms increase the chi-squared as the value of the oscillation parameters used to generate the prediction move away from their central values. The chi-squared minima therefore corresponds to a trade-off between a better fit of a ‘shifted from truth’ prediction and the magnitude of the penalty terms preventing the prediction from being shifted ‘indefinitely’ to optimise the match to observation.

So which oscillation parameters are most likely to introduce degeneracies? The dominant terms in equation 6.1 reveal the most problematic oscillation parameters for DUNE to be: θ_{13}, Δm^2_{32} and θ_{23}. Degenerate solutions will decouple with enough statistics and this occurs at low exposures for the mass hierarchy degeneracies due to the extremely long, 1300km baseline of DUNE.

The θ_{13} and θ_{23} octant degeneracies are more persistent. The results from reactor experiments can be used to reduce the uncertainties on θ_{13}, which limits the freedom of the fits and reduces the degeneracies arising from θ_{13} and from θ_{23}, to which it is strongly coupled. Degeneracies can also be resolved by considering data in which the behaviour of the degenerate contribution differs from that of the CP-phase. This is appropriate for the θ_{23} octant degeneracy, where the impact of the octant to the ν_e and $\bar{\nu}_e$ appearance is the same, but the impact of the CP contribution is different.

6.2.4 Sensitivity Estimate with Systematic Uncertainties

To further increase the realism of the sensitivity estimate, systematic uncertainties can be included. The discussion of the impact of the systematics to the sensitivity is addressed in section 9.1.

6.3 The $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ Selection

This section will detail the CC $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ Pandora-based selection procedure used to build the neutrino reconstructed energy spectra. The ultimate focus of the author's resources in this PhD is to optimise an existing CP-violation analysis and as such the aim of this section is not to motivate improvements to these tools.

6.3.1 Overview

The placement of the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection procedure in the CP-violation analysis is illustrated in figure 6.6.

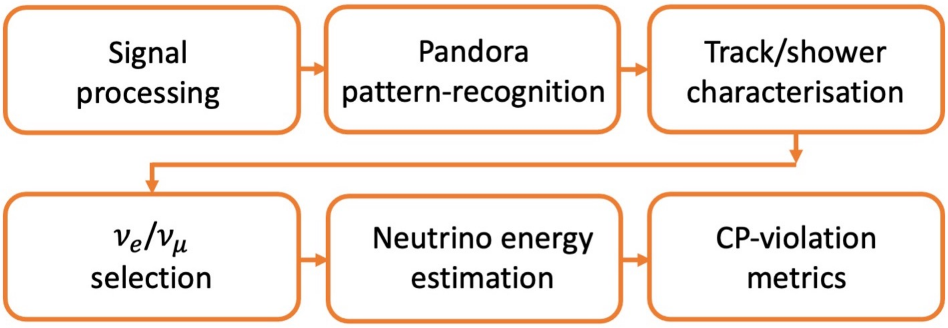


Figure 6.6: The workflow of the Pandora-based CP-violation analysis where the output of one stage is the input to subsequent stages.

The aim of the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection [114] is to select true CC $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ interactions. Here, a true interaction is defined to be one in which a $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ undergoes a CC interaction on liquid argon, i.e. not between detector chambers, and with a true neutrino vertex inside the DUNE FD Fiducial Volume (FV) defined in table 6.4.

The selection procedure is designed to operate on the output of the Pandora reconstruction and events are selected based on the suspected identity of the reconstructed leading leptons, if they exist. The workflow is illustrated in figure 6.7 and a description of each key stage will follow.

Coordinate	Detector edge [cm]		FV distance from edge [cm]	
	lower	upper	lower	upper
x	-360.0	360.0	50.0	50.0
y	-600.0	600.0	50.0	50.0
z	0.0	1394.0	50.0	150.0

Table 6.4: For module I, the distance of the FV boundaries in the x, y and z dimensions from the detector edges. The definition of the FV is inherited from the initial Pandora-based CP-violation analysis [114].

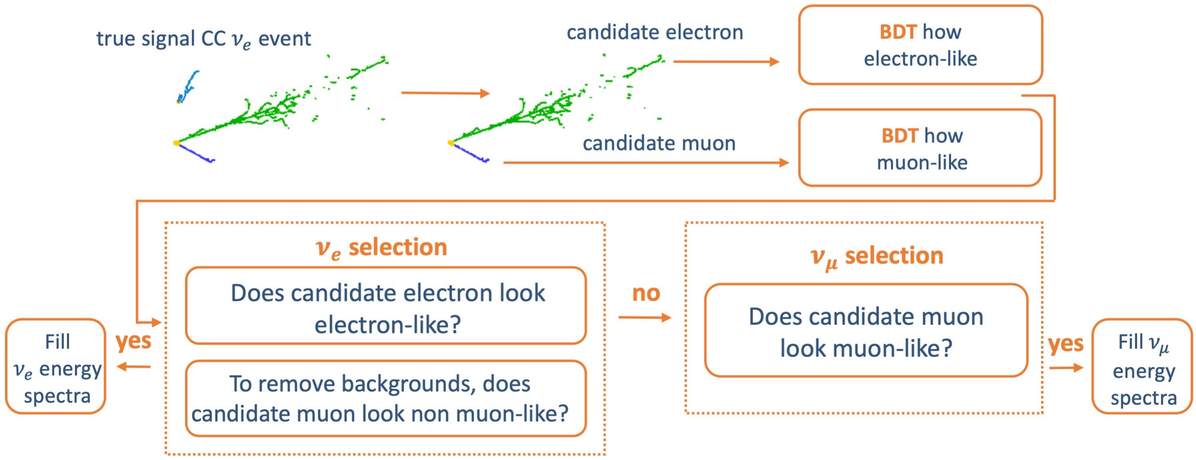


Figure 6.7: A flow chart outlining the Pandora-based selection procedure.

1. Select Leading Leptons

First the candidate electron shower and candidate muon track, should they exist, are identified. The candidate electron is chosen to be the highest energy primary shower in the event and the candidate muon the longest primary track: these are typical choices that work well. The candidate leading leptons are the only objects in the reconstructed event to be considered by the selection.

2. Evaluate the Leading Lepton Likeness

The candidate muon(electron) is passed into a BDT, which determines its muon(electron)-likeness. The muon(electron)-like BDT is discussed in section 6.4 (6.5).

3. Perform $\overset{(-)}{\nu}_e$ Selection

The $\overset{(-)}{\nu}_e$ selection is performed. This selection immediately rejects events if no candidate electron is found and selects events if:

- the electron candidate is sufficiently electron-like,
- the candidate muon is not sufficiently muon-like, and

- the reconstructed neutrino vertex passes the fiducial cut defined in table 6.4.

A muon(electron) candidate is said to be ‘sufficiently muon(electron)-like’ if its muon(electron)-like score exceeds a predetermined threshold. The determination of this threshold is detailed in subsection 6.3.2.

The first requirement of the selection criteria is motivated by the extremely likely presence of a primary (anti)electron in CC $\overset{(-)}{\nu}_e$ interactions. With only this requirement, background events with genuine or mistaken electron showers may be selected, hindering the selection purity. In an attempt to reduce the selected background, the second condition is introduced to reject CC $\overset{(-)}{\nu}_\mu$ events with primary (anti)muons. Selected events are used to construct the $\overset{(-)}{\nu}_e$ reconstructed spectrum.

4. Perform $\overset{(-)}{\nu}_\mu$ Selection

If an event is rejected by the $\overset{(-)}{\nu}_e$ selection, it is considered by the $\overset{(-)}{\nu}_\mu$ selection. This selection runs second to maximise the efficiency of the $\overset{(-)}{\nu}_e$ selection, as it is typically much harder to select $\overset{(-)}{\nu}_e$ events. Events are immediately rejected if no candidate muon was found and are selected if:

- the muon candidate is sufficiently muon-like, and
- the reconstructed neutrino vertex passes the fiducial cut defined in table 6.4.

Selected events are used to construct the $\overset{(-)}{\nu}_\mu$ reconstructed spectrum.

6.3.2 Tuning the Selection

The $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection thresholds are tuned such that the selection performance, quantified by

$$\text{selection performance} = \text{selection efficiency} \times \text{selection purity}, \tag{6.10}$$

is maximised. Here, the selection efficiency is defined to be the fraction of signal events that are selected and the selection purity is defined as the fraction of selected events that are signal. More detailed definitions of these quantities are given in section 6.6.

6.3.3 Initial Modifications

In order to integrate the inherited $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection into the CP-violation analysis, a number of initial adjustments were motivated and applied.

Muon/Electron-like Threshold Determination

The choice of the muon/electron-like thresholds used in the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection depends on one’s motivations and in a CP-violation analysis it is most appropriate to tune the selection thresholds with respect to the analysis goal i.e. to optimise the sensitivity to

CP-violation. The selection thresholds were therefore chosen to be tuned with respect to the highest sensitivity achieved across the δ_{CP} phase space.

There are three thresholds used in the selection procedure, and since the output of the $\overset{(-)}{\nu}_e$ selection provides the input to the $\overset{(-)}{\nu}_\mu$ selection, they would ideally be tuned simultaneously. This requires time and computational resources far exceeding those available to the author and so several simplifications are made:

1. The simplest estimate of the sensitivity is considered; that of subsection 6.2.2.

2. Since the $\overset{(-)}{\nu}_e$ contribution dominates the sensitivity, tune the $\overset{(-)}{\nu}_e$ selection to optimise the $\overset{(-)}{\nu}_e$ contribution to the sensitivity and the $\overset{(-)}{\nu}_\mu$ selection to maximise the selection metric of equation 6.10. This reduced the threshold phase space from 3D to 2D.

3. Calculate the oscillation weights once per test δ_{CP} value. For this, a 4D histogram is built with axes: electron-like score, muon-like score, total weight and reconstructed $\overset{(-)}{\nu}_e$ energy. For each point in the 2D threshold phase space, the histogram can be truncated in the electron-like score and muon-like score axes and project onto the $\overset{(-)}{\nu}_e$ reconstructed energy spectra axis such that the sensitivity can be obtained.

With these simplifications, the sensitivity, as a function of the CP-phase, is first sampled across the 2D threshold phase space. The electron-like and muon-like thresholds used in the $\overset{(-)}{\nu}_e$ selection are those that maximise the peak value of the sensitivity. The muon-like threshold used in the $\overset{(-)}{\nu}_\mu$ selection is then found and is equal to that which maximises the selection performance as defined by equation 6.10.

Candidate Leading Lepton Selection

In the inherited selection the primary track with the longest length is assumed to be the leading muon. With this assumption, it was thought that in $\overset{(-)}{\nu}_\mu$ interactions with broken or genuinely short muon tracks the candidate muon identification would be incorrect. This would likely result in the rejection of such events in the $\overset{(-)}{\nu}_\mu$ selection.

An alternative, of choosing the candidate muon to be the primary track with the highest muon-like score, was investigated. However, it is seen in table 6.5 that the new approach was found to be only slightly better than the standard choice.

The candidate electron/muon choice is the foundation of the selection and with the foresight that future optimisation work would focus on improvements to the electron/muon-like BDTs, it was decided that the candidate electron/muon would be chosen to be the most electron/muon-like primary shower/track. This ensured that the reliance on the BDTs is maximised and that knock-on improvements to the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection are achieved.

Assumption	Correct Candidate Electron	Correct Candidate Muon
Highest energy shower/longest track	86.5%	91.1%
Most electron/muon-like	88.1%	92.9%

Table 6.5: The percentage of correctly identified candidate electrons/muons in signal CC ν_e events.

6.4 The Muon-like BDT

The muon-like BDT takes, as an input, reconstructed particles that have been classified as track-like by the Pandora pattern-recognition software (see subsection 4.3.1). The signal is muon particles, and the background is track-like particles (such as pions), however, the background will also include showers (such as low-energy electrons) that have been incorrectly classified as track-like by the Pandora pattern-recognition software.

This section will begin with an overview of the muon-like BDT variables, which will be ordered by their discriminating power and, for clarity, will be grouped with respect to the type of muon characteristic they examine. Following this, the BDT variable correlations will be detailed and the performance of the BDT will be characterised. The results of this section are obtained from a sample of approximately 20,000 neutrino events of each file type, detailed in table 6.1.

6.4.1 BDT Variables

Topological Variables

The topological variables are motivated by the typically longer and straighter trajectories of muon tracks compared to those of the background. They probe a feature that differs in muons and charged pions which is important as charged pions are responsible for the selection of the dominant background in the $\overset{(-)}{\nu}_\mu$ selection (see subsection 6.6.2). The topological BDT variables are:

1. **Track Length (T1):** the distance between the track start and end point.

2. **Deviation from Straightness (T2):** the estimated standard deviation of the angular deviations along the fitted track calculated using the same method as MicroBooNE, detailed in reference [115].

Figure 6.8 presents the topological variable distributions of the events from which the BDT learns. The upper bin, is an overflow bin, which contains the tracks that lie beyond the histogram axis. All training events must be contained within these histograms, and so, if it is possible that a variable cannot be evaluated for some tracks (as is the case for

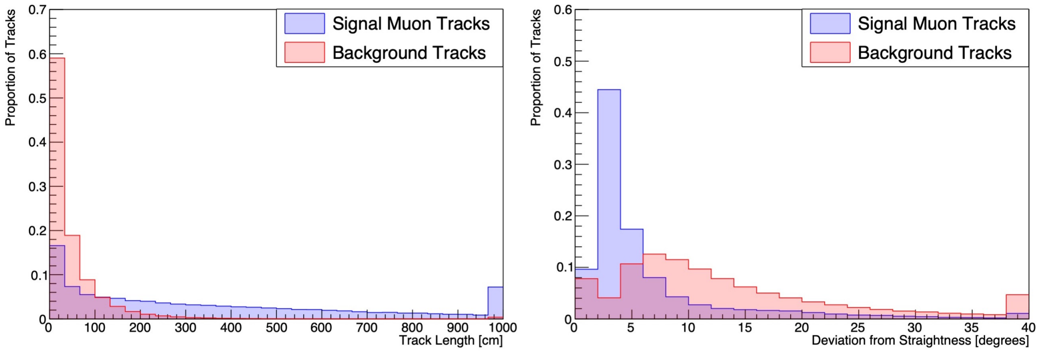


Figure 6.8: The signal (blue) and background (red) track length (left) and deviation from straightness (right) distributions.

the Michel electron variables), an underflow bin is created to hold these events, which bounds the negative region. The topological BDT variables are seen to provide excellent signal-background separation and dominate the muon-like BDT. This has the consequence that shorter, deviating muon tracks will obtain poor muon-like scores, which explains the lack of improvements seen when the candidate muon choice was replaced in subsection 6.3.3.

Calorimetric Variables

The calorimetric variables are motivated by the well defined and somewhat unique calorimetric behaviour of track-like particles, demonstrated in the dE/dx distribution of figure 6.9. The calorimetric BDT variables are:

1. **Initial dE/dx (C1):** the dE/dx of the initial 5% of the track.
2. **End dE/dx (C2):** the dE/dx of the final 10% of the track.
3. **End Region dE/dx Ratio (C3):** the ratio of the dE/dx of the end region and the penultimate region, where the penultimate region is defined to be the section between the 80% and 90% track length boundaries and the end region is the remaining 10% of the track.

As seen in figure 6.10 the calorimetric BDT variables provide good signal-background separation. The signal/background distributions are somewhat smeared as a result of reconstruction errors, such as track merges, which can throw off the, already difficult dE/dx calculation.

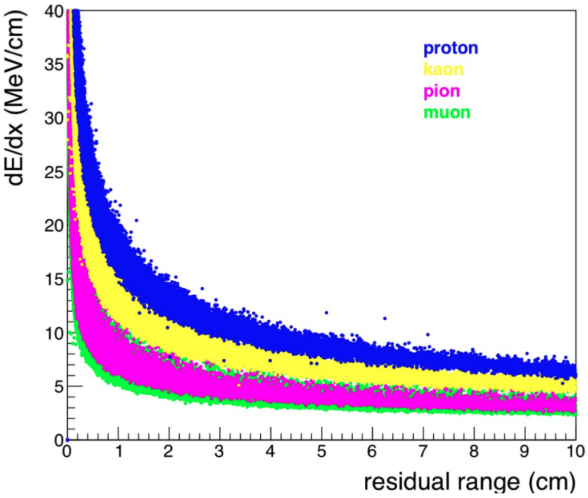


Figure 6.9: Theoretical dE/dx distributions, as a function of the residual range in liquid argon, obtained using the Geant4 simulation. Figure reproduced from reference [116].

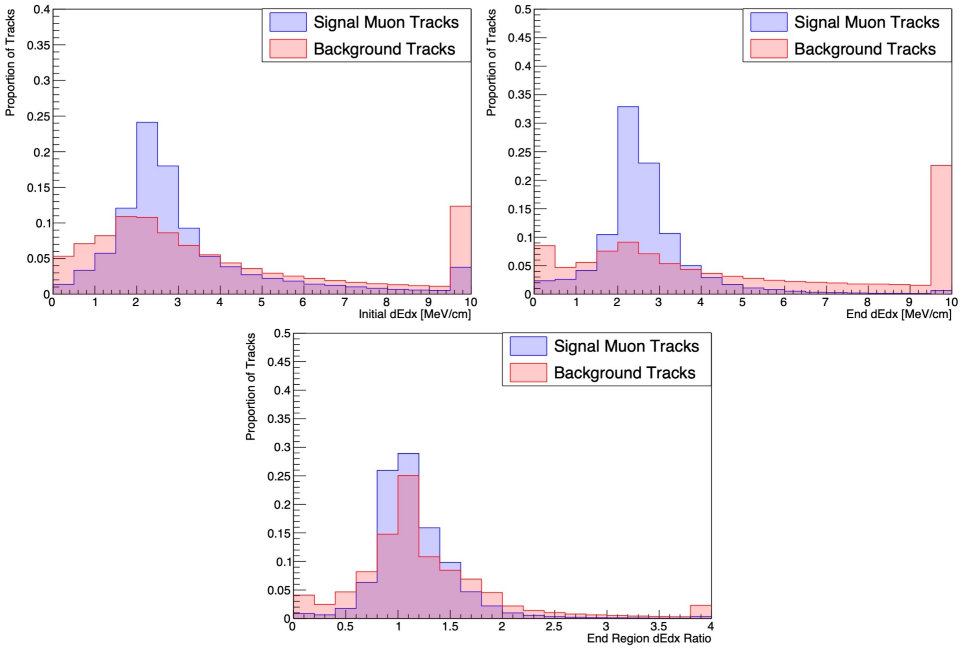


Figure 6.10: The signal (blue) and background (red) initial dE/dx (top-left), end dE/dx (top-right) and end region dE/dx ratio (bottom) distributions.

Michel Electron Variables

Stopping muons can decay to produce a Michel electron which, because the muon decays at rest, have a unique energy distribution. Several variables have been created which characterise the children of the primary tracks in order to exploit this unique signature. The Michel electron BDT variables are:

1. **Number of 2D Hits (ME1):** the total number of 2D hits of the Michel candidate.

2. **Electron-like MVA score (ME2):** the electron-like score of the Michel candidate

given by the Warwick MVA [117].

3. **Reconstructed Energy (ME3):** the total reconstructed energy of the Michel candidate on the collection plane.

A Michel candidate is defined to be any child of the primary track that has a vertex within 4cm of its endpoint; it is not required to be shower-like. Not all muons stop in the detector nor do all stopping muons decay to Michel electrons and in these cases, the Michel electron BDT variables are given default values, which accounts for the large underflow peaks present in figure 6.11. As seen, there are more background events with candidate Michels than there are signal events, and so it is believed that the Michel electron BDT variables are actually learning the non-electron nature of the hadronic system of background events. Some separation power is thought to be achieved albeit not through the nature envisaged.

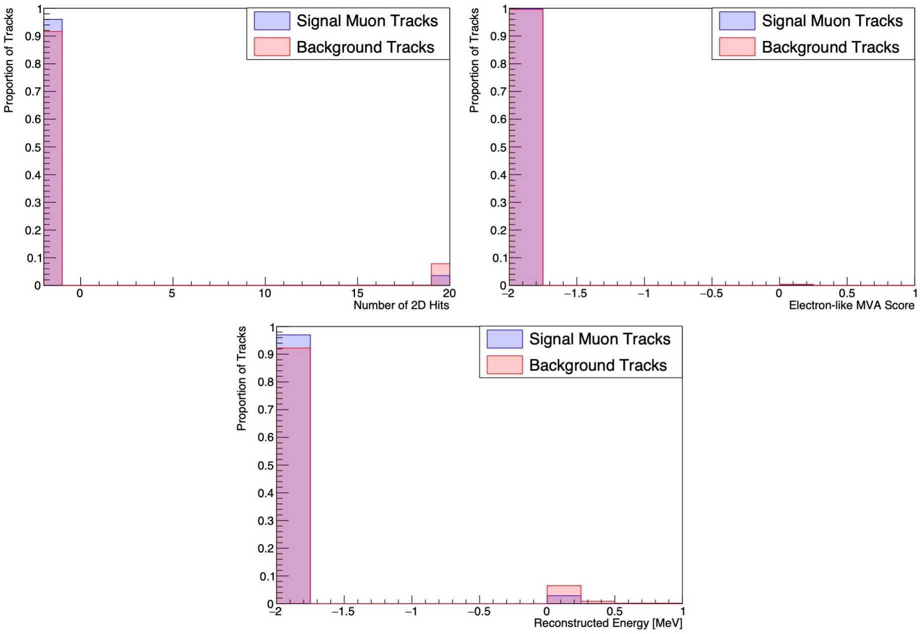


Figure 6.11: The signal (blue) and background (red) number of 2D hits (top-left), electron-like MVA score (top-right) and reconstructed energy (bottom) distributions.

Track/Shower BDT Variables

The muon-like BDT is an improvement on an older, much more general, particle identification BDT developed at the University of Warwick [117, 118] and has inherited several BDT variables that are intended to provide separation between tracks and showers. However, since the muon-like BDT uses the Pandora track/shower classification score to perform track/shower separation at the input level, these variables are somewhat redundant. The track/shower variables, in this BDT, probe the differences in the track-like behaviour of the signal muons and background track-like particles. This may result in some separating power, as some background tracks (such as charged pions) are expected to be curvier, with sparse halos that contribute to a larger transverse extent. The track/shower BDT variables are:

1. **Concentration (TS1):** the Coloumb potential energy of a track about its trajectory, where the trajectory is estimated to be the best-fit axis of a linear fit of the 3D track hits. It is calculated by

   $$\text{concentration} = \frac{\sum_{i=0}^{n} \frac{q_i}{r_i}}{\sum_{i=0}^{n} q_i}, \tag{6.11}$$

   where i is the enumeration of the track hits, n is the number of hits in the track, q_i is the charge of the i'th hit and r_i is the shortest distance between the i'th hit and the track trajectory and is restricted to be no less than 0.01cm. Here, and in all of the track/shower variable calculations, the charge corresponds to an ADC count.

2. **Halo-Core Ratio (TS2):** an examination of the charge distribution of the particle given by

   $$\text{Halo-Core ratio} = \frac{\text{total charge in halo}}{\text{total charge in core}}, \tag{6.12}$$

   where the core(halo) is defined to be the region inside(outside) a cylinder that is centred on the trajectory of the track with a radius of 20% of the molière radius in liquid argon.

3. **Conicalness (TS3):** a comparison of the charge-weighted transverse extent at the start and end of the particle trajectory. The inherited implementation, upon close inspection, proved to be rather ill defined. The 'charge-weighted transverse extent' is calculated to be

   $$\text{`charge-weighted transverse extent'} = \frac{\sqrt{\sum_{i=0}^{n} q_i r_i^2}}{\sum_{i=0}^{n} q_i}, \tag{6.13}$$

   which has dimensions of length/$\sqrt{\text{charge}}$. If the square-root was removed in this equation one could view this as an estimate of the charge-weighted mean cross-sectional area, on the other hand if the q_i term in the numerator was squared one could view this as a normalised quadrature sum of the charge-weighted radial distance of the track hits. It is argued that the exact definition is unimportant and that the ratio

   $$\frac{\text{end `charge-weighted transverse extent'}}{\text{start `charge-weighted transverse extent'}} \tag{6.14}$$

   sufficiently encapsulates the intended physical phenomena.

4. **Eigenvalue Ratio (TS4):** an examination of the spatial extent of a particle in terms of the eigenvalues obtained from a PCA of the particle's 3D hits. The eigen-

value ratio is given by:

$$\text{eigenvalue ratio} = \frac{\sqrt{e_2^2 + e_3^2}}{e_1}, \tag{6.15}$$

where e_1 is the eigenvalue of the eigenvector that points in the direction that has the largest extent, and e_2 and e_3 are the eigenvalues of the remaining eigenvectors. This variable is therefore a measure of the elongation of the 3D ellipsoidal shape.

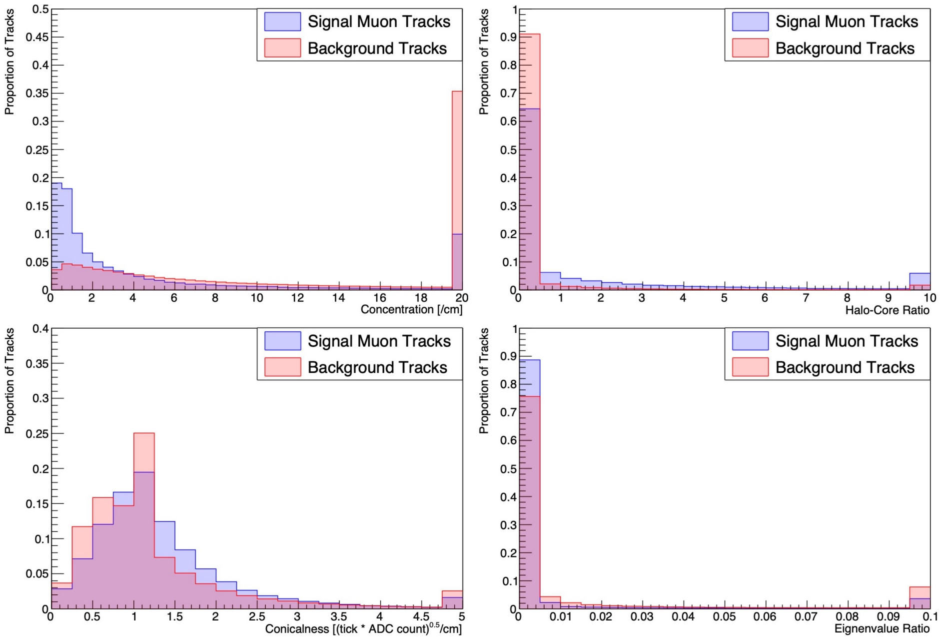


Figure 6.12: The signal (blue) and background (red) concentration (top-left), halo-core ratio (top-right), conicalness (bottom-left) and eigenvalue ratio (bottom-right) distributions.

Figure 6.12 demonstrates evidence that some signal-background track separation has been achieved. This is seen in the eigenvalue ratio signal-background distribution where the background creeps into the higher eigenvalue ratio region: a consequence of the aforementioned sparse halo and a significantly shorter lengths of the background.

However, as result of a simplistic parameterisation of the track trajectory, the potential signal-background track separation power is in general not achieved. The track trajectory is estimated by the best-fit axis of a linear fit to the 3D track hits. Since muon tracks are long and not exactly straight, global fits fail to follow the true trajectory of the track such that the hits are ‘seen’ in the above calculations to lie some distance away from the trajectory resembling the halo characteristic of the background. This could be solved by performing a local linear fit of segments of the track.

6.4.2 BDT Variable Correlations

Correlation matrices illustrate how linked the variables used in a BDT are to one another. Variables that are highly (anti-)correlated probe the same underlying phenomena, which can bias the BDT as it gives more weight to the probed characteristic.

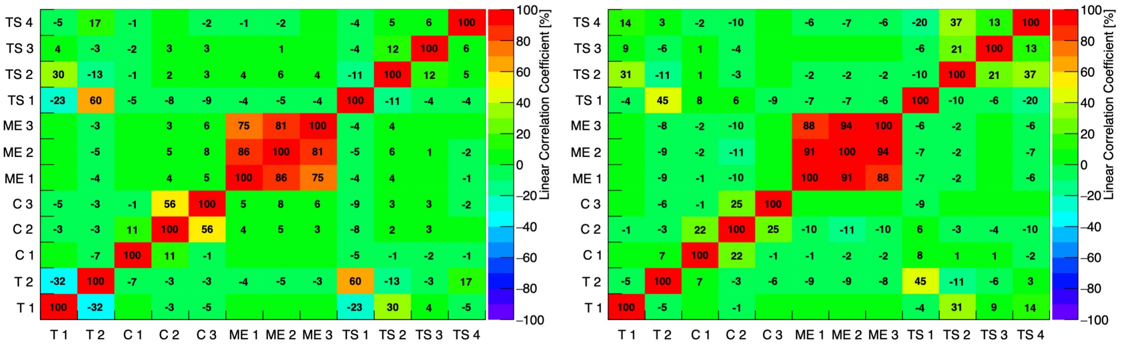


Figure 6.13: The signal (left) and background (right) correlation matrices of the muon-like BDT. The extent of correlation is indicated by the number and colour of the entry, where higher numbers and a red(purple) fill corresponds to high (anti-)correlation.

The most notable correlations seen in the signal and background correlation matrices of figure 6.13 are:

- Between the Michel electron variables. The Michel variables can only be calculated if a Michel candidate exists, which is only true in the minority of cases. When a candidate is not found, these variables are set to a default value which introduces a shared dependence on a common condition.

- Between the concentration and the straightness deviation variables. When the deviation from straightness is high, a linear fit becomes less appropriate and the track hits are interpreted to lie some distance away from the fitted trajectory which results in an artificially large concentration.

- Between the end dE/dx and the end region dE/dx ratio. This correlation is seen only in the variables of the signal muons which, since Michel decays represent a small proportion of muon final states, can be assumed to relate to stopping and through-going muons. Figure 6.9 demonstrates that the end dE/dx and end region dE/dx ratio are both highly dependent on the residual range and thus are highly correlated.

The removal of each correlated variable was investigated and in each case, performance slightly lessened. As it is not the aim of this thesis to fine-tune the already existing selection tools, it was decided to leave the correlated variables.

6.4.3 BDT Performance

The direct performance of the muon-like BDT can be understood by the signal/background classification distribution and its achieved Receiver Operating Characteristic (ROC) curve shown in figure 6.14. The ROC curve shows the selection signal efficiency vs background rejection, and is built by sliding a cut across the classification distribution such that tracks with scores exceeding the cut are identified as muon-like. If the threshold is set to the maximum muon-like value, no tracks are classified as muon-like and the classification purity is maximised whilst the efficiency is minimised. As the threshold decreases, the efficiency increases at the cost of an lessened purity until eventually the purity is minimised whilst efficiency is maximised.

In general the aim is to have narrowly peaked signal and background distributions with small, short tails such that they are highly separated. This results in a rapid increase in efficiency as the threshold is lowered and purity degradation only occurs when the threshold reaches its minimum, this would correspond to a ROC curve with an apex as close to the top right hand side of figure 6.14.

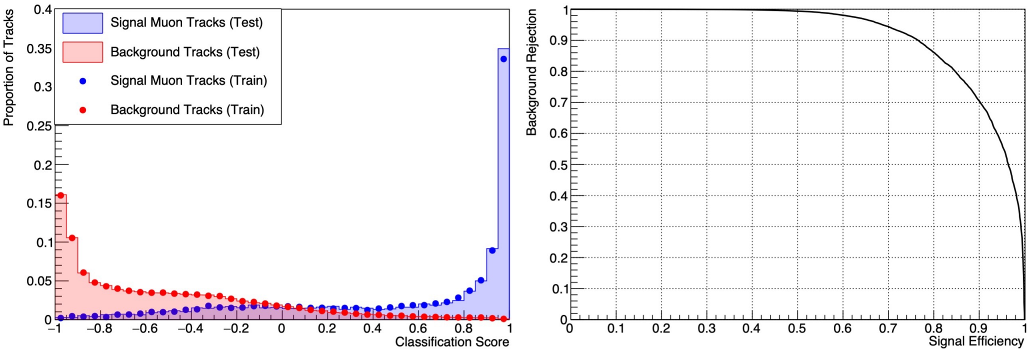


Figure 6.14: Left: the signal (blue) and background (red) classification distributions of the training (dotted) and test (fill) MC data sets. Right: the ROC curve.

In figure 6.14, the signal is seen to have a high and narrow peak in the extreme signal region but a long tail containing a significant number of tracks in the background region. Comparatively, the background distribution has a much smaller peak with more events in the tail, where the tail contamination in the extreme signal region is negligible. This is demonstrated in figure 6.14, where an initial high purity is maintained over a significant range of efficiency values reaching a good efficiency before rapidly decreasing. The performance of the BDT can be quantified by the area under the ROC curve, which is found to be 0.92. This supports the statement that the muon-like BDT is understood to perform well.

Additionally, the training and validation signal/background classification distributions can be compared to check for BDT over-training. Over-training means that the BDT has ‘remembered’ the specific training sample and not learnt the general signal/background characteristics. This will lead to very different classification distributions of the training and validation set. Over-training of muon-like BDT can be ruled out as, in figure 6.14, the training/validation classification distributions are seen to agree.

6.5 The Electron-like BDT

The electron-like BDT takes, as an input, reconstructed particles that have been classified as shower-like by the Pandora pattern-recognition software (see subsection 4.3.1). Electron showers constitute the signal, whilst the background is comprised of photons (87%) and hadronic particles that have been incorrectly classified as shower-like by the Pandora pattern-recognition software. This can happen if the hadronic particle has a moderate halo around its main trajectory or, if it has been merged together with a descendent shower-like particle, e.g. in pion decay. This will be of importance in the discussion of the track/shower BDT variables in subsection 6.5.1.

The layout of this section will mirror the muon-like BDT section.

6.5.1 BDT Variables

The biggest challenge in the electron-like BDT is to distinguish electron from photon showers. The success of this is incredibly important as photon showers are responsible for the selection of the dominant background of the $\overset{(-)}{\nu}_e$ selection: NC interactions (see subsection 6.6.1). Consequently, the majority of the BDT variables examine characteristics that are expected to differ between electrons and photons. As future work will focus on improving the electron/photon separation, these variables will be discussed in detail.

1. Displacement (D)

Electrons(photons) are electrically charged(neutral) particles and therefore (do not) ionise the liquid argon of the detector. Electrons are thus seen to originate from the neutrino interaction vertex whereas photon initiated showers are only seen when the photon converts into an electron-positron pair (the dominant photon interaction in the energy range of DUNE). The displacement BDT variable characterises this and is equal to the distance between the neutrino and shower vertices, identified during the event reconstruction.

It is seen in figure 6.15 that a high separation power is achieved but that the signal-background distributions are smeared: this is caused by reconstruction errors. On the side of the signal, the region around the neutrino vertex can often be very dense and so the initial trajectory of the signal electron may overlap with those of other particles (see left panel of figure 6.16). This means that an ‘electron hit’ may also have charge contributions from other particles making the true owner of the hit ambiguous. In the reconstruction, the ambiguous hits may not be clustered into the electron, truncating the initial region of the electron and resulting in a photon-like gap. Furthermore, electron

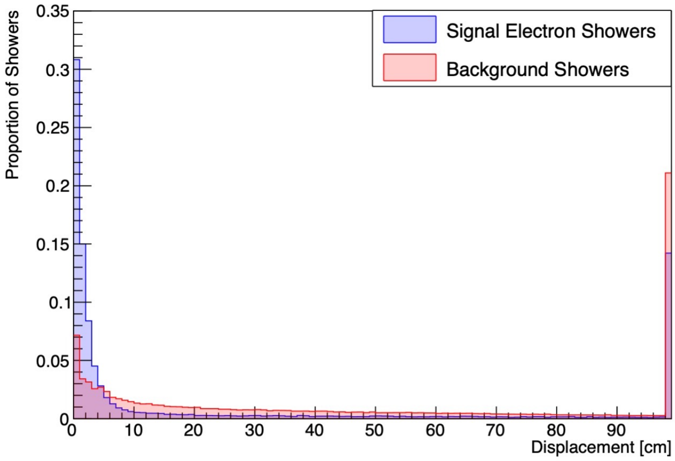


Figure 6.15: The signal (blue) and background (red) displacement distribution.

showers can be split in the reconstruction such that their initial track-like region and subsequent shower-region are separated (see right panel of figure 6.16). This also results in a non-zero displacement.

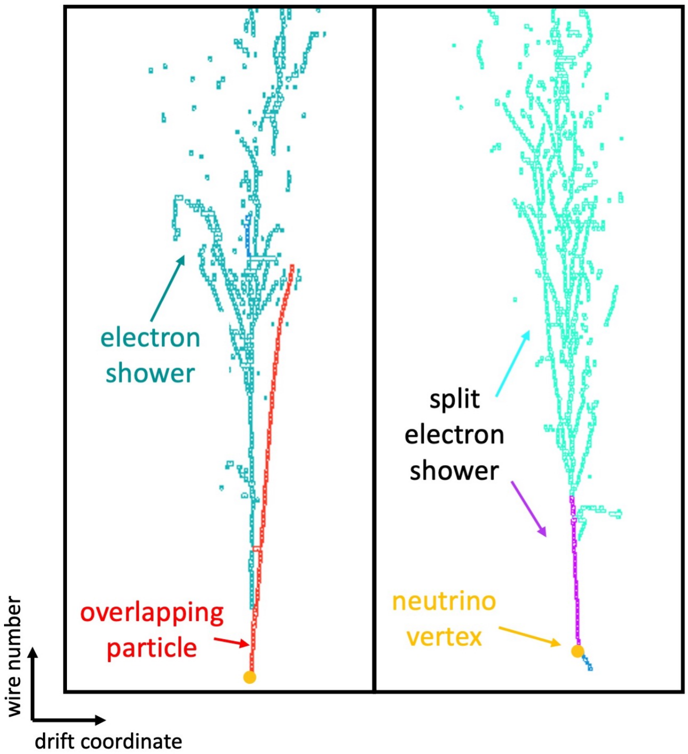


Figure 6.16: Examples of CC ν_e events in which the primary electron trajectory overlaps with that of another particle in the event (left), and the primary electron is split into its track-like and shower-like regions (right).

Photon trajectories can also coincide with those of other particles, such that the the photon shower may look as though it is an extension of the coincident particle. In the reconstruction, the two may be merged together and the photon shower will be seen to come directly from the neutrino vertex, alike an electron (see right panel of figure 6.17).

It is also true, as seen in figure 6.18, that a significant number of photons convert very close to the neutrino vertex and may, within the resolution of the hit reconstruction and the wire spacing, have a ‘gap’ that is consistent with zero (see left panel of figure 6.17).

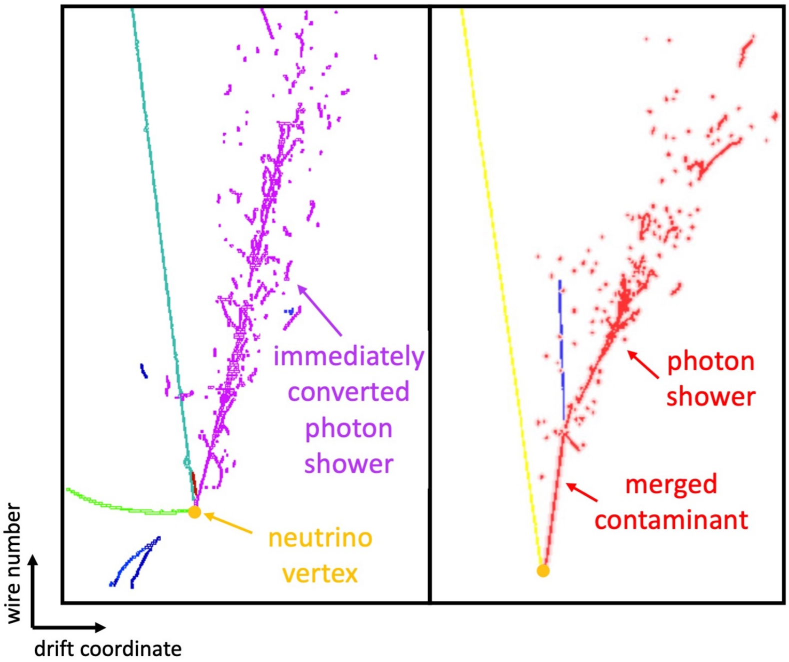


Figure 6.17: Examples of CC ν_μ events in which a primary photon immediately showers (left), and a contaminant track-like particle is merged into a primary photon-induced shower such that the latter looks to originate from the neutrino vertex (right).

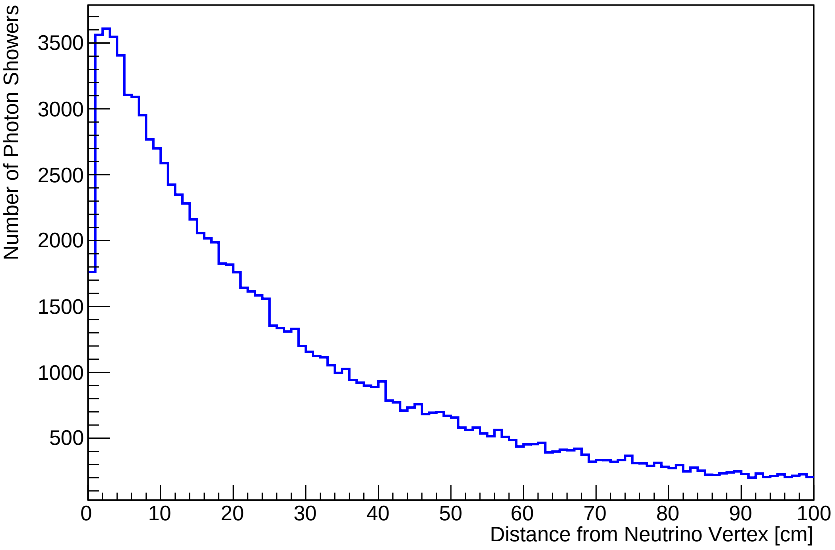


Figure 6.18: The distance between the true neutrino interaction point and the position of the first energy deposit that, in truth, belongs to the photon-induced shower. This distribution can be interpreted as the photon conversion distribution.

2. Initial dE/dx

The dE/dx of the initial region of a photon shower has contributions from the electron-positron pair created when the photon converts, and is therefore expected to be twice that of the initial region of an electron shower. This motivates the initial dE/dx BDT variable which is calculated from the energy deposits of the collection plane.

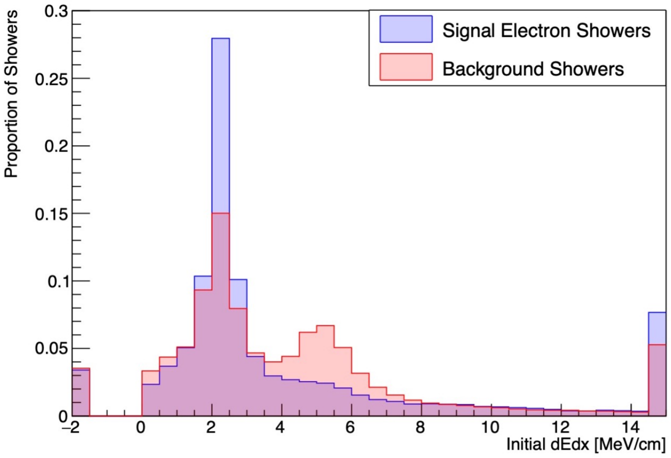


Figure 6.19: The signal (blue) and background (red) initial dE/dx distribution.

The achieved signal-background separation is shown in figure 6.19 where smearing of the signal/background distributions is once again observed. In general both dE/dx distributions are smeared by contaminants being merged into the initial shower region and the signal distribution is further smeared by ambiguous hits having contaminant charge contributions. Electron truncation and immediate electron showering can result in a dE/dx that resembles that expected of a photon. On the other hand, the dE/dx of photons can resemble that expected of electrons if a single-electron shower is instead induced by Compton scattering or, in the case of pair-production, if the opening angle between the electron-positron pair is large enough that each electron is reconstructed as a separate shower.

3. Distance of Closest Approach (DCA)

Photon-induced showers are seen in the detector only after a secondary process. The photon may be scattered in the interim and the observed shower may pick up an angular deviation from the photon trajectory in the secondary process. The observed shower would therefore look as though it did not originate from the neutrino vertex.

One can use the shower direction to trace back the shower trajectory from the shower start position obtaining the Distance of Closest Approach (DCA) as the perpendicular distance between this extrapolation and the neutrino vertex.

In general, the expected behaviour is seen in figure 6.20 but there is a significant presence of the background in the signal region. This is likely to arise from photon

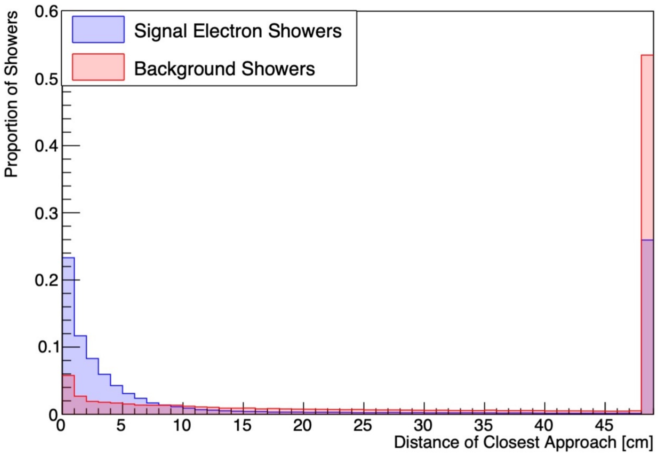


Figure 6.20: The signal (blue) and background (red) DCA distribution.

showers that convert early. Reconstruction errors can also distort this estimator; for example, incorrect particle merges can skew the fitted direction and consequently the DCA, which is believed to account for the signal events in the background region.

4. Wideness (W) and Energy Density (ED)

The physical development of electron and photon showers are equivalent and only differs by the start position in the pair-production/Bremsstrahlung chain. It is difficult to predict if this could have an effect on the reconstructed shape and energy distribution of the showers. This is probed by the wideness and energy density variables where the wideness is equal to the opening angle of the shower divided by its length and the energy density is the total energy of the shower divided by the volume.

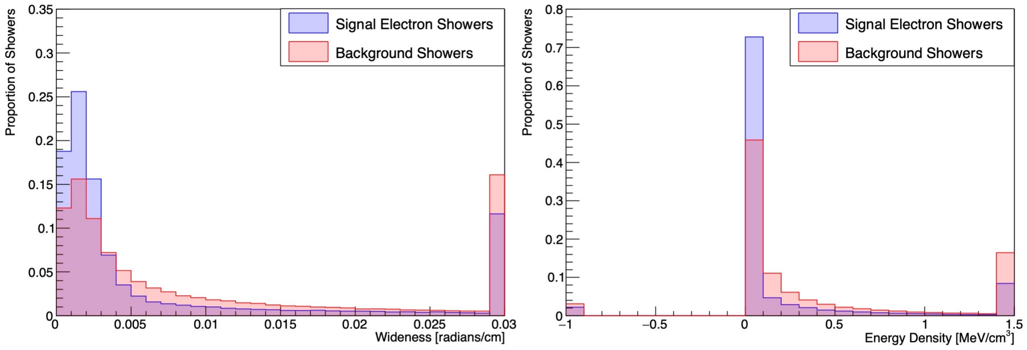


Figure 6.21: The signal (blue) and background (red) wideness (left) and energy density (right) distributions.

The signal-background distributions are presented in figure 6.21 where it is found that the dominant photon background showers widen more per unit length than the electron signal and that they have a higher energy density.

5. Track/Shower BDT Variables

The track/shower variables detailed in subsection 6.4.1 were also inherited by the electron-like BDT. These variables examine the shower-like behaviour of the signal electron and background showers.

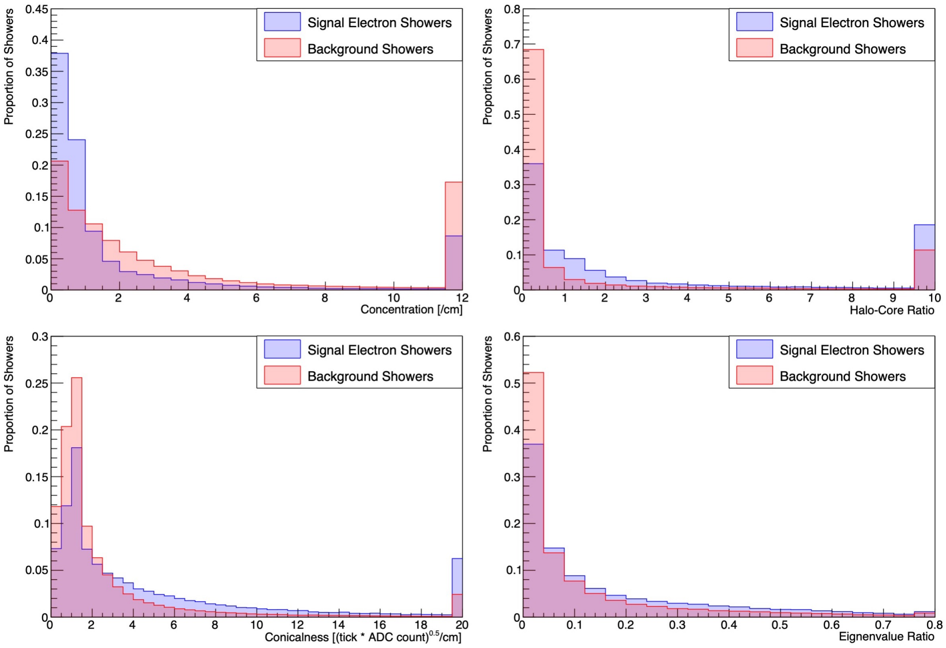


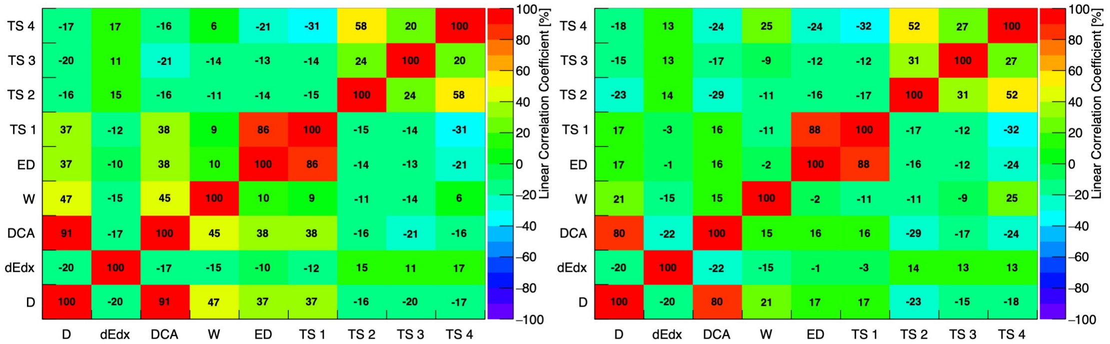
Figure 6.22: The signal (blue) and background (red) concentration (top-left), halo-core ratio (top-right), conicalness (bottom-left) and eigenvalue ratio (bottom-right) distributions.

Figure 6.22 demonstrates the separation of a significant track-like background from the signal. It is remembered that the background is composed of all reconstructed particles that are classified as showers by Pandora i.e. not only those that belong to photons. It is concluded that the track/shower variables provide separation between electrons and hadrons but contribute little to electron/photon separation.

6.5.2 BDT Variable Correlations

The most notable correlations seen in the signal and background correlation matrices of figure 6.23 are:

- Between the energy density and the concentration, which were found to depend linearly on one another. The calculation of concentration (equation 6.11) can be interpreted to be the charge-weighted average of the inverse of the radial distance of a shower hit from the shower core and so, a smaller concentration represents a wider transverse spread of the shower's energy i.e. a lower energy density. Therefore, both variables probe the transverse energy distribution of the shower and a linear relationship between the two is expected.



	D	dEdx	DCA	W	ED	TS 1	TS 2	TS 3	TS 4
TS 4	-17	17	-16	6	-21	-31	58	20	100
TS 3	-20	11	-21	-14	-13	-14	24	100	20
TS 2	-16	15	-16	-11	-14	-15	100	24	58
TS 1	37	-12	38	9	86	100	-15	-14	-31
ED	37	-10	38	10	100	86	-14	-13	-21
W	47	-15	45	100	10	9	-11	-14	6
DCA	91	-17	100	45	38	38	-16	-21	-16
dEdx	-20	100	-17	-15	-10	-12	15	11	17
D	100	-20	91	47	37	37	-16	-20	-17

	D	dEdx	DCA	W	ED	TS 1	TS 2	TS 3	TS 4
TS 4	-18	13	-24	25	-24	-32	52	27	100
TS 3	-15	13	-17	-9	-12	-12	31	100	27
TS 2	-23	14	-29	-11	-16	-17	100	31	52
TS 1	17	-3	16	-11	88	100	-17	-12	-32
ED	17	-1	16	-2	100	88	-16	-12	-24
W	21	-15	15	100	-2	-11	-11	-9	25
DCA	80	-22	100	15	16	16	-29	-17	-24
dEdx	-20	100	-22	-15	-1	-3	14	13	13
D	100	-20	80	21	17	17	-23	-15	-18

Figure 6.23: The signal (left) and background (right) correlation matrices of the electron-like BDT distributions. The extent of correlation is indicated by the number and colour of the entry, where higher numbers and a red(purple) fill corresponds to the high (anti-)correlation.

- Between the displacement and the DCA. If the displacement is large then the distance over which deviations in the shower's direction can occur is lengthened, increasing the value of the DCA. Of course if the displacement is minimal then so too will be the value of the DCA.

6.5.3 BDT Performance

As with the muon-like BDT, the performance of the electron-like BDT can be understood in terms of the signal/background classification distribution and by the achieved ROC curve. These are shown in figure 6.24.

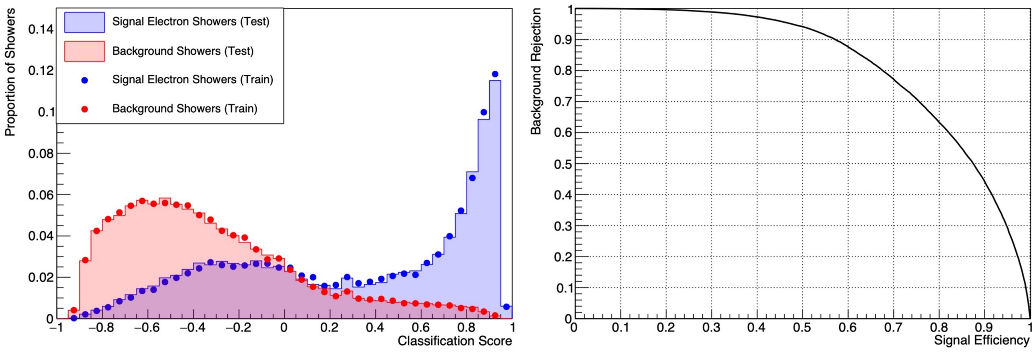


Figure 6.24: Left: the signal (blue) and background (red) classification distributions of the training (dotted) and test (fill) MC data sets. Right: the ROC curve.

In figure 6.24, the signal distribution peaks where expected, with little background contamination, but is seen to be quite broad. This is a result of the previously discussed reconstruction errors, which reduce the resemblance of the reconstructed electrons to the true signal profile. This in turn distorts the signal profile learnt by the BDT and reduces the number of electrons that match the distorted profile: a two-fold reduction of the signal peak.

The background has a much broader peak suggesting that the background profile is less clear than the signal profile. This is because photons often resemble electrons and the characterisation of photons is less discrete than that of electrons i.e. photons can shower immediately or after some distance and their initiation occurs via several different processes. Furthermore, the background also includes hadronic track-like showers and miss-reconstructed tracks that have shower contaminants merged into them, and these will have different characteristics to the photon showers further smearing the background profile.

Again it is seen that, the training and validation signal/background classification distributions agree well, ruling out BDT over-training.

In comparison to the muon-like BDT, the performance of the electron-like BDT is much lower. This is seen by the area under the ROC curve, which is 0.82(0.92) in the electron(muon)-like BDT case. The reason for the lesser performance is simply because showers have a more complicated structure, which makes them harder to reconstruct, and have a very prominent background that closely resembles the signal.

6.6 $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ Selection Performance

With an understanding of the operation of the selection procedure and of the underlying electron and muon-like BDTs, the performance of the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection can now be quantified. In the following, only the FHC selection is examined - the performance of the RHC selection is expected to be the same. The presented results are obtained from a sample of approximately 50,000 neutrino events of each file type, detailed in table 6.1:

- 20% of the events are used to train the electron and muon-like BDTs,
- 20% of the events are used to validate the training of the electron and muon-like BDTs, and
- the remaining events are used to create a FD fake data prediction, which corresponds to a 3.5 year exposure (1.36×10^{23} POT) in FHC mode, and the NO values of the oscillation parameters of table 6.2 with $\delta_{\mathrm{CP}} = 0$.

The performance of the selection procedure is understood in terms of the:

Efficiency: The proportion of true signal events that were selected.

Purity: The proportion of selected events that are true signal flavour events, where the extension to 'true signal flavour' is made as for example, CC ν_e and $\bar{\nu}_e$ interactions, look indistinguishable.

Background Rejection: The proportion of background events that were rejected in the selection, where background is defined to be all events other than the true signal flavour events.

As the ν_e selection is performed first, it will first be discussed.

6.6.1 ν_e Selection Performance

The selection metrics, integrated over all parameters, are detailed in table 6.6. The selection efficiency and purity are both found to be quite low as a result of the limited signal-background separation achieved by the electron-like BDT.

Efficiency	Purity	Background Rejection
61.8%	64.9%	98.4%

Table 6.6: The integrated selection metrics of the ν_e selection procedure.

The selection metrics, as a function of energy, are presented in figure 6.25. General behaviours cannot be discerned in the extreme energy regions as the number of true and selected events are small: the latter is evidenced in the decomposed selected energy spectra of figure 6.25. However, it can be said that the efficiency is consistently poor in the low-energy region as, at these energies, electrons exhibit little shower behaviour, which, if managed to be reconstructed, will lead to poor electron-like scores. In the region of the oscillation peak, the efficiency remains fairly constant. The purity however, is seen to be much higher on the lower energy side of the oscillation peak than on the higher energy side, which is a consequence of the increased selection of NC background events. In general, NC events are found to be the dominant background in the ν_e selection, suggesting that the performance of the electron-photon separation in the electron-like BDT is insufficient. The electron-like BDTs are trained on all signal/background showers, regardless of their energy. As the electron/photon characteristics vary as a function of energy, the electron-like BDT performance could be improved by restricting the training to different energy regions. However, this is not the focus of this PhD and is left as future work.

So how does the ν_e selection performance impact the sensitivity to CP-violation? To answer this, one first interprets DUNE as a counting experiment in which the aim is to distinguish the observed number of CC ν_e interactions from that predicted by a CPC hypothesis. In this simplification:

- A higher efficiency will result in a larger difference between observed and expected counts.
- A higher purity will result in a more significant difference with respect to the scale of the counts.

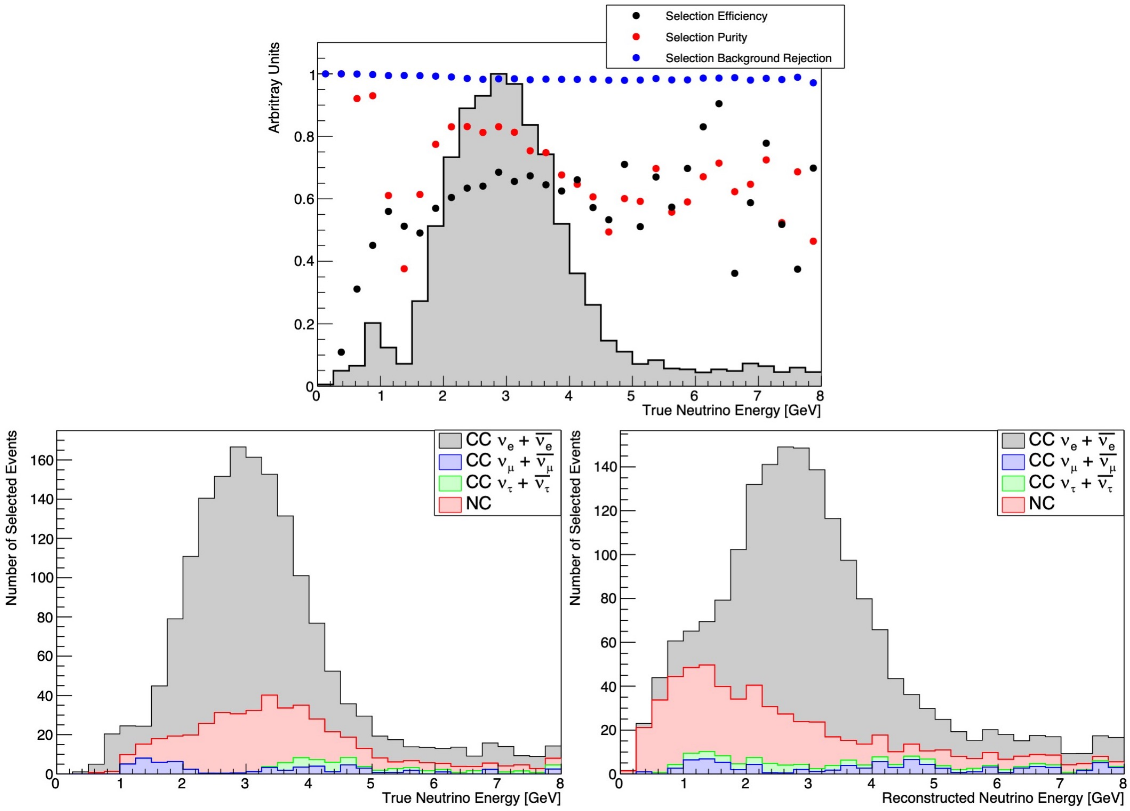


Figure 6.25: Top: The ν_e selection metrics, as a function of true neutrino energy, where the normalised true energy spectrum, of all signal events, has been added for context. In the high-energy region (> ∼5GeV), the number of true and selected events is small resulting in large uncertainties on efficiency and purity. Bottom: The selected ν_e true (left) and reconstructed (right) energy spectra. These are stacked histograms where each colour denotes a selected component type.

Following the conclusions drawn in subsection 6.2.2, a higher efficiency/purity therefore results in a larger sensitivity to CP-violation.

But DUNE isn't a counting experiment, and instead searches for differences between the observed and expected selected neutrino reconstructed energy spectra. In this approach one performs a counting experiment for each energy bin. As mentioned in subsection 6.2.2, some energy regions provide a higher discriminating power than others, and it is important that the efficiency and purity of such regions is maximised. To identify which energy bins have the highest potential discriminating power, we take the true selected spectrum (that with perfect efficiency, purity and energy resolution), and examine each bin's contribution to the sensitivity. This is shown in figure 6.26, where the ∼0.25GeV, ∼1GeV and ∼2-3GeV energy bands are found to have an above-average contribution to the sensitivity. Their discriminating power is largest at the maximally violating phases as, here, deviations from CPC are maximal.

As we move from true to reconstructed energy, the discriminating power of the three identified energy regions is reduced as a result of the smearing of the ν_e spectrum (seen by

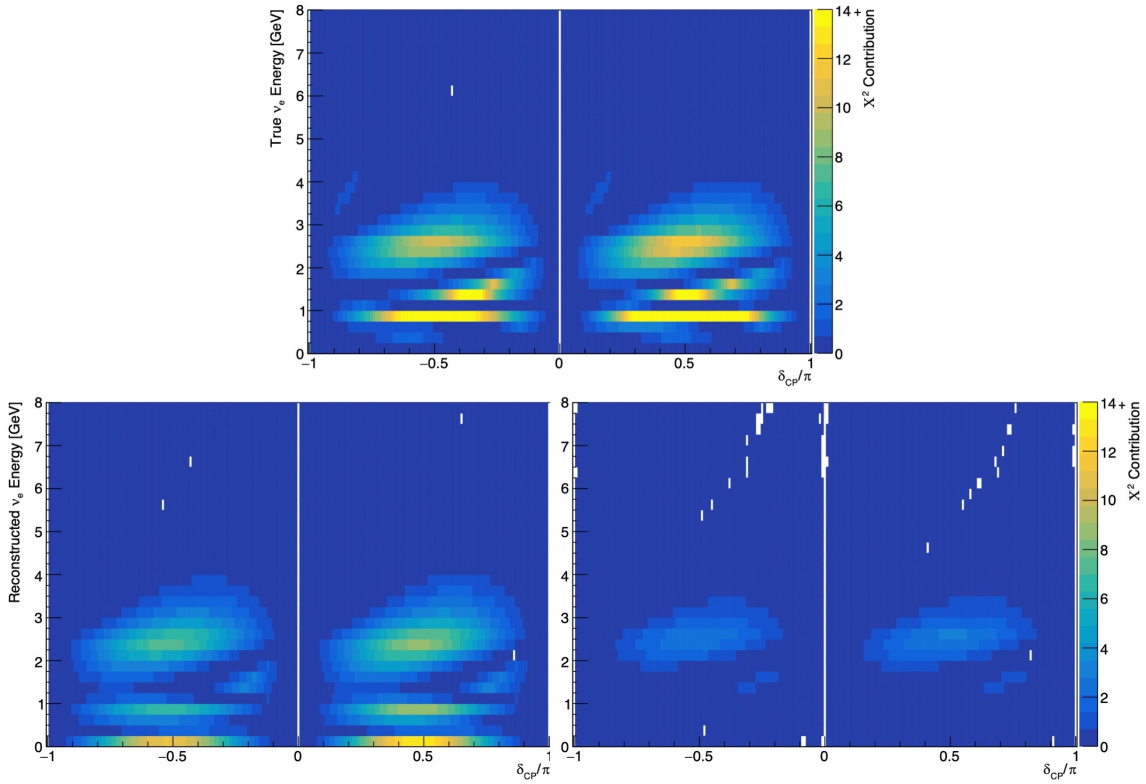


Figure 6.26: The sensitivity contribution of each bin of the ν_e selected spectrum, where the spectrum is obtained using: top) a perfect ν_e selection and energy estimators, bottom-left) a perfect ν_e selection and reconstruction energy estimators, and bottom-right) the Pandora-based ν_e selection and reconstruction energy estimators.

comparing the true and reconstructed energy spectra of figure 6.25). More significantly, the discriminating power is lessened by the inefficiencies and impurities introduced when we move to the Pandora-based ν_e selection. In this regime, only the ~2-3GeV energy band continues to achieve an above-average sensitivity contribution. The discrimination power of the ~1GeV energy band is washed away as a result of the large NC contamination - shown in figure 6.25 - of this area. The ~0.25GeV energy band is completely lost as a consequence of the poor reconstruction and selection efficiency of low-energy events. It is concluded that, as these energy regions are not reaching their potential discriminating power, the obtained sensitivity to CP-violation is limited by the efficiency and purity of the ν_e selection.

6.6.2 ν_μ Selection Performance

The integrated selection metrics of the ν_μ selection are given in table 6.7. The achieved efficiency and purity are respectively ~28% and ~20% greater than the equivalent metrics of the ν_e selection, which supports the earlier statement that the ν_μ selection is easier to perform.

The selection metrics, as a function of energy, are presented in figure 6.27. The

Efficiency	Purity	Background Rejection
88.2%	87.3%	94.4%

Table 6.7: The integrated selection metrics of the ν_μ selection procedure.

efficiency is seen to increase as a function of true neutrino energy until it stabilises at approximately 4GeV. This is understood to be a consequence of the improvement of the muon-like score with increased track length, and verifies that the muon-like BDT is dominated by the idea that muons typically have long lengths. The purity is highest at low energies, where background track lengths are often short and easily rejected, and exhibits a sharp dip in the region of the second oscillation minima as a consequence of a small number of selected events. At higher energies, the purity degrades from its low-energy value as a result of an increased selection of the NC background, which is the dominant background of the ν_μ selection.

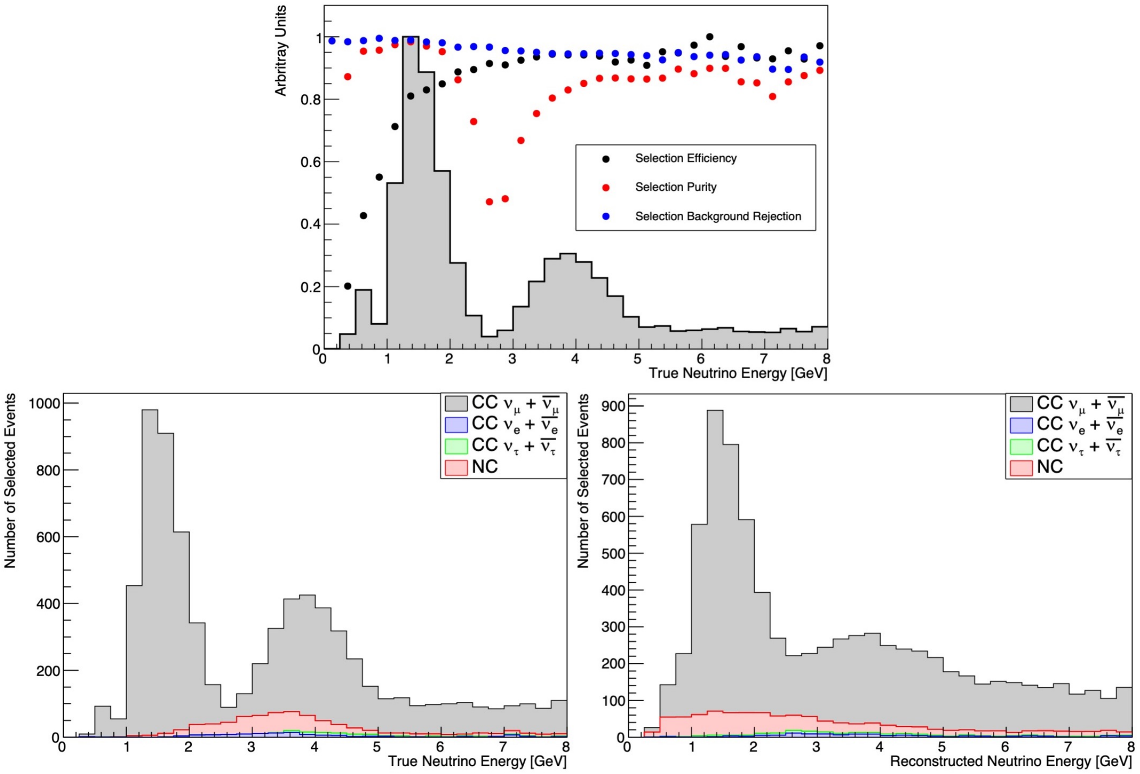


Figure 6.27: Top: The ν_μ selection metrics, as a function of true neutrino energy, where the normalised true energy spectrum, of all signal events, has been added for context. Bottom: The selected ν_μ true (left) and reconstructed (right) energy spectra. These are stacked histograms where each colour denotes a selected component type.

Now to think about what this means for the CP-violation analysis. The ν_μ contribution to the CP-violation analysis is not expected to be limited by the selection efficiency and/or purity, as is the case for the ν_e contribution, but rather by the performance of the neutrino energy estimators. This is seen by comparing the selected ν_μ true and reconstructed energy spectra presented in figure 6.27. A ν_μ hypothesis is used to estimate the neutrino energy of selected events, which results in an underestimation of the energy of NC events. The energy of high-energy ν_μ events is also underestimated, as a MCS approach is used, which is much less accurate in comparison to the range-based calculation used for contained muon tracks (see section 4.5). These shifted events wash away the shape of the second oscillation minima, which may impact the CP-violation analysis.

The significance of the shape of the ν_μ spectrum is verified by figure 6.28, in which the first plot presents the sensitivity contribution of each bin of the true ν_μ selected spectrum (that with perfect efficiency, purity and energy resolution). The ∼0.5GeV, ∼1.25GeV, ∼1.75-2.75GeV and ∼3-4GeV energy regions are found to achieve an above-average contribution to the sensitivity. Figure 6.27 tells us that these energy bands correspond to each side of the first and second oscillation maxima. As earlier mentioned, the energy spectrum smears as we move from true to reconstructed energy and this is confirmed to diminish the sensitivity contributions. The ν_μ sensitivity contribution is therefore believed to be limited by the performance of the ν_μ energy estimators rather than the selection performance. For completeness, the introduction of the Pandora-based selection lessens the sensitivity contributions further, such that the contribution from the lower-energy oscillation maxima region almost vanishes. This is explained by the reconstructed energy spectrum of figure 6.27, where it is seen that the shape of the first oscillation maxima is completely washed away.

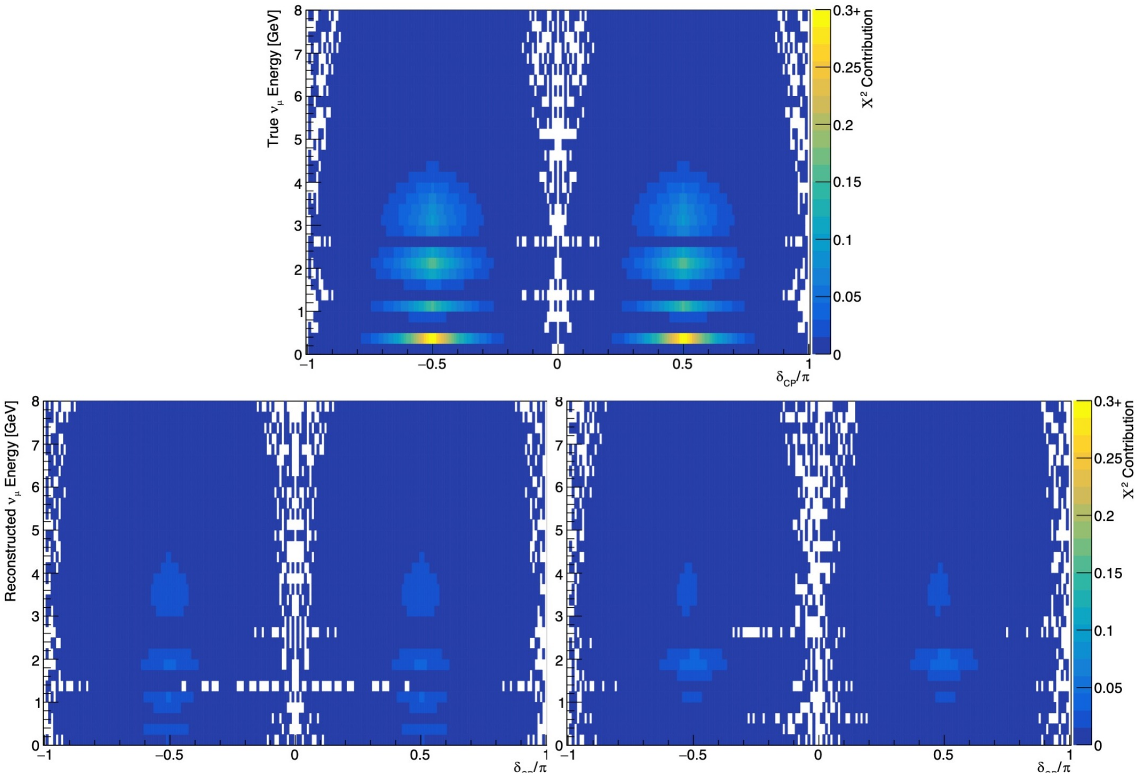


Figure 6.28: The sensitivity contribution of each bin of the ν_μ selected spectrum, where the spectrum is obtained using: top) a perfect ν_μ selection and energy estimators, bottom-left) a perfect ν_μ selection and reconstruction energy estimators, and bottom-right) the Pandora-based ν_μ selection and reconstruction energy estimators.

Chapter 7

The Search for Improvements

The goal of this analysis is to develop the reconstruction and/or the event selection in order to optimise the performance of the CP-violation analysis outlined in chapter 6. To achieve this, we must first identify the leading limitation of the reconstruction with respect to the performance of the CP-violation analysis.

This can be achieved via 'cheating studies' in which an aspect of the reconstruction is perfected, such that the corresponding CP-violation analysis gains are understood, and the importance of the reconstruction improvement determined. MC truth information is used to perfect, or cheat, a reconstruction procedure. For example, to cheat the neutrino vertex placement, the reconstruction vertexing methods are replaced with a cheated approach in which the reconstructed vertex is simply assigned its true position.

To navigate such a vast and complex workflow, cheating studies will be performed in an iterative manner such that the results of each iteration are used to narrow an initial, broad hypothesised reconstruction improvement until a specific, well-defined goal is identified. This chapter illustrates that process. A cheating study will be the subject of each section, which will be divided into subsections that outline the motivation of the cheating study, the details of the cheating configuration and the results of the study.

7.1 Quantifying the Improvements

The potential gains to the CP-violation analysis will be understood in terms of the sensitivity to CP-violation. It is chosen to use the 'simple sensitivity' estimate; that with no oscillation parameter or systematic uncertainties included (see subsection 6.2.2). This choice has two motivations. Firstly, the simple sensitivity estimate is the quickest to perform. This is desirable as each cheating study requires a significant length of time and many cheating study iterations are expected as a result of the expanse of the workflow. Secondly, degeneracies do not occur in the simplest estimate which eases the interpretation of the improvements.

The fake data FD prediction, used to obtain the results of this section, corresponds to a 3.5 year run time in FHC mode (1.36×10^{23} POT) and assumes the normal-ordering oscillation parameter values (given in table 6.2). It was chosen to consider a full FHC run-time in order to reduce the time taken to perform each cheating study, to avoid exceeding storage quotas, and to simplify an already nearly-overwhelming analysis procedure. This is not expected to affect the conclusions reached as FHC and RHC neutrino interactions are, to a large extent, indistinguishable.

Finally, the electron-like and muon-like BDTs (which guide the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection) were retrained in each cheating study in order to optimise the sensitivity achieved.

7.2 Cheating the Pandora Pattern-Recognition

7.2.1 Motivation

The track/showers that inform the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection are built by the Pandora pattern-recognition software. If their quality is poor, the performance of the electron/muon-like BDTs, the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection, and the CP-violation analysis will be limited. So this investigation starts at the foundations of the Pandora-based CP-violation analysis, with cheating the Pandora pattern-recognition software.

7.2.2 Cheating Details

To cheat the Pandora pattern-recognition is to cheat the:

- **2D Clustering:** 2D clusters are built by grouping the hits of the same owner in each 2D view, where 2D hits are assigned to the particle that contributes most to the hit's charge. Shower hierarchies are grouped as one particle.

- **2D → 3D Matching:** 2D clusters with the same owner are grouped together across the three views.

- **3D Spacepoint Creation:** 2D hits are combined to create 3D spacepoints, this step is cheated in the sense that the inputs are perfect.

- **3D Particle Vertex Placement:** particle vertices are assigned to their true positions. The vertices of primary particles will lie on the true neutrino vertex except those of photons which, since they are not immediately seen in the detector, are placed on the 3D position of the first associated energy deposition.

- **Track/Shower ID:** 3D particles that are, in truth, electrons or photons are classified as showers and all other particles are classified as tracks.

- **Neutrino Hierarchy:** the correct parent-to-child links of the neutrino hierarchy are made.

- **Neutrino Vertex Placement:** the neutrino vertex is placed on its true position.

The multitude of the now reconstructed low-hit photon showers were found to distort the signal/background BDT variables and so the training of the electron-like BDT was restricted to showers that have more than 20 2D hits.

7.2.3 Results

Figure 7.1 presents the sensitivity gains obtained when the Pandora reconstruction is perfected. The peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) increases from 5.3σ (obtained in the initial reconstruction/selection configuration) to 7.3σ. But where do these gains come from?

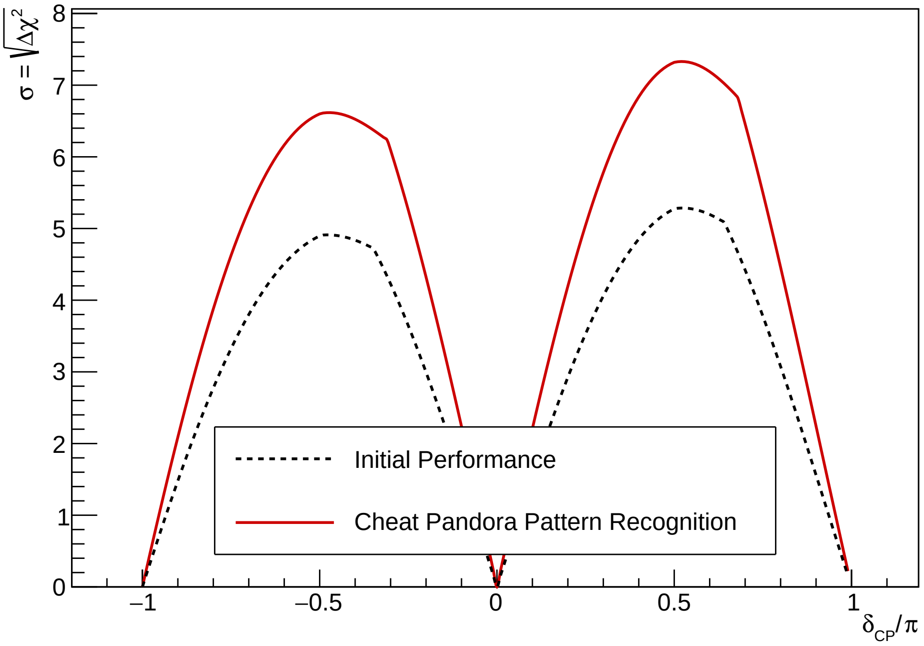


Figure 7.1: The estimated sensitivity to CP-violation achieved by fully cheating the Pandora pattern-recognition software (solid dark red), compared to that obtained with no cheating applied (dashed black).

In general, with a perfected reconstruction, the BDT variable distributions become much cleaner, resulting in well-defined signal and background profiles. The perfectly reconstructed tracks and showers align with these profiles, leading to highly separated signal and background classification distributions, a more efficient and pure selection, and, ultimately, large gains to the sensitivity. But let's dig into the specifics of each selection.

Table 7.1 reveals that the ν_e selection efficiency improves substantially. This is a result of the aforementioned reasons but also because, in a cheated pattern-recognition, all particles have associated reconstructed objects i.e. electrons will not be merged into overlapping showers and those that are very sparse will be fully reconstructed.

	ν_e Selection			ν_μ Selection		
Config.	Efficiency	Purity	BG Rejection	Efficiency	Purity	BG Rejection
Initial	61.8%	64.9%	98.4%	88.2%	87.3%	94.4%
Cheat Pandora	98.9%	85.9%	99.2%	91.2%	91.2%	96.1%

Table 7.1: The integrated efficiency, purity and background (BG) rejection of the ν_e and ν_μ selection, where 'initial' denotes the performance with no cheating applied.

Comparison of the selected ν_e reconstructed energy spectra of figures 6.25 and 7.2, demonstrates the reduction of the NC contamination of the low-energy region, which accounts for the observed purity improvements. This energy region was identified, in subsection 6.6.1, to have an above-average sensitivity to CP-violation, and thus the purity improvements are likely the catalyst of the sensitivity gains.

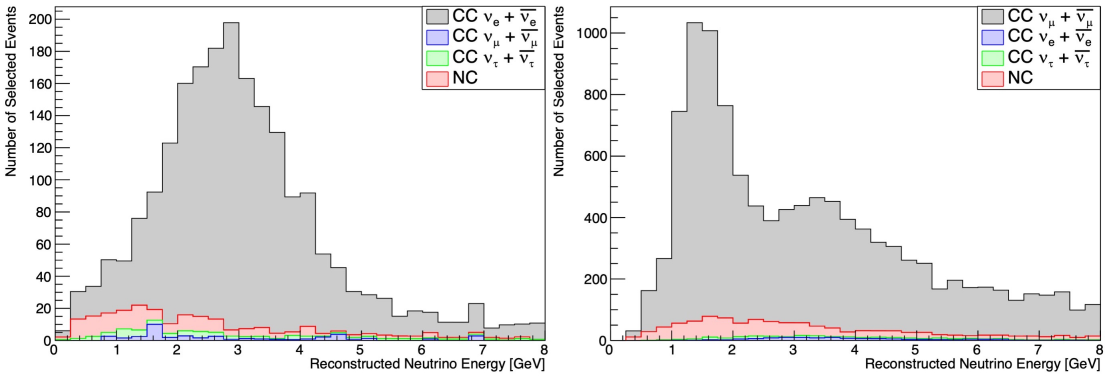


Figure 7.2: The ν_e (left) and ν_μ (right) selected reconstructed energy spectra obtained when the Pandora pattern-recognition is fully cheated. These are stacked histograms, where each colour denotes a selected component type.

However, there is still a significant NC background contamination which is believed to be a consequence of the limited ability to separate electrons and photons in the electron-like BDT. This can be improved by expanding the signal and background profiles i.e. by adding orthogonal information that examines the differences between electrons and photons to the electron-like BDT.

On the side of the ν_μ selection, efficiency and purity gains are much less significant, which demonstrates that the ν_μ selection is not limited by the quality of the input. This suggests that the muon-like BDT has a rather unforgiving signal profile such that muon tracks which do not resemble a typical muon, for example those that are not straight and long, obtain poor scores. To improve this, an alternative BDT could be created and tailored to this subset.

The ν_μ contribution to the sensitivity has been understood to be limited by the performance of the neutrino energy estimator (see subsection 6.6.2) and this is further evidenced by the comparison of figures 6.27 and 7.2, which demonstrate that, even with a perfect reconstruction, the shape of the reconstructed energy spectrum is smeared by the underestimation of the neutrino energy.

It is concluded that the current $\overset{(-)}{\nu}_e$ contribution to the CP-violation analysis is limited by the poor efficiency and purity of the $\overset{(-)}{\nu}_e$ selection whilst the $\overset{(-)}{\nu}_\mu$ contribution is limited by the neutrino energy estimator. The former is improved with a perfected Pandora reconstruction whilst the latter is not. Since the $\overset{(-)}{\nu}_e$ contribution dominates the sensitivity, the $\overset{(-)}{\nu}_e$ selection improvements are of higher priority and thus will be focused on.

7.3 Cheating the Neutrino Vertex

7.3.1 Motivation

The Pandora pattern-recognition cannot be completely perfected, so what specifically do we need to improve upon? The position of the reconstructed neutrino vertex guides the reconstruction and, if placed incorrectly, can cause the particles to be broken up or merged together, and the neutrino hierarchies to be completely mis-reconstructed. This can severely impact the $\overset{(-)}{\nu}_e$ and $\overset{(-)}{\nu}_\mu$ selections: poorly reconstructed signal particles will smear the BDT variable distributions, reducing the selection efficiency and purity, and an incorrect particle hierarchy or particle track/shower ID can prevent the signal particles from even being considered in the selection, further reducing the efficiency.

7.3.2 Cheating Details

- **Neutrino Vertex Placement:** see subsection 7.2.2.

7.3.3 Results

Figure 7.3 reveals that negligible gains are achieved when the neutrino vertex placement is cheated. The peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) increases from 5.3σ (obtained in the initial reconstruction/selection configuration) to 5.4σ. This suggests that the foremost limitation of the CP-violation analysis is the reconstruction of the particles themselves.

Moving forwards, the placement of the neutrino vertex will be continued to be cheated for the following reasons:

- The BDT used in the Pandora vertexing algorithms could not be adapted to handle a cheated 2D cluster input.

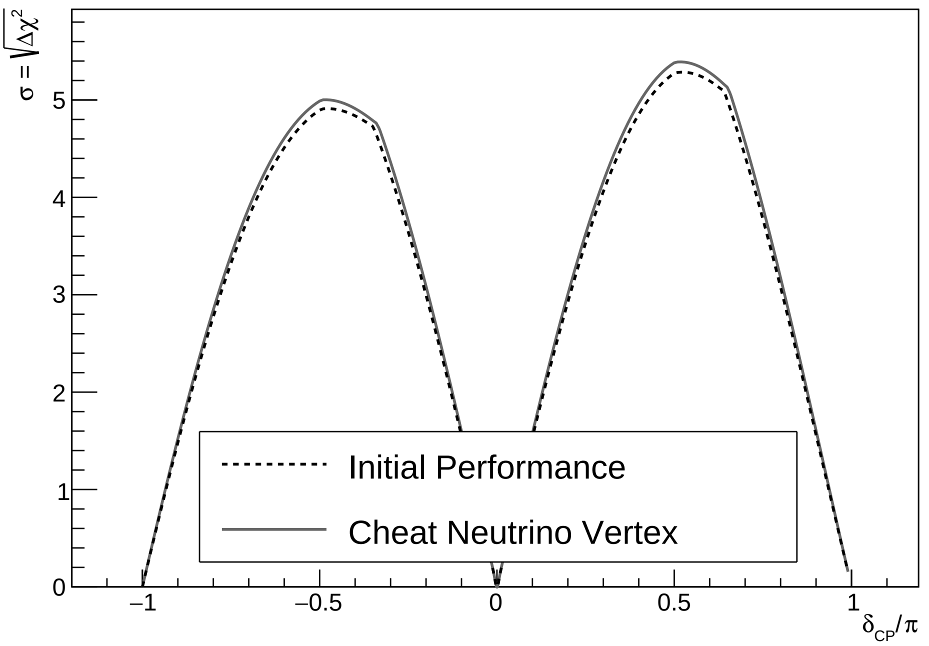


Figure 7.3: The estimated sensitivity to CP-violation achieved by fully cheating the placement of the neutrino vertex (solid grey) compared to that obtained with no cheating applied (dashed black).

- Consider an event in which both a vertexing failure and the true leading reconstruction failure occur. If the true leading reconstruction issue is cheated, the vertexing failure could prevent the event being 'fixed' such that the corresponding CP-violation improvements are not seen, making it harder to identify the leading reconstruction issue.

The continuation of the neutrino vertexing cheat must be remembered as, although it only results in minor sensitivity gains, its removal may significantly diminish the CP-violation analysis improvements of a future cheating configuration. This will be returned to in subsection 8.5.3.

7.4 Cheating the Signal Electrons

7.4.1 Motivation

The $\overset{(-)}{\nu}_e$ contribution to the CP-violation analysis is limited by the $\overset{(-)}{\nu}_e$ selection efficiency and purity, which is improved by a better ability to separate electrons and photons in the electron-like BDT. But how is this increased separation to be achieved? Is it dominated by better reconstructing the signal or the background? Or are both required to improve?

At this stage, we are limited in our ability to answer these questions. A cheated shower will look perfect to an extent that is unattainable with a realistic reconstruction:

there will be no impurities and all hits, even the very outer halo of sparse hits, will be clustered into the shower. So if one cheats only the electron or photon reconstruction, the electron-like BDT will learn the difference between truth and reconstructed showers and not the difference between realistic electron and photon characteristics, as reconstructed in a LArTPC.

We are therefore only able to cheat the reconstruction of all showers. The question of whether it is the electron or photon reconstruction that dominates the improvements will be revisited in section 7.7.

7.4.2 Cheating Details

- **Neutrino Vertex Placement:** see subsection 7.2.2.

- **Cheated Leading Electron and Photon Showers:** the reconstruction of the true primary electrons in CC $\overset{(-)}{\nu}_e$ interactions is cheated as well as the reconstruction of **primary** photons in all events. This includes the 2D clustering, 2D → 3D matching, 3D spacepoint creation, 3D shower vertex placement, the parent-to-child link between the neutrino and shower, and the track/shower ID.

- **Training Hit Cut:** see subsection 7.2.2.

7.4.3 Results

Figure 7.4 reveals that the gains to the CP-violation analysis obtained in section 7.2 arise from a perfected shower reconstruction, which confirms that an improved electron-photon separation drives the sensitivity improvements. The peak sensitivity (obtained at the $\delta_{\mathrm{CP}} = \pi/2$ peak) is 7.4σ, which agrees with the 7.3σ obtained when the full Pandora pattern-recognition is cheated.

The electron-photon separation is mainly achieved by the displacement and dE/dx variables, their distributions with and without a perfected shower reconstruction are shown in figure 7.5. Here it is revealed that the signal displacement distribution is almost completely separated from the background as result of the overlap of primary electron shower vertices with the true neutrino vertex in this cheating configuration (see figure 7.6). The electron-like BDT is therefore able to learn that: a (non-)zero displacement corresponds to a signal(background) event, i.e. it has learnt to separate events by learning the operation of the cheating configuration, rather than by using the characteristics of electrons and photons. This makes it ‘easy’ to select(reject) signal(background) events, and results in a falsely high efficiency, purity, and sensitivity. The zero displacement is real physics, but is not achievable with practical reconstruction methods.

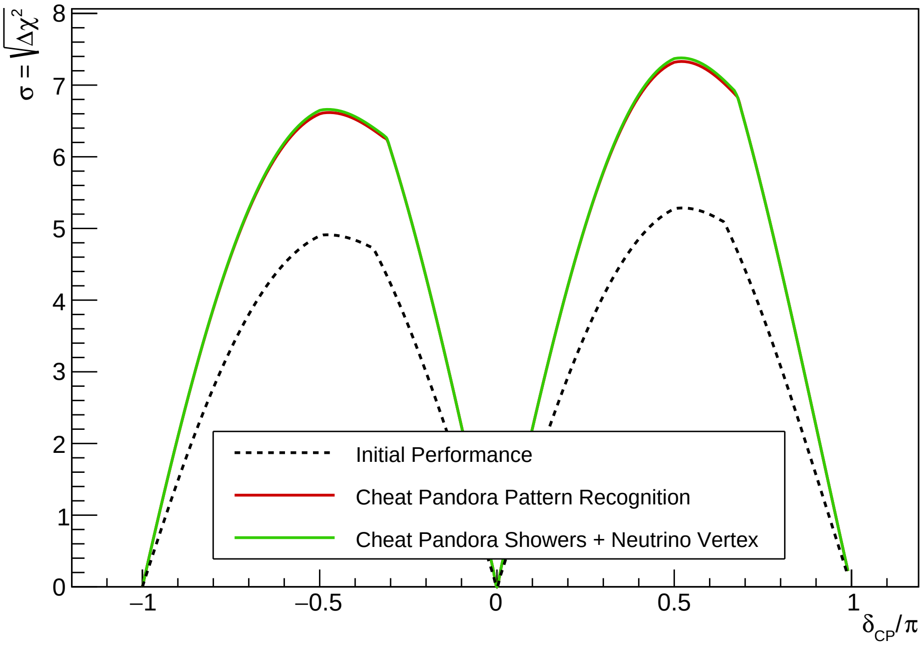


Figure 7.4: The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the shower reconstruction (solid green), compared to that obtained when the Pandora pattern-recognition is perfected (solid red), and that with no cheating applied (dashed black).

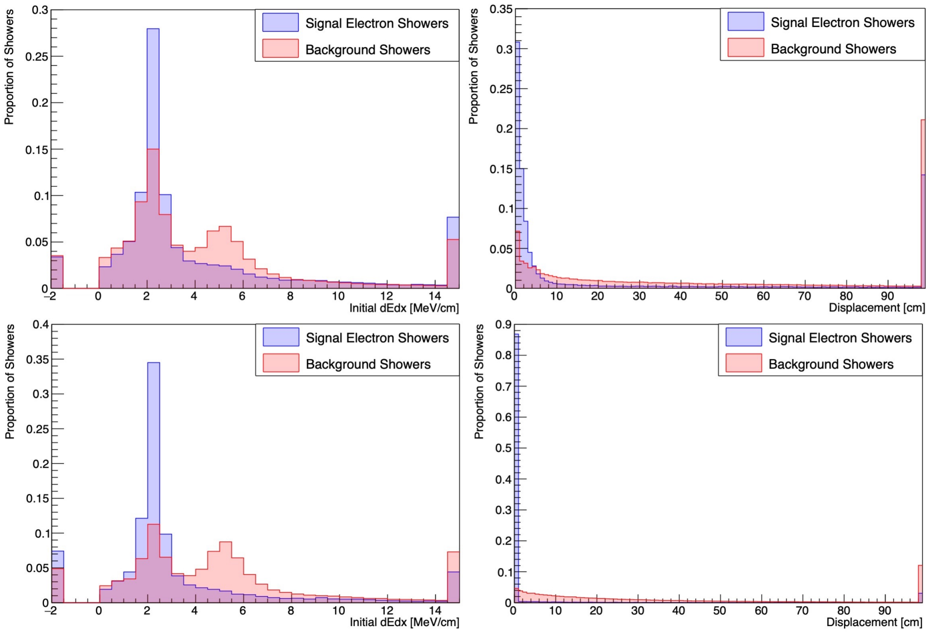


Figure 7.5: The signal (blue) and background (red) dE/dx (left column) and displacement (right column) distributions used to train the electron-like BDT in the standard Pandora reconstruction configuration (top row) and in the cheating configuration of this section (bottom row).

7.5 Visible Vertexing

7.5.1 Motivation

Our next step is to understand the CP-violation analysis gains that can be obtained with a more realistic handling of the perfected electron vertexing.

A realistic aim for a 'perfect' reconstruction is to achieve that which can be discerned by eye (and maybe with a bit of head scratching). The cheating configuration of the previous section is therefore modified such that electron vertices are placed at the position of the first 3D hit that undoubtedly belongs to the electron, this is termed the 'visible vertex'. A good estimator of the visible vertex is the closest 3D position to the neutrino vertex obtained from the 2D hits for which the electron is the dominant contributor.

Two examples are shown in figure 7.6, where the right panel indicates a common scenario in which the initial electron trajectory overlaps with another particle in the event. In this case, both particles will contribute to the charge of the initial hits and the 'sole owner' of the hit, which is demanded by the reconstruction since there is no functionality for hit sharing, is ambiguous. This can result in a visible vertex that lies some distance from the neutrino vertex.

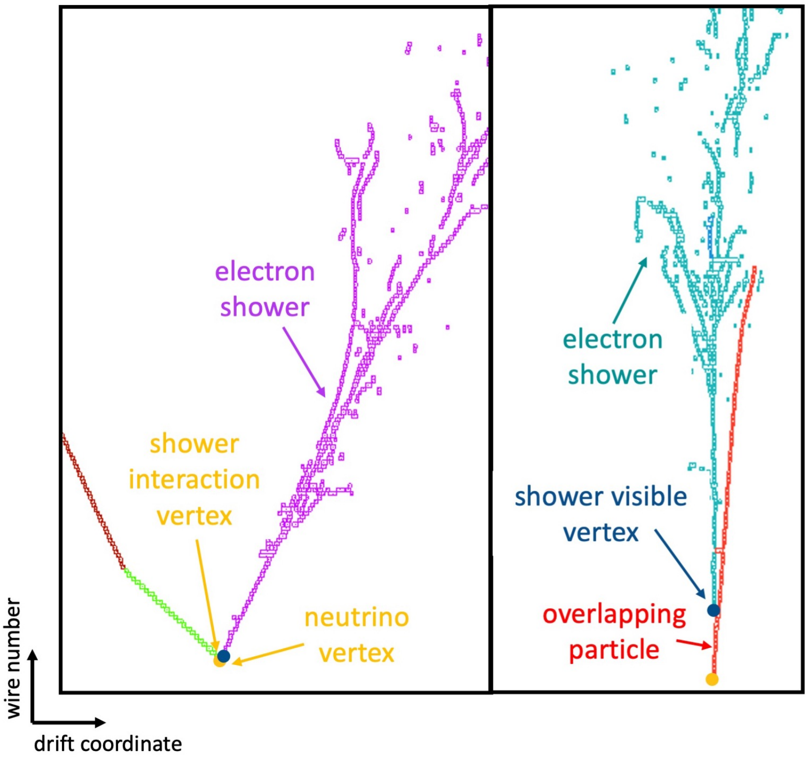


Figure 7.6: An illustration of the 'interaction' and 'visible' shower vertex in two CC ν_e interactions, where the electron shower in the left(right) panel is seen to have unambiguous(ambiguous) energy deposits that lead out from the neutrino vertex, resulting in a visible and interaction vertex that almost(do not) coincide.

7.5.2 Cheating Details

- **Neutrino Vertex Placement:** see subsection 7.2.2.
- **Cheated Leading Electron and Photon Showers:** see subsection 7.4.2, with the modification that the shower vertex is assigned the visible vertex rather than the interaction vertex.
- **Training Hit Cut:** see subsection 7.2.2.

7.5.3 Results

Figure 7.7 demonstrates that the sensitivity to CP-violation falls significantly when moving to a realistic shower vertex placement method: the peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) decreases from 7.4σ (achieved in the interaction vertex configuration of section 7.4) to 6.2σ. The catalyst of the deficit is the broadening of the signal displacement distribution, which reduces the variable's electron-photon separation power such that the BDT is no longer aware that a cheating configuration has been applied. Consequently - in figure 7.8 - the signal and background classification distributions are seen to smear, causing the signal peak to decrease and the background contamination of the signal region to increase. The former results in a lower selection efficiency whilst the latter causes the purity to fall as detailed in table 7.2. Figure 7.9 demonstrates that the NC background is sufficiently removed, which suggests that the sensitivity deficit arises from the efficiency degradation.

Configuration	Efficiency	Purity	BG Rejection
Initial	61.8%	64.9%	98.4%
Cheat Showers (Interaction Vertex)	98.9%	88.2%	99.3%
Cheat Showers (Visible Vertex)	75.0%	74.8%	98.8%

Table 7.2: The integrated efficiency, purity and background (BG) rejection of the ν_e selection achieved with the initial reconstruction/selection and the cheating of the neutrino vertex placement and shower reconstruction.

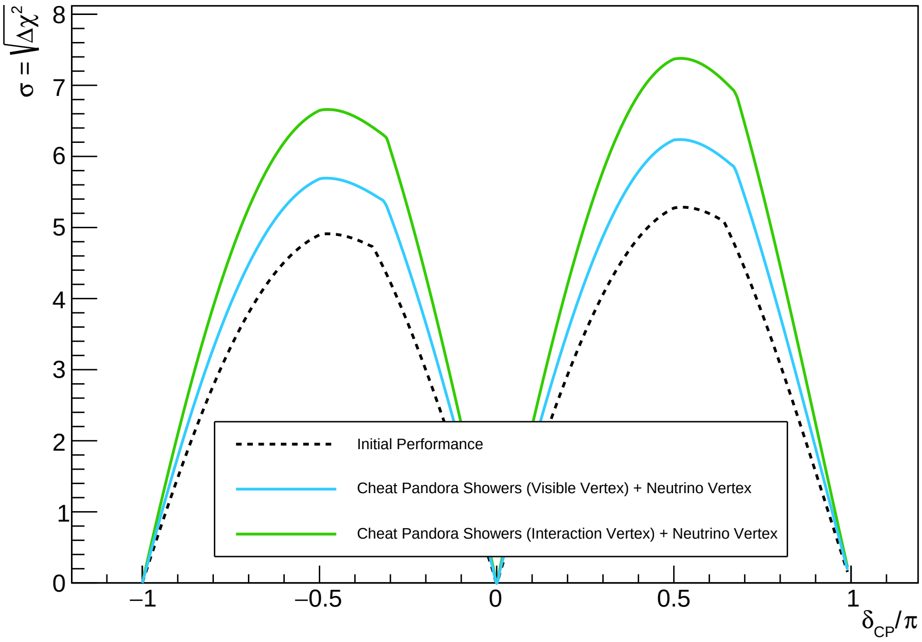


Figure 7.7: The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the shower reconstruction, where both the visible vertex (blue) and interaction vertex configuration (green) are shown. The initial sensitivity, before any cheating, (black) is added for reference.

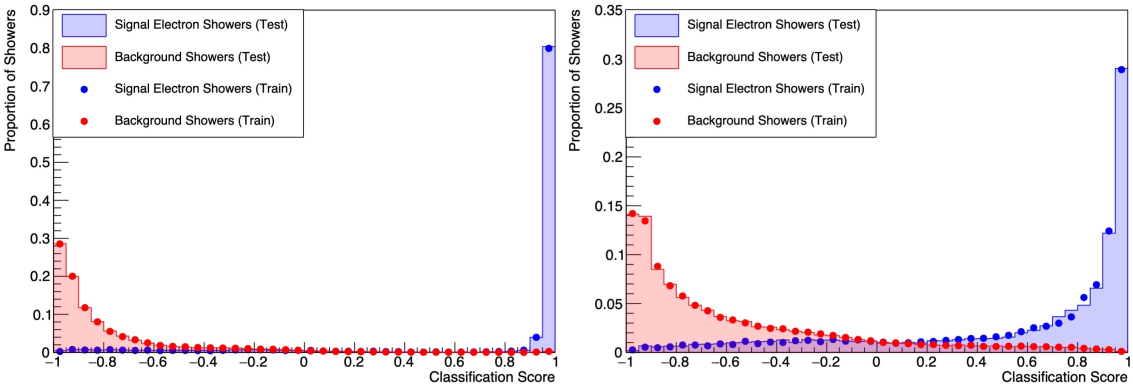


Figure 7.8: The signal (blue) and background (red) classification distributions of the electron-like BDT obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction, where the interaction vertex (left) and visible vertex (right) configurations are shown.

7.6 Shifted Vertexing

7.6.1 Motivation

Is visible vertexing really the best that we expect from the reconstruction? Let's return to the right panel of figure 7.6, where, by eye, one could postulate that the electron trajectory overlaps with that of another particle. This could be confirmed with the use

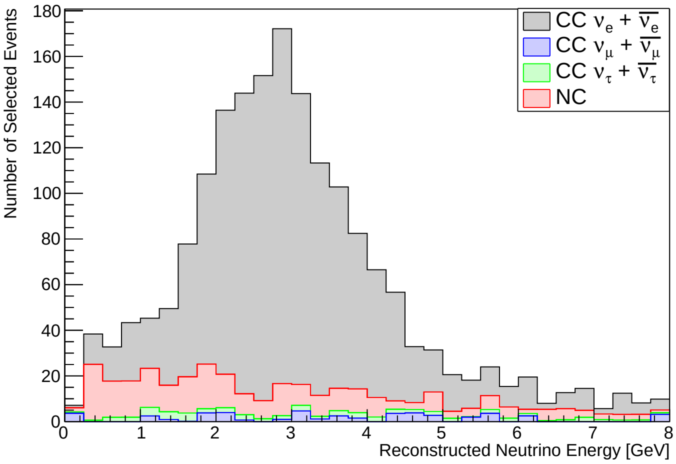


Figure 7.9: The selected ν_e reconstructed energy spectrum obtained when the shower reconstruction is cheated, such that the electron vertex is equal to the visible vertex. This is a stacked histogram where each colour denotes a selected component type.

of calorimetric information, and would motivate the shifting of the electron vertex closer to the neutrino vertex.

To investigate the potential of an algorithm that follows this logic, the cheating configuration of section 7.5 is modified such that electron vertices are placed at the position of the first 3D hit that is, in truth, shared with the electron. This is termed the 'shifted vertex'. A good estimator of the shifted vertex is the closest 3D position to the neutrino vertex, obtained from the 2D hits, to which the electron has a charge contribution i.e. the electron does not have to be the dominant contributor.

7.6.2 Cheating Details

- **Neutrino Vertex Placement:** see subsection 7.2.2.
- **Cheated Leading Electron and Photon Showers:** see subsection 7.4.2, with the modification that the shower vertex is assigned the shifted vertex rather than the interaction vertex.
- **Training Hit Cut:** see subsection 7.2.2.

The 2D hits for which the electron is a sub-dominant contributor are not clustered into the electron as this could distort the signal dE/dx distribution.

7.6.3 Results

Figure 7.10 reveals that the gains to the sensitivity to CP-violation are largely recovered in the shifted vertex configuration: the peak sensitivity (obtained at the $\delta_{\mathrm{CP}} = \pi/2$ peak) now reaches 7.0σ, which is an improvement from 6.2σ (attained in the visible vertex configuration) but falls short of 7.4σ (achieved in the interaction vertex configuration).

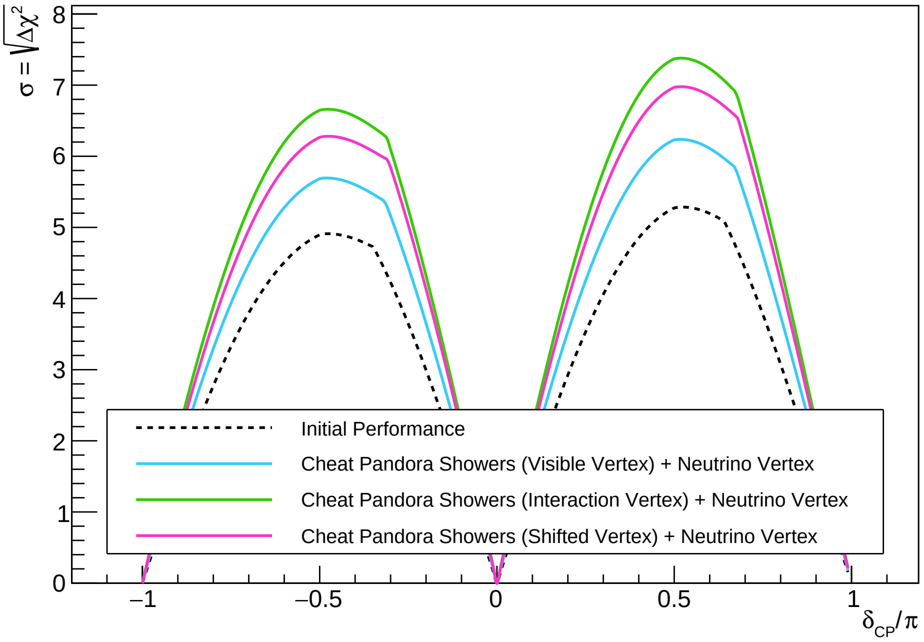


Figure 7.10: The estimated sensitivity to CP-violation achieved by cheating the placement of the neutrino vertex alongside the shower reconstruction, where the shifted vertex (pink), visible vertex (blue) and interaction vertex configuration (green) are shown. The initial sensitivity, before any cheating, (black) is added for reference.

From figure 7.11, we infer that the sensitivity gains are recovered as a result of the narrowing of the signal displacement distribution. This improves the electron-photon separation power of the displacement variable, which increases the peak of the signal classification and almost completely removes the background contamination of the signal region, as demonstrated in figure 7.12. This means that a high efficiency can be achieved with an electron-like BDT threshold that lies close to unity, at which an equally large purity is obtained (see table 7.3).

But, has the electron-like BDT learnt the cheating configuration, like we saw section 7.4? Table 7.3 demonstrates that the achieved purity is similar to that obtained when the electron vertex is placed on the neutrino vertex, which suggests that the electron-like BDT is once again reduced to a single consideration of 'if the displacement is consistent with zero then the shower is an electron'. This is supported by the drop in efficiency which originates from the rejection of the reconstructed electrons that now have a small, but not insignificant, displacement.

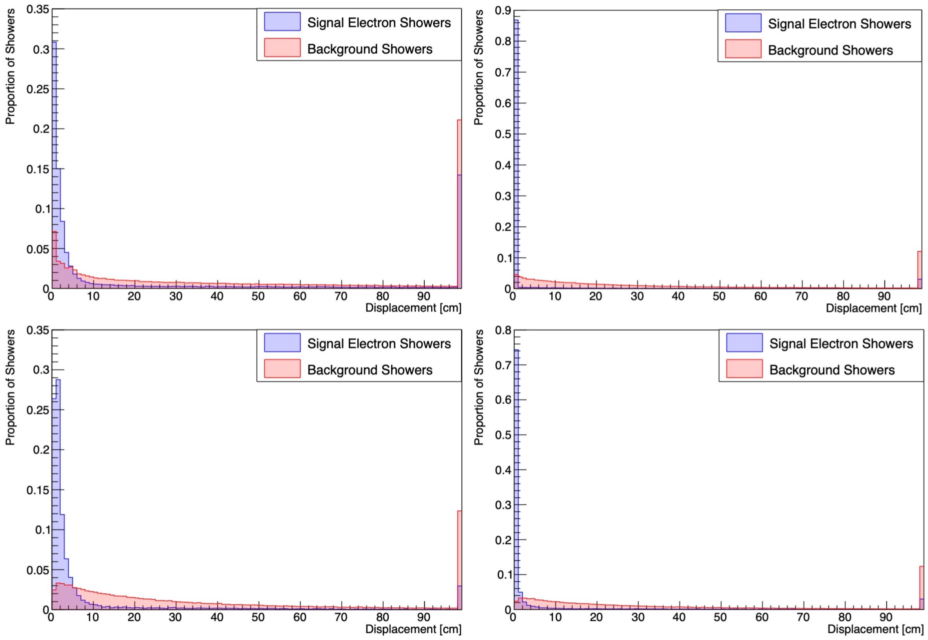


Figure 7.11: The signal (blue) and background (red) and displacement distributions used to train the electron-like BDT in the standard Pandora reconstruction configuration (top left) and in the cheating shower cheating configurations where the shower vertex is placed in its true (top right), visible (bottom left) and shifted (bottom right) position.

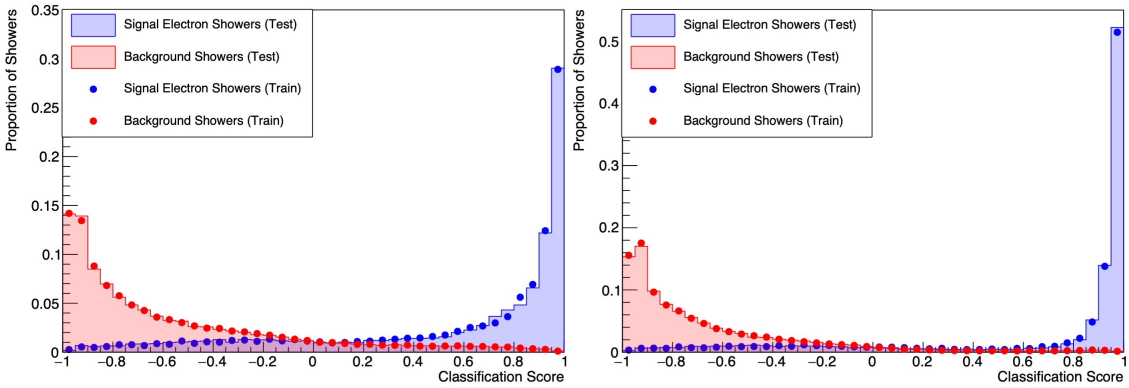


Figure 7.12: The signal (blue) and background (red) classification distributions of the electron-like BDT obtained when cheating the placement of the neutrino vertex alongside the shower reconstruction, where the visible vertex (left) and shifted vertex (right) configurations are shown.

7.7 Cheating the Initial Shower Region

7.7.1 Motivation

Despite an achieved spread in the signal displacement distribution, achieved by using the shifted vertex, the electron-like BDT was still able to learn the cheating configuration. This is believed to be a consequence of the fact that, in these cheating configurations, **all** targeted reconstruction errors are fixed. To work to a more realistic estimate of potential sensitivity gains, the efficiency with which a real reconstruction algorithm can fix a given

Configuration	Efficiency	Purity	BG Rejection
Initial	61.8%	64.9%	98.4%
Cheat Showers (Interaction Vertex)	98.9%	88.2%	99.3%
Cheat Showers (Visible Vertex)	75.0%	74.8%	98.8%
Cheat Showers (Shifted Vertex)	86.7%	90.1%	99.5%

Table 7.3: The integrated efficiency, purity and background (BG) rejection of the ν_e selection achieved with the initial reconstruction/selection and the cheating of the neutrino vertex placement and shower reconstruction.

problem needs to be injected into the cheating study. This is very difficult to do, and so we must be certain that the CP-violation analysis is limited by the electron vertex placement before taking this path. However, we are not yet certain of this, as the entire shower reconstruction is also here perfected. So now let's remove the cheating of the shower region, to determine whether it is the reconstruction of the initial shower region, and thus the shower vertex placement, which dominates the CP-violation analysis improvements.

7.7.2 Perfecting the Initial Electron Region Algorithm

The initial region of an electron can be mis-reconstructed if the reconstructed shower has been split at the point where the shower begins, or, if the initial region hits are assigned to another particle because their owner is ambiguous (as illustrated in figure 6.16). In each case the shower vertex will be placed downstream of its true position, resulting in a photon-like gap. The initial dE/dx will be measured from this point and, as it is closer to (or even inside) the shower region, the initial dE/dx may resemble that of a photon shower. Furthermore, if the lost initial region hits form a significant object, the shower particle may be tagged as its child. This will make the shower a secondary particle, and thus prevent its selection in the $\overset{(-)}{\nu}_e$ selection.

The *CheatingCCElectronRefinement* algorithm was developed to intercept the reconstructed signal electrons of CC $\overset{(-)}{\nu}_e$ interactions and extend their pathways to the neutrino vertex where necessary. It is placed at the end of the Pandora reconstruction chain, at which point the electron reconstruction is nominally complete. This ensures final control of the electron reconstruction. The algorithm has the following workflow:

1. In CC $\overset{(-)}{\nu}_e$ interactions, all MC primary electrons are identified.
2. In each 2D view, 2D hits, whose charge is dominated by an identified MC electron, are mapped to that MC particle.
3. Where possible, each MC primary electron is matched to a reconstructed particle

that has a purity and completeness above 50% and 30% respectively. The former threshold makes certain that the reconstructed particle unambiguously belongs to the matched MC particle, whilst the latter ensures that we only modify particles that are likely to be fixed in reality. If the electron has been split, the reconstructed particle that is furthest from the neutrino vertex is matched to the MC electron. This ensures that the shower region of the electron will be connected to the neutrino vertex, reflecting the aim of a real reconstruction algorithm.

4. For each reconstructed particle that has been matched to a MC electron, an extension cone is defined between the MC neutrino vertex and the reconstructed shower vertex with an axis equal to the MC direction of the shower and an opening angle of 10°. This number was tuned via interactive hand scanning. Mapped MC hits that fall within this cone are added into the matched reconstructed particle. The shower vertex in each view is estimated to be the position of the 2D hit that has the shortest longitudinal displacement to the neutrino vertex.

7.7.3 Perfecting the Initial Photon Region Algorithm

When it comes to photons, the main reconstruction issue is the merging of other particles into the photon shower, such that the photon looks to originate from the neutrino interaction vertex (as illustrated in the right panel of figure 6.17). In this case, the shower vertex will be placed at the vertex of the contaminant particle resulting in an electron-like displacement and an incorrect estimation of the initial dE/dx.

The *CheatingPhotonRefinement* algorithm was developed to intercept the reconstructed primary photons and truncate their pathways where necessary. The algorithm runs at the same point in the Pandora reconstruction chain as the *CheatingCCElectronRefinement* algorithm so that no other changes to the photon reconstruction are subsequently made. The algorithm has the following workflow:

1. Hits are mapped to the MC particle that contributes most to their charge forming the ‘hit ownership’ map.

2. Reconstructed showers with over 100 hits are identified.

3. Matches between reconstructed particles and true primary photons are identified. A match is made if the main contributor to the total charge of the reconstructed particle is a photon. However, this algorithm targets reconstructed photons that have merged other particles into them and thus the photon may not be the main charge contributor. Therefore, a match can also be made if the particle contains a large proportion of hits that belong to a significant true primary photon in one view where, ‘large’ and ‘significant’ correspond to a 50% completeness and 100 hit

threshold respectively. In the case where multiple true primary photons satisfy this criteria, the match is made to the MC photon that contributes the highest number of hits.

4. In each 2D view, the hits contained within the matched reconstructed shower, that lie upstream of the true photon vertex with respect to the MC direction of the photon, are removed. So to not underestimate the neutrino energy, the removed hits with the same owner are grouped together to create new particles. However, the existence of low-hit particles around the neutrino vertex may have undesired consequences in the reconstruction and so a 50 hit threshold is imposed for particle creation.

7.7.4 Cheating Details

- **Neutrino Vertex Placement:** see subsection 7.2.2.
- **Cheated Initial Shower Region:** The *CheatingCCElectronRefinement* and *CheatingPhotonRefinement* algorithms are active in the final stages of the Pandora reconstruction workflow.

In this configuration, the electron initial region is cheated such that the shower vertex is placed on the 'visible' vertex position. This will give a cautious estimate of the potential sensitivity, but ensures that the BDT operates as desired.

7.7.5 Results

In this study, several initial region cheating configurations were investigated:

- Only cheat electron initial regions.
- Only cheat photon initial regions.
- Cheat both electron and photon initial regions.

The corresponding CP-violation sensitivities are shown in figure 7.13 alongside the goal configuration i.e. that achieved when the shower reconstruction is fully cheated with visible vertex placement.

With both initial region improvements, a good efficiency and purity are achieved (see table 7.4) resulting in substantial CP-violation gains: the peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) reaches:

- 6.0σ when only cheating the electron initial regions,
- 5.8σ when only cheating the photon initial regions, and

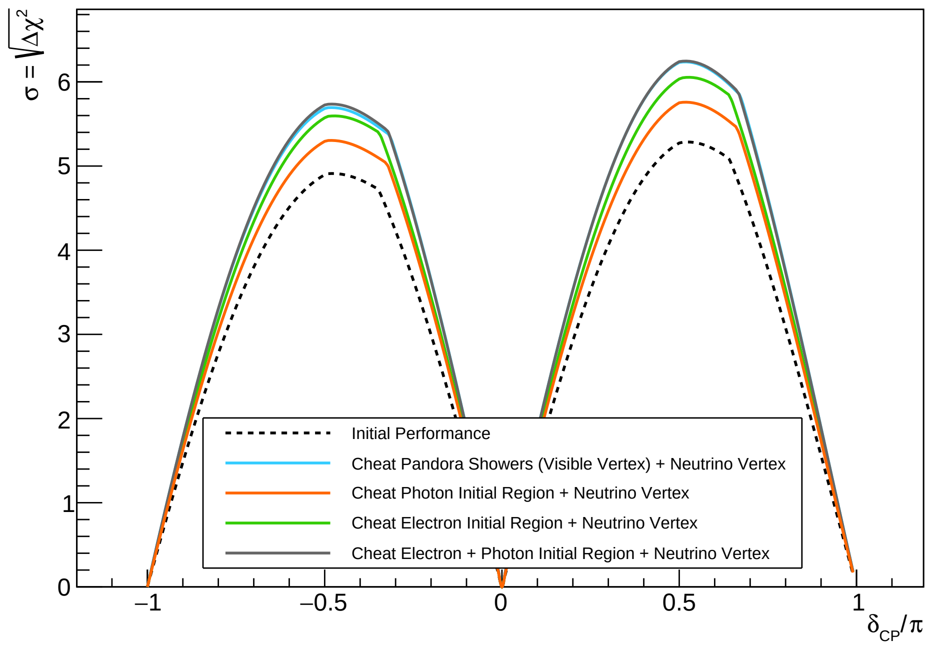


Figure 7.13: The estimated sensitivity to CP-violation achieved by cheating the reconstruction of the initial electron (green), photon (orange) and shower (grey) region alongside the placement of the neutrino vertex. For comparison, the sensitivity achieved when the neutrino vertex placement and the full shower reconstruction is cheated (blue) is shown as well as that achieved with no cheating applied (dashed black).

- 6.2σ when only cheating both electron and photon initial regions,

These numbers are compared to 6.2σ, which is attained in the goal visible vertex configuration (of section 7.5). The sensitivity gains achieved when cheating all shower initial regions, coincide with those obtained when the entire shower reconstruction is perfected. This confirms that the reconstruction of the initial shower region is the leading limitation on the CP-violation analysis.

Additionally, figure 7.13 allows us to answer the question posed at the beginning of section 7.4; it seems that it is more important to improve the electron reconstruction than the photon reconstruction. The reasons for this are as follows. The displacement variable of signal electrons is single valued, which means that the signal displacement distribution can be narrowed more than the background distribution (see figure 7.14), and a higher electron-photon separation can be achieved. This has the consequence - as shown in figure 7.15 - that when cheating the electron initial region, as opposed to the photon initial region:

- fewer events of the signal classification distribution lie in the background region,
- more events lie in the signal region, whilst
- the background contamination remains similar.

Configuration	Efficiency	Purity	BG Rejection
Initial	61.8%	64.9%	98.4%
Cheat Showers (Visible Vertex)	75.0%	74.8%	98.8%
Cheat Electron Initial Region	73.2%	67.0%	98.3%
Cheat Photon Initial Region	65.2%	74.6%	99.0%
Cheat Shower Initial Region	77.4%	75.9%	98.8%

Table 7.4: The integrated efficiency, purity and background (BG) rejection of the ν_e selection achieved with the initial reconstruction/selection; the cheating of the neutrino vertex placement and shower reconstruction; and the cheating of the neutrino vertex placement and the electron initial region, photon initial region, and shower initial region.

The electron-like BDT threshold of the ν_e selection can therefore lie at a lower value in the electron initial region cheat, such that a higher efficiency is obtained at the cost of a lower purity (table 7.4). This demonstrates that once the purity is above a certain threshold, efficiency improvements dominate the gains to the sensitivity.

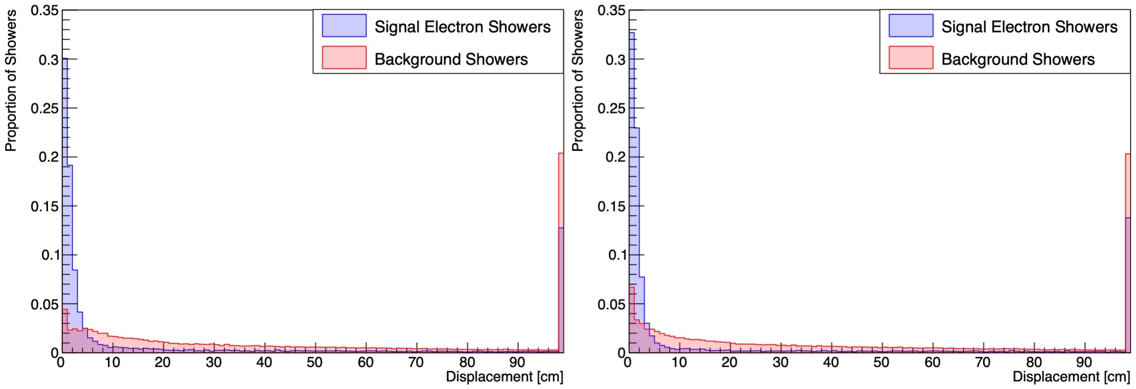


Figure 7.14: The signal (blue) and background (red) displacement distributions used to train the electron-like BDT obtained when cheating the neutrino vertexing alongside only the photon initial regions (left) and only the electron initial regions (right).

7.8 Outlook

The cheating studies of this chapter have identified the leading limitation of the CP-violation analysis to be the reconstruction of the initial shower region. However, the hypothesised sensitivity improvements rely on correcting the initial shower region with 100% efficiency, which a realistic reconstruction algorithm will not obtain. So what sensitivity improvements are we realistically able to achieve? Is it possible that an efficiency beyond that which can be achieved by realistic reconstruction methods is needed to achieve these

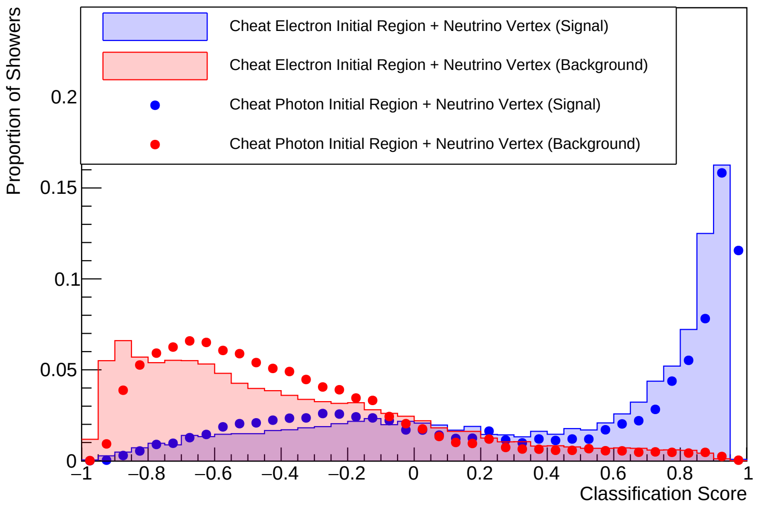


Figure 7.15: The signal (blue) and background (red) classification distributions of the electron-like BDT obtained when cheating the neutrino vertexing alongside only the photon initial regions (dots) and only the electron initial regions (shaded).

postulated gains?

One can theoretically understand the behaviour of the CP-violation gains as a function of the efficiency of a reconstruction algorithm by only applying the cheat to a certain proportion of events. But, how does one choose the events? Should all events have the same chance of being cheated? Surely not, as higher-energy neutrino interactions present much more complex topologies. But picking a subset of events to cheat risks biasing our analysis, which heavily relies on the neutrino interactions kinematics; for example, via the shape of the reconstructed neutrino energy spectra and through the signal and background profiles of the electron and muon-like BDTs.

One could pursue these cheating studies and move ever closer to a more realistic estimate. However, at this point, it seems sensible to take a very well motivated leap of faith and to develop a reconstruction algorithm to better reconstruct the initial shower region. We will proceed with caution, however, and perform cheating studies at key milestones in the reconstruction algorithm's development to understand the impact of its inefficiencies on the estimated CP-violation analysis gains and stop development if necessary.

Chapter 8

Optimising the Sensitivity to CP-Violation at DUNE

The CP-violation analysis has been found to be limited by the reconstruction of the initial region of showers in the Pandora pattern-recognition software:

- Electron showers can be disconnected from their initial track-like region, such that they exhibit a photon-like separation from the neutrino vertex and a photon-like initial dE/dx.
- Photon showers can have contaminants merged into them, such that they exhibit electron-like connections to the neutrino vertex and dE/dx values that are not photon-like.

Following the Pandora multi-algorithm approach, a 'shower refinement' algorithm was developed to specifically target these issues, such that improvements to the CP-violation analysis are achieved. The workflow of this algorithm is outlined in section 8.1; then the development and performance of an intermediate version of the algorithm, which relies on truth information, are presented in sections 8.2 and 8.3 respectively. The progression to the final *ShowerRefinement* algorithm, and the performance of the CP-violation analysis with its implementation, are outlined in sections 8.4 and 8.5 respectively.

8.1 The Shower Refinement Algorithm

The aim of the shower refinement algorithm is to:

- extend reconstructed electron showers that have an incomplete initial region and,
- truncate reconstructed photons that have an impure initial region.

As with the *CheatingPhotonRefinement* and *CheatingCCElectronRefinement* algorithms of section 7.7, the algorithm runs after the shower reconstruction is complete. The shower refinement algorithm considers all significant showers in the event and, for each:

1. Finds **all** pathways that look to connect the shower to the neutrino vertex. These pathways are termed ‘connection pathways’.
2. Determines whether each connection pathway belongs to the shower, and hence whether the shower is an electron/photon shower.
3. Adds(Removes) the hits of the connection pathway into(from) the reconstructed particle if it is believed to be an electron(photon).
4. Re-clusters any removed connection pathways so as to not affect the neutrino energy estimation.

The connection pathway is a 3D object, but the reconstruction issues that truncate electrons and extend photons malform the corresponding 3D hits, which were created in response to the 2D pattern-recognition. Step one of the above workflow is therefore composed of two steps:

(a) Find the 2D hits of the connection pathway in each 2D view.

(b) Construct the 3D connection pathway by matching the 2D pathways across the three views.

8.2 The Hybrid Shower Refinement Algorithm

From a development point of view, the shower refinement algorithm can be factorised into two elements:

1. the algorithm mechanics, and
2. the decision making, i.e. whether the connection pathway belongs to an electron/photon.

This allows for an intermediate stage to be defined in the development of the shower refinement algorithm. This is where one develops the reconstruction methods to perform all tasks except for the decision making (and the re-clustering of removed photon connection pathways), which is then cheated, using truth information from the simulation. This intermediate stage algorithm is referred to as the ‘hybrid algorithm’ and it enables the impact of real reconstruction inefficiencies on the predicted sensitivity of section 7.7 to be understood at an immediate stage, so that we can decide whether to continue its development. The remainder of this section is devoted to demonstrating its operation with the use of the CC ν_e event of figure 8.1.

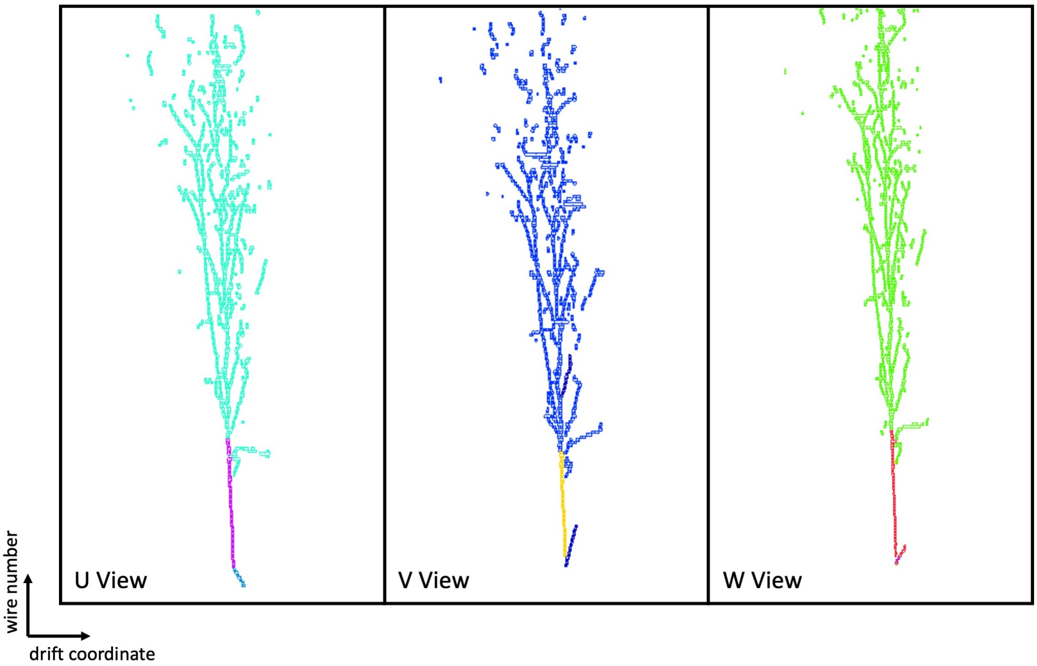


Figure 8.1: The reconstruction output in the U, V and W views of a CC ν_e event in which the primary electron shower is reconstructed as two separate particles i.e. it is split in the reconstruction. Each colour denotes an individual cluster, some colours may be repeated.

8.2.1 Finding the 2D Connection Pathways

First, the connection pathways are identified in each 2D view, via the enumerated procedure:

1. Find the Initial Directions of the Potential Connection Pathways

(a) Define a 'region of interest' (RoI), which is given by a 2D cone that is bounded by the angular extrema of the reconstructed shower and the neutrino vertex.

(b) Construct the angular distribution of the hits that lie within the RoI, and close to (within 10cm of) the neutrino vertex, and apply to it a simple smoothing algorithm, which averages each bin by its two neighbours. Here, the 'close to' distance of 10cm was tuned by eye rather than by an automated minimisation procedure - this approach motivates all thresholds used in this algorithm.

(c) Obtain an initial direction from each peak identified in the distribution. There is no threshold, to ensure that the signal connection pathway is found.

These concepts are illustrated, via our example CC ν_e event, in figure 8.2. In the V view, in this figure, we have identified the initial direction of two pathways that could lead to the electron shower: one belongs to the sought electron, the other to a proton.

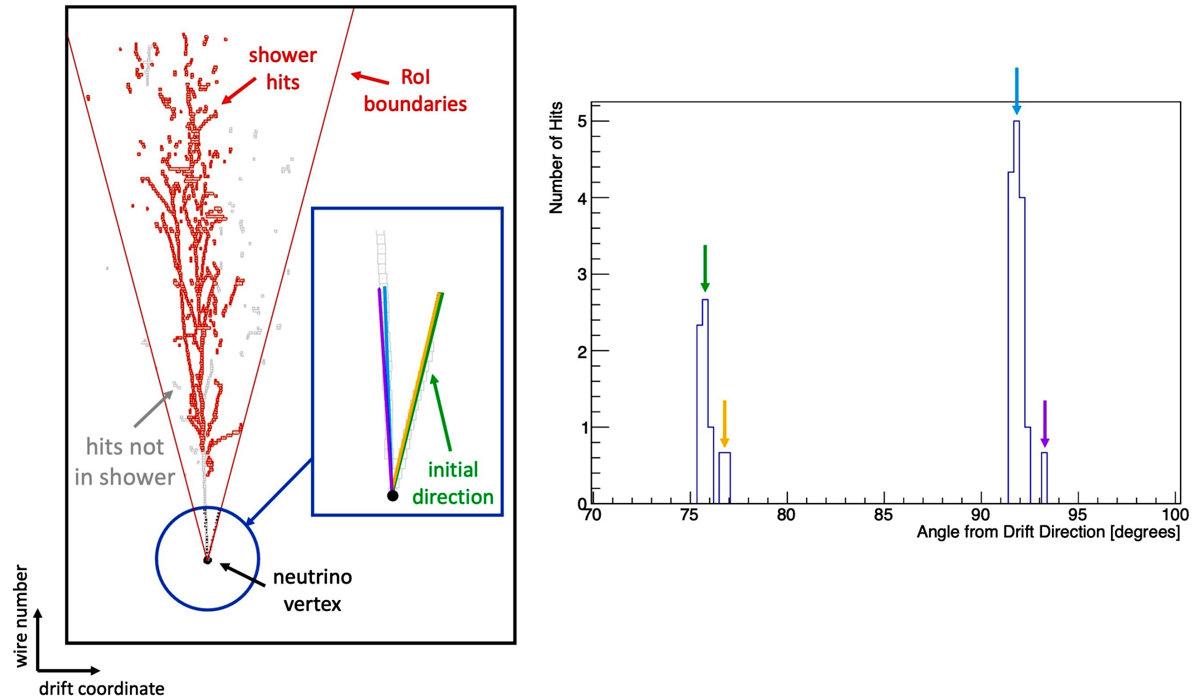


Figure 8.2: The initial directions (left panel zoom) out of the neutrino vertex, as obtained from the angular distribution (right panel) of the hits within the RoI boundaries. The colours of the initial direction vectors match the peaks of the angular distribution histogram from which they were obtained. The leftmost(rightmost) initial direction vector corresponds to the path of an electron(a proton).

2. Find the Shower Spines

Each initial direction is now followed and a pathway of hits collected. For the sought connection pathway these hits should lead into the shower and so this pathway is termed the 'shower spine'. The initial pathway directions are investigated in order of their corresponding peak height, starting from the largest, to make sure that the completeness of the sought connection pathway is optimised.

- For each found initial direction, the shower spine finder walks along the corresponding shower spine, its steps defined by a start position, direction, and length. In each step, hits are collected if:

  (a) their projection onto the step's path lies between the step's start and end points, and

  (b) they lie within 0.75cm of the step's path.

- The initial step start position and direction are equal to the neutrino vertex position and the found initial direction, but, in later steps, these are obtained from a sliding linear fit to the collected hits. The parameters of the fit, and the hits that it considers, are tuned to allow the shower spine finder to follow the natural bends and kinks of the shower spine, and to prevent it from falling down dead-end offshoots.

- In order to not replicate any found shower spines, collected hits are protected such that they cannot be assigned to another shower spine.

Figure 8.3 illustrates the steps taken by the shower spine finder and the found 'shower spines' of the electron and proton in the example CC ν_e event.

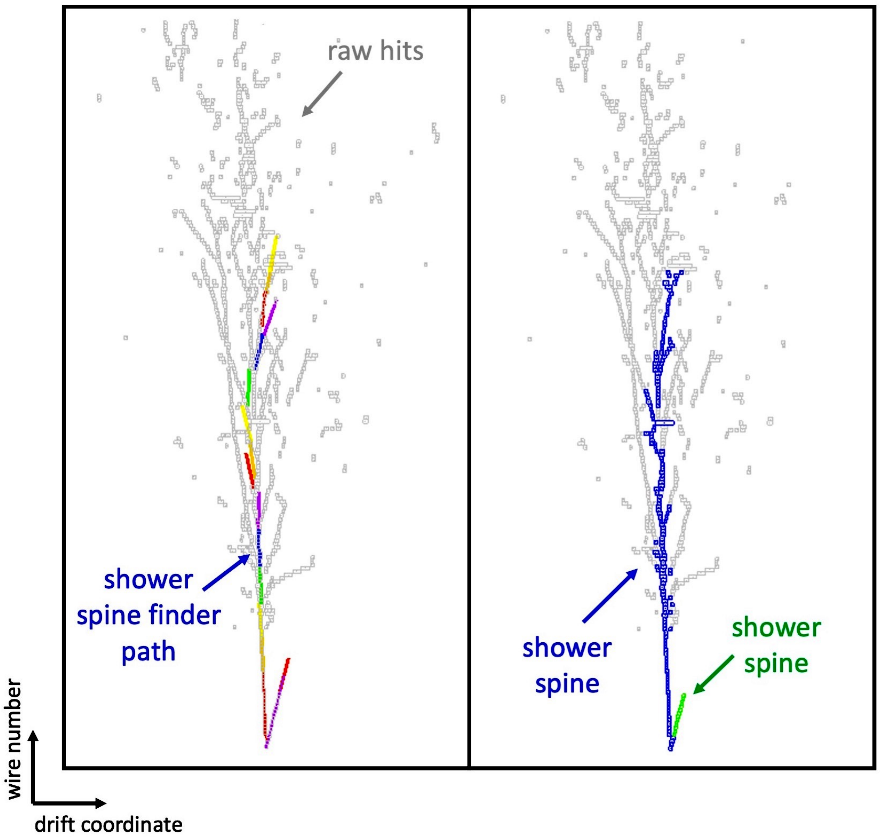


Figure 8.3: The left panel shows the two paths taken by the shower spine finder to obtain the electron and proton 'shower spines' displayed in the right panel. The steps of the shower path finder are denoted by the coloured lines, some colours may be repeated.

3. Filter Connection Pathways

In this stage, one has a first attempt at rejecting the shower spines that do not belong to the shower under investigation. For instance, in the example CC ν_e event we want to reject the proton 'shower spine'. Here, only shower spines that extend past the shower vertex are checked, in order to prevent the loss of the sought connection pathways of the shower, which may have stopped prematurely or have gaps. It is demanded that:

- 70% of the shower spine hits that lie downstream of the shower vertex belong to the shower, i.e. the shower spine must lie inside the shower once it has passed the shower vertex.

In the example CC ν_e event, we are unable to remove the proton 'shower spine' from consideration, but we will have another chance in the 2D→3D matching stage and the electron/photon identification stage.

4. Finding the Shower Start Position

We now identify the shower start position, this is the position at which the shower cascade looks to begin i.e. the position associated with the initial Bremsstrahlung radiation from an electron, or the initial pair production from a photon. It will be used to truncate the shower spine to obtain the hits of the connection pathway. To identify the shower start:

(a) Candidate shower start positions are identified as positions where the dE/dx deviates significantly from its initial value.

(b) The candidate positions are investigated, in order of proximity to the neutrino vertex, and the shower start position is identified as the first position at which a 'sensible' downstream shower can be found. An example of a 'sensible' shower is illustrated in figure 8.4, and is defined to be one where:

    (i) the shower boundaries are within 45° of the local shower spine direction, i.e are forward-going,

    (ii) the extrapolation of the shower spine from the shower start position lies between either the first or last boundary points and,

    (iii) the opening angle of the shower exceeds 2°, where the opening angle is the angle between the shower boundaries.

If no shower start position can be found (which is expected for low energy signal showers, shower spines that are incomplete, and connection pathways with true gaps), the shower start will be placed at the end of the shower spine.

Figure 8.4 shows the found shower start positions in the example CC ν_e event. In the electron(proton) case, a 'sensible shower' can(cannot) be found at one of the candidate shower start positions and so the shower start is placed along(at the end of) the shower spine.

8.2.2 Matching the Connection Pathways

With the 2D connection pathways obtained, triplets of 2D connection pathways (one in each view) are sought. A match is made if the 2D shower start positions and/or the initial 2D connection pathway directions are found to be consistent.

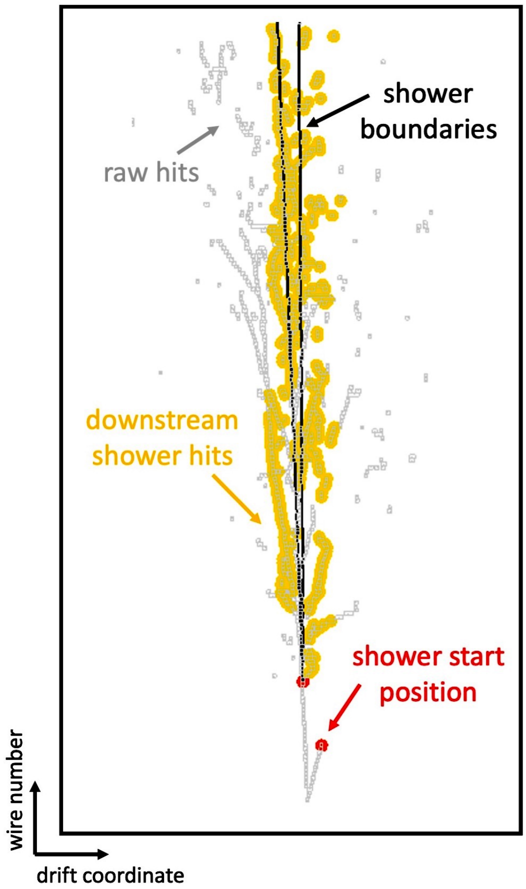


Figure 8.4: The shower start positions of the electron and proton shower spines. A 'sensible shower' is found at the shower start position of the electron, its hits are denoted by the yellow dots and its boundaries by the black lines.

Shower Start Consistency

The connection pathways are deemed to match if:

1. the difference between the shower starts in the drift-coordinate does not exceed 5cm and,
2. the average separation of the projected and actual shower start positions is less than 5cm, where 'projected' refers to the prediction in a given (third) view, based on observations in the other two views.

Initial Connection Pathway Direction Consistency

The connection pathways are deemed to match if:

1. the average opening angle between the projected and actual initial connection pathway directions is less than 5°, where the projected initial connection pathway direction of each view is calculated from the initial connection pathway directions of the other two views.

Figure 8.5 shows the actual/projected shower start positions and initial directions of the matched electron 2D shower spines in the example CC ν_e event. It is seen that the match is made via the consistency of the initial connection pathway direction, rather than the shower start position. Although not shown, a triplet of matched 2D shower spines is found for the proton 'shower spine' i.e. it is not rejected at this stage.

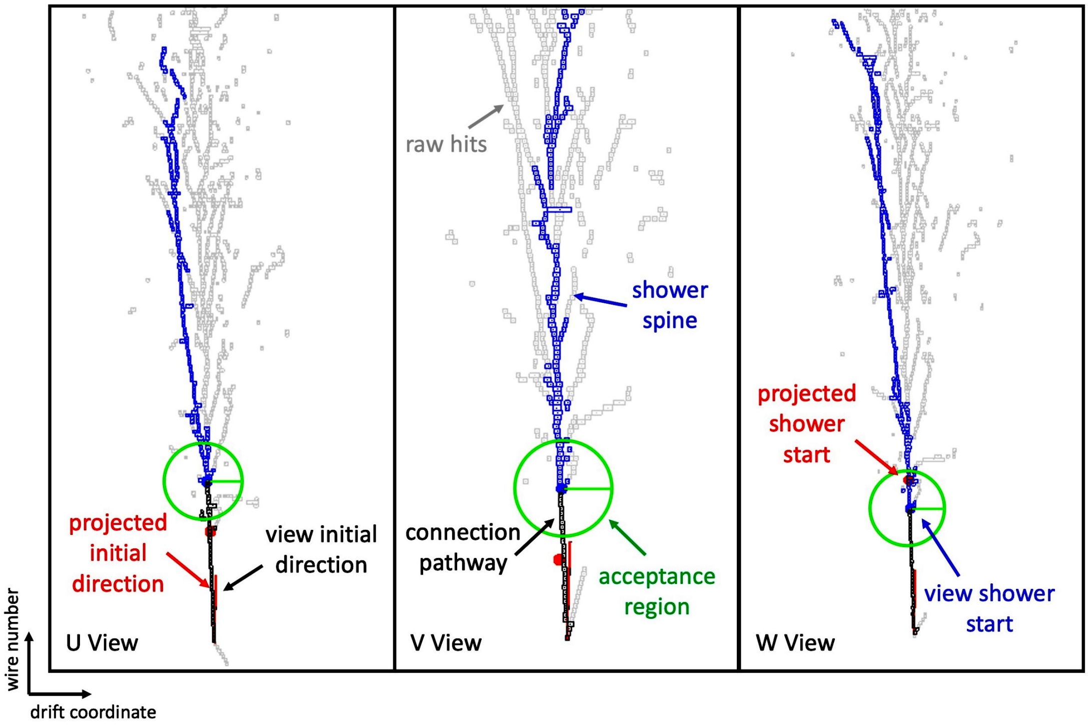


Figure 8.5: The actual/projected shower start positions and initial directions of the 2D shower spines of the electron. The radius of the green circle in each view is 5cm, which corresponds to the maximum average actual-projected shower start position separation demanded by the matching criteria. The actual initial direction vectors (black lines) are hidden beneath the projected initial direction vectors (red lines).

8.2.3 Identification of the Connection Pathway

It is now decided as to whether the 3D connection pathway belongs to an electron/photon. In the hybrid algorithm, truth information is used to cheat this decision and a 3D connection pathway is deemed to be:

- **Electron-like** if the reconstructed shower in consideration can be matched to a MC electron with a completeness and purity exceeding 33% and 50%, where hits are assigned to the MC electrons that contribute most to their charge.
- **Photon-like** if the reconstructed shower in consideration can be matched to a MC photon with the same rules as in the *CheatingPhotonRefinement* algorithm of subsection 7.7.3, and with the additional requirement that the closest reconstructed

3D position of the photon lies at least 3cm upstream of the true photon shower vertex.

The flaw of this approach is that all found 3D connection pathways are assumed to belong to the shower. So, in the example CC ν_e event, both the electron and proton 3D connection pathways will be considered to be electron-like. This is improved upon in the future iteration of this algorithm by considering the topological association between the shower and connection pathway (see section 8.4). However, this level of sophistication is sufficient for our current needs - it will be seen that only one connection pathway is only ever added to the electron shower (which, because of the ordering in which the initial directions are considered, is often the one that we seek).

8.2.4 Electron Modification

If a 3D connection pathway is deemed to be electron-like then its associated 2D hits are added into the shower. However, if the shower has multiple electron-like connection pathways, only the first matched pathway is added, so as to not damage the initial dE/dx. In each 2D view, the hit addition mechanics proceed as follows:

1. The 2D hits of the shower spine that lie upstream of the shower region are identified as possible 'hits to add'.
2. Ambiguous hits near the neutrino vertex, deemed to have a charge contribution from another particle in the event, are removed from the list of 'hits to add', in order to protect the initial dE/dx of the electron shower.
3. A continuous pathway of hits from the list of 'hits to add' is identified.
4. The identified continuous pathway hits are removed from the particles to which they currently belong, and are added into the electron shower. If a significant proportion of hits originate from a single particle, and the particle looks to be topologically associated to the electron, it is entirely merged into the electron shower.

These concepts are illustrated in figure 8.6, which shows the hits added to the electron shower in the example CC ν_e event. It is seen that only the hits of the sought electron-like connection pathway are added to the shower. This event is ultimately processed successfully.

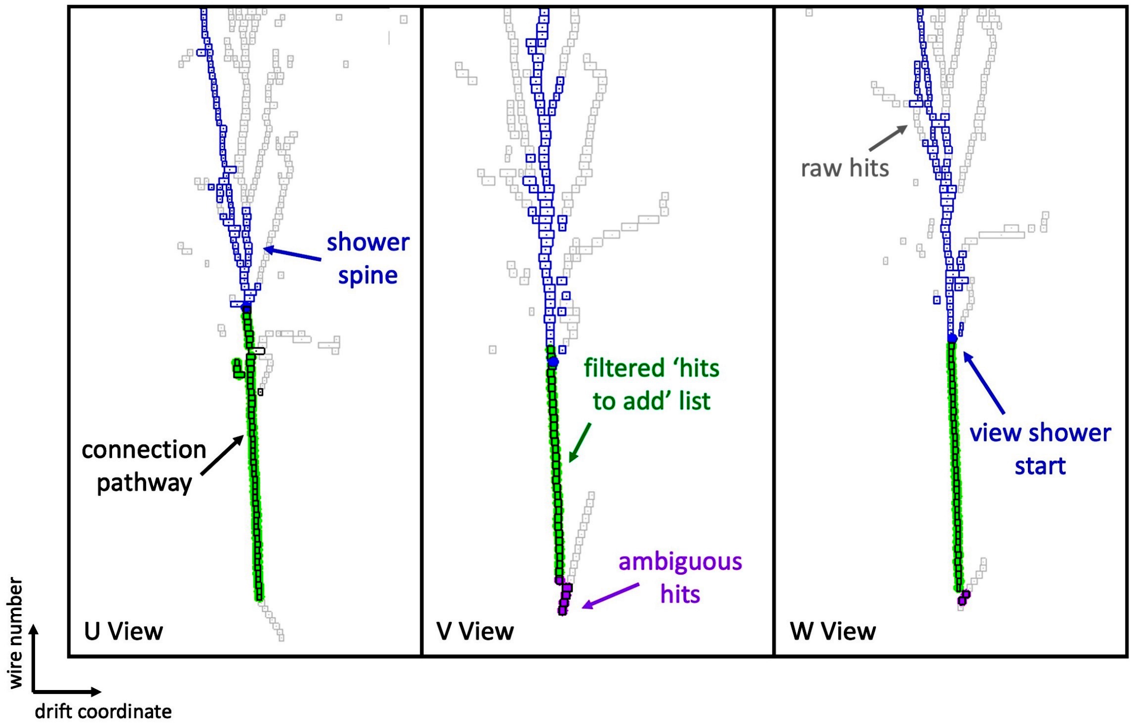


Figure 8.6: The 2D hits added (green) to the U, V and W views of the electron shower in the example CC ν_e event. The ambiguous hits, with charge contributions from both the electron and proton, are shown in purple and are not added to to electron shower.

8.2.5 Photon Modification

Finally, if at least one 3D connection pathway has been found to be photon-like:

1. The hits of all 2D connection pathways (matched and unmatched) that lie upstream of the associated 2D shower start position are removed.

2. The removed hits are re-clustered using the cheated particle creation procedure of the *CheatingPhotonRefinement* algorithm (see subsection 7.7.3).

8.2.6 Hybrid Algorithm Results

In this cheating study, several configurations of the algorithm were investigated:

- Only allow electrons to be extended.

- Only allow allow photons to be truncated.

- Allow electrons to be extended and photons to be truncated.

In each cheating study the neutrino vertex placement was also cheated. The comparison of the estimated achieved sensitivity of the hybrid configurations with the analogous cheating configurations is shown in figures 8.7, 8.8, and 8.9.

It is seen in figure 8.7 that the achieved sensitivity of the hybrid photon algorithm configuration fall just short of those of the target cheating configuration, which suggests that the connection pathway finder and photon truncation method are working well.

On the side of the hybrid electron algorithm, figure 8.8 reveals that the achieved sensitivity is significantly lower than that of the analogous cheating configuration; the peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) is 5.6σ, which is significantly below 6.0σ, obtained when the initial electron region is cheated. This is a consequence of the cautious approach to the electron extension, which is evidenced by the displacement distributions of figure 8.11. Here it is seen that the signal distribution is improved with respect to the initial Pandora performance, but not to the extent achieved by the electron initial region cheating configuration. As demonstrated in figure 8.9, this problem is not solved when both hybrid algorithms are active.

To solve this, one could prevent the removal of the ambiguous hits from the hit addition stage of the hybrid algorithm but then would risk impacting the initial dE/dx measurement. The ideal solution is to create a hit-splitting procedure in which multiple hits can be created from a single hit's charge. This is technically simple to do in the Pandora reconstruction but complications arise when the reconstruction flow moves from Pandora back into LArSoft as there is no functionality for 2:1 hit mapping.

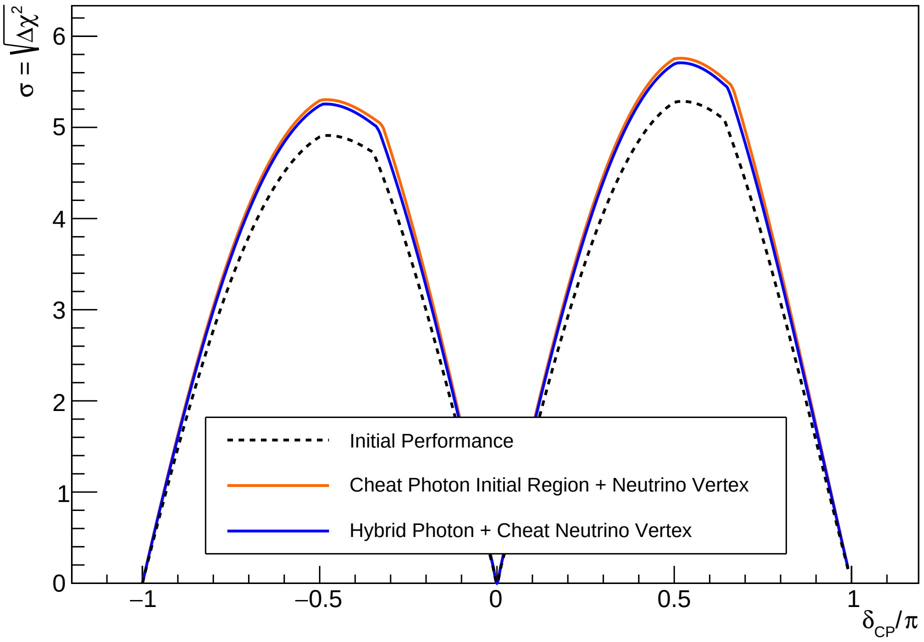


Figure 8.7: The estimated sensitivity to CP-violation achieved with the cheating of the neutrino vertex alongside: the use of the hybrid photon refinement algorithm (blue), the cheating of the reconstruction of photon initial regions (orange) as described in subsection 7.7.3. The sensitivity achieved in the standard reconstruction/selection is also shown (black).

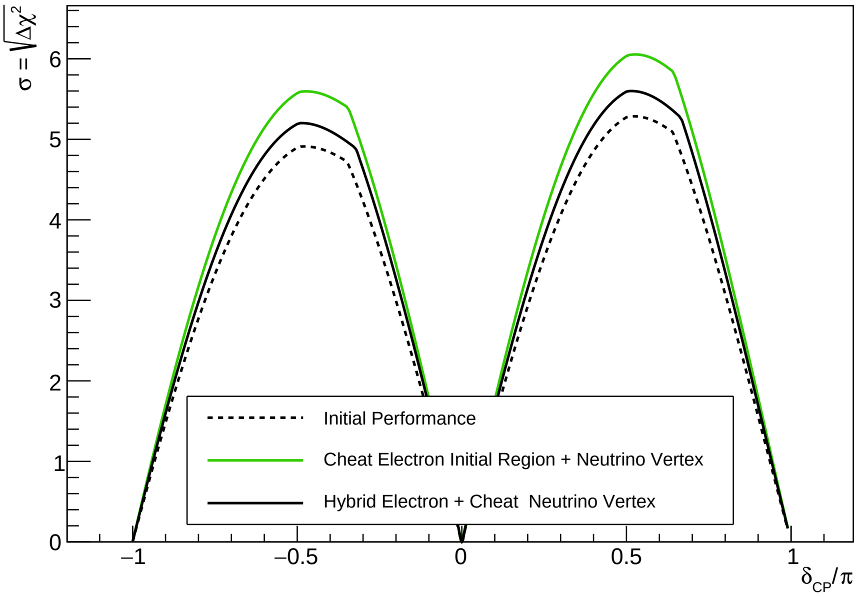


Figure 8.8: The estimated sensitivity to CP-violation achieved with the cheating of the neutrino vertex alongside: the use of the hybrid electron refinement algorithm (solid black), the cheating of the reconstruction of electron initial regions (green) as described in subsection 7.7.2. The sensitivity achieved in the standard reconstruction/selection is also shown (dashed black).

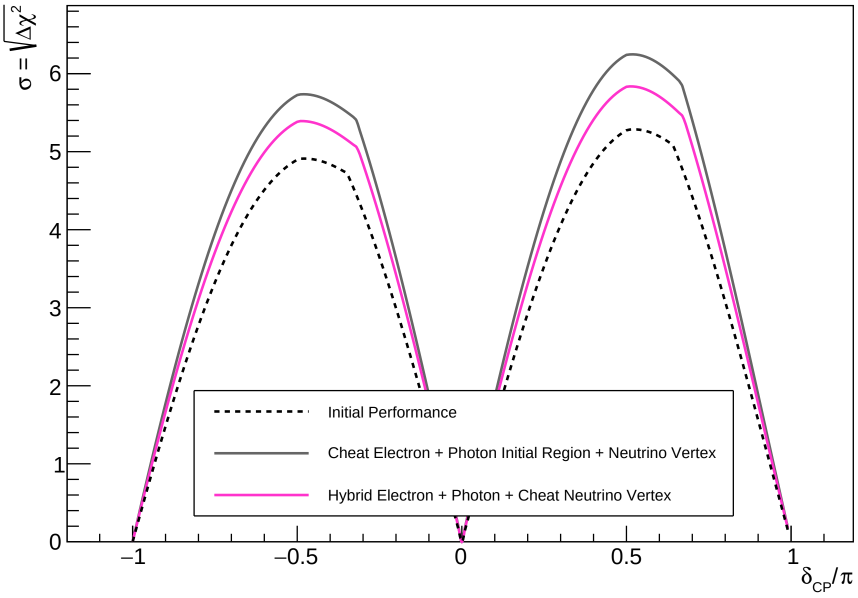


Figure 8.9: The estimated sensitivity to CP-violation achieved with the cheating of the neutrino vertex alongside: the use of the hybrid electron and photon refinement algorithms (pink), the cheating of the reconstruction of electron and photon initial regions (grey) as described in subsections 7.7.2 and 7.7.3. The sensitivity achieved in the standard reconstruction/selection is also shown (black).

8.3 Optimising the Hybrid Algorithms

The displacement variable, in the electron-like BDT, could be manually set to the electron-like value of zero for all showers that are thought to be electrons. Additionally, since electrons have a well defined dE/dx distribution, one could also set the dE/dx value to an electron-like value. This is chosen to be 2.3MeV which is the approximate peak of the signal distribution of figure 6.19. The gains achieved by doing this are shown in figure 8.10, and are discussed in the following.

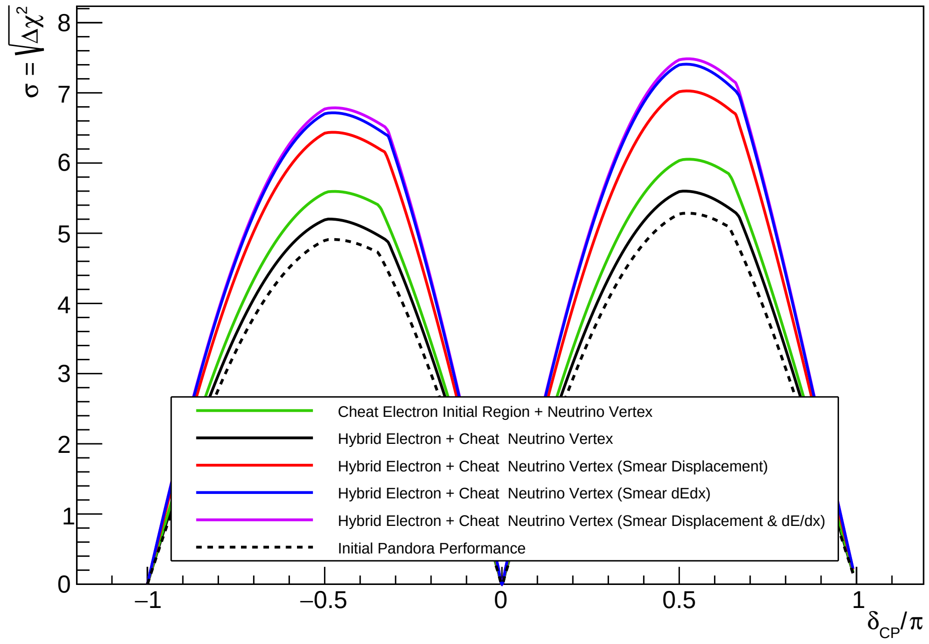


Figure 8.10: The estimated sensitivity to CP-violation achieved with the cheating of the neutrino vertex alongside: the cheating of the reconstruction of electron initial regions (green), the use of the hybrid electron refinement algorithm (solid black). The sensitivity achieved in the hybrid electron configuration when the displacement (red) and dE/dx (blue) variables are individually, and together (purple), smeared is also presented. For reference, the sensitivity achieved in the standard reconstruction/selection is shown (dashed black).

8.3.1 Smearing the Displacement

It is seen in figure 8.10, that by smearing the displacement of the thought-to-be electrons, the algorithm can achieve the CP-violation sensitivity goal of the cheated electron initial region configuration. This is because, as observed in figure 8.11, the displacement smear has recovered the separation of the displacement variable seen when cheating the initial electron region. In fact, the separation is exceeded because the smear results in a zero displacement, whereas the cheat is purposefully designed so that cheated showers obtain a 'visible vertex' and thus will have a small, but non-zero displacement.

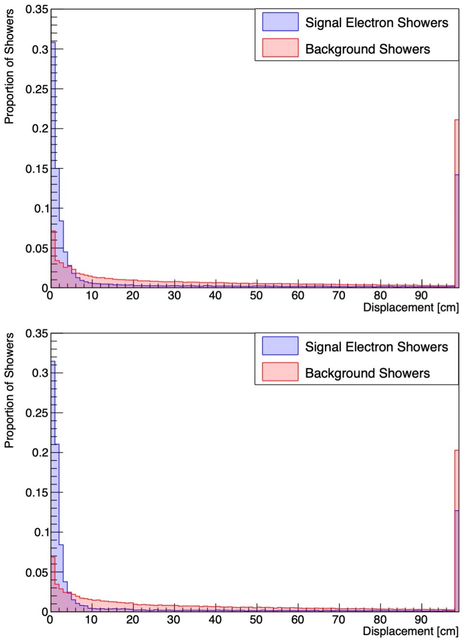


Figure 8.11: The signal (blue) and background (red) displacement distributions obtained in the standard Pandora reconstruction (top-left) and when cheating the neutrino vertexing alongside: cheating only the electron initial regions (top-right), the use of the hybrid electron refinement algorithm (bottom-left) and, the use of the hybrid electron refinement algorithm with smeared displacement (bottom-right).

8.3.2 Smearing the dE/dx

It is seen in figure 8.10 that the CP-violation improvements achieved when smearing the dE/dx are substantial and exceed that obtained by smearing the displacement and even by cheating the initial electron region. This occurs because there are many other reasons for an incorrect initial dE/dx, aside from an incomplete extension of the electron pathway. For example, the inability to share hit charge between particles could damage the dE/dx, the dE/dx calculation could be inaccurate, the charge information could be malformed and the choice of using the dE/dx of a single plane could be too simplistic. This is evidenced by figure 8.12 in which, despite the cheating of the initial electron region, a significant proportion of the signal showers lie outside the signal region.

Configuration	Efficiency	Purity
Smear Displacement	70.8%	78.4%
Smear dE/dx	69.0%	87.0%

Table 8.1: The ν_e selection efficiency and purity achieved with the cheating of the neutrino vertex placement and the use of the hybrid electron refinement algorithm, where the displacement(dE/dx) is set to 0(2.3MeV) for signal electrons.

It is interesting to relate the CP-violation gains of figure 8.10 to the integrated selection metrics detailed in table 8.1. Here it is seen that the two smearing configurations achieve the same efficiency and that the difference arises from a better purity as a result of the higher electron-photon separation power of electron-like BDT in the smeared dE/dx configuration.

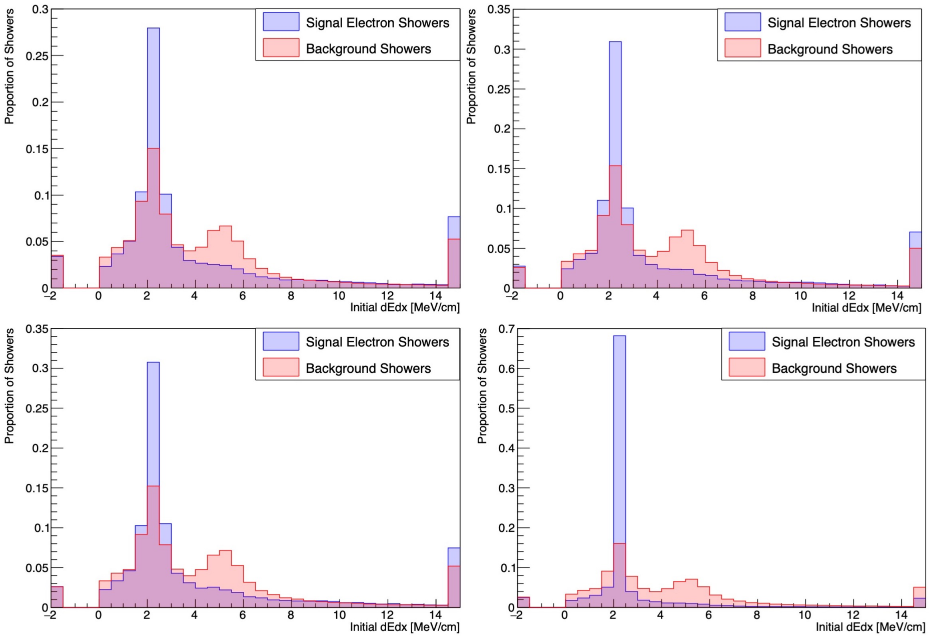


Figure 8.12: The signal (blue) and background (red) dE/dx distributions used to train the electron-like BDT obtained in the standard Pandora reconstruction (top-left) and when cheating the neutrino vertexing alongside: cheating only the electron initial regions (top-right), the use of the hybrid electron refinement algorithm (bottom-left) and, the use of the hybrid electron refinement algorithm with smeared dE/dx (bottom-right).

8.3.3 Trustworthiness of the Smearing Configuration Results

Setting a BDT variable, of identified signal showers, to the same exact value may result in the electron-like BDT picking up on the operation of the cheating configuration (as first seen in subsection 7.4.3). This should unease the reader.

In the smeared configuration, a significant proportion of signal showers are not smeared because their 3D connection pathway cannot be found by the refinement algorithms. This means that the smeared displacement and dE/dx variables of figures 8.11 and 8.12 have an observable spread. But is this enough?

The CP-violation sensitivity estimates of figure 8.10 suggest that it is. The showers that are smeared in each configuration are the same, so if the BDT had learnt the cheating configuration (that signal = smeared and background = not-smeared), one would expect the same events to be selected and rejected, and the same CP-violation results to be obtained. This is not the case. To further confirm this, one can compare the efficiency results of table 8.1 with that of a selection in which only events that have been smeared are selected. The corresponding efficiency is 79.0% which exceeds that of the smearing configurations.

8.3.4 Outlook

We've learnt that the hybrid photon algorithm is successful but that the hybrid electron algorithm falls short of the cheating electron initial region goal. However, these gains can be recovered and even exceeded if one smears the dE/dx and displacement of thought-to-be electrons. But we recall that the determination of the connection pathway as electron/photon-like is cheated and that a real reconstruction determination will mistake electrons as photons, and vice versa, and reduce the improvements demonstrated in this section. So one must now ask, is there enough motivation to continue the development of this algorithm?

A BDT is a natural choice for the replacement of the cheated connection pathway identification process. This BDT would answer the question of 'is this connection pathway electron-like in nature?' and this answer should be equal to that of the question 'is this shower an electron'. The latter is answered by the electron-like BDT that informs the $\overset{(-)}{\nu}_e$ selection. Therefore, any variables developed for the connection pathway BDT can be added into electron-like BDT, which would probably increase the electron-photon separation power of the electron-like BDT and achieve sensitivity improvements. This means that even if the accidental truncation(extension and/or smearing) of the signal electrons(background photons) washes away the improvements of the correctly extended and smeared electrons and truncated photons, sensitivity gains may still be achieved. It is therefore concluded that there is ample motivation to continue this development.

8.4 Full Reconstruction Shower Refinement Algorithm

To progress to an algorithm that does not rely on cheating, we need to replace the cheated identification of the connection pathway and the cheated re-clustering of removed connection pathways. The former task is the focus of this section. The development of the 'connection pathway BDT' is outlined in subsection 8.4.1 and its performance is presented in subsection 8.4.5.

8.4.1 The Connection Pathway BDT

The connection pathway BDT determines the likelihood that a connection pathway belongs to an electron. Signal(background) connection pathways are therefore those that are composed of charge depositions that, in truth, (do not) belong to electrons; signal(background) connection pathways are found for reconstructed shower-like particles that, in truth, are electrons (mostly photons). The interpretation of the output score is made with respect to a threshold, above(below) which connection pathways are thought to be (not) electron-like. We note that 'not electron-like' does not equate to photon-like, so additional checks are made to prevent the undesired truncation of the corresponding particles (see subsection 8.5.1).

8.4.2 Shower Parameterisation

The majority of the connection pathway BDT variables examine the 2D projections of the combined connection pathway and shower object. The probed regions will now be illustrated via the electron connection pathway found in the example CC ν_e event of section 8.2. Items in bold will be referred back to in subsection 8.4.3.

1. **The Initial Region:**
    - Detailed in figure 8.13, it is the area closest to the neutrino vertex.
    - It is parameterised, in each 2D view, by the **initial region axis**, which originates from the neutrino vertex and extends in the **initial direction** of the connection pathway. The obtainment of the initial direction was discussed in subsection 8.2.1.
2. **The Connection Pathway:**
    - Detailed in figure 8.13, it is the region between the 3D neutrino vertex and 3D shower start position, where the latter is obtained from the combination of the found 2D shower start positions.
    - It is parameterised, in each 2D view, by the **connection pathway central axis** which originates from the neutrino vertex and extends in the **connection pathway direction**.
    - The **connection pathway direction** is obtained from a Pandora sliding fit of the 2D shower spine hits and is equal to the direction at the fitted position closest to the projected neutrino vertex.
3. **The Shower Region:**
    - Detailed in figure 8.14, it is the region downstream of the 3D shower start position that resembles an electromagnetic shower.
    - It is parameterised, in each 2D view, by the **shower central axis** and the **shower boundaries**.
    - The **shower central axis** is obtained from a linear fit to the 2D shower region hits: it originates from the fitted position that is closest to the projected neutrino vertex and extends in the direction of the fit.

- The **shower boundaries** are obtained from a linear fit to the shower hits, which lie at the edges of the shower and are parameterised as straight lines, defined by the extremal points of the boundary fit.

4. **The Overlap Region:**

- Detailed in figure 8.14, it is the region around the neutrino vertex where particle trajectories overlap and hit ownership can become ambiguous.
- It is parameterised, in each 2D view, by the ambiguous hits and the **trajectories** of the particles to whom they may belong.
- The **trajectories** are parameterised as straight lines that originate from the neutrino vertex, with directions given by the particle's **initial region direction**.

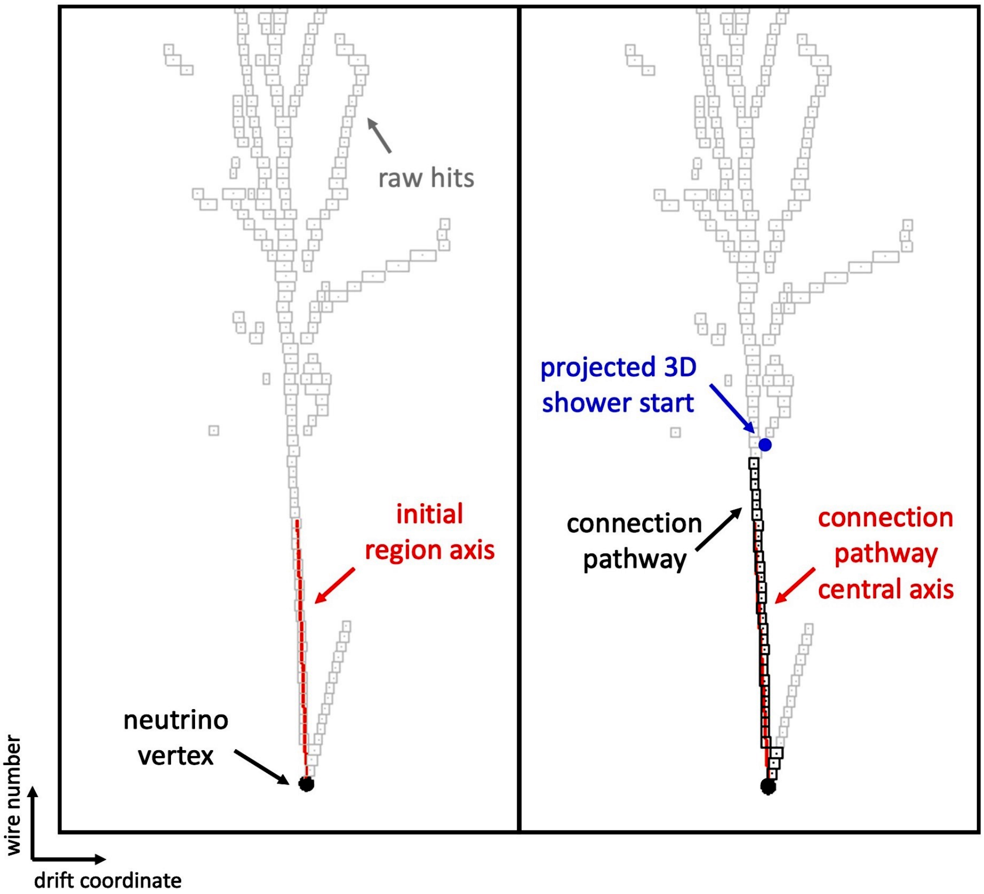


Figure 8.13: The initial region (left) and connection pathway (right) parameterisations, exampled by the CC ν_e event of section 8.2.

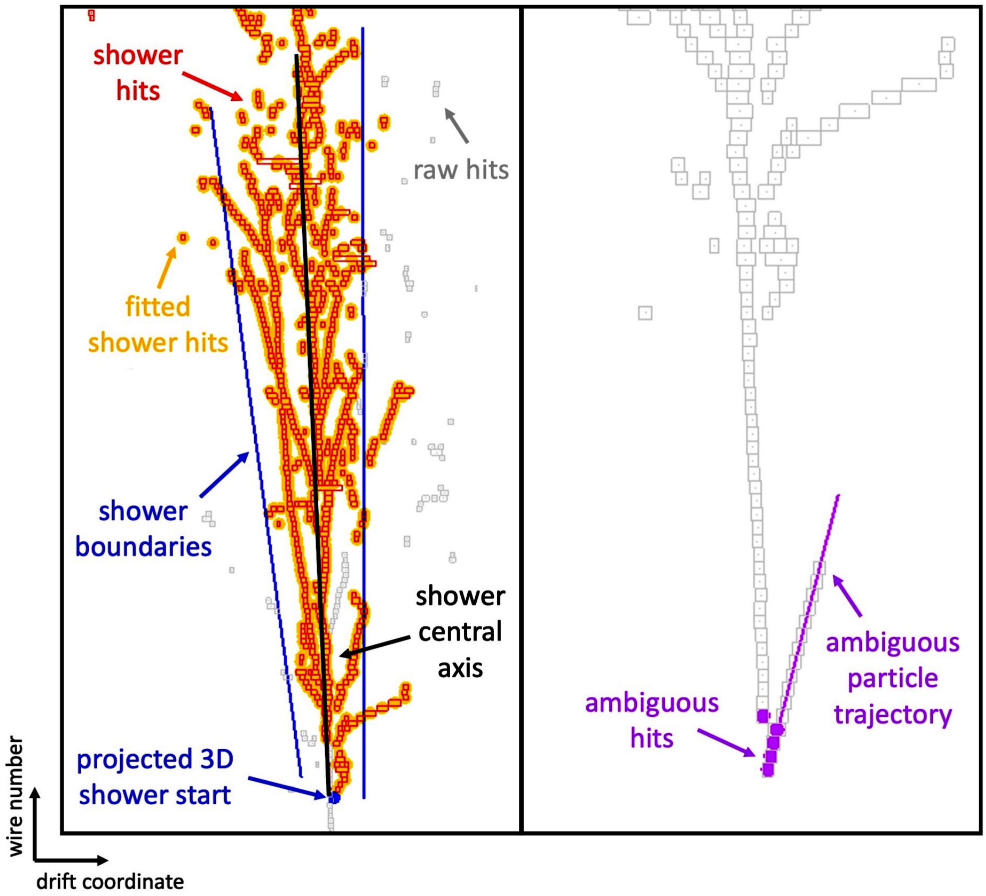


Figure 8.14: The shower region (left) and ambiguous region (right) parameterisations, exampled by the CC ν_e event of section 8.2.

8.4.3 BDT Variables

The connection pathway BDT variables are now introduced. Variables which examine an aspect of the 2D projection of the combined connection pathway and shower object will have three values (one for each view). Only the minimum; median; or maximum value is considered by the BDT, where that chosen is the one that optimises the performance of the BDT. To demonstrate the relative importance of each variable in the connection pathway BDT, this subsection will conclude with table 8.2, which lists the BDT variables in order of their separation power.

The Initial Region

These variables are motivated by the expectation that the connection pathways of electrons are continuous will be seen to originate from the neutrino vertex. This is in contrast to those of the background photons as they are likely to be built from an amalgamation of particles.

1. **2D Initial Region Gap Size (IR 1):** The largest projected gap between consecutive spine hits in the initial region, where the projection is onto the **initial region**

axis.

2. **2D Initial Gap Size (IR 2):** The distance between the neutrino vertex and the closest projected spine hit.

Figure 8.15 presents the initial region variable distributions from which the BDT learns. The upper bin, is an overflow bin, which contains the entries that lie beyond the histogram axis. All training events must be contained within these histograms, and so, if it is possible that a variable cannot be evaluated, an underflow bin is created to hold these events, which bounds the negative region.

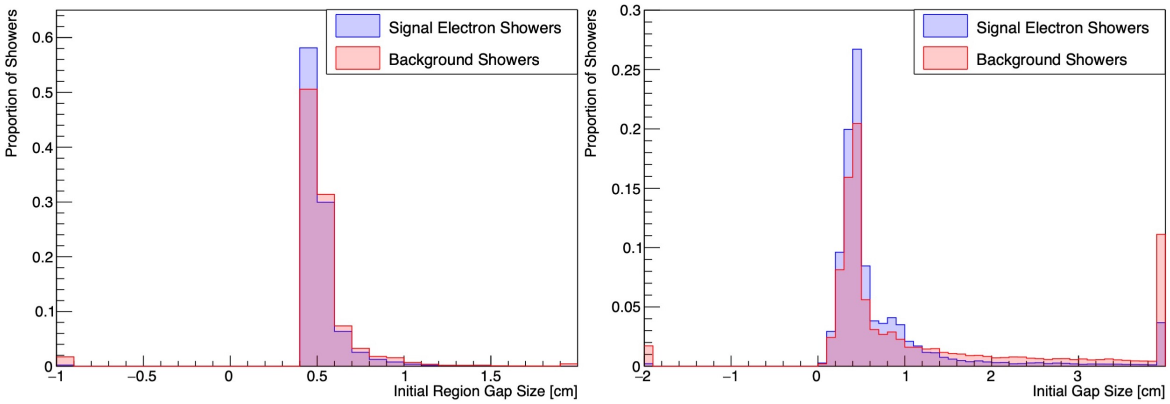


Figure 8.15: The signal (blue) and background (red) initial region gap size (left) and initial gap size (right) distributions.

The Connection Pathway

These variables are inspired by the expectation that the connection pathways of photons, that in truth do not originate from the neutrino vertex, follow the trajectories of other particles in the event and therefore are expected to be longer than the connection pathways of electrons and exhibit kinks.

1. **3D Pathway Length (CP 1):** The separation between the 3D shower start position and the neutrino vertex.

2. **2D Connection Pathway Deviation (CP 2):** To obtain this, the shower spine is fitted between the neutrino vertex and the projected shower start position and the fitted trajectory is split into equal sized segments. One considers triplets of consecutive segments such that a kink is searched for in the middle segment and the outer segments are used to smooth out kinks that correspond to ‘bumps’ rather than a net direction change. The smoothed opening angle of each segment is calculated and the connection pathway deviation is the maximum across all segments.

The achieved separation is shown in figure 8.16.

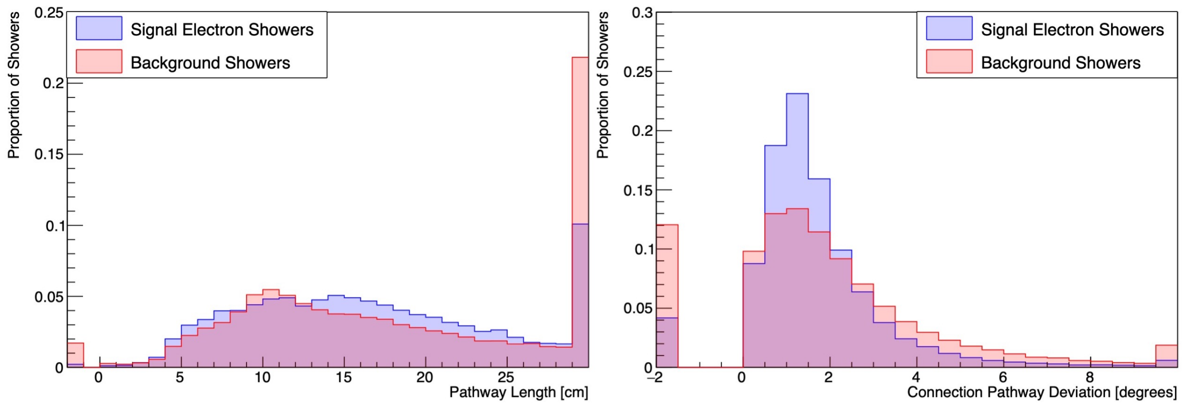


Figure 8.16: The signal (blue) and background (red) pathway length (left) and connection pathway deviation (right) distributions.

The Shower Region

The following variables are motivated by the expectation that the reconstructed shower region of a true electron should resemble a shower topology. These variables separate the reconstructed showers of electron and photons from those of particles that exhibit a more track-like structure, like charged pions.

1. **Number of 2D Shower Hits (SR 1):** The number of hits in the shower region.
2. **2D Opening Angle (SR 2):** The angle between the **shower boundaries**.
3. **2D Shower Energy Asymmetry (SR 3):** The absolute difference in the fraction of energy that lies either side of the **shower central axis**.
4. **2D Molière Radius (SR 4):** The radius of the cylinder, centred on the **shower central axis**, that contains 90% of the shower region energy.

The achieved separation is shown in figure 8.17. Here, the peak of the 2D Molière radius signal distribution is seen to disagree with the known 10cm Moliere radius for liquid Argon [119]. This occurs because this variable is calculated in 2D, rather than 3D.

These variables also provide some separation between electrons and photons as the shower region of the latter is more likely to contain contaminants and thus look nonsensical. This affect is probed more directly with the:

5. **Found 2D Hit Ratio (SR 5):** The ratio:

$$\frac{\text{Number of hits in connection pathway and shower region}}{\text{Number of hits in the reconstructed shower particle}}. \tag{8.1}$$

The achieved separation is shown in figure 8.18.

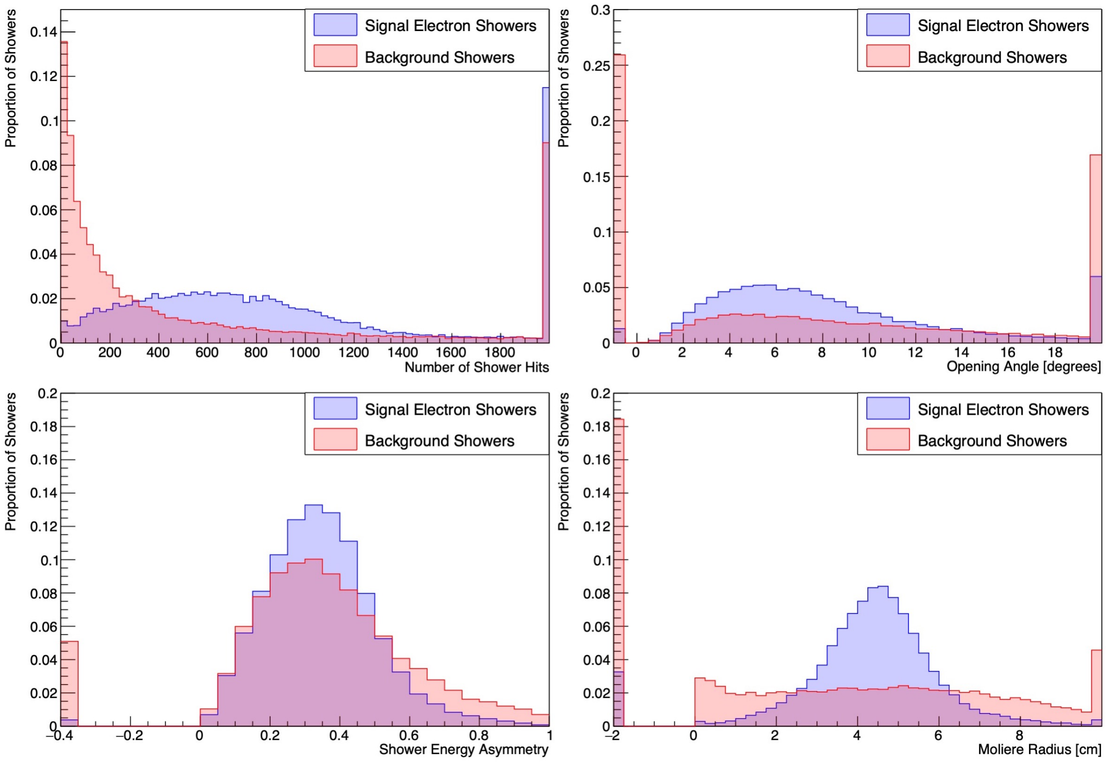


Figure 8.17: The signal (blue) and background (red) number of shower hits (top-left), opening angle (top-right), shower energy asymmetry (bottom-left) and Molière radius (bottom-right) distributions.

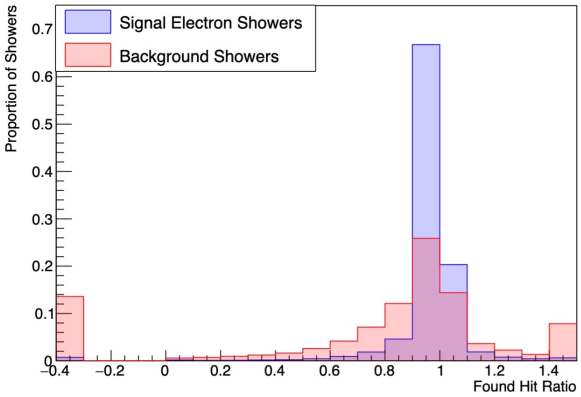


Figure 8.18: The signal (blue) and background (red) found hit ratio distributions.

The remaining shower region BDT variables are inspired by the expectation that electrons will look as though they originate from the neutrino vertex whereas mis-reconstructed photons will likely not.

6. **2D Scatter Angle (SR 6):** The angle between the direction of the **connection pathway central axis** and the **shower central axis**.
7. **2D Connection Pathway Energy Asymmetry (SR 7):** The absolute difference in the fraction of energy that lies either side of the **connection pathway central axis**.

8. **2D Connection Pathway Energy Weighted Mean Radial Distance (SR 8):** The mean radial distance of the shower region hits from the **connection pathway central axis** where the contribution of each hit is weighted by its charge.

The achieved separation of these variables is shown in figure 8.19.

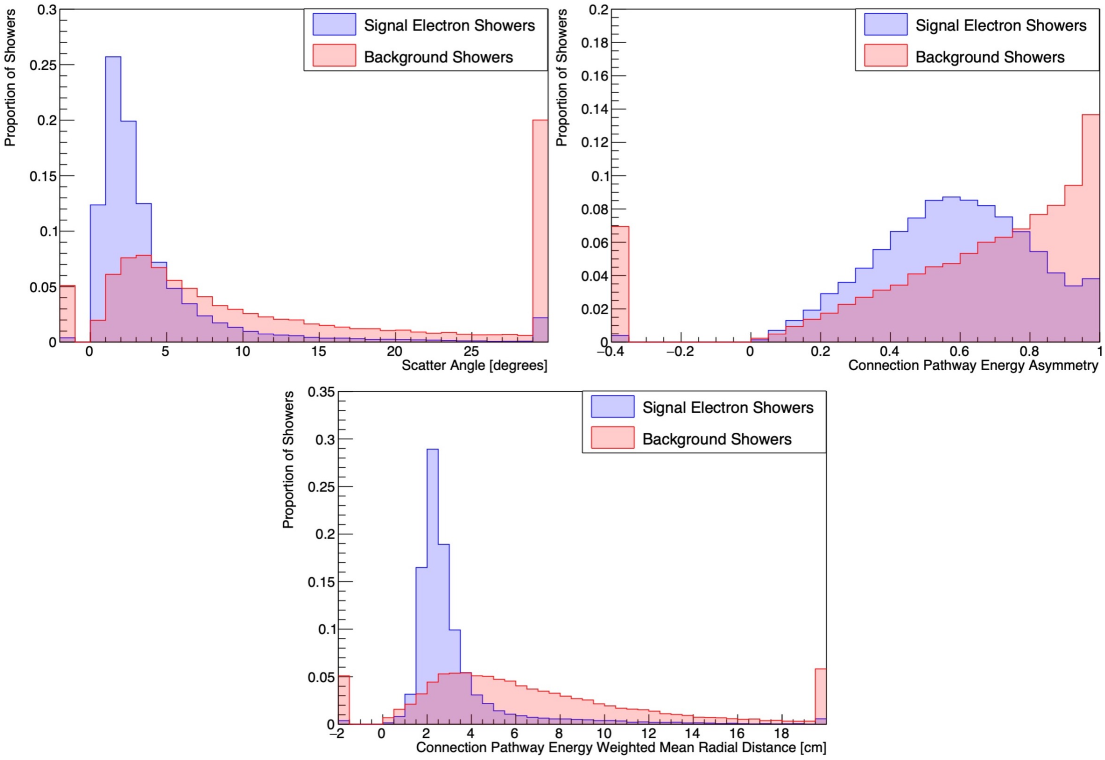


Figure 8.19: The signal (blue) and background (red) scatter angle (top-left), connection pathway energy asymmetry (top-right) and connection pathway energy weighted mean radial distance (bottom) distributions.

The Overlap Region

These variables are motivated by the expectation that the ambiguous hits of true electron connection pathways will have a charge contribution from the electron whereas those of photons, that in truth do not originate from the neutrino vertex, will not have a charge contribution from the photon.

1. **Number of Ambiguous Hit Views (OR 1):** The number of 2D views with ambiguous hits.

2. **Ambiguous 2D Hit Unaccounted Energy (OR 2):** The remaining ambiguous hit energy after the energy of the contributing particles has been removed, it is calculated by:

$$\text{ambiguous hit mean energy} - \sum \text{contributing particle mean energy}, \tag{8.2}$$

where the former is calculated with the consideration of all ambiguous hits, and the latter is calculated from a found path of continuous hits which lie on the ambiguous particle direction.

The achieved separation of these variables is shown in figure 8.20.

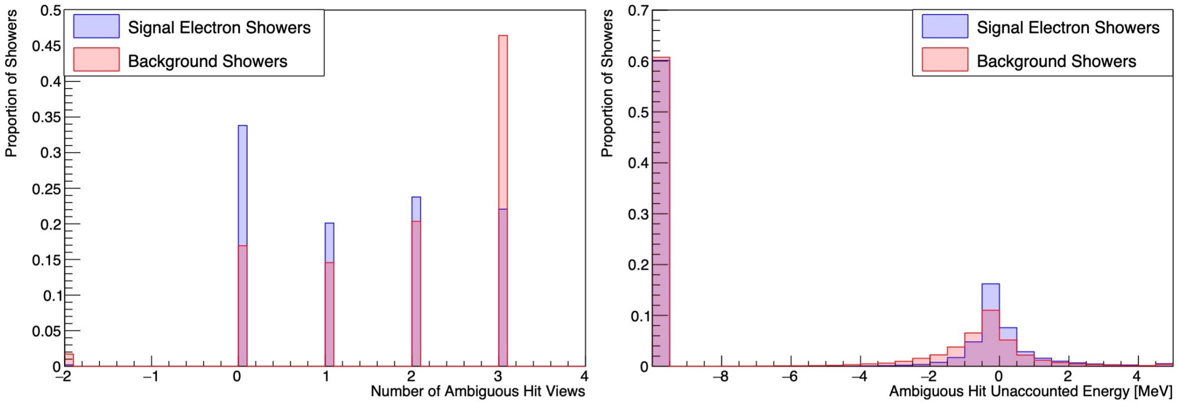


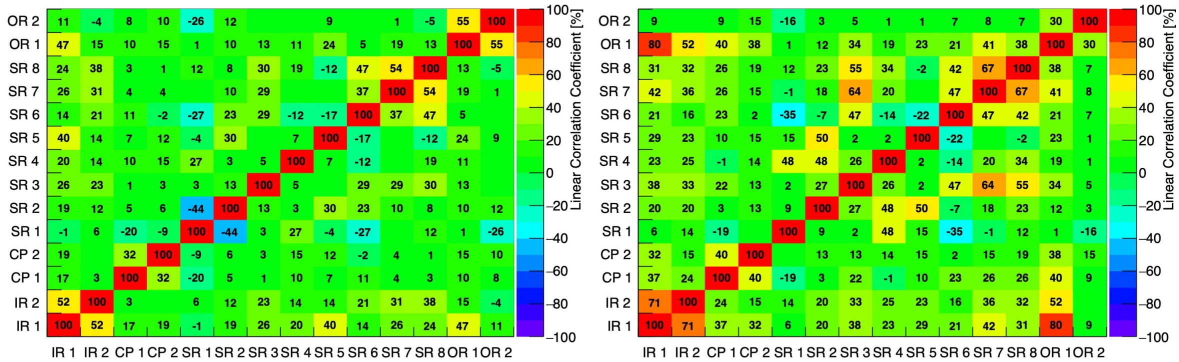
Figure 8.20: The signal (blue) and background (red) number of ambiguous hit views (left) and ambiguous hit unaccounted energy (right) distributions.

Importance Ranking	Connection Pathway BDT Variable
1	Number of 2D Shower Hits
2	2D Connection Pathway Energy Weighted Mean Radial Distance
3	3D Pathway Length
4	2D Molière Radius
5	2D Initial Gap Size
6	2D Opening Angle
7	2D Shower Energy Asymmetry
8	Found 2D Hit Ratio
9	2D Initial Region Gap Size
10	2D Connection Pathway Energy Asymmetry
11	2D Connection Pathway Deviation
12	2D Scatter Angle
13	Number of Ambiguous Hit Views
14	Ambiguous 2D Hit Unaccounted Energy

Table 8.2: The relative importance of the connection pathway BDT variables, where a lower ranking corresponds to a higher importance.

8.4.4 Connection Pathway BDT Correlations

The signal and background correlation matrices of the connection pathway BDT variables are shown in figure 8.21.



	IR 1	IR 2	CP 1	CP 2	SR 1	SR 2	SR 3	SR 4	SR 5	SR 6	SR 7	SR 8	OR 1	OR 2
OR 2	11	-4	8	10	-26	12			9		1	-5	55	100
OR 1	47	15	10	15	1	10	13	11	24	5	19	13	100	55
SR 8	24	38	3	1	12	8	30	19	-12	47	54	100	13	-5
SR 7	26	31	4	4		10	29			37	100	54	19	1
SR 6	14	21	11	-2	-27	23	29	-12	-17	100	37	47	5	
SR 5	40	14	7	12	-4	30		7	100	-17		-12	24	9
SR 4	20	14	10	15	27	3	5	100	7	-12		19	11	
SR 3	26	23	1	3	3	13	100	5		29	29	30	13	
SR 2	19	12	5	6	-44	100	13	3	30	23	10	8	10	12
SR 1	-1	6	-20	-9	100	-44	3	27	-4	-27		12	1	-26
CP 2	19		32	100	-9	6	3	15	12	-2	4	1	15	10
CP 1	17	3	100	32	-20	5	1	10	7	11	4	3	10	8
IR 2	52	100	3		6	12	23	14	14	21	31	38	15	-4
IR 1	100	52	17	19	-1	19	26	20	40	14	26	24	47	11

	IR 1	IR 2	CP 1	CP 2	SR 1	SR 2	SR 3	SR 4	SR 5	SR 6	SR 7	SR 8	OR 1	OR 2
OR 2	9		9	15	-16	3	5	1	1	7	8	7	30	100
OR 1	80	52	40	38	1	12	34	19	23	21	41	38	100	30
SR 8	31	32	26	19	12	23	55	34	-2	42	67	100	38	7
SR 7	42	36	26	15	-1	18	64	20		47	100	67	41	8
SR 6	21	16	23	2	-35	-7	47	-14	-22	100	47	42	21	7
SR 5	29	23	10	15	15	50	2	2	100	-22		-2	23	1
SR 4	23	25	-1	14	48	48	26	100	2	-14	20	34	19	1
SR 3	38	33	22	13	2	27	100	26	2	47	64	55	34	5
SR 2	20	20	3	13	9	100	27	48	50	-7	18	23	12	3
SR 1	6	14	-19		100	9	2	48	15	-35	-1	12	1	-16
CP 2	32	15	40	100		13	13	14	15	2	15	19	38	15
CP 1	37	24	100	40	-19	3	22	-1	10	23	26	26	40	9
IR 2	71	100	24	15	14	20	33	25	23	16	36	32	52	
IR 1	100	71	37	32	6	20	38	23	29	21	42	31	80	9

Figure 8.21: The signal (left) and background (right) correlation matrices of the connection pathway BDT distributions. The extent of correlation is indicated by the number and colour of the entry, where higher numbers and a red(purple) fill corresponds to high (anti-)correlation.

Significant correlations are seen between the BDT variables of the same group, this is expected since they probe the same region and thus are indirectly linked. For example, if the shower region does (not) align with the connection pathway then the scatter angle (SR 6), connection pathway energy asymmetry (SR 7) and connection pathway energy weighted mean radial distance (SR 8) variables will be small (large).

However, a cross-group significant correlation is seen between the initial region gap size (IR 1) and number of ambiguous hit views (OR 1) variables. A higher number of ambiguous views suggests that the connection pathway follows the trajectory of a contaminant in all views. One would expect to see a gap between the contaminant and shower particle trajectories and it is thought that this would lead to higher values of the IR 1 variable resulting in a positive correlation.

8.4.5 Connection Pathway BDT Performance

The connection pathway BDT performance is understood in terms of the BDT classification distributions and ROC curve introduced in subsection 6.4.3. The performance metrics are presented in figure 8.22.

The signal and background classification distributions are seen to be very well separated, with their peaks occurring at the extremal scores and little background(signal) contamination in the signal(background) region. The high performance of the connection pathway BDT is further demonstrated by the ROC curve whose apex lies close to the top-right corner of the figure, as desired.

This is compared to the muon-like BDT, the most performant BDT of the CP-violation analysis thus far. The area under the ROC curve is found to be 0.94 in the connection

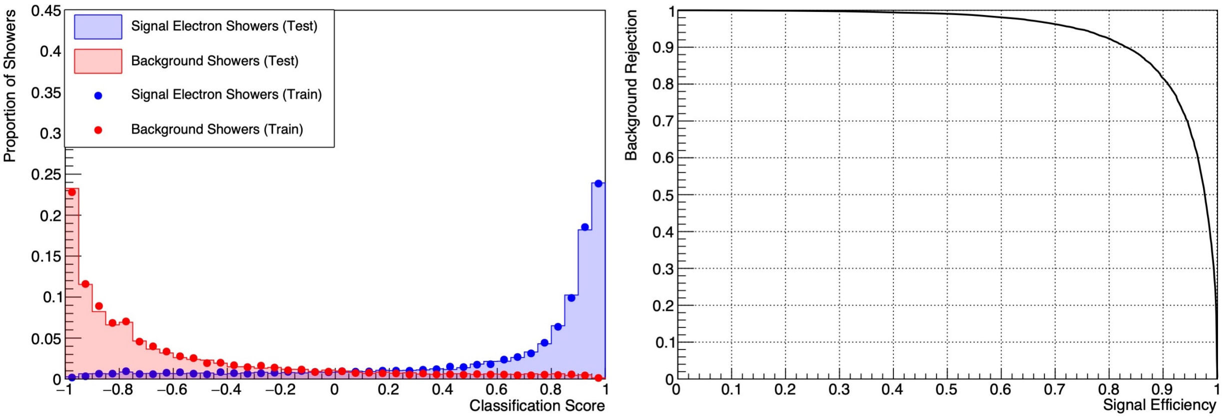


Figure 8.22: Left: the signal (blue) and background (red) classification distributions of the training (dotted) and test (fill) MC data sets. Right: the ROC curve.

pathway BDT case and 0.92 in the muon-like BDT case, which demonstrates that the connection pathway BDT outperforms the muon-like BDT. By comparing figures 6.14 and 8.22, we see that the connection pathway BDT achieves a lesser signal classification distribution peak. A better performance is ultimately achieved as a result of the lesser contamination in the signal and background regions.

8.5 Analysis Improvements

The connection pathway BDT has been seen to achieve good electron-photon separation and can now be implemented into the CP-violation analysis workflow. This section details this process and demonstrates the achieved sensitivity analysis gains.

8.5.1 Hybrid Algorithm Cheating Method Removal

The cheated electron/photon identification of the connection pathways in the hybrid algorithm is now replaced with a method that uses the connection pathway BDT. This version of the algorithm is referred to as the *ShowerRefinement* algorithm. As a reminder, the algorithm performs two tasks:

1. merges in the connection pathway of a reconstructed shower if thought to be electron-like,

2. removes the connection pathway of a reconstructed shower if thought to be photon-like.

We will first investigate a version of the algorithm in which only electron extensions can be made.

Algorithm Version: Only Electron Extension

In this algorithm, the identification of a connection pathway as electron-like requires the connection BDT score to exceed a threshold, which is tuned with respect to the $\overset{(-)}{\nu}_e$ contribution to the sensitivity, as in subsection 6.3.3. Several values, in 0.1 intervals, were tested and the optimal threshold was found to be 0.2.

Figure 8.23 demonstrates the success of the *ShowerRefinement* algorithm. With its implementation, and with the neutrino vertex placement **still cheated**, the hybrid electron refinement configuration goal is met. In fact, it is slightly exceeded, which is thought to be a consequence of the resolution of the tuning procedure.

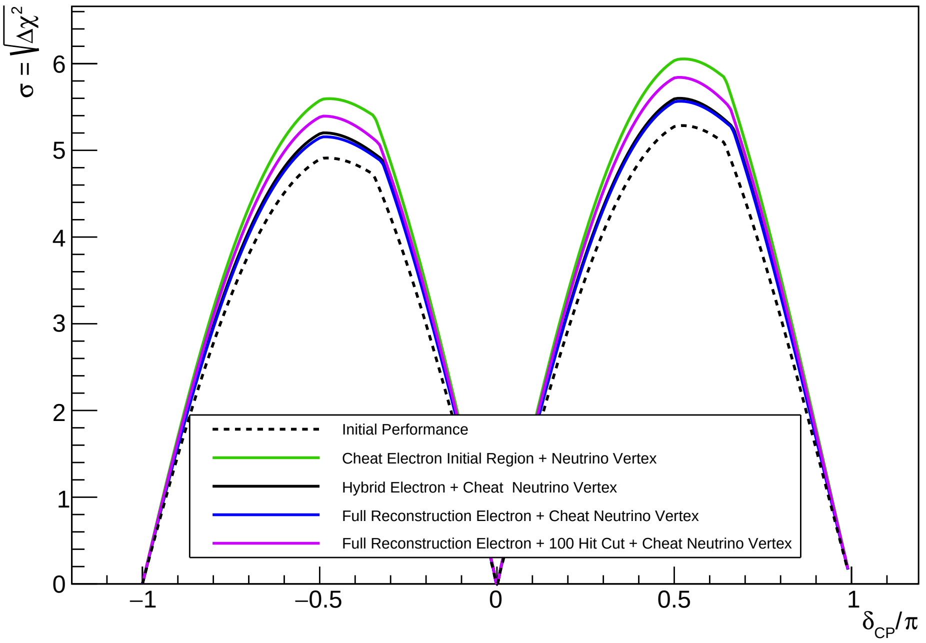


Figure 8.23: The estimated sensitivity to CP-violation achieved with the cheating of the neutrino vertex alongside: the cheating of the reconstruction of electron initial regions (green), the use of the hybrid electron refinement algorithm (solid black), the use of the *ShowerRefinement* algorithm in electron extension mode with (purple) and without (blue) a 100 hit threshold applied to the selection. The sensitivity achieved in the standard reconstruction/selection is also shown (dashed black).

At this stage it was realised that showers with a low number of 2D hits were smearing the electron-like BDT distributions, reducing the achieved electron-photon separation. A hit threshold was therefore imposed such that only showers with over 100 2D hits are used to train the BDT and can be selected, the resulting gains are shown in figure 8.23. It is worth mentioning that this modification does not move us closer to our ‘cheat electron initial region’ goal configuration, which also improves, but is still an improvement on the initial Pandora selection performance.

It was seen, in section 8.3, that significant sensitivity gains could be obtained by smearing the dE/dx and displacement BDT variables of showers with electron-like connection pathways. Unfortunately, despite the investigation of a range of electron-like thresholds, these gains could not be realised with the connection pathway BDT method. This is because the improvements gained by the smearing of correctly identified electron connection pathways are washed away by a reduced selection purity introduced by incorrect identifications.

Algorithm Version: Electron Extension and Photon Truncation

But what about a version of the *ShowerRefinement* algorithm in which both extension and truncation methods are implemented? In this version, a connection pathway is deemed to be photon-like, and thus removed, if:

- its connection pathway BDT score falls below the electron-like threshold, and
- there is evidence of a significant shower region.

The latter requirement prevents the truncation of non-photon backgrounds i.e. charged pions.

This version of the algorithm was found to result in lesser improvements despite the investigation of different electron-like thresholds. This is because the accidental truncation of true electrons reduced the efficiency of the selection, suppressing any gains achieved from the improved background rejection.

Outcome

It is concluded that the full realisation of the improvements of this algorithm are prevented by the mistakes the connection pathway BDT makes. This is not to say that the BDT performance is poor but rather that the postulated gains require an almost perfect connection pathway identification.

Working to a near-perfect performance is not the optimal route forwards as there is a more tantalising alternative: that the electron-photon separation of the electron-like BDT can be improved by merging in the variables of the connection pathway BDT. We will therefore continue this investigation, keeping the electron extension version of the *ShowerRefinement* algorithm active in the Pandora workflow.

8.5.2 Merging the Electron-like and Connection Pathway BDTs

The Enhanced Electron-like BDT

The connection pathway BDT variables are merged into the electron-like BDT to create the ‘enhanced electron-like BDT’.

The connection pathway finder can sometimes fail, and so some signal electrons will not have an associated connection pathway. If, for these cases, the connection pathway variables are assigned a default value, the connection pathway BDT variables are found to become significantly correlated. To prevent this, the training and application of the enhanced electron-like BDT is restricted to showers with a connecting pathway. This means that signal electrons, that do not have a found connection pathway, are rejected in the selection. The performance of the enhanced electron-like BDT is demonstrated in figure 8.24, in which it is compared to the electron-like BDT performance when trained on the same data set. The enhanced electron-like BDT is seen to outperform the standard electron-like BDT.

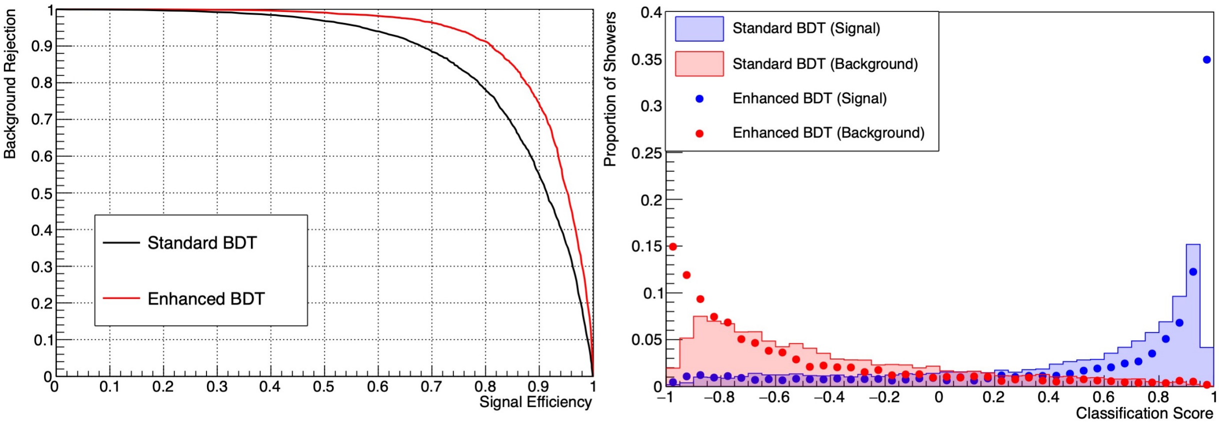


Figure 8.24: Left: the ROC curve of the standard and enhanced electron-like BDTs. Right: the signal (blue) and background (red) classification distributions of the standard (fill) and enhanced (dotted) electron-like BDTs.

The Backup Electron-like BDT

A second electron-like BDT, referred to as the ‘backup electron-like BDT’, was created to ensure the selection of the signal electrons that do not have an associated connection pathway. Since this BDT is intended as a ‘catch all’, the hit threshold is lowered from 100 to 25.

All reconstructed showers are fitted and characterised by the Pandora Modular Shower module (see subsection 4.4.2). During the fitting process, the initial track-like region of the shower is found, which is comparable to the connection pathway and can be used to evaluate the connection pathway BDT variables. Through this method, the connection pathway BDT variables were therefore implemented into the backup electron-like BDT. It was found that the addition of the pathway length, number of shower hits and connection pathway energy weighted mean radial distance BDT variables reached a saturated BDT performance and so only these variables were retained.

The performance of the backup electron-like BDT is demonstrated in figure 8.25, in which it is compared to the electron-like BDT performance when trained on the same data set. The backup electron-like BDT is seen to have slight improvements on the electron-like BDT.

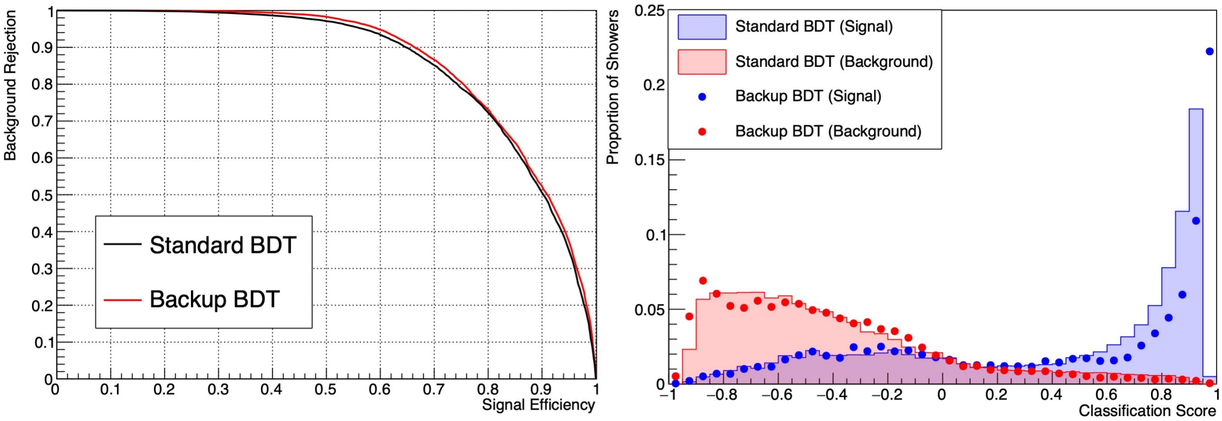


Figure 8.25: Left: the ROC curve of the standard and backup electron-like BDTs. Right: the signal (blue) and background (red) classification distributions of the standard (fill) and backup (dotted) electron-like BDTs.

CP-Violation Analysis Improvements

So what improvements does this bring to the $\overset{(-)}{\nu}_e$ selection, and by extension the CP-violation analysis? Firstly, we look at the integrated selection metrics of table 8.3.

Configuration	Efficiency	Purity	BG Rejection
Initial	61.8%	64.9%	98.4%
Improved	65.1%	85.3%	99.5%

Table 8.3: The ν_e selection efficiency, purity and background (BG) rejection achieved with the initial reconstruction/selection, compared to that achieved in the 'improved' configuration; the configuration in which the neutrino vertexing is cheated, the *Shower-Refinement* algorithm is introduced, and the two-pass selection procedure is implemented.

The efficiency of the ν_e selection is seen to increase by ∼3%, whilst the purity increases by ∼20%, which suggests that the resulting sensitivity gains are driven by purity improvements. But are these purity improvements in the ∼1GeV and ∼2-3GeV energy bands, which have been identified to be especially important when discriminating against CPC? These bands were shown, in subsection 6.6.1, to be limited by the selection of a NC background in the initial reconstruction/selection.

Comparison of the selected reconstructed energy spectrum of figure 6.25 with that of figure 8.26 demonstrates that the NC background that plagues the left-hand side of the oscillation peak has been significantly reduced. Figure 8.27 shows that this increases the discriminating power of the ∼1GeV and ∼2-3GeV energy bands and, in the case of the former, means that an enhanced sensitivity to CP-violation is now seen across a wider range of CP-phase values. As a result, significant sensitivity gains are observed in figure 8.28; the peak sensitivity (obtained at the $\delta_{\mathrm{CP}} = \pi/2$ peak) increases from 5.3σ (obtained in the initial reconstruction/selection configuration) to 6.4σ, which exceeds the 6.0σ achieved in the cheated electron initial region configuration goal.

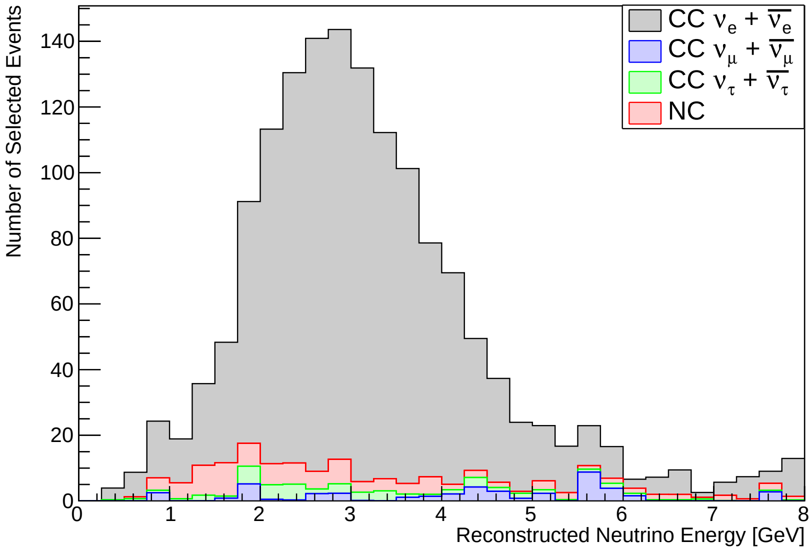


Figure 8.26: The selected ν_e reconstructed energy spectrum as a stacked histogram, where each true contribution is identified by its shading.

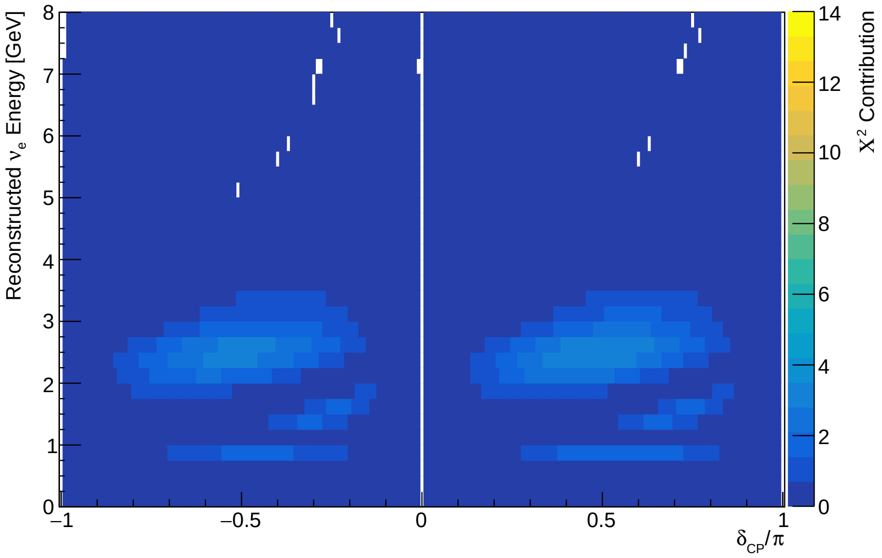


Figure 8.27: The sensitivity contribution of each bin of the ν_e selected spectrum, where the spectrum is obtained using the improved reconstruction (with cheated neutrino vertexing), two-pass ν_e selection, and reconstruction energy estimators. The range of the z-axis is chosen to enable comparisons to figure 6.26.

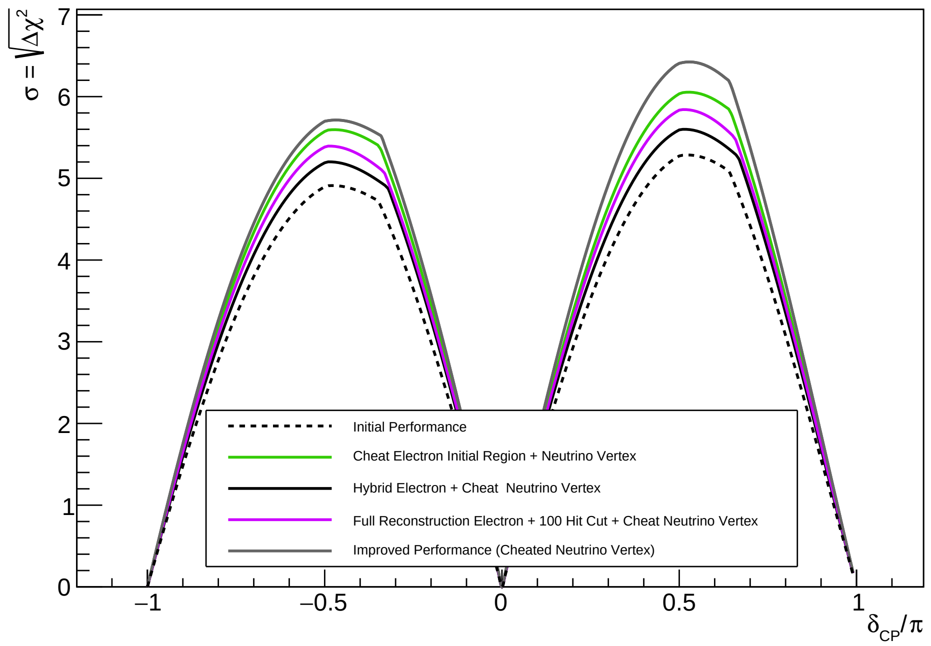


Figure 8.28: The estimated sensitivity to CP-violation achieved with a configuration which includes: the introduction of the *ShowerRefinement* algorithm, the cheating of the neutrino vertex and the implementation of the two-pass ν_e selection procedure (grey). For comparison, the sensitivities achieved in the standard reconstruction/selection (black), the hybrid electron refinement (solid black), and the cheated electron initial region (green) configurations are also shown.

8.5.3 Neutrino Vertex Cheat Removal

Hitherto, the placement of the neutrino vertex has been cheated in order to ease the development of the algorithm and the interpretation of the CP-violation analysis gains. The neutrino vertex placement will now be removed and the CP-violation analysis performance losses understood.

First the connection pathway BDT electron-like threshold was re-evaluated: the optimised threshold was found to be 0.1. The reconstruction was performed again and the enhanced and backup electron-like BDTs were retrained. The comparison of the enhanced electron-like BDT performance with respect to the electron-like BDT, when trained on the same data set, and the cheated neutrino vertex results of section 8.5.2 is presented in figure 8.29.

The enhanced electron-like BDT is still seen to significantly outperform the electron-like BDT, but achieves a lesser peak height than that obtained when cheating the neutrino vertex. This has the implication that, with the removal of the vertexing cheat, the optimal sensitivity to CP-violation is achieved with reduced enhanced electron-like and backup electron-like BDT thresholds in the $\overset{(-)}{\nu}_e$ selection. This results in a higher $\overset{(-)}{\nu}_e$ selection efficiency but a lower purity as demonstrated in table 8.4.

As seen in figure 8.30, the purity degradation is a result of the partial return of the NC background which, despite the efficiency gains, reduces the sensitivity contributions of the ∼1GeV and ∼2-3 GeV energy bands. This reduces the achieved sensitivity which is presented in figure 8.31; the peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) falls from 6.4σ to 5.8σ, with the removal of the neutrino vertexing cheat.

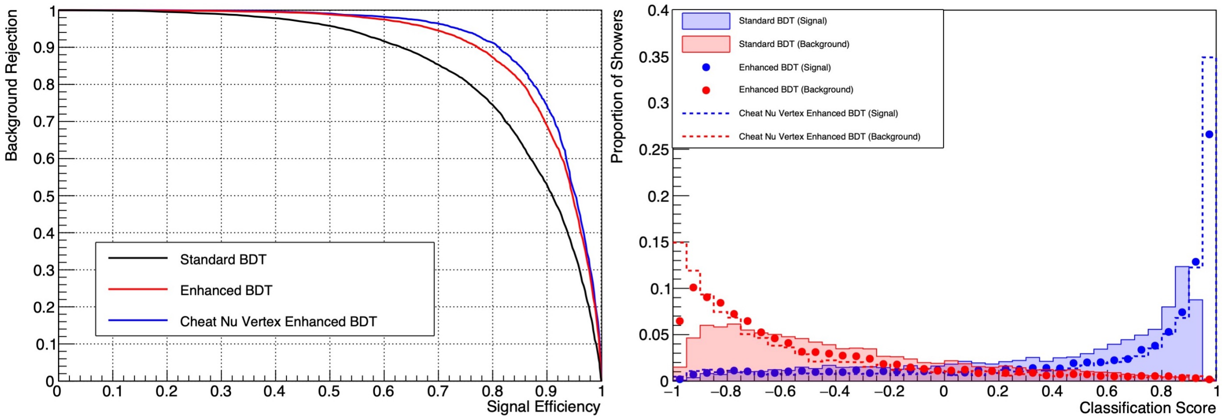


Figure 8.29: Left: the ROC curve of the standard and enhanced electron-like BDTs, where the latter is shown with and without the cheating of the neutrino vertex. Right: the signal (blue) and background (red) classification distributions of the standard (fill) and enhanced electron-like BDTs, with (dashed) and without (dotted) a cheated neutrino vertex.

Configuration	Efficiency	Purity	BG Rejection
Initial	61.8%	64.9%	98.4%
Improved (Cheated Neutrino Vertex)	65.1%	85.3%	99.5%
Improved (Standard Neutrino Vertex)	69.9%	71.1%	98.7%

Table 8.4: The ν_e selection efficiency, purity and background (BG) rejection achieved with the initial reconstruction/selection, compared to that achieved in the ‘improved’ configuration, where the *ShowerRefinement* algorithm is introduced and the two-pass selection procedure is implemented.

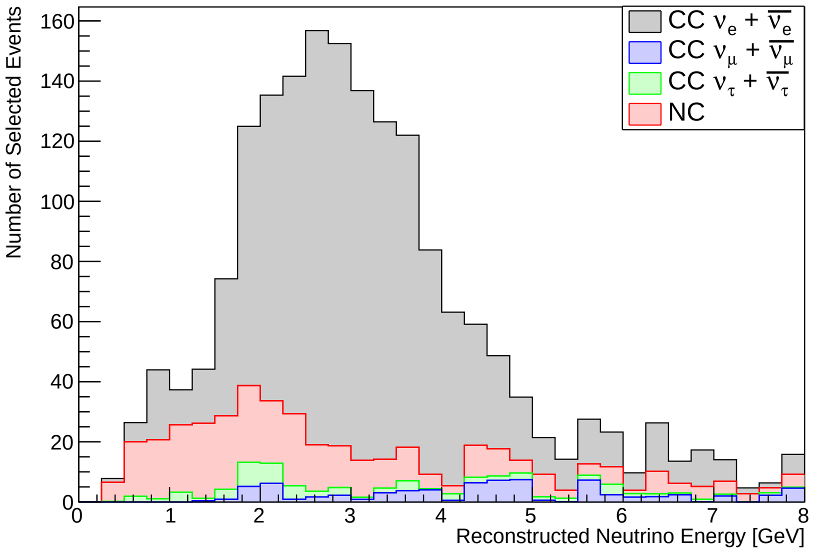


Figure 8.30: The selected ν_e reconstructed energy spectrum as a stacked histogram, where each true contribution is identified by its shading.

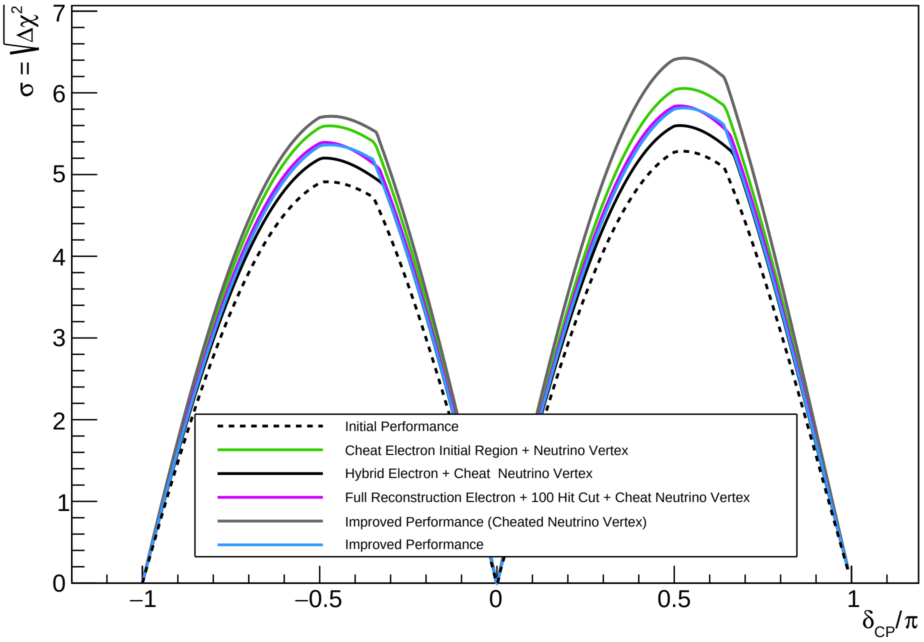


Figure 8.31: The estimated sensitivity to CP-violation achieved in a configuration which includes: standard neutrino vertexing, the introduction of the *ShowerRefinement* algorithm and the implementation of the two-pass ν_e selection procedure (cyan). This is compared to a similar configuration in which the neutrino vertex is cheated (grey). The sensitivities obtained in the standard reconstruction/selection (black), the hybrid electron refinement (solid black), and the cheated electron initial region (green) configurations are also shown.

8.5.4 Concluding Remarks

The implementation of the connection pathway BDT into the CP-violation analysis workflow has been a huge success, resulting in significant gains to the achieved sensitivity to CP-violation. These gains have the potential to be furthered with a more accurate neutrino vertex placement procedure, but this is an extremely difficult and time intensive task and has consequently not been pursued in this PhD. However, a machine-learning approach to neutrino vertexing has been developed by the Pandora team, and its incorporation into this analysis is postponed for future work.

Chapter 9

Validity of Results

The presented improvements to the CP-violation analysis have so far assumed that the underlying MC model is a perfect reflection of reality. In this chapter, we move away from this assumption.

In section 9.1, the uncertainties of the MC model will be incorporated into the CP-violation analysis metric estimates. This introduces the challenges of real-world oscillation parameter degeneracies and allows the sensitivity to CP-violation to be quantified across the distribution of possible universes given by the MC model uncertainties.

Following this, in section 9.2, we will attempt to compare the performance of the CP-violation analysis in MC and data in order to verify the extension, to data, of the conclusions reached in this thesis with a MC simulation of the DUNE FD.

9.1 Systematic Uncertainties

9.1.1 MC Model Uncertainties

The systematics detailed in this thesis are those identified by the DUNE collaboration for the sensitivity studies presented in the TDR [53]. At the time of writing, these systematics constitute the best formulation of the MC model uncertainties available to the collaboration. A significant effort is currently underway to improve upon them.

A description of the three types of systematic uncertainties to be incorporated into the CP-violation analysis is now given.

Flux Uncertainties

G4LBNF is an implementation of the beam-line model within GEANT4 and is used to predict the flux seen at the FD. The flux systematics arise from:

- The uncertainties in the quantity of, and kinematic distributions of, the hadrons that are produced in the interactions of the proton beam with the target.

- The limited precision with which the beam-line parameters, such as horn currents and target positioning, are known.

There are 30 significant flux systematics that were considered in the TDR analysis, for more information see reference [53].

Cross Section Uncertainties

Cross section systematics quantify the uncertainties within the model used to describe the interactions between the incoming neutrinos and the nuclei of the liquid argon. Some of the systematics used in this analysis exist within GENIE, in which the interaction model is implemented [90], whilst others were developed by the DUNE collaboration. A full description of the cross section parameters can be found in reference [53].

Energy Uncertainties

The energy systematics quantify the uncertainty in the estimation of the energy, both at the global scale and on the level of the individual particles. The energy uncertainties considered are:

- The energy scale and particle response uncertainties, which result in an energy shift equal to
$$E^{'}_{reco} = E_{reco}\Big(p_0 + p_1\sqrt{E_{reco}} + \frac{p_2}{\sqrt{E_{reco}}}\Big), \tag{9.1}$$
where $E^{'}_{reco}$ and E_{reco} denote the shifted and original reconstructed energy respectively and, the relative uncertainties on the parameters p_0, p_1 and p_2 are given in table 9.1. The first term is energy independent and corresponds to calibration uncertainties and the second and third terms quantify the uncertainties present at the high and low energy scales respectively [120].

- The energy resolution uncertainties, which are taken to be 2% for protons, muons and electromagnetic showers; and 10% for neutrons.

Particle Type	p_0	p_1	p_2
Total, except muons	2%	1%	2%
$p, \pi^{\pm}$	5%	5%	5%
μ	2%	0.5%	2%
n	20%	30%	30%
e, γ, π^0	2.5%	2.5%	2.5%

Table 9.1: The relative uncertainties of the energy scale and particle response parameters. Table adapted from [53]; these values match those used in the DUNE TDR CP-violation analysis detailed in reference [64].

9.1.2 Behaviour of the Systematics

Before blindly implementing these systematics into the CP-violation analysis, let's take some time to understand the behaviour of each systematic, and how it can impact our ability to measure the true value of the CP-phase. To do this, for each type of systematic, we:

1. Generate observations across the δ_{CP} phase space where, to keep things simple, we assume the central values of the oscillation and systematic parameters.
2. Apply a systematic shift to each observation, such that each systematic of the systematic type is shifted with a magnitude drawn from a Gaussian distribution of an appropriate width.
3. Search for the best-fit position of a hypothesis which assumes the central values of the oscillation and systematic parameters but allows the CP-phase to vary.
4. Compare the best-fit CP-phase value to the true CP-phase value.

The results of the execution of the above procedure are presented in this section. The improved reconstruction/selection of chapter 8 (that with the use of the *ShowerRefinement* algorithm, two-pass selection procedure, and standard neutrino vertexing) was used to create 500 observations at each of 100 equally spaced positions across the δ_{CP} phase space.

Flux Systematics

For each throw, the aforementioned CPV fit is performed simultaneously over the observed $\nu_e, \nu_\mu, \bar{\nu}_e$ and $\bar{\nu}_\mu$ spectra. It will be most strongly influenced by the dominant contributor to the sensitivity: the ν_e spectrum. Hence we will focus on the ν_e spectrum to gain our insight on how the systematic can 'throw off the fit'. Examples of thrown flux shifts are given in figure 9.1, where the ν_e spectrum is evaluated at $\delta_{\mathrm{CP}} = \pi/4$. This value is chosen as it lies away from the CPC, and maximally violating CP-phases. Here, the flux systematics are seen to scale the reconstructed energy spectra; positive(negative) systematic shifts increase(decrease) the total number of selected events.

The response of the fit is also illustrated in figure 9.1. It is seen that positive net shifts will mimic the spectrum of a CP-phase value closer to the negative maximally violating phase whilst negative net shifts mimic the spectrum of a CP-phase value closer to the positive maximally violating phase. The resulting separation between the true and best-fit CP-phase values will depend on the magnitude of the net shift, resulting in a range best-fit CP-phase positions, as shown in figure 9.2.

The fit is seen to get better at finding the true value of the CP-phase as we approach the maximally violating phases, is this a real effect? The maximally violating phases correspond to the largest positive and negative deviations from the CPC spectra and so, if the systematic shift takes the spectrum past one of these boundaries, the best-fit position will be that of the closest spectrum i.e. that of the maximally violating phase. The high accuracy at the maximally violating phase is therefore an artefact of the assumed oscillation model in the fit.

This indicates that the interpretation of figure 9.2 is nuanced but we can, with caution, use it to infer the effect on the achieved sensitivity, as done in subsection 9.1.3.

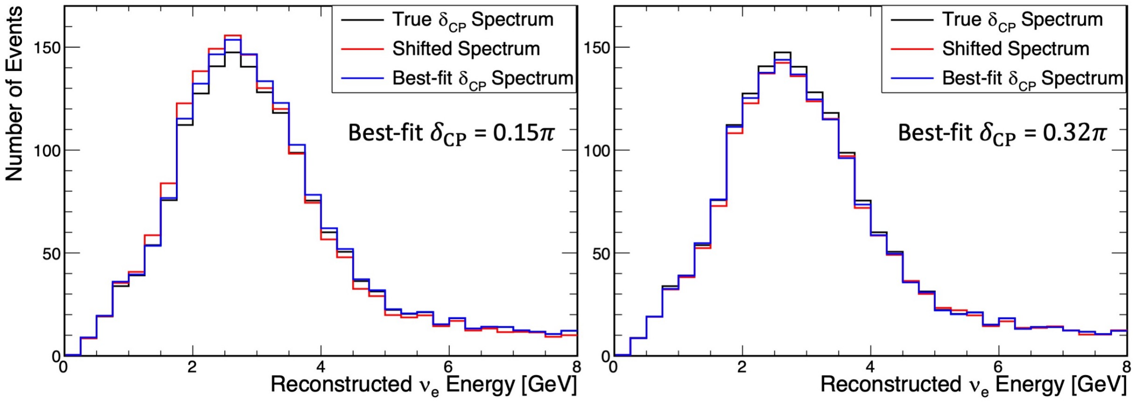


Figure 9.1: The selected ν_e reconstructed energy spectrum without any systematic shifts applied (black), where $\delta_{\text{CP}} = \pi/4$. The ν_e reconstructed energy spectrum of example throws of the flux systematics (red), and that of the corresponding best-fit CP-phase value (blue) are also shown.

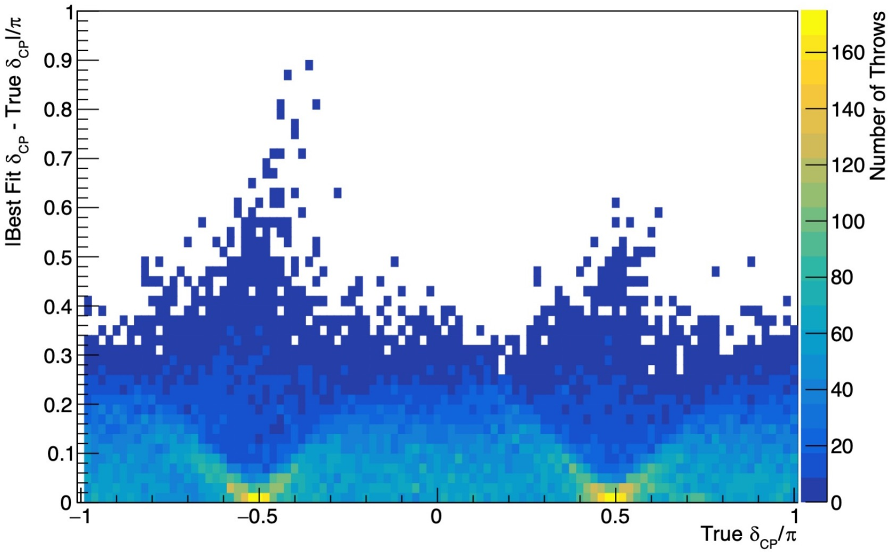


Figure 9.2: The absolute difference between the true CP-phase value, with which the spectra were generated, and the best-fit CP-phase value, where 500 throws of the flux systematics were made at each of 100 equally-spaced CP-phase values.

Cross Section Systematics

Examples of thrown cross section systematics are given in figure 9.3, where the ν_e spectrum is evaluated at $\delta_{\text{CP}} = \pi/4$. Alike the the flux systematics, the cross section systematics are seen to scale the reconstructed energy spectra and the fit responds in the same way.

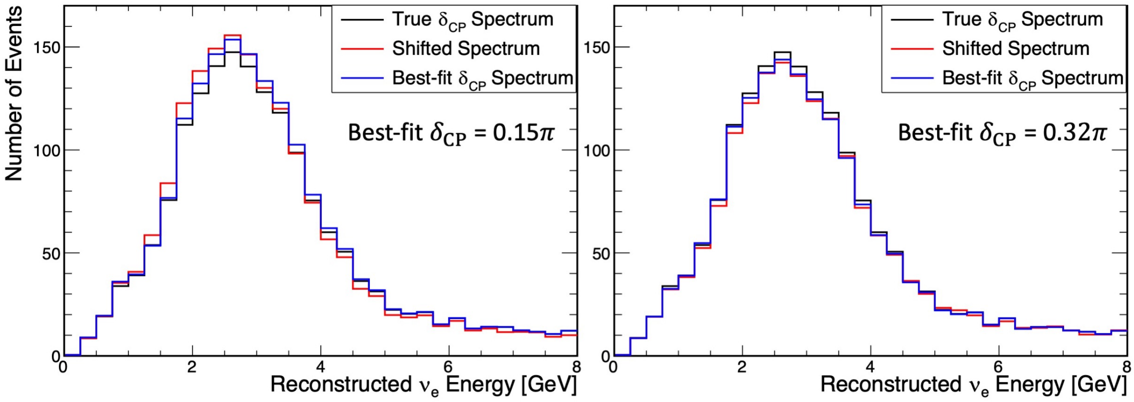


Figure 9.3: The selected ν_e reconstructed energy spectrum without any systematic shifts applied (black), where $\delta_{\text{CP}} = \pi/4$. The ν_e reconstructed energy spectrum of example throws of the cross section systematics (red), and that of the corresponding best-fit CP-phase value (blue) are also shown.

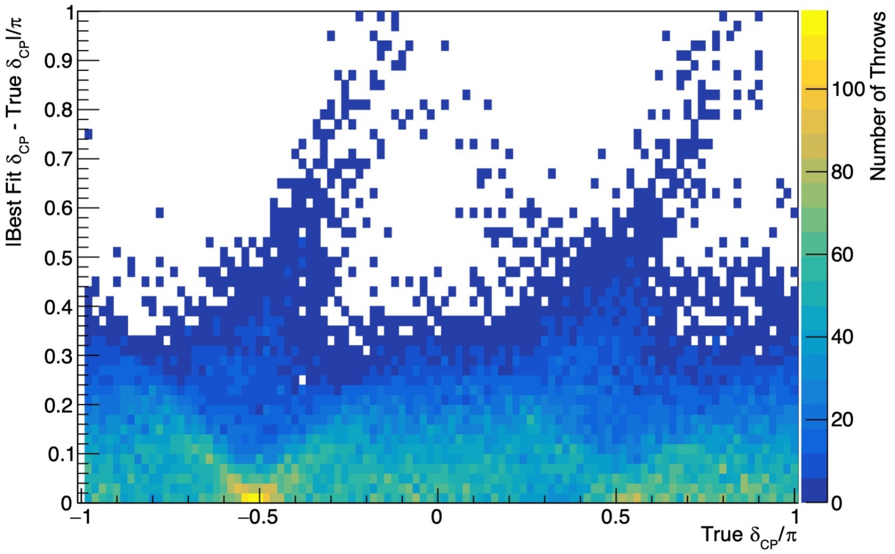


Figure 9.4: The absolute difference between the true CP-phase value, with which the spectra were generated, and the best-fit CP-phase value, where 500 throws of the cross section systematics were made at each of 100 equally-spaced CP-phase values.

The cross section systematic shifts result in a much larger deviation to the spectra than seen in the flux systematic case, and an asymmetry in the magnitude of the positive and negative shifts is observed. Figure 9.4 shows that these larger deviations result in a wider spread in the accuracy across the δ_{CP} phase space. This includes the maximally violating CP-phases, where, as a result of the asymmetry of the deviations, the impact is larger for the positive maximally violating phase.

Energy Systematics

Examples of thrown energy systematics are given in figure 9.5, where the ν_e spectrum is evaluated at $\delta_{\mathrm{CP}} = \pi/4$. The energy systematics are seen to smear the spectra; positive(negative) systematic shifts move events to higher(lower) energy bins.

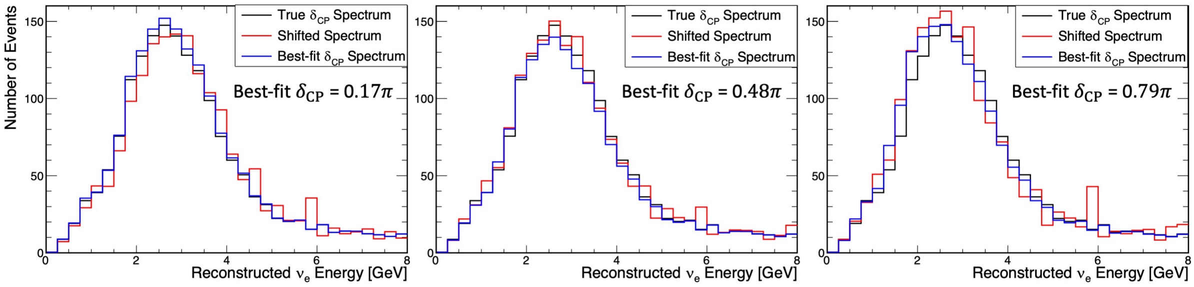


Figure 9.5: The selected ν_e reconstructed energy spectrum without any systematic shifts applied (black), where $\delta_{\mathrm{CP}} = \pi/4$. The ν_e reconstructed energy spectrum of example throws of the energy systematics (red), and that of the corresponding best-fit CP-phase value (blue) are also shown.

When the net impact of the systematic shifts is significant, the shift results in a general excess on one side of the oscillation peak and a deficit on the other. The spectrum of a CP-phase closer to the positive maximally violating phase will have less events than that of the true spectrum, and will better match one side of the shifted spectrum. The spectrum of CP-phase closer to the negative maximally violating phase meanwhile, will increase the number of events in the spectrum, such that it better matches the other side. The best-fit position will correspond to the case with the lowest chi-squared value. Again, this results in a spread of the best-fit position around the true position, shown by the red box in figure 9.6. The spread is much larger at the maximally violated values of the CP-phase, which is a consequence of the sinusoidal behaviour of the spectral deviation with the CP-phase, as explained in subsection 6.2.2.

The spectra that correspond to CP-phase values equidistant from the maximally violating phases, are shown in figure 9.7. They are seen to look similar, but to shifted in energy with respect to one another. This accounts for the results in the green boxed region of figure 9.6, where the best-fit CP-phase values correspond to the point, equidistant from, but on the other side of, the closest maximally violating phase.

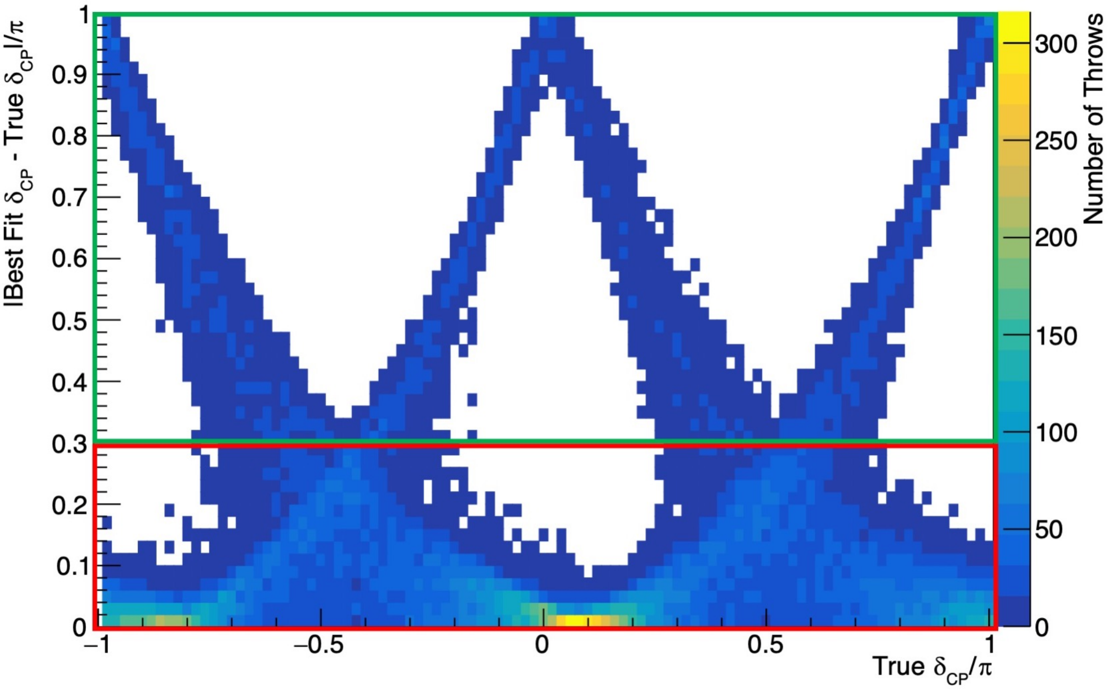


Figure 9.6: The absolute difference between the true CP-phase value, with which the spectra were generated, and the best-fit CP-phase value, where 500 throws of the energy systematics were made at each of 100 equally-spaced CP-phase values.

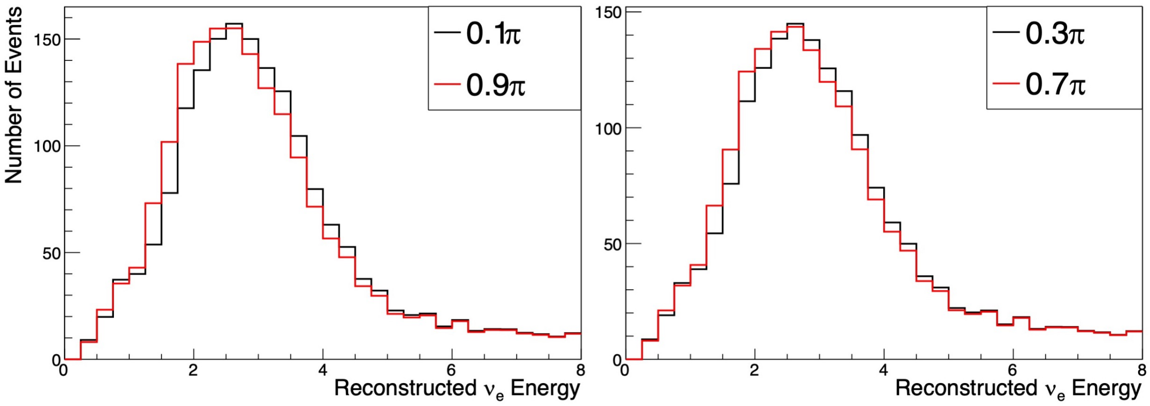


Figure 9.7: Two examples of the selected ν_e spectra, evaluated at CP-phase values that are equidistant from the maximally violating CP-phase value $\delta_{\mathrm{CP}} = \pi/2$. CP-phase values less(more) than the maximally violating phase are shown in black(red).

9.1.3 Impact on the CP-Violation Analysis

So how does this impact DUNE's sensitivity to CP-violation? Subsection 9.1.2 has demonstrated that systematic shifts can move CPC spectra closer to observations that, in truth, violate CP. This was seen before, in subsection 6.2.3, which investigated the impact of the implementation of oscillation parameter uncertainties on the sensitivity. It was then concluded that the attained sensitivity was lessened by introduced degeneracies. With this in mind, we can draw the following conclusions:

- The cross section systematics result in the largest shifts of the best-fit CP-phase value from its true value, and are therefore expected to be the most detrimental to our sensitivity. This is followed by the flux, and then the energy systematics.
- The energy systematics have the ability to shift predictions such that they resemble that of a CP-phase that is equidistant from, but on the other side of, the nearest CP-violating maxima. This does not significantly change the ability to rule out CPC and thus isn't expected to affect the sensitivity.

These postulations can be validated by a sensitivity estimate that allows the fits of equation 6.4 to explore the phase space defined by the systematic uncertainties. Again, each systematic will be investigated in isolation and to keep things simple:

- Observations will be generated assuming the central values of the oscillation and systematic parameters.
- The fits of equation 6.4 will explore the phase space of the oscillation parameters (as in subsection 6.2.3), but will assume the central values of the systematic parameters, except for the systematic under investigation, which is allowed to vary but is restrained by a penalty term.

The results are shown in figure 9.8, and are seen to agree with that predicted above.

9.1.4 Implementing the Systematics

We are now ready to implement the systematics into our sensitivity estimates. To do this, we extend the sensitivity estimate of subsection 6.2.3, where oscillation parameter uncertainties were implemented, by:

1. Generating observations across the phase space defined by the uncertainties of the oscillation parameters and systematics by 'throwing' the oscillation parameter set and applied systematic shifts.
2. Allowing the fits of equation 6.4 to explore the CPC predictions allowed by the oscillation parameter and systematic uncertainties. The systematics are implemented

in the fits as nuisance parameters and the oscillation parameters are implemented as described in subsection 6.2.3.

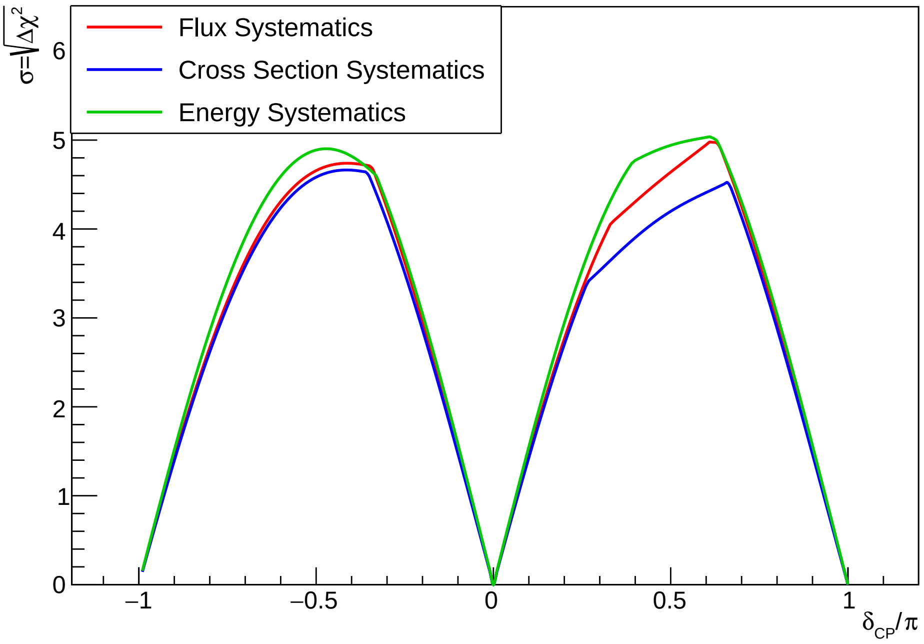


Figure 9.8: The sensitivity achieved by the improved Pandora-based CP-violation analysis (that with the *ShowerRefinement* algorithm, two pass selection procedure, and standard neutrino vertexing), with a 3.5 year exposure (1.36 $\times 10^{23}$ POT) in each beam mode (FHC and RHC). Oscillation parameter uncertainties are included in the estimate and normal-ordering of the neutrino mass hierarchy is assumed. The inclusion of the cross section, energy and flux systematics into the fit are shown.

Furthermore, to make this sensitivity estimate as complete as possible, Poisson statistical fluctuations are applied to the generated observations.

We will first consider the sensitivity achieved with the initial Pandora reconstruction and Pandora-based $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection i.e. before the *ShowerRefinement* algorithm and two pass $\overset{(-)}{\nu}_e$ selection procedure were implemented. The median sensitivity and 68% contours calculated from 600 throws at each of 200 δ_{CP} values (20 were used in the analysis of reference [64]) are shown in figure 9.9. The computational time required to perform the fits of each throw is not insignificant and, since many values of the CP-phase were sampled, the time taken to obtain the sensitivity curve is on the order of weeks. This, alongside grid system changes, ultimately capped the number of throws that were made and resulted in a lack of smoothness to the sensitivity curves however, this is not believed to hinder the results that are to be drawn. In order to estimate what one would see with a higher number of throws, a smoothing algorithm was applied, which averaged each sensitivity value with its two nearest neighbours. The smoothing algorithm is seen - in figure 9.9 - to be behaving sensibly. By comparing figure 9.9 to that attained by the

DUNE collaboration in reference [64], the key features of the sensitivity curve (its shape, the relative heights of the sensitivity peaks and the non-zero sensitivity attained at CPC values) are seen to be reproduced. This observation supports the statement made in subsection 6.2.3 that, with the cheated seeds, the fits are now correctly converging.

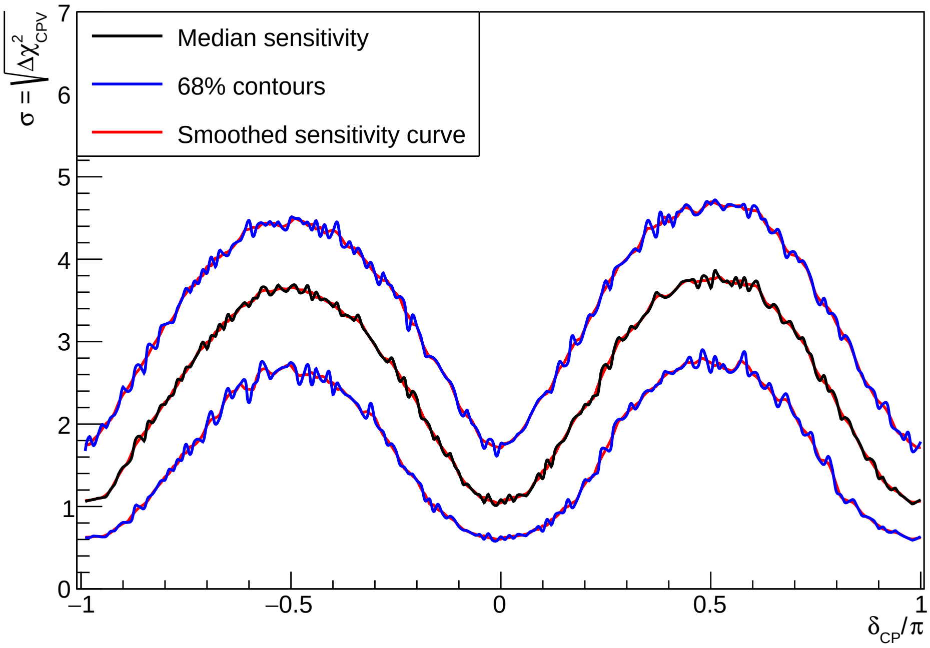


Figure 9.9: The median sensitivity (black line) and 68% contours (blue) achieved - before smoothing - with the initial Pandora reconstruction and Pandora-based selection procedure. Oscillation parameter uncertainties, systematic uncertainties and statistical fluctuations are included in the estimate and normal-ordering of the neutrino mass hierarchy is assumed. The smoothed sensitivity curves are shown in red.

The peak of the smoothed median sensitivity (obtained from the $\delta_{\mathrm{CP}} = \pi/2$ peak) achieved with the initial Pandora reconstruction and Pandora-based $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ is $3.8\sigma^{+0.9\sigma}_{-1.1\sigma}$. This is to be compared to:

- $5.1\sigma^{+0.4\sigma}_{-0.5\sigma}$ in the equivalent estimate, where only oscillation parameter uncertainties are considered (as explained in subsection 6.2.3).
- 5.8σ in the equivalent estimate, where no uncertainties are considered (as explained in subsection 6.2.2).

These results demonstrate the degradation of DUNE's sensitivity to CP-violation as a result of the degeneracies introduced by oscillation parameter and systematic uncertainties.

Figure 9.10 presents the smoothed sensitivity achieved with the implementation of the *ShowerRefinement* algorithm and the two pass $\overset{(-)}{\nu}_e$ selection procedure (and hit thresholds). Here, the median sensitivity and 68% contours are calculated from 300-600 throws at

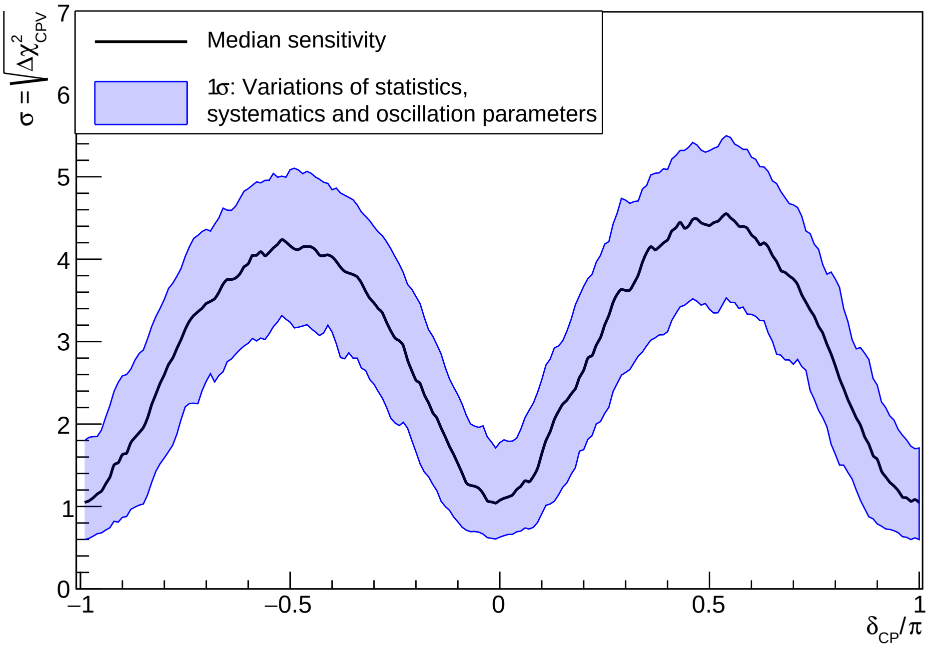


Figure 9.10: The smoothed median sensitivity (black line) and central 68% of throws (blue shaded) achieved with the implementation of the *ShowerRefinement* algorithm and the two pass selection procedure (and hit thresholds). Oscillation parameter uncertainties, systematic uncertainties and statistical fluctuations are included in the estimate and normal-ordering of the neutrino mass hierarchy is assumed.

each of 200 δ_{CP} values. The peak of the smoothed median sensitivity (obtained from the $\delta_{\text{CP}} = \pi/2$ peak) is $4.6\sigma^{+0.9\sigma}_{-1.0\sigma}$. When compared to the initial value of $3.8\sigma^{+0.9\sigma}_{-1.1\sigma}$, this demonstrates that the improvements to the CP-violation analysis (made in chapter 8) 'survive' the systematics and increase the success of the Pandora-based CP-violation analysis with real beam neutrino data. For completeness, these results are also compared to

- 6.4σ, obtained from an equivalent estimate in which no uncertainties are considered (as explained in subsection 6.2.2),

where the impact of degeneracies is once again seen.

With respect to the future of the Pandora-based CP-violation analysis, it is worthwhile to see how these improvements further with the addition of the neutrino vertexing cheat. The smoothed sensitivity curves for this configuration are presented in figure 9.11, where the median sensitivity and 68% contours are calculated from 600 throws at each of 200 δ_{CP} values. The peak of the smoothed median sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) is $5.1\sigma^{+1.0\sigma}_{-1.1\sigma}$. When compared to $4.6\sigma^{+0.9\sigma}_{-1.0\sigma}$, the importance of the neutrino vertex placement accuracy to the operation of the *ShowerRefinement* algorithm is demonstrated. As discussed in chapter 10, this will be an avenue of exploration in the future. Once again, for completeness, this is compared to

- 6.8σ, obtained from an equivalent estimate in which no uncertainties are considered (as explained in subsection 6.2.2).

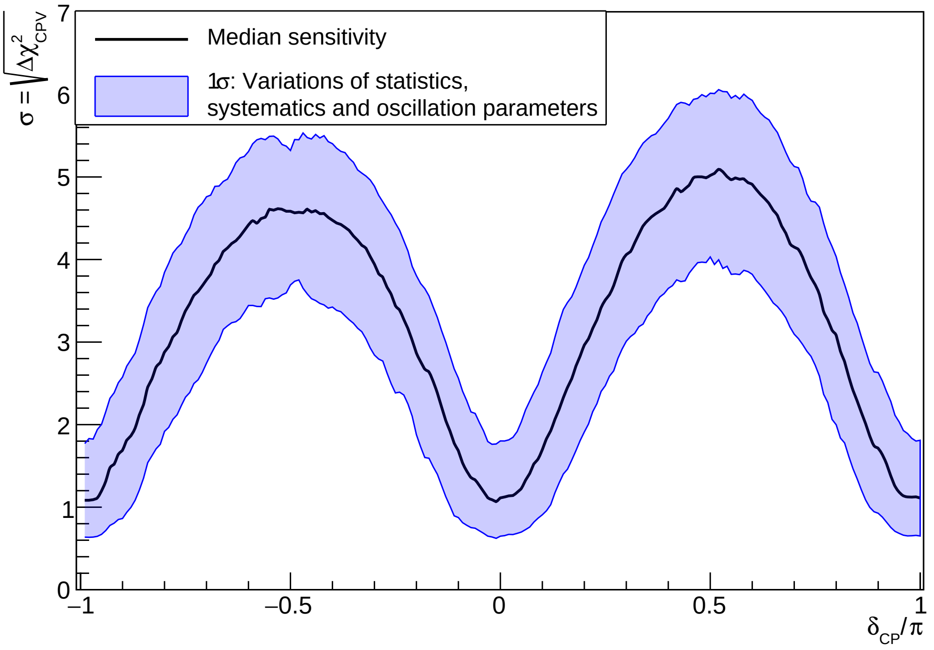


Figure 9.11: The smoothed median sensitivity (black line) and central 68% of throws (blue shaded) achieved with the implementation of the *ShowerRefinement* algorithm, the two pass selection procedure (and hit thresholds) and a cheated neutrino vertex placement. Oscillation parameter uncertainties, systematic uncertainties and statistical fluctuations are included in the estimate and normal-ordering of the neutrino mass hierarchy is assumed.

9.2 MC/Data Comparison

9.2.1 ProtoDUNE Data

The author's access to LArTPC data is via ProtoDUNE-SP. As the ProtoDUNE-SP detector is serviced by the CERN test-beam, which is not a neutrino beam, we cannot directly study the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection performance in data. However, ProtoDUNE's excellent PID capabilities allow us to infer its performance by examining the signal and background distributions of the muon and electron-like BDTs that drive the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection.

The BDT distributions will differ in shape and scale from those of DUNE as a result of the difference in beam energy, and the particle hierarchy tier the examined showers and tracks belong to. However, as the detector technologies, software, and analysis methods of ProtoDUNE-SP will be inherited by DUNE, a good agreement between MC and data is likely to be extendable.

Unlike DUNE, ProtoDUNE-SP is a surface detector and is exposed to a large cosmic-ray flux. This gives rise to the SCE (see section 3.5) which must be accounted for in this study.

9.2.2 Data Sample Selection

It has been chosen to examine data corresponding to a beam momentum of 1GeV/c. This data-set has been studied extensively by the collaboration, so validated analysis methods can be inherited, and significant MC/data discrepancies are known.

The signal muon-like BDT distributions will be obtained from a pure sample of beam stopping muons and the background from a pure sample of beam stopping protons and pions.

The signal electron-like BDT distributions will be obtained from a pure sample of beam positrons and the background from a sample of beam pion child showers. The background sample will be mostly composed of photons, as is the goal, but will also contain other decay products, such as positron showers. The creation, or implementation, of a photon selection procedure is unnecessary and it is sufficient to compare this ‘background’ sample to get a sense of the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection performance in data.

The selection procedure used to obtain these samples is heavily based on that presented in reference [83] and will be outlined below.

1. Bad Event Removal

Events are removed if:

- the data was taken during periods of high voltage instability,
- there is no readout from the beamside detector chamber, or if
- there is less than 10 reconstructed 3D hits in the beam region.

2. Beam Particle Identification

The reconstructed beam particle is identified by the Pandora reconstruction software to be the primary particle that has the most test-beam-like particle hierarchy, as determined by a BDT [99].

To remove beam particles that are in truth cosmic-rays or that are test-beam particle children, the following cuts are imposed:

- $\vec{r}_{\text{Beamline}} \cdot \vec{r}_{\text{Beam Particle}} < 0.93$, where $\vec{r}_{\text{Beamline}}$ and $\vec{r}_{\text{Beam Particle}}$ denote the direction of the beamline and beam particle respectively.

- $0\text{cm} < \Delta x < 10\text{cm}$, $-5\text{cm} < \Delta y < 10\text{cm}$ and $30\text{cm} < \Delta z < 35\text{cm}$, where $\Delta x, \Delta y$ and Δz represent the difference in the x, y and z coordinates between the beam particle start position and the entrance point of the projected beamline into the TPC.

2. Beam Particle Identification

Beam PID is provided by the beam instrumentation discussed in section 3.5. At 1GeV, the existence of a signal in the low-pressure Cherenkov (XCET-L) detector and the Time of Flight (TOF) between bookend trigger counters, illustrated in figure 3.14, is used to identify positrons and protons as seen in table 9.2.

Particle	XCET-L	Time of Flight [ns]	CSDA Cut
e	Y	0 - 105	N/A
μ	N	0 - 110	[0.9, 1.1]
π	N	0 - 110	[0.0,0.9], [1.1, ∞]
p	N	110 - 160	[0.74, 1.09]

Table 9.2: The criteria used to identify different particle types, where the 'XCET-L' column refers to whether a signal is seen in the low-pressure Cherenkov detector of the ProtoDUNE-SP beamline and the 'CSDA cut' column details the accepted range of CSDA values (see equation 9.2).

Muons and pions are not separated by this information. However, since the interaction length of pions is approximately a quarter of that of muons [83], they can be separated by a cut on the ratio:

$$\frac{\text{Reconstructed Track Length}}{\text{CSDA Stopping Range}}, \tag{9.2}$$

where the reconstructed track length is the distance between the SCE corrected track endpoints, and the value of the stopping range is calculated by the Continuous-Slowing-Down-Approximation (CSDA) [121]. The CSDA stopping range is predicted using the measured beam momentum and the assumption that the beam particle is a muon. For muons this ratio is required to be within $[0.9, 1.1]$ and the tracks that fall outside this range are assumed to be pions.

3. Sample Selection

Data muon, pion and proton samples are obtained from Run 5387 whilst positron samples are collected from Run 5809.

- **Signal Muon-like BDT Sample**: A selected muon sample in which reconstructed beam particles:
  - Are reconstructed as tracks.

  - Satisfy the muon criteria of table 9.2.
  - Have a reconstructed track length exceeding 350cm.

  The latter is required in response to the malfunctioning of the electron diverters between the first and second APAs of the beamside detector chamber (see figure 3.13), which had the consequence of distorting the electric field in this region, leading to the splitting of tracks in the reconstruction. The effect was overestimated in the simulation resulting in large MC/data discrepancies, which we sought to avoid.

- **Background Muon-like BDT Sample:** Composed of a selected pion sample and a selected proton sample. In the selected pion sample, reconstructed beam particles:
  - Are reconstructed as tracks.
  - Satisfy the pion criteria of table 9.2.

  In the selected proton sample, reconstructed beam particles:
  - Are reconstructed as tracks.
  - Satisfy the proton criteria of table 9.2.
  - Have a track length to CSDA stopping range ratio in the interval $[0.74, 1.09]$, where the CSDA stopping range is calculated with the assumption that the beam particle is a proton.

  The CSDA cut removes interacting protons from the sample.

- **Signal Electron-like BDT Sample**: A selected positron sample in which reconstructed beam particles:
  - Are reconstructed as showers.
  - Satisfy the electron criteria of table 9.2.
  - Shower within the detector rather than before.

  The final item is enforced by demanding that the ratio of the total shower energy to the reconstructed beam momentum exceeds 80%.

- **Background Electron-like BDT Sample**: The child showers of the selected pion tracks.

9.2.3 MC Sample Selection

The selection of the MC samples follows the procedure detailed in subsection 9.2.2 with the following adjustments:

- The reconstructed beam particle start positions cuts are: $-3\text{cm} < \Delta x < 7\text{cm}$, $-8\text{cm} < \Delta y < 7\text{cm}$ and $27.5\text{cm} < \Delta z < 32.5\text{cm}$.
- The beam instrumentation PID methods are replaced by accessing the true identity of the beam particle. The CSDA, muon track length and shower energy cuts are retained.
- The true beam momentum is used in place of the reconstructed beam momentum.

9.2.4 Muon-like BDT Variable Comparison

With our signal and background samples in hand, we can now investigate the agreement of the muon-like BDT variable distributions in data and MC. The implementation of the muon-like BDT variables in ProtoDUNE-SP requires:

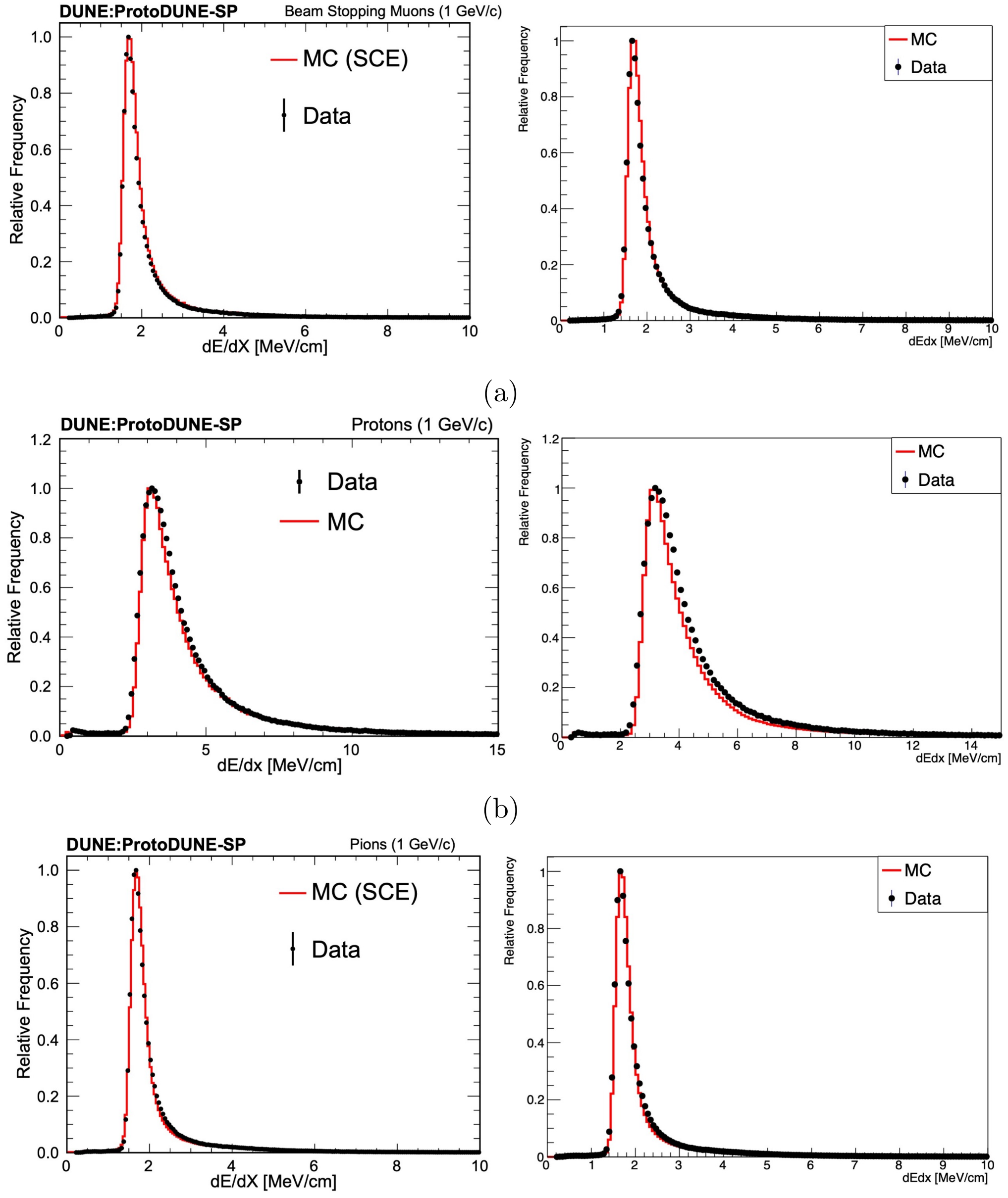
- the track trajectories to be obtained from a fit to the SCE corrected 3D space points, and
- the track's charge deposition to be calibrated with respect to the ProtoDUNE-SP detector.

The main corrections applied to convert between charge deposition and energy are:

- A space-charge correction.
- An electron-lifetime correction, which accounts for the loss of ionisation electrons due to impurities in the liquid argon.
- A YZ correction, which encapsulates the corrections perpendicular to the drift direction that arise from e.g. transverse diffusion, unresponsive readout wires etc.
- An X correction, which encapsulates the corrections in the drift direction that arise from e.g. diffusion, attenuation etc.

The ProtoDUNE collaboration performed the calibration for the data but not the MC. Using existing methods and predetermined calibration factors, the MC was calibrated by the author. Figure 9.12 compares the dE/dx distributions of the selected tracks in this study with the equivalent figures of reference [83]. The dE/dx distributions are normalised such that the highest bin height is unity. Good agreement is seen, demonstrating that the calibration was applied correctly.

It was found that the implementation of the SCE correction and energy calibration at the hit level was non-trivial, which made it challenging to implement the track/shower BDT variables of subsection 6.4.1. However, as these variables contribute very little to the separating power of the muon-like BDT, their implementation is not essential to acquire an understanding of the performance of the BDT in reality. For this reason they, and the Michel electron variables of subsection 6.4.1, were not considered further.



(a)

(b)

(c)

Figure 9.12: The dE/dx distribution of a) selected muons b) selected protons, and c) selected pions obtained in this study (right) compared to the equivalent figures of reference [83] (left).

Topology Variables

The topology muon-like BDT variables were implemented as follows:

- **Track Length:** the distance between the SCE corrected track start and end point.
- **Deviation from Straightness:** calculated from the SCE corrected trajectory following the procedure outlined in subsection 6.4.1.

Their distributions are are shown in figure 9.13. It was chosen to normalise the histograms such that the total number of tracks/showers sum to unity, this is so that the discriminating power of the BDT variables can be seen. In each distribution, a negative bin corresponds to cases in which the BDT variable could not be calculated, and the final bin is an overflow bin which contains the tracks/showers that lie beyond the histogram scale. Error bars are calculated assuming that the MC/data follows a Poisson distribution.

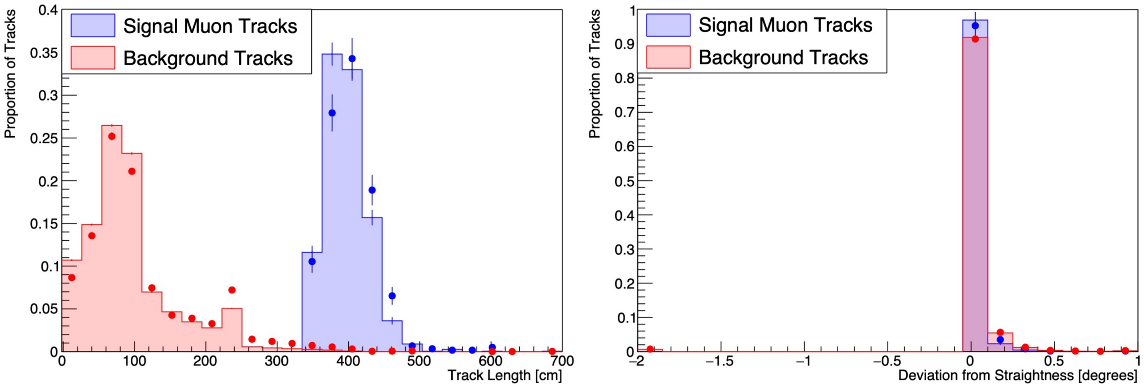


Figure 9.13: The signal (blue) and background (red) track length (left) and deviation from straightness (right) distributions. The MC sample distribution is shaded whilst the data distribution is dotted.

The separation achieved by these variables reproduce that seen in subsection 6.4.1; muon tracks are longer and straighter than background tracks. Good agreement is seen between the MC and data in the topology-based BDT variable distributions of figure 9.13. There is a slight discrepancy observed in the track length background distribution around 225cm. This is believed to arise from the incorrect pion selection, in the data, of muon tracks that have been broken by the electron diverters.

Calorimetry Variables

The calorimetry muon-like BDT variables were implemented as follows:

- **Initial dE/dx:** the median calibrated dE/dx of the first 4cm of the track trajectory.
- **End dE/dx:** the median calibrated dE/dx of the last 4cm of the track trajectory.

- **End Region dE/dx Ratio:** the ratio of the median calibrated dE/dx of the end region and the penultimate region, where the penultimate region is defined to be the 4cm section before the end region.

Their distributions are are shown in figure 9.14.

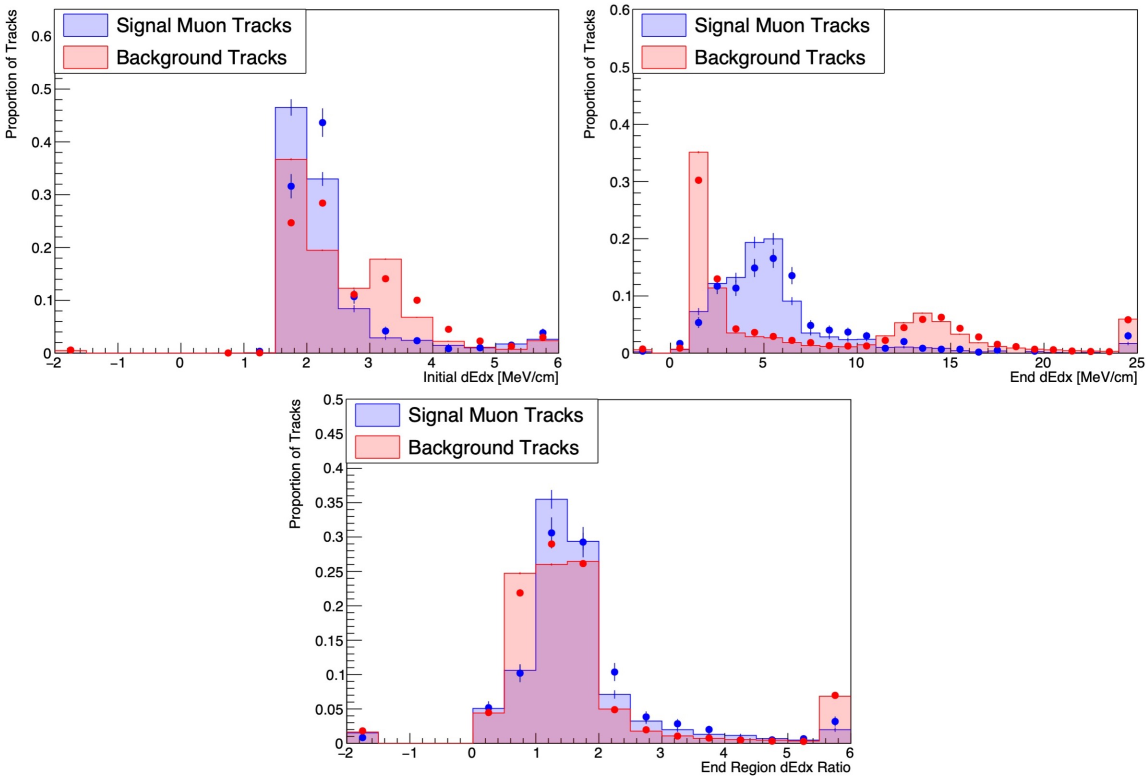


Figure 9.14: The signal (blue) and background (red) initial dE/dx (top-left), end dE/dx (top-right) and end region dE/dx ratio (bottom) distributions. The MC sample distribution is shaded whilst the data distribution is dotted.

The separation achieved by these variables reproduce that seen in section 6.4.1; the calorimetry variables succeed(fail) to separate the signal from protons(pions). Again, good MC/data agreement is seen. However the muon, proton and pion peaks are found to have a wider spread in data, suggesting that the energy resolution is over-estimated in the MC. This was suppressed in figure 9.12, where the distributions are filled by each dE/dx value along the track's trajectory, as opposed to a single dE/dx value representing a particular region. Figure 9.16 examines the initial dE/dx of positron showers where such a discrepancy is also seen. It is suggested in reference [83] that this arises from 'imperfect modelling of physics processes and/or detector effects'.

9.2.5 Electron-like BDT Variable Comparison

The implementation of the electron-like BDT variables in ProtoDUNE-SP requires:

- the shower fits to be performed on the SCE corrected 3D space points, and
- the shower's charge deposition to be calibrated with respect to the ProtoDUNE-SP detector.

Showers have a complex structure, which demands a tailored energy calibration. A collaboration-wide shower calibration procedure is yet to be implemented. So, as it is not within the scope of this study to create a shower calibration procedure, showers were calibrated under a track hypothesis. To do this, the SCE corrected 3D space points of showers also need to be fitted with a track hypothesis.

With this assumption, an estimate of the total shower energy can be obtained and one can identify the positron showers that shower inside the detector. Since the shower region treatment is too simplistic, the total shower energy estimate will be inaccurate but it is sufficient to remove the positrons that shower prematurely.

Calorimetry Variables

The calorimetry electron-like BDT variables were implemented as follows:

- **Initial dE/dx:** the median calibrated dE/dx of the hits that are within a 4cm longitudinal and 1cm transverse distance of the projected hypothesised track trajectory, which is defined by the start position and direction obtained from the track fit.
- **Energy Density:** the ratio of the calibrated total energy and the volume calculated from the SCE corrected shower fit.

Their distributions are shown in figure 9.15, where, to demonstrate the region occupied by the photons, the component of the MC 'background' distribution that belongs, in truth, to the photon background has been added.

Firstly, a comment on the appropriateness of the use of the track calibration procedure. Figure 9.16 demonstrates that the positron initial dE/dx distribution agrees with that obtained in reference [83], suggesting that the track calibration procedure works well. This is expected, as the initial shower region is track-like. The energy density distributions of figure 9.15 have the same shape as those seen in the context of DUNE. This is evidence that, although far from perfect, the total energy estimate is sensible.

The dE/dx distributions achieve a similar separation to that expected for DUNE; the distributions peak at different dE/dx values and are smeared by reconstruction errors. This is true too for the energy density distributions, although a better separation is

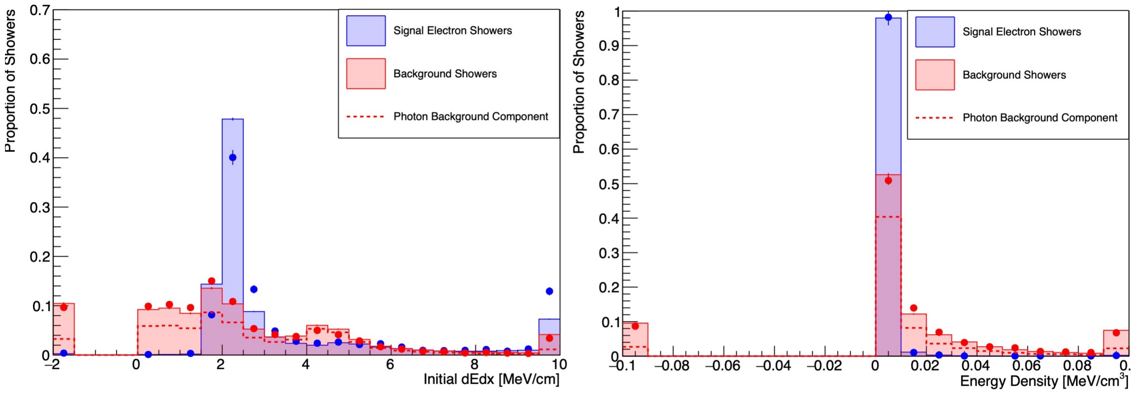


Figure 9.15: The signal (blue) and background (red) initial dE/dx (left) and energy density (right) distributions. The MC sample distribution is shaded and the true photon component of the background is given by a dashed line, the data distribution is dotted.

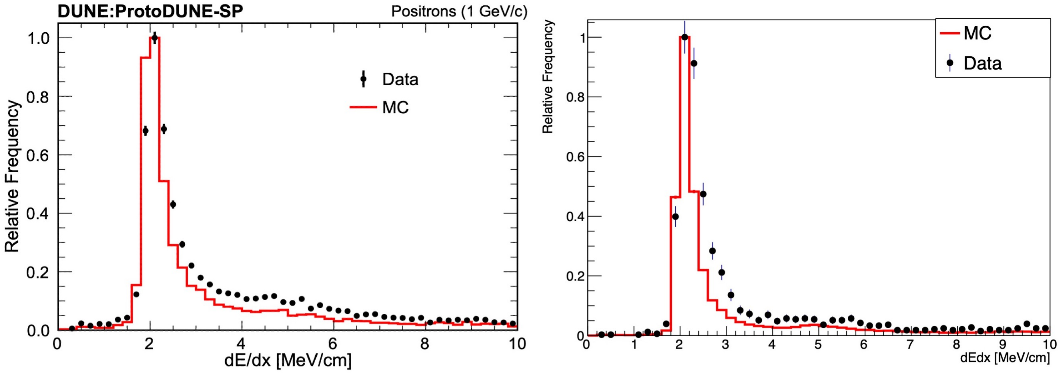


Figure 9.16: The initial dE/dx distribution of selected positrons obtained in this study (right) compared to the equivalent figures of reference [83] (left).

achieved because incorrect shower merges (resulting in larger volumes) are less likely for beam particles than for beam particle secondaries. A good MC/data agreement is observed in figure 9.15 where slight discrepancies are believed to arise from the over-estimated energy resolution in the MC.

Topology Variables

The topology electron-like BDT variables were implemented as follows:

- **Wideness:** the opening angle of the SCE corrected fitted shower, divided by its length.
- **Distance of Closest Approach (DCA):** calculated from the SCE corrected shower fit with respect to the beam-particle end point, following the procedure outlined in subsection 6.4.1.
- **Displacement:** the distance between the SCE corrected beam-particle end point and the SCE corrected shower start point.

For DUNE, the displacement and DCA variables are calculated with respect to the neutrino vertex. The equivalent vertex for ProtoDUNE is the position where the beam particle interacts. Since the signal sample is built from beam particles it is only worthwhile examining the ‘background’ BDT variable distributions.

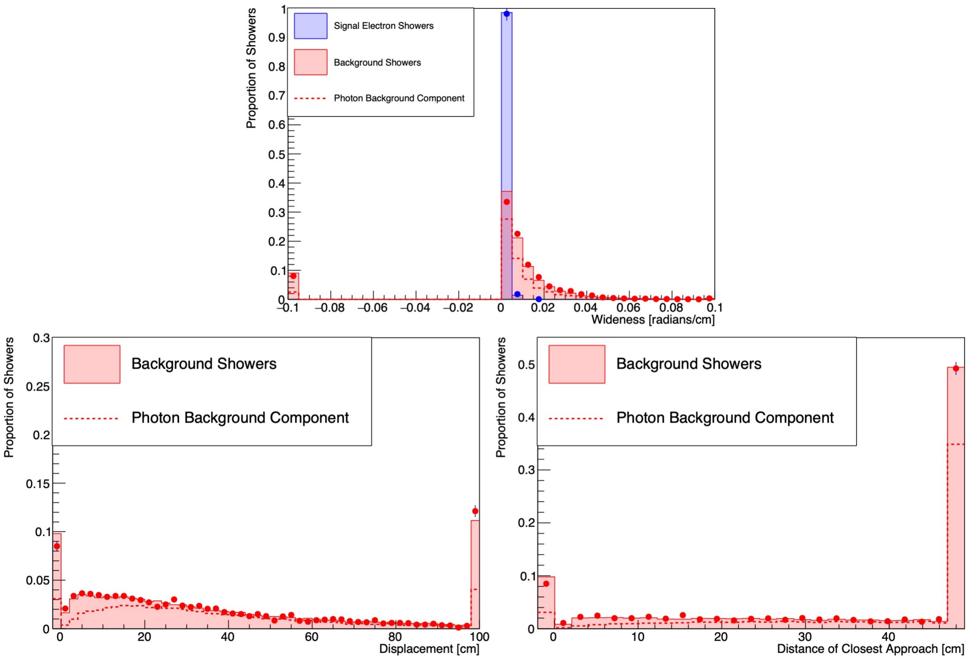


Figure 9.17: The signal (blue) and background (red) wideness (top), displacement (bottom-left) and distance of closest approach (bottom-right) distributions. The MC sample distribution is shaded and the true photon component of the background is given by a dashed line, the data distribution is dotted.

As with the energy density variable, the achieved separation in the wideness distribution is likely exaggerated as a result of the different hierarchy tiers of the signal and background. The displacement and DCA variables reproduce that seen for DUNE: that photons will mostly travel some distance from their creation vertex before showering with a significant proportion of them showering immediately. A good MC/data agreement is seen in figures 9.17 where the slight discrepancies are thought to arise from the SCE.

9.2.6 Conclusion

In this section, the signal and background distributions of the variables that are most important to the electron and muon-like BDTs have been examined. It is argued that, because of DUNE’s inheritance of ProtoDUNE’s analysis, software, and technology, the seen agreement between MC and data is likely to be extendable to DUNE. Therefore, to the extent allowed using the currently-available ProtoDUNE-SP data, we can draw confidence that the analysis developed in this thesis, using a simulation of the DUNE FD, will be able to be deployed with similar success to real beam neutrino data.

Chapter 10

Conclusions and Outlook

This thesis has presented an analysis that targets the flagship physics goal of DUNE: the study of CP-violation. The analysis has assumed the predicted LBNF beam flux, as described in chapter 3, and a 3.5 year exposure (1.36×10^{23} POT) in each beam mode (FHC and RHC). Only the predicted data of the far detector modules has been used; near detector samples have not been included in this analysis.

The workflow of the CP-violation analysis was outlined in chapter 6. At its core is the Pandora-based $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection procedure, which, as its name suggests, operates on the output of the Pandora pattern-recognition software. An electron-like and a muon-like BDT are used to infer the identity of the leading lepton in a given reconstructed neutrino interaction, and, by proxy, determine whether the event is selected. The BDTs exploit the fine granularity of the images produced by LArTPCs to separate electrons and muons from their backgrounds via their key characteristics. To give an (extremely relevant) example, electrons and photons are separated by the examination of their initial region:

- because photons are electrically neutral, photon-induced showers will often exhibit a gap between the neutrino and shower vertex whereas, electron-induced showers will not;

- photon-induced showers are likely to have an initial dE/dx that is twice that of electron-induced showers, because their first energy deposits typically arise via the production of an electron-positron pair.

With the assumption of the NO values of the oscillation parameters given in table 6.2, the $\nu_e(\nu_\mu)$ selection was found to achieve a 61.8%(88.2%) and 64.9%(87.3%) efficiency and purity respectively, where uncertainties are negligible as a result of the scale of the MC statistics used. The selected ν_e, $\bar{\nu}_e$, ν_μ and $\bar{\nu}_\mu$ energy spectra were used to obtain a 'simple estimate' of the sensitivity, in which the oscillation parameter and systematic uncertainties were not considered. It was verified that the FHC contributions to the sensitivity exceeded the RHC contributions (as a result of the larger cross section for neutrinos) and that appearance channels contributed more than disappearance channels. For the ν_e selection, the $\sim$0.25GeV, $\sim$1GeV and $\sim 2-3$GeV energy bands were seen to be the most sensitive to CP-violation, but the discriminating power of all bands was found to be limited by a low selection efficiency, with the $\sim$1GeV and $\sim 2-3$GeV energy bands also impacted by the selection of a NC background.

In chapter 7, efforts turned to optimising the sensitivity to CP-violation. Investigations focused on the Pandora pattern-recognition software, as it determines the quality of leading leptons that guide the selection. Reconstruction improvement hypotheses were tested by performing cheating studies, in which one uses the truth information of a generated MC event to perfect a targeted aspect of the reconstruction. The importance of the improvements were assessed in terms of their impact on a 'simple estimate' of the dominant ν_e sensitivity contribution, in which the NO values of the oscillation parameters in table 6.2 were assumed. Cheating studies were performed using an iterative approach: in each iteration, the sources of any improvements were identified and used to focus subsequent cheating configurations until a well-defined reconstruction goal was determined. From these studies, it was found that the sensitivity was limited by the reconstruction of the initial region of showers:

- reconstructed electrons could be incomplete (their hits had been stolen such that the electron had no pathway to the neutrino vertex),
- reconstructed photons could be impure (contaminant hits had been merged in, providing a pathway to the neutrino vertex).

Cheating studies suggested that, with these issues addressed, and the neutrino vertexing perfected, the ν_e selection efficiency(purity) that can be achieved is 77.4%(75.9%) and the simple estimate peak sensitivity to CP-violation (obtained at the $\delta_{\mathrm{CP}} = \pi/2$ peak) would increase from 5.1σ (initial performance) to 6.1σ. The reason for the cheated vertexing was to ease the interpretation of the CP-violation analysis gains.

Chapter 8 focused on realising the postulated improvements of chapter 7. It introduced the *ShowerRefinement* algorithm, which identified the hits needed/used to connect showers to the neutrino vertex and added(removed) the 'connection pathway' of hits if the shower was thought to be an electron(a photon). The connection pathway BDT was created to determine the electron-likeness of a given connection pathway, and was found to perform well. The accidental truncation of electrons was found to harm the sensitivity, so the *ShowerRefinement* algorithm was implemented such that only electron extensions could be made. With the addition of the perfected neutrino vertexing, the simple estimate peak sensitivity (obtained at the $\delta_{\mathrm{CP}} = \pi/2$ peak) reached 5.4σ. Further sensitivity gains were realised by incorporating the connection pathway BDT variables into the electron-like BDT, used in the $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ event selection, to form the 'enhanced electron-like BDT'. This introduced the requirement that showers must have an associated connection pathway in order to be selected, which resulted in the loss of signal in cases where the connection pathway finder failed. The ν_e selection was consequently performed in two passes with the corresponding BDTs:

1. The enhanced electron-like BDT.
2. The backup electron-like BDT, a slightly modified version of the original electron-like BDT.

The addition of a 100(25) hit threshold for selection in the first(second) pass was found to further increase the obtained sensitivity. With perfected neutrino vertexing retained, the selection efficiency(purity) achieved was 65.1%(85.3%) and the simple estimate peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) reached 6.1σ. Without cheating the neutrino vertex, the selection efficiency(purity) achieved was 69.9%(71.1%) and the simple estimate peak sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) reached 5.7σ.

The results of chapter 8 assume a MC model and so chapter 9 considered the impact on the analysis of the uncertainties inherent in the simulation. In the first half of the chapter we moved away from the simple sensitivity estimate and incorporated the following into our generated MC samples and spectrum fits:

- Oscillation parameter uncertainties.
- Cross section, energy, and flux systematics.
- Statistical fluctuations.

This allowed the sensitivity to be understood across a distribution of universes, and the impact of degeneracies to be quantified. The median sensitivity and 68% contours of this 'realistic estimate' were calculated with the inclusion of all selected spectra (in both FHC and RHC modes) and the assumption of NO. This led to the following results:

1. The peak of the median sensitivity (obtained at the $\delta_{\text{CP}} = \pi/2$ peak) achieved with the initial Pandora reconstruction and Pandora-based $\overset{(-)}{\nu}_e/\overset{(-)}{\nu}_\mu$ selection was $3.8\sigma^{+0.9\sigma}_{-1.1\sigma}$ (i.e. this was before the *ShowerRefinement* algorithm and the two-pass $\overset{(-)}{\nu}_e$ selection procedure, with hit thresholds). This is to be compared with 5.8σ in the equivalent 'simple estimate'.
2. With the addition of the *ShowerRefinement* algorithm, the two-pass $\overset{(-)}{\nu}_e$ selection procedure (and hit thresholds), and cheated neutrino vertexing, the peak of the median sensitivity reaches $5.1\sigma^{+1.0\sigma}_{-1.1\sigma}$. This is to be compared with 6.8σ in the equivalent 'simple estimate'.
3. Without cheating the neutrino vertexing, the peak of the median sensitivity reaches $4.6\sigma^{+0.9\sigma}_{-1.0\sigma}$. This is to be compared with 6.4σ obtained in the equivalent 'simple estimate'.

In the latter half of chapter 9, a MC/data study was performed in which the electron-like and muon-like BDT variable distributions were compared in the ProtoDUNE-SP MC simulation and data. Good agreement was observed, so, because of the inherent similarity of ProtoDUNE-SP and the DUNE SP FD, it was concluded that the postulated CP-violation analysis performance is an accurate reflection of that which can be achieved with real neutrino-beam data at DUNE.

The work of this thesis has demonstrated the power of the reconstruction-to-analysis continuum - the idea that by connecting the analysis performance to that of the LArTPC event reconstruction, we can target **'meaningful'** reconstruction improvements. This is essential to meet the physics goals of DUNE via Pandora-based analyses and marks a paradigm shift in the approach taken by Pandora developers. Projects are planned to cover the remaining physics goals of DUNE: sensitivity to supernovae neutrinos [122] and beyond the standard model physics.

This thesis is not the end for the Pandora-based CP-violation analysis, which is to be developed further by the author, and the Pandora team, in the years to come. Several future steps have already been suggested or hinted at and will now be discussed. On the analysis side, the key next step will be to incorporate near detector samples into the CP-violation fit, such that systematics are reduced. It is vital that the fitting procedure be developed such that one does not rely on truth information to find the minima of the chi-squared surface. A complete overhaul of the electron-like and muon-like BDTs has also been suggested, in which a convolutional neural network could be developed to learn from the LArTPC representation of the reconstructed tracks and showers. In terms of the reconstruction, since the improvements of this work rely on the identification of the connection pathway, it may prove useful to investigate a more sophisticated approach to finding the connection pathways e.g. via a machine-learning approach. Improvements to the neutrino interaction vertexing will be of high priority, in order to realise the gains indicated in chapter 9. The first step will likely involve tuning the recently released machine-learning vertexing [123] specifically for the signal events of the CP-violation analysis. Furthermore, it is believed that, at this stage, the CP-violation analysis is limited by the ν_e selection efficiency. This will be a focus of this work in the future, and the implementation of a newly developed shower splitting algorithm [124] will likely be a good starting point. Finally, we recall that the ν_μ selection was limited by poor energy estimation. This will be investigated, and development of the neutrino energy estimators may be pursued.

With continued development, the Pandora-based CP-violation analysis will become a serious competitor to the 'official' DUNE CP-violation analysis [64] (which is based on a convolutional visual network [125]) and that of Hyper-Kamiokande (HK), the natural successor of T2K. HK will operate 300t (near) and 260kt (far) water Cherenkov detectors, separated by a ∼295km baseline, and serviced by a 1.3MW neutrino beam [62]. With a shorter baseline than DUNE, HK will not be sensitive to the mass hierarchy, and so will need to incorporate external constraints from other experiments (e.g. DUNE or JUNO [126]) in order to have sufficient sensitivity to CP-violation. Both experiments will ultimately determine the value of the CP-phase, and as a result of their differing detector designs, their measurements will complement one another, such that, together, they will propel us into the precision era of neutrino physics.

Bibliography

[1] Pauli W. Dear radioactive ladies and gentlemen. Phys Today. 1978;31N9:27.

[2] Ellis CD, Wooster WA. The average energy of disintegration of radium E. Proc Roy Soc Lond A. 1927;117(776):109-23.

[3] Chadwick J. Possible Existence of a Neutron. Nature. 1932;129:312.

[4] Fermi E. An attempt of a theory of beta radiation. 1. Z Phys. 1934;88:161-77.

[5] Bethe H, Peierls R. The 'neutrino'. Nature. 1934;133:532.

[6] Franklin AD. Are There Really Neutrinos?: An Evidential History. 2nd ed. Frontiers in Physics. Taylor & Francis Group; 2020.

[7] Cowan CL, et al. Detection of the free neutrino: A Confirmation. Science. 1956;124:103-4.

[8] Danby G, et al. Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos. Phys Rev Lett. 1962;9:36-44.

[9] Perl ML, et al. Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation. Phys Rev Lett. 1975;35:1489-92.

[10] Perl ML, et al. Properties of the Proposed τ Charged Lepton. Phys Lett B. 1977;70:487-90.

[11] Kodama K, et al. Observation of tau neutrino interactions. Phys Lett B. 2001;504:218-24.

[12] Akrawy MZ, et al. Measurement of the Z^0 Mass and Width with the OPAL Detector at LEP. Phys Lett B. 1989;231:530-8.

[13] Ahmad QR, et al. Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory. Phys Rev Lett. 2001;87:071301.

[14] Ahmad QR, et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. Phys Rev Lett. 2002;89:011301.

[15] Fukuda Y, et al. Evidence for oscillation of atmospheric neutrinos. Phys Rev Lett. 1998;81:1562-7.

[16] Lee TD, Yang CN. Parity Nonconservation and a Two Component Theory of the Neutrino. Phys Rev. 1957;105:1671-5.

[17] Wu CS, et al. Experimental Test of Parity Conservation in β Decay. Phys Rev. 1957;105:1413-4.

[18] Acero MA, et al. Improved measurement of neutrino oscillation parameters by the NOvA experiment. Phys Rev D. 2022;106(3):032004.

[19] Abe K, et al. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. Nature. 2020;580(7803):339-44. [Erratum: Nature 583, E16 (2020)].

[20] Haxton WC, et al. Solar Neutrinos: Status and Prospects. Ann Rev Astron Astrophys. 2013;51:21-61.

[21] Bahcall JN. Solar neutrinos: Where we are, where are we going. Astrophys J. 1996;467:475-84.

[22] Davis R Jr, et al. Search for neutrinos from the sun. Phys Rev Lett. 1968;20:1205-9.

[23] Bahcall JN, Bahcall NA, Shaviv G. Present status of the theoretical predictions for the Cl-36 solar neutrino experiment. Phys Rev Lett. 1968;20:1209-12.

[24] Cleveland BT, et al. Measurement of the solar electron neutrino flux with the Homestake chlorine detector. Astrophys J. 1998;496:505-26.

[25] Fukuda Y, et al. Solar neutrino data covering solar cycle 22. Phys Rev Lett. 1996;77:1683-6.

[26] Abdurashitov JN, et al. Measurement of the solar neutrino capture rate with gallium metal. III: Results for the 2002–2007 data-taking period. Phys Rev C. 2009;80:015807.

[27] Hampel W, et al. GALLEX solar neutrino observations: Results for GALLEX IV. Phys Lett B. 1999;447:127-33.

[28] Gaisser TK, Honda M. Flux of atmospheric neutrinos. Ann Rev Nucl Part Sci. 2002;52:153-99.

[29] Hirata KS, et al. Experimental Study of the Atmospheric Neutrino Flux. Phys Lett B. 1988;205:416.

[30] Casper D, et al. Measurement of atmospheric neutrino composition with IMB-3. Phys Rev Lett. 1991;66:2561-4.

[31] Ambrosio M, et al. Measurement of the atmospheric neutrino induced upgoing muon flux using MACRO. Phys Lett B. 1998;434:451-7.

[32] Allison WWM, et al. Measurement of the atmospheric neutrino flavor composition in Soudan-2. Phys Lett B. 1997;391:491-500.

[33] Giganti C, Lavignac S, Zito M. Neutrino oscillations: The rise of the PMNS paradigm. Prog Part Nucl Phys. 2018;98:1-54.

[34] Bilenky SM. Neutrinos: Majorana or Dirac? Universe. 2020;6(9).

[35] Pontecorvo B. Inverse beta processes and nonconservation of lepton charge. Zh Eksp Teor Fiz. 1957;34:247.

[36] Maki Z, Nakagawa M, Sakata S. Remarks on the unified model of elementary particles. Prog Theor Phys. 1962;28:870-80.

[37] Kayser B. On the Quantum Mechanics of Neutrino Oscillation. Phys Rev D. 1981;24:110.

[38] Nunokawa H, Parke SJ, Valle JWF. CP Violation and Neutrino Oscillations. Prog Part Nucl Phys. 2008;60:338-402.

[39] Boyd S. Neutrino Oscillations; 2020. `https://warwick.ac.uk/fac/sci/physics/staff/academic/boyd/stuff/neutrinolectures/lec_oscillations.pdf`.

[40] Sakharov AD. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. Pisma Zh Eksp Teor Fiz. 1967;5:32-5.

[41] Cooper NG. Los Alamos Science: Celebrating the neutrino; 1997. `https://www.osti.gov/biblio/569122`.

[42] Workman RL, et al. Review of Particle Physics. PTEP. 2022;2022:083C01.

[43] Wolfenstein L. Neutrino Oscillations in Matter. Phys Rev D. 1978;17:2369-74.

[44] Mikheev SP, Smirnov AY. Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy. Nuovo Cim C. 1986;9:17-26.

[45] De Salas PF, et al. Neutrino Mass Ordering from Oscillations and Beyond: 2018 Status and Future Prospects. Front Astron Space Sci. 2018;5:36.

[46] Decowski MP. KamLAND's precision neutrino oscillation measurements. Nucl Phys B. 2016;908:52-61.

[47] de Salas PF, et al. 2020 global reassessment of the neutrino oscillation picture. JHEP. 2021;02:71.

[48] Aartsen MG, et al. Measurement of Atmospheric Neutrino Oscillations at 6–56 GeV with IceCube DeepCore. Phys Rev Lett. 2018;120(7):071801.

[49] Abe K, et al. Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV. Phys Rev D. 2018;97(7):072001.

[50] Adamson P, et al. Combined analysis of ν_μ disappearance and $\nu_\mu \rightarrow \nu_e$ appearance in MINOS using accelerator and atmospheric neutrinos. Phys Rev Lett. 2014;112:191801.

[51] Abe K, et al. Search for CP Violation in Neutrino and Antineutrino Oscillations by the T2K Experiment with 2.2×10^{21} Protons on Target. Phys Rev Lett. 2018;121(17):171802.

[52] Acero MA, et al. First Measurement of Neutrino Oscillation Parameters using Neutrinos and Antineutrinos by NOvA. Phys Rev Lett. 2019;123(15):151803.

[53] Abi B, et al. Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics. 2020.

[54] Gando A, et al. Constraints on θ_{13} from A Three-Flavor Oscillation Analysis of Reactor Antineutrinos at KamLAND. Phys Rev D. 2011;83:052002.

[55] Rigolin S. Physics reach of β-beams and ν-factories: the problem of degeneracies. Nucl Phys B Proc Suppl. 2006;155:33-7.

[56] Abe K, et al. Evidence of Electron Neutrino Appearance in a Muon Neutrino Beam. Phys Rev D. 2013;88(3):032002.

[57] Adamson P, et al. First measurement of electron neutrino appearance in NOvA. Phys Rev Lett. 2016;116(15):151806.

[58] Abe Y, et al. Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector. JHEP. 2014;10:086. [Erratum: JHEP 02, 074 (2015)].

[59] Bak G, et al. Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO. Phys Rev Lett. 2018;121(20):201801.

[60] Adey D, et al. Measurement of the Electron Antineutrino Oscillation with 1958 Days of Operation at Daya Bay. Phys Rev Lett. 2018;121(24):241805.

[61] Abi B, et al. Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE. JINST. 2020;15(08):T08008.

[62] Abe K, et al. Hyper-Kamiokande Design Report. 2018.

[63] Bass M, et al. Baseline Optimization for the Measurement of CP Violation, Mass Hierarchy, and θ_{23} Octant in a Long-Baseline Neutrino Oscillation Experiment. Phys Rev D. 2015;91(5):052015.

[64] Abi B, et al. Long-baseline neutrino oscillation physics potential of the DUNE experiment. Eur Phys J C. 2020;80(10):978.

[65] Abud Abed A, et al. Low exposure long-baseline neutrino oscillation sensitivity of the DUNE experiment. Phys Rev D. 2022;105(7):072006.

[66] Abi B, et al. Prospects for beyond the Standard Model physics searches at the Deep Underground Neutrino Experiment. Eur Phys J C. 2021;81(4):322.

[67] Abi B, et al. Supernova neutrino burst detection with the Deep Underground Neutrino Experiment. Eur Phys J C. 2021;81(5):423.

[68] Abed Abud A, et al. Snowmass Neutrino Frontier: DUNE Physics Summary. 2022.

[69] DUNE at LBNF; 2023. `https://vms.fnal.gov/gallery/view?id=39`.

[70] Nygren DR. Proposal to investigate the feasibility of a novel concept in particle detection. 1974.

[71] Rubbia C. The Liquid Argon Time Projection Chamber: A New Concept for Neutrino Detectors. 1977 5.

[72] Acciarri R, et al. Design and Construction of the MicroBooNE Detector. JINST. 2017;12(02):P02017.

[73] Abi B, et al. Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume IV: Far Detector Single-phase Technology. JINST. 2020;15(08):T08010.

[74] Amerio S, et al. Design, construction and tests of the ICARUS T600 detector. Nucl Instrum Meth A. 2004;527:329-410.

[75] Anderson C, et al. The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab. JINST. 2012;7:P10019.

[76] Cavanna F, et al. LArIAT: Liquid Argon In A Testbeam. 2014.

[77] Abi B, et al. The Single-Phase ProtoDUNE Technical Design Report. 2017.

[78] Abi B, et al. The DUNE Far Detector Interim Design Report, Volume 3: Dual-Phase Module. 2018.

[79] Lantwin O. The DUNE vertical drift TPC. PoS. 2022;ICHEP2022:332.

[80] Sacerdoti S. A LArTPC with Vertical Drift for the DUNE Far Detector. PoS. 2022;NuFact2021:173.

[81] Abed Abud A, et al. Deep Underground Neutrino Experiment (DUNE) Near Detector Conceptual Design Report. Instruments. 2021;5(4):31.

[82] LBNF/DUNE Collaboration. Long-Baseline Neutrino Facility (LBNF)/DUNE Conceptual Design Report; 2017. `https://docs.dunescience.org/cgi-bin/private/RetrieveFile?docid=4559&filename=CDR_Optimized_Beam_Oct02.pdf&version=12`.

[83] Abi B, et al. First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN Neutrino Platform. JINST. 2020;15(12):12004.

[84] Abed Abud A, et al. Scintillation light detection in the 6-m drift-length ProtoDUNE Dual Phase liquid argon TPC. Eur Phys J C. 2022;82(7):618.

[85] ArgonCube Collaboration. DUNE Near-Detector; 2023. `https://argoncube.org/duneND.html`.

[86] Fermilab. Assembly of test stand for DUNE near detector; 2022. `https://news.fnal.gov/2022/11/assembly-of-test-stand-for-dune-near-detector`.

[87] Charitonidis N, Efthymiopoulos I. Low energy tertiary beam line design for the CERN neutrino platform project. Phys Rev Accel Beams. 2017;20(11):111001.

[88] *LArSoft*; 2023. `https://larsoft.org/`.

[89] Snider EL, Petrillo G. LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors. J Phys Conf Ser. 2017;898(4):042057.

[90] Andreopoulos C. The GENIE neutrino Monte Carlo generator. Acta Phys Polon B. 2009;40:2461-75.

[91] Kudryavtsev VA. Muon simulation codes MUSIC and MUSUN for underground physics. Comput Phys Commun. 2009;180:339-46.

[92] Agostinelli S, et al. GEANT4–a simulation toolkit. Nucl Instrum Meth A. 2003;506:250-303.

[93] Allison J, et al. Geant4 developments and applications. IEEE Trans Nucl Sci. 2006;53:270.

[94] Allison J, et al. Recent developments in Geant4. Nucl Instrum Meth A. 2016;835:186-225.

[95] *WireCell*; 2023. `https://lar.bnl.gov/wire-cell/`.

[96] Adams C, et al. Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation. JINST. 2018;13(07):P07006.

[97] Baller B. Liquid argon TPC signal formation, signal processing and reconstruction techniques. JINST. 2017;12(07):P07010.

[98] Acciarri R, et al. The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector. Eur Phys J C. 2018;78(1):82.

[99] Abed Abud A, et al. Reconstruction of interactions in the ProtoDUNE-SP detector with Pandora. 2022.

[100] *PandoraTrackCreation_module*; 2023. `https://github.com/PandoraPFA/larpandora/blob/develop/larpandora/LArPandoraEventBuilding/LArPandoraTrackCreation_module.cc`.

[101] *CalorimetryAlg*; 2023. `https://internal.dunescience.org/doxygen/CalorimetryAlg_8cxx_source.html`.

[102] *PandoraModularShowerCreation_Module*; 2023. `https://github.com/PandoraPFA/larpandora/blob/develop/larpandora/LArPandoraEventBuilding/LArPandoraShower/LArPandoraModularShowerCreation_module.cc`.

[103] Barker D. Developments Towards a ν_e CC Sterile Appearance Sensitivity in the Short-Baseline Neutrino Programme. University of Sheffield; 2021.

[104] Tyley E. Reconstructing and Selecting Electron Neutrino and Anti-Neutrino Interactions on Argon in the Short-Baseline Near Detector. University of Sheffield; 2023.

[105] Brailsford D. PandoraModularShower in the DUNE far detector; 2021. `https://indico.fnal.gov/event/47321/contributions/206216/attachments/139072/174415/PandoraModularShower_210121FDSimRecoMeeting_DBrailsford.pdf`.

[106] *PandoraShowerCreation_Module*; 2023. `https://github.com/PandoraPFA/larpandora/blob/develop/larpandora/LArPandoraEventBuilding/LArPandoraShower/LArPandoraShowerCreation_module.cc`.

[107] *NeutrinoEnergyRecoAlg*; 2023. `https://github.com/DUNE/dunereco/blob/develop/dunereco/FDSensOpt/NeutrinoEnergyRecoAlg/NeutrinoEnergyRecoAlg.cc`.

[108] Grant N, Yang T. Neutrino energy reconstruction in the DUNE far detector; 2019. DUNE-doc-2278-v5.

[109] Ankowski A, et al. Measurement of through-going particle momentum by means of multiple scattering with the ICARUS T600 TPC. Eur Phys J C. 2006;48:667-76.

[110] Adams C, et al. Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons. JINST. 2020;15(03):P03022.

[111] Bonetti S, et al. A Study of the Electron Image Due to Ionizing Events in a Two-dimensional Liquid Argon TPC With a 24cm Drift Gap. Nucl Instrum Meth A. 1990;286:135.

[112] Mooney M. DUNE Calibrations Using Delta Rays; 2020. `https://indico.fnal.gov/event/44295/contributions/191165/attachments/131115/160268/DUNE_CalibrationMeeting_20_07_09.pdf`.

[113] Baker S, Cousins RD. Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms. Nuclear Instruments and Methods in Physics Research. 1984;221(2):437-42.

[114] Brailsford D. Selecting CC ν_μ and ν_e events in the DUNE far detector using Pandizzle: a PFP-based muon PID; 2019. `https://indico.fnal.gov/event/21096/contributions/60844/attachments/38074/46225/PandoraPandizzlePIDSelection_SimRecoMeeting080719_DBrailsford.pdf`.

[115] Abratenko P, et al. Determination of muon momentum in the MicroBooNE LArTPC using an improved model of multiple Coulomb scattering. JINST. 2017;12(10):P10010.

[116] Selection of ν_μ charged–current induced interactions with $N > 0$ protons and performance of events with $N = 2$ protons in the final state in the MicroBooNE detector from the BNB. 2018 10.

[117] *MVAAlg.cc*; 2023. `https://github.com/LArSoft/larana/blob/develop/larana/ParticleIdentification/MVAAlg.cc`.

[118] Back JJ, et al. Electron-Hadron shower discrimination in a liquid argon time projection chamber. Eur Phys J C. 2013;73(3):2369.

[119] Amsler C, et al. Review of Particle Physics. Phys Lett B. 2008;667:1-1340.

[120] Jones S. Using charged particle test beams to constrain systematic uncertainties for the DUNE experiment. University Coll. London; 2022.

[121] Groom DE, Mokhov NV, Striganov SI. Muon stopping power and range tables 10MeV to 100TeV. Atom Data Nucl Data Tabl. 2001;78:183-356.

[122] Osbsiton M. Tuning Pandora for supernova neutrinos; 2023. `https://indico.fnal.gov/event/58757/contributions/261526/attachments/164863/218779/March%202023%20Low%20Energy%20Talk.pdf`.

[123] Chappell A. Neutrino interaction vertex reconstruction at DUNE; 2022. `https://warwick.ac.uk/fac/sci/physics/research/epp/events/seminars/deep_learning.pdf`.

[124] Brunetti MB. A reclustering approach in Pandora; 2023. `https://indico.fnal.gov/event/53965/contributions/257915/attachments/163303/216102/CollaborationMeetingJan23_SimReco_reclustering.pdf`.

[125] Abi B, et al. Neutrino interaction classification with a convolutional neural network in the DUNE far detector. Phys Rev D. 2020;102(9):092003.

[126] Djurcic Z, et al. JUNO Conceptual Design Report. 2015.