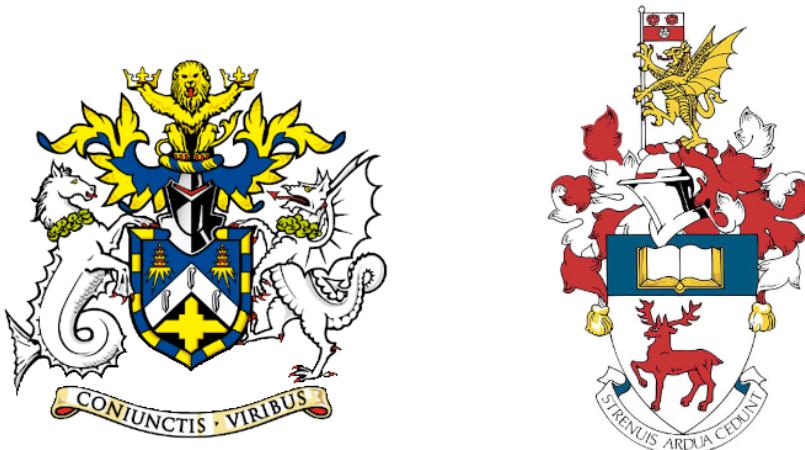


¹ ADVANCING NEUTRINO
² DETECTION AND TRIGGERING IN
³ DUNE



⁵ Francisco Martínez López

⁶ Submitted in partial fulfillment of the requirements
⁷ of the Degree of Doctor of Philosophy

⁸ School of Physical and Chemical Sciences
⁹ Queen Mary University of London

¹⁰ School of Physics and Astronomy
¹¹ University of Southampton

¹² December 2024

Statement of originality

- 14 I, Francisco Martínez López, confirm that the research included within this thesis is my
15 own work or that where it has been carried out in collaboration with, or supported by
16 others, that this is duly acknowledged below and my contribution indicated. Previously
17 published material is also acknowledged below.
- 18 I attest that I have exercised reasonable care to ensure that the work is original, and
19 does not to the best of my knowledge break any UK law, infringe any third party's
20 copyright or other Intellectual Property Right, or contain any confidential material.
- 21 I accept that the University has the right to use plagiarism detection software to check
22 the electronic version of the thesis.
- 23 I confirm that this thesis has not been previously submitted for the award of a degree
24 by this or any other university.
- 25 The copyright of this thesis rests with the author and no quotation from it or information
26 derived from it may be published without the prior written consent of the author.
- 27 Signature: [can be digital signature]
- 28 Date:
- 29 Details of collaboration and publications:
30 [insert details here if applicable]

Abstract

31

32 Work in progress ...

¡Oh memoria, enemiga mortal de mi descanso!

El ingenioso hidalgo don Quijote de la Mancha

MIGUEL DE CERVANTES SAAVEDRA

Acknowledgements

34 Work in progress ...

Contents

Statement of originality	3
Abstract	5
Acknowledgements	9
Contents	15
List of Figures	17
List of Tables	33
List of Abbreviations	35
1 Introduction	37
2 Neutrino physics	39
2.1 Neutrinos in the SM	39
2.2 Trouble in the neutrino sector	43
2.2.1 The solar neutrino problem	43
2.2.2 The atmospheric neutrino problem	45
2.3 Massive neutrinos	47
2.4 Neutrino oscillation formalism	49
2.4.1 Oscillations in vacuum	51
2.4.2 Oscillations in matter	54
2.4.3 Current status of neutrino oscillations	56
2.5 Open questions in the neutrino sector	57
2.6 Neutrino interactions	58

CONTENTS

3 The Deep Underground Neutrino Experiment	63
3.1 Overview	63
3.2 Physics goals of DUNE	65
3.3 LBNF beamline	67
3.4 Near Detector	69
3.4.1 ND-LAr	71
3.4.2 TMS/ND-GAr	72
3.4.3 PRISM	73
3.4.4 SAND	74
3.5 A More Capable Near Detector	74
3.5.1 Requirements	75
3.5.2 Reference design	75
3.5.3 R&D efforts	78
3.6 Far Detector	81
3.6.1 Horizontal Drift	82
3.6.2 Vertical Drift	84
3.6.3 FD Data Acquisition System	86
4 Matched Filter approach to Trigger Primitives	89
4.1 Motivation	89
4.2 Signal-to-noise ratio definition	91
4.3 Low-pass FIR filter design	93
4.4 Matched filters	97
4.5 Monte Carlo studies	103
4.5.1 Angular dependence	109
4.5.2 Distortion and peak asymmetry	111
4.5.3 Hit sensitivity	114
4.6 VD ColdBox data taking	122

CONTENTS

5 DM searches with neutrinos from the Sun	125
5.1 Gravitational capture of DM by the Sun	125
5.2 Neutrino flux from DM annihilations	132
5.3 Computing limits from solar neutrino fluxes	133
5.4 Example: Kaluza-Klein Dark Matter	137
5.5 High energy DM neutrino signals	141
5.5.1 DIS events	143
5.5.2 Single proton QEL events	148
5.5.3 Results	151
5.6 Example: Leptophilic Dark Matter	153
5.7 Systematic uncertainties	158
5.7.1 Systematic uncertainties in the solar WIMP signal	158
5.7.2 Systematic uncertainties in the atmospheric background	160
5.7.3 Common systematic uncertainties	160
6 Particle identification in ND-GAr	161
6.1 GArSoft	161
6.1.1 Event generation	162
6.1.2 Detector simulation	163
6.1.3 Reconstruction	164
6.2 dE/dx measurement in the TPC	166
6.2.1 Energy calibration	168
6.2.2 Truncated dE/dx mean	178
6.2.3 Mean dE/dx parametrisation	181
6.2.4 Particle identification	184
6.3 Muon and pion separation in the ECal and MuID	184
6.3.1 Track-ECal matching	187
6.3.2 Classification strategy	193
6.3.3 Feature selection and importance	196

CONTENTS

6.3.4	Hyperparameter optimisation	207
6.3.5	Probability calibration	211
6.3.6	Performance	213
6.4	ECal time-of-flight	213
6.4.1	Arrival time estimations	215
6.4.2	Proton and pion separation	217
6.5	Charged pion decay in flight	218
6.5.1	Track breakpoints	220
6.6	Neutral particle identification	227
6.6.1	ECal clustering	227
6.6.2	π^0 reconstruction	230
6.7	Integration in GArSoft	232
7	Event selection in ND-GAr	235
7.1	Data sample	235
7.2	ν_μ CC selection	237
7.2.1	Selection optimisation	240
7.2.2	Selection performance	246
7.2.3	Primary muon kinematics	248
7.3	Charged pion identification	252
7.3.1	ν_μ CC $1\pi^\pm$ selection	260
7.4	Neutral pion identification	264
7.5	Neutrino energy reconstruction	266
7.6	Systematic uncertainties	269
7.6.1	Flux uncertainties	269
7.6.2	Cross section uncertainties	270
7.6.3	Detector uncertainties	273
8	Conclusion and outlook	275

CONTENTS

A An appendix	277
Bibliography	279

List of Figures

2.1	Solar neutrino fluxes for the solar model BS05(OP).	44
2.2	Zenith angle distributions for the selected ν_e and ν_μ events in the SK detector.	46
2.3	$K^0 \rightleftharpoons \bar{K}^0$ mixing through W^\pm exchange.	50
2.4	Feynman diagrams of the four most relevant CC interactions to long baseline neutrino oscillation experiments.	59
2.5	Total muon neutrino CC cross section per nucleon as a function of the neutrino energy.	60
2.6	Schematic representation of a ν_μ CCQE interaction with a neutron inside a nucleus.	61
3.1	Schematic diagram of the DUNE experiment and the LBNF beamline. .	64
3.2	Schematic longitudinal section of the LBNF beamline at Fermilab. . .	67
3.3	Predicted neutrino fluxes at the FD in FHC mode and RHC mode. . .	68
3.4	Representation of the ND hall in Phase II, showing the different subcomponents.	69
3.5	Schematic representation of the external components of ND-LAr, including the cryostat and the PRISM movable system and detailed drawing of one ArgonCube module.	70
3.6	Schematic view of the TMS detector, highlighting its main parts. . . .	71
3.7	Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet.	72
3.8	Predicted beam muon neutrino flux at the ND location for different off-axis positions.	73

LIST OF FIGURES

3.9	Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers.	76
3.10	Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers.	77
3.11	Photographs of the TOAD pressure vessel at RHUL.	79
3.12	Electron microscope image and schematic diagram of a GEM electrode.	80
3.13	Schematic diagram showing the operating principle of a LArTPC with wire readout.	81
3.14	Proposed design for the FD-1 and FD-2 modules following the HD principle.	82
3.15	Schematic representation of an APA frames showing the U, V, X and G wires.	83
3.16	A PDS module containing 24 X-ARAPUCAs and the location of the modules on the APAs.	84
3.17	Proposed design for the FD-3 module following the VD principle.	85
3.18	Schematic representation of the electrode strip configuration for a top and bottom CRU.	86
3.19	Detailed diagram of the DUNE FD DAQ system.	87
4.1	Schematic representation of the Trigger Primitive Generation chain in the DUNE FD.	90
4.2	<i>Left panel:</i> Zoomed unfiltered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i> (blue line). The green dashed lines mark the region $\pm 3\sigma_{raw}$. The resulting noise waveform is also shown (red line). <i>Top right panel:</i> ADC distribution for channel 7840, where the green shaded region represents $\pm \sigma_{raw}$. <i>Bottom right panel:</i> noise ADC distribution for channel 7840, where the green shaded region represents $\pm \sigma_{noise}$	92

LIST OF FIGURES

- 4.3 *Left panel:* Zoomed filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The filter used was the current implementation of the low-pass FIR filter in `dtp-firmware`. The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$. The resulting noise waveform is also shown (red line). *Top right panel:* ADC distribution for channel 7840 after filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$. *Bottom right panel:* noise ADC distribution for channel 7840 after filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$ 94
-
- 4.4 Power spectrum in decibels for the current implementation of the low-pass FIR filter in `dtp-firmware` (blue line), compared to the response of an optimal filter obtained using the Parks-McClellan algorithm for the same pass-band (red line). Also for comparison I include the spectrum of the optimal filter when taking only the integer part of the coefficients (red dashed line). 95
-
- 4.5 Relative change in the S/N for the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44`, using different values of the cutoff frequency f_c and the transition width δf . The optimal Chebyshev filters were applied using just the integer part of the coefficients given by the Parks-McClellan algorithm. 96
-
- 4.6 Distribution of the relative change of the S/N on the different wire planes from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` after the optimal Chebyshev filter was applied. The filter was computed with the Parks-McClellan algorithm using a cutoff of $f_c = 0.068 \text{ ticks}^{-1}$ and a transition width $\delta f = 0.010 \text{ ticks}^{-1}$ 97

LIST OF FIGURES

- 4.7 *Left panel: Zoomed match filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The filter used was directly extracted from the data, being the 32 values around the first peak in the original waveform. The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$. The resulting noise waveform is also shown (red line). Top right panel: ADC distribution for channel 7840 after match filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$. Bottom right panel: noise ADC distribution for channel 7840 after match filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$* 98
- 4.8 *Relative improvement in the S/N for the raw data capture `felix-2020-07-17-21:31:44`, using the matched filter following the parametrisation in Eq. (4.16). The black crosses in both panels denote the location of the maximum ratio value.* 101
- 4.9 *Left panel: Optimal matched filter coefficients for the U (blue line) and V (red line) planes. The filters were computed with our parametrisation in Eq. (4.16) for the parameter values $\delta = 0.035$, $\sigma = 0.191$ and $\delta = 0.018$, $\sigma = 0.191$ respectively. Right panel: Distribution of the relative change of the S/N on the two induction wire planes from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` after their respective optimal matched filters were applied.* 103
- 4.10 *Left panel: distributions of the particles track length in the liquid argon for the generated $E_k = 100$ MeV monoenergetic samples, electrons (blue), muons (red), protons (green) and neutral pions (purple). Right panel: distribution of the length of the longest photon in the neutral pion sample after the decay process $\pi^0 \rightarrow \gamma\gamma$.* 104

LIST OF FIGURES

- 4.11 *Left panel: schematic representation of the two new rotated reference frames used in this analysis (denoted as prime and double prime), viewed from the yz plane. The magenta stack of lines represent the wires in the U plane, whereas the green lines correspond to the wires in the V plane. Right panel: 3D representation of the momentum of one of the generated monoenergetic muons (red arrow) in the original reference frame (black lines), along with the new reference frame used for the U plane waveforms (blue lines). In the yz plane I added the projection of these three.* 106
- 4.12 *Distributions of the mean S/N improvement per event for the corresponding sample after applying the matched filters. Here I separated the change in the U plane (blue) and the V plane (red) channels. From top left to the right: muon, electron, proton and neutral pion. All the events have a fixed kinetic energy of $E_k = 100$ MeV.* 107
- 4.13 Angular dependence of the mean S/N and the S/N improvement, for the different monoenergetic samples considered (from top to bottom: electrons, muons, protons and neutral pions). The two columns on the left represent the values for the U plane waveforms. The top subplots show the mean S/N for raw (green) and filtered (red) waveforms whereas the bottom subplots depict the averaged S/N improvement (black). 108
- 4.14 Angular dependence of the mean S/N and the S/N improvement, for the different monoenergetic samples considered (from top to bottom: electrons, muons, protons and neutral pions). The two columns on the left represent the values for the U plane waveforms. The top subplots show the mean S/N for raw (green) and filtered (red) waveforms whereas the bottom subplots depict the averaged S/N improvement (black). 110

LIST OF FIGURES

4.15 Selected consecutive waveforms corresponding to two monoenergetic $E_k = 100$ MeV muon events, one is parallel to the APA and to the wires in the U plane (left panel) and the other is normal to the APA plane and perpendicular to the U plane wires (right panel). The solid lines represent the raw waveforms whereas the dashed lines correspond to the waveforms after the matched filter was applied. The waveforms on the left panel have been scaled by a factor of 0.15 to have similar amplitude to the ones on the right panel.	111
4.16 Left panel: peak asymmetry distribution for the case of the monoenergetic $E_k = 100$ MeV muon sample. Each value corresponds to a single bipolar signal peak from a channel in any event. The blue distribution represents the peaks on U plane channels, whereas the red corresponds to signal peaks in V wires. Right panel: relation between the mean peak asymmetry per event with the S/N for U channel waveforms from the $E_k = 100$ MeV muon sample. The top subplot shows the decimal logarithm of the mean S/N for the raw (red) and the matched filtered (blue) waveforms. The bottom subplot contains the mean S/N improvement ratio after the matched filter was applied.	112
4.17 Raw data display in the plane time (in firmware ticks) vs. offline channel number for an $E_k = 100$ MeV electron event. The produced true hits are superimposed (black boxes) as well as the hits comming from the standard hit finder chain (blue circles) and the hit finder using the matched filter (green triangles).	115

LIST OF FIGURES

<p>4.18 Dependence of the precision (blue), sensitivity (red) and F_1 (green) scores on the threshold values used in the hit finder, for the FIR (left panel) and matched filter (right panel) cases. The results were obtained after matching the hits to the true hits in the case of the isotropic muon sample with kinetic energy in the range 5 to 100 MeV, taking only into account the induction plane channels. The points represent the mean value while the error bars indicate one standard deviation around that mean value.</p> <p>4.19 Dependence of the averaged hit sensitivity on the kinetic energy of the events for the matched filter (blue) and standard (red) hits, for the case of the muon (left panel) and electron (right panel) samples, separated between U (top plots) and V (bottom plots) induction wire planes. The top subplots contain the hit sensitivities for the two hit finder alternatives, while the bottom subplots show the ratio between the two. The horizontal lines sit at the mean value and represent the size of the energy bins, while the vertical error bars indicate one standard deviation around that mean value.</p> <p>4.20 Distributions of the hit sensitivity in the U (top panels) and V (bottom panels) planes versus the hit sensitivity in the X plane, both for the standard hits (left panels) and the matched filter hits (right panels), in the case of the electron sample and a threshold of 30 ADC.</p> <p>4.21 Top panels: standard residual plots of the hit sensitivities between the X and U planes. Bottom panels: quantile-quantile plots of the hit sensitivity standard residuals between the X and U planes. In all cases, the left panel corresponds to the standard hits while the right panel represents the matched filter case, all from the electron sample with a 30 ADC threshold.</p>	<p>117</p> <p>119</p> <p>120</p> <p>121</p> <p>122</p> <p>123</p> <p>124</p>
--	--

LIST OF FIGURES

5.1	Input solar parameters used in the capture rate computation as a function of the solar radius.	128
5.2	Capture rates as a function of the DM mass for the DM-electron interactions, SD DM-nucleons interactions, and SI DM-nucleons interactions.	130
5.3	NuWro computed $\nu_\mu - {}^{40}\text{Ar}$ charged-current scattering cross section as a function of the neutrino energy.	134
5.4	Expected atmospheric neutrino flux as a function of the neutrino energy E_ν at Homestake at solar minimum.	136
5.5	Feynman diagrams for B^1B^1 annihilation into SM fermions.	138
5.6	Feynman diagrams for B^1B^1 annihilation into a Higgs boson pair.	138
5.7	Computed spectra of muon neutrinos at the DUNE FD site from B^1 annihilations in the Sun for various values of M_{LKP}	139
5.8	Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent B^1 -proton scattering cross section as a function of M_{LKP}	140
5.9	Computed spectra of muon neutrinos at the DUNE FD site from $\tau^+\tau^-$ and $b\bar{b}$ annihilations in the Sun for different DM masses.	141
5.10	Distribution of the muon neutrino energies from the $\tau^+\tau^-$ and $b\bar{b}$ annihilation channels.	143
5.11	Angular distributions for the $b\bar{b}$ DIS sample with $m_{\text{DM}} = 10$ GeV and the atmospheric background.	144
5.12	Signal efficiency and background rejection for the $b\bar{b}$ sample with $m_{\text{DM}} = 10$ GeV as a function of the angular cuts.	145
5.13	Signal efficiencies for the $\tau^+\tau^-$ and $b\bar{b}$ DIS samples as functions of the DM mass.	147
5.14	Angular distributions for the $\tau^+\tau^-$ QEL sample with $m_{\text{DM}} = 5$ GeV and the atmospheric background.	148
5.15	Performance of the MLP classifier for the $\tau^+\tau^-$ QEL signal with $m_{\text{DM}} = 5$ GeV.	149

LIST OF FIGURES

5.16 Signal efficiencies for the $\tau^+\tau^-$ and $b\bar{b}$ single proton QEL samples as functions of the DM mass.	150
5.17 Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent DM-nucleon scattering cross section as a function of m_{DM} , for the annihilation channels $\tau^+\tau^-$ and $b\bar{b}$ separated by interaction mode	152
5.18 Projected 90% confidence level sensitivity of DUNE (400 kT yr) to the scale Λ of an EFT containing only leptophilic DM axial-axial interactions.	156
6.1 Schematic diagram showing the different modules involved in the ND-GAr production.	165
6.2 Distribution of the fraction of energy deposits with residual range less than 20% of the total track length, and distribution of the ionisation per unit length after removing the tracks with less than 30% of their energy deposits in the last 20% of the track.	170
6.3 Distribution of the reconstructed ionisation charge per unit length for different reclustering values, and distribution of the median change in dQ/dx per track for the $N_{\text{group}} = 4$ reclustering.	171
6.4 Distribution of the Geant4-simulated energy losses per unit length versus residual range for the stopping proton sample.	172
6.5 Fitted most probable dQ/dx values for each dE/dx bin, together with best fit to the logarithmic calibration function.	173
6.6 Fitted most probable dQ/dx values for each dE/dx bin for three different ADC bit limits.	175
6.7 Area normalised dE/dx distributions for the true and the reconstructed energy deposits in the stopping proton sample, both after applying the calibration and the calibration and the normalisation correction.	176

LIST OF FIGURES

6.8 Fractional residuals between the true and the corrected dE/dx means and the 60% truncated means, and fractional residuals between the true and the uncorrected, corrected and uncalibrated dE/dx 60% truncated means.	178
6.9 Estimated values of the mean dE/dx bias and resolution for the stopping proton sample at different values of the truncation factor.	179
6.10 Examples of the truncated mean dE/dx LanGauss fits for various $\beta\gamma$ bins, from a simulated FHC neutrino sample.	181
6.11 Resulting one and two dimensional projections of the posterior probability distributions of the ALEPH $\langle dE/dx \rangle$ parameters obtained by fitting the 60% truncated mean dE/dx values from a FHC neutrino sample.	183
6.12 Truncated mean dE/dx obtained for the FHC neutrino sample as a function of the $\beta\gamma$ product, together with the fitted most probable values for each $\beta\gamma$ bin and the best fit obtained using the ALEPH parametrisation.	184
6.13 Distribution of the 60% truncated mean dE/dx versus reconstructed momentum for the FHC neutrino sample.	185
6.14 Estimated values of the mean dE/dx bias and resolution obtained for the true protons in a FHC neutrino sample.	186
6.15 True momentum distribution for the primary muon in ν_μ CC $N\pi^\pm$ interactions inside the fiducial volume of ND-GAr, compared to the post FSI charged pion spectrum.	187
6.16 Distributions of energy deposits in the ECal for a muon and a charged pion with similar momenta.	188
6.17 Left panel: comparison between the precision (blue), sensitivity (yellow) and F_1 score (red) obtained for the default (horizontal lines) and new algorithms, both with the χ^2 -based direction estimator (squares) and cheating the directions (circles), for different values of the χ^2 cut. Right panel: comparison of the performance of the new algorithm when applying the cluster t_0 correction (squares) and when (circles).	189

LIST OF FIGURES

6.18 Schematics of a possible option to deal with track-ECal associations in non-zero t_0 neutrino interaction events, trying to correct for the drift direction uncertainty in a cluster-by-cluster basis using the cluster time, $t_{cluster}$	191
6.19 Momentum distribution for the reconstructed muons (top panel) and charged pions (bottom panel) in a FHC neutrino sample, together with the fraction of them reaching the ECal (red) and MuID (blue). Each entry corresponds to a reconstructed track, backtracked to a true muon or pion which has not produced any other reconstructed track.	193
6.20 Predicted truncated mean dE/dx versus momentum, for electrons, muons, charged pions and protons, obtained using the ALEPH parametrisation. The vertical dashed lines represent the boundaries of the six regions used for the muon and pion classification training.	194
6.21 Example ECal feature distributions for muons and charged pions in the five different momentum ranges considered.	198
6.22 Example MuID feature distributions for muons and charged pions in the three different momentum ranges considered.	199
6.23 Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$	201
6.24 Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range $0.8 \leq p < 1.5 \text{ GeV}/c$	202

LIST OF FIGURES

6.25 Evolution of the SHAP importance for the top six most important features across all five momentum ranges.	205
6.26 Permutation importances for the ten most important features in the different momentum ranges (from left to right, top to bottom, in increasing momentum order). The bars indicate the effect that permutations of each feature have on the purity (blue) and the sensitivity (yellow), the translucent regions representing one standard deviation around the central value.	206
6.27 Values of the precision and sensitivity obtained for 10000 BDT hyperparameter configurations, for the momentum regions I, III and V.	209
6.28 Reliability diagrams for the BDT classifier used in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$, both for the original (blue circles) and calibrated (yellow squares) responses. For reference, the response of a perfectly calibrated classifier is also shown (black dashed line).	212
6.29 Uncalibrated (left panel) and calibrated (right panel) predicted probabilities assigned by the BDT classifiers for true muons (blue) and charged pions (red) in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$	213
6.30 Schematic of the hit selection used for the ToF measurement. The grid represents the layers of the inner ECal, with coloured squares indicating the tiles with hits. Green squares indicate the selected hits.	214
6.31 Particle velocity versus momentum measured with different ECal arrival time estimations. From left to right: earliest hit time, average hit time, and fitted hit time. In all cases the time resolution is $\Delta\tau = 0.1 \text{ ns}$	216
6.32 Mass spectra for p (blue) and π^\pm (yellow) particles, using different ECal time resolution values (from top to bottom, in ascending order), and arrival time estimates. From left to right: earliest hit time, average hit time, and fitted hit time. The dashed lines indicate the true masses of the particles.	218

LIST OF FIGURES

6.33 Efficiency (top panel) and purity (bottom panel) for the proton selection as a function of the momentum, for $\Delta\tau = 0.10$ ns.	219
6.34 Distributions of the velocities measured by ToF with the inner ECal, for different momentum bins, in a FHC neutrino interaction sample. The Gaussian fits are performed around the maxima for each particle species.	220
6.35 Left panel: number of non-decaying, decaying and decaying in the fiducial volume pions for a MC sample of 100000, $p = 500$ MeV/c isotropic positively charged pions inside the TPC. Right panel: event display for a positive pion decaying inside the fiducial volume, with a single reconstructed track for the pion and muon system.	221
6.36 Values of $\chi_k^{2(FB)}$ (top left panel), F_k (top right panel), $D_k^{1/R}$ (bottom left panel) and D_k^ϕ (bottom right panel) versus position along the drift direction for a reconstructed track in a positive pion decay event. The vertical red dashed line indicates the true location of the decay point.	222
6.37 Fractional residual distributions of the true and reconstructed decay position along the drift coordinate, using the position of the maximum of $\chi_k^{2(FB)}$ (left panel) and F_k (right panel) as estimates of the decay position. Also shown are double Gaussian fits to these points (red lines).	223
6.38 Distributions of the extreme values of $\chi_k^{2(FB)}$ (top left panel), F_k (top right panel), $D_k^{1/R}$ (bottom left panel) and D_k^ϕ (bottom right panel) for non-decaying reconstructed pion tracks (blue) and tracks which include the decay inside the fiducial volume (red).	225
6.39 Left panel: distributions of the predicted probabilities assigned by the BDT classifier to a test sample of decaying pion+muon tracks (blue) and non-decaying pion tracks (red). Left: signal efficiency versus background acceptance (ROC curve) obtained from the BDT for the test sample.	226

LIST OF FIGURES

6.40 Left panel: dependence of the decay position finding resolution on the true value of the decay angle for the χ_k^2 (FB) (red) and F_k (blue) methods. Right panel: signal efficiency (blue line) and background rejection (red line) from the BDT classifier versus true decay angle.	226
6.41 Mean values of the F_1 -score marginal distributions for the different free parameters of the new clustering algorithm, with the error bars representing one standard deviation around the mean. The F_1 -score values were computed for the 6561 possible parameter configurations using 1000 ν_μ CC interaction events.	229
6.42 Left panel: distributions of the number of ECal clusters per photon from π^0 decays for the standard (red) and new (blue) clustering algorithms. Right panel: reconstructed invariant mass distributions for photon pairs from single π^0 events using the standard (red) and new (blue) ECal clustering algorithms.	231
6.43 Distributions of proton dE/dx and ToF scores for a sample of 100000 FHC neutrino interactions in the HPgTPC.	233
7.1 Schematic diagram of the HPgTPC including the fiducial volume.	239
7.2 True positive, false positive, and false negative true neutrino energy distributions for a ν_μ CC selection.	240
7.3 Efficiency and purity for the ν_μ CC selection as a function of the different cuts.	242
7.4 Significance for the ν_μ CC selection as a function of the different cuts.	242
7.5 Normalised 2D distributions of efficiency, purity and significance for the ν_μ CC selection. The $S/\sqrt{S+B}$ is normalised to the highest value achieved. The vertical dashed line indicates a purity value of 0.85, whereas the horizontal one corresponds to an efficiency of 0.80.	243
7.6 Cumulative efficiency and purity for the ν_μ CC selection.	244
7.7 True neutrino energy spectra for the ν_μ CC selection.	245

LIST OF FIGURES

7.8 True neutrino energy spectra for the ν_μ CC selection.	246
7.9 True neutrino energy spectra for the ν_μ CC selection with (blue) and without (red) sign selection. The selected events are broken down by true positives (signal) and false positives (background). The true distribution is also shown (black data points). The bottom panel shows the ratios between the number of false positives and total selected events per bin. .	247
7.10 Efficiency 2D distributions for the ν_μ CC selection given the true position of the interaction vertex.	248
7.11 Distributions for the reconstructed versus truth primary muon momentum, longitudinal momentum, transverse momentum, and angle.	249
7.12 Efficiency and purity of the ν_μ CC selection as a function of the primary muon true momentum and beam angle.	250
7.13 Fractional residual distributions for the position of the primary vertex in the ν_μ CC selection.	251
7.14 Distribution of events given their true and reconstructed π^\pm multiplicity, for a given selection.	253
7.15 Efficiency and purity for the ν_μ CC $0\pi^\pm$ selection as a function of the different cuts.	255
7.16 Efficiency and purity for the ν_μ CC $1\pi^\pm$ selection as a function of the different cuts.	256
7.17 Row and column normalised distributions of events given their true and reconstructed π^\pm multiplicity, for the selection that maximises the significance of the ν_μ CC $1\pi^\pm$ selection.	258
7.18 Purity versus efficiency achieved for the different cut configurations explored separated by the various ν_μ CC $N\pi^\pm$ selections.	259
7.19 Efficiency of the ν_μ CC $N\pi^\pm$ selections as a function of the hadronic invariant mass.	260
7.20 Reconstructed kinematic distributions for the pion candidate in the ν_μ CC $1\pi^\pm$ selection.	261

LIST OF FIGURES

7.21 Distributions for the reconstructed versus truth pion kinetic energy and angle between the pion and muon in the ν_μ CC $1\pi^\pm$ selection.	262
7.22 Efficiency of the ν_μ CC $1\pi^\pm$ selection as a function of the true pion kinetic energy and pion-muon angle.	263
7.23 Invariant mass distribution obtained for the unassociated ECal cluster pair in the event with the value closest to the true π^0 mass.	265
7.24 Reconstructed neutrino energy spectra for the ν_μ CC selection obtained using the truth and reconstructed information	268
7.25 Neutrino energy residuals distributions for the ν_μ CC selection obtained using the truth and reconstructed information	269

List of Tables

2.1	Values of T_3 and $Y/2$ assigned to the first generation of fermions.	41
2.2	Neutral current couplings.	42
2.3	Summary of neutrino oscillation parameters determined in the Neutrino Global Fit of 2020 [1].	57
3.1	Summary of the two-phased plan for DUNE.	65
3.2	Exposure and time required to achieve the different physics milestones of the two phases.	66
4.1	<i>Characteristic parameters of the two monoenergetic muon events selected, relative to the U plane: projected angles in the xz' and $y'z'$ planes, S/N values for the raw and filtered waveforms, mean improvement of the S/N and peak asymmetry.</i>	112
5.1	Systematic uncertainties for the solar WIMP signal events.	159
5.2	Systematic uncertainties for the solar WIMP atmospheric background events.	159
6.1	Calibration parameters obtained from the fit of the ND-GAr simulated stopping proton sample to the calibration function, for different ADC limits.	175
6.2	Momentum ranges and description of the PID approach assumed for the muon and pion classification task.	195
6.3	Optimal values of the hyperparameters used by the BDT, for each momentum range.	210

LIST OF TABLES

6.4	Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges.	210
6.5	Summary of parameters and sampled values used in the optimisation of the clustering algorithm.	229
7.1	Estimated event rates in ND-GAr.	237
7.2	Step-by-step ν_μ CC selection cuts and cumulative passing rates.	244
7.3	Neutrino interaction systematic parameters implemented in GENIE used in the DUNE TDR LBL analysis.	271
7.4	Neutrino interaction systematic parameters used in the DUNE TDR LBL analysis not present in GENIE	272

List of Abbreviations

ADC	Analog to Digital Converter	HPgTPC	High Pressure gaseous Time Projection Chamber
ALEPH	Apparatus for LEP PHysics		
ALICE	A Large Ion Collider Experiment	LBL	Long BaseLine
BDT	Boosted Decision Tree	MuID	Muon IDentification system
CC	Charged Current	NC	Neutral Current
DM	Dark Matter	ND	Near Detector
DUNE	Deep Underground Neutrino Experiment	ND-GAr	Near Detector Gaseous Argon
ECal	Electromagnetic Calorimeter	ND-LAr	Near Detector Liquid Argon
FD	Far Detector	PDG	Particle Data Group
FHC	Forward Horn Current	RHC	Reverse Horn Current

1

35

Introduction

36

2

37

38

Neutrino physics

39 *Little particles of inspiration sleet through the universe all the time traveling
40 through the densest matter in the same way that a neutrino passes through a
41 candyfloss haystack, and most of them miss.*

42

– Terry Pratchett, *Sourcery*

43 Ever since they were postulated in 1930 by Wolfgang Pauli to explain the continuous
44 β decay spectrum [2] and later found by F. Reines and C. Cowan at the Savannah
45 River reactor in 1953 [3], neutrinos have had a special place among all other elementary
46 particles. They provide a unique way to probe a wide range of quite different physics,
47 from nuclear physics to cosmology, from astrophysics to colliders. Moreover, there is
48 compelling evidence to believe that the study of neutrinos may be key to unveil different
49 aspects of physics beyond the SM, difficult to test elsewhere.

50 In this Chapter, I will review the basics of neutrino physics, from its role within the
51 SM to the main open questions related to the neutrino sector, paying special attention
52 to the phenomenology of neutrino oscillations.

53 2.1 Neutrinos in the SM

54 The SM of fundamental interactions was initially proposed in 1967 by S. Glashow,
55 S. Weinberg and A. Salam[4–6]. This theoretical framework describes the dynamics
56 of leptons and quarks, by introducing a collection of mediating gauge vector bosons
57 and one scalar particle, known as the Higgs boson. It assumes that the local $SU(3) \times$

CHAPTER 2. NEUTRINO PHYSICS

58 SU(2)_L × U(1)_Y gauge symmetry is an internal symmetry of the system, with SU(3)
 59 describing quantum chromodynamics, and SU(2)_L × U(1)_Y being the gauge groups of
 60 the electroweak sector. For a detailed overview of the SM of electroweak interactions,
 61 see Ref. [7].

62 In the SM, neutrinos appear in three flavours, namely ν_e , ν_μ , and ν_τ . These are
 63 associated with the corresponding charged leptons, e , μ , and τ . Neutrinos exist only
 64 as left-handed particles, grouped in doublets with the charged leptons, while the later
 65 come in both chirality states:

$$\begin{pmatrix} \nu_e \\ e_L^- \end{pmatrix}, \quad \begin{pmatrix} \nu_\mu \\ \mu_L^- \end{pmatrix}, \quad \begin{pmatrix} \nu_\tau \\ \tau_L^- \end{pmatrix}, \quad e_R^-, \quad \mu_R^-, \quad \tau_R^-. \quad (2.1)$$

66 Similarly, quarks also exist in both chirality states, and are grouped as:

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad \begin{pmatrix} s_L \\ c_L \end{pmatrix}, \quad \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \quad u_R, \quad d_R, \quad s_R, \quad c_R, \quad t_R, \quad b_R. \quad (2.2)$$

67 The fact that there are no right-handed neutrino fields implies that neutrinos are
 68 strictly massless within the SM. This restriction follows from the experimental observation
 69 that all neutrinos produced via weak interactions are pure left-handed helicity states
 70 (and similarly antineutrinos are pure right-handed states). The hypothetical existence
 71 of right-handed neutrinos could be indirectly inferred from the observation of non-zero
 72 neutrino masses, nevertheless the existence of neutrino masses is not a sufficient condition
 73 for the existence of such fields.

74 Left and right-handed fermions transform differently under SU(2)_L × U(1)_Y rotations,
 75 as the right-handed particles are singlets under SU(2)_L. Applying a local transformation,
 76 they change as:

$$\begin{aligned} \psi_L &\longrightarrow e^{-iY\beta(x)/2} e^{-iT_a\alpha_a(x)} \psi_L, \\ \psi_R &\longrightarrow e^{-iY\beta(x)/2} \psi_R, \end{aligned} \quad (2.3)$$

77 where $Y/2$ and T_a are the generators of SU(2)_L and U(1)_Y, respectively, and $\beta(x)$ and

2.1. NEUTRINOS IN THE SM

Table 2.1: Values of T_3 and $Y/2$ assigned to the first generation of fermions.

	e_L	ν_e	e_R	u_L	d_L	u_R	d_R
T_3	-1/2	1/2	0	1/2	-1/2	0	0
$Y/2$	-1/2	-1/2	-1	1/6	1/6	2/3	-1/3

78 $\alpha_a(x)$ are the parameters of the rotation.

79 The values of the quantum numbers $Y/2$ and T_3 , the third component of the weak
80 isospin, have to be assigned to the different particles. The values of T_3 follow from the
81 commutation relations of the generators of $SU(2)$. After the spontaneous symmetry
82 breaking $SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}$, one finds the relation which determines the electric
83 charge:

$$Q = T_3 + \frac{Y}{2}, \quad (2.4)$$

84 Setting the electric charge to -1 for electrons, we can find the values of the hypercharge
85 for the rest of the fermions. The resulting values for the first generation of leptons and
86 quarks are shown in Tab. 2.1.

87 It is clear that the free Lagrangian of the theory is not be invariant under the gauge
88 transformations, as the kinetic terms contain derivatives. Therefore, to make it invariant,
89 one needs to introduce a set of gauge bosons. They appear in the so-called covariant
90 derivative, which replaces the common derivative and transforms in the same way as the
91 fermion fields under local rotations. This constrain fixes completely the transformations
92 of the spin-1 fields. For left and right-handed particles, the covariant derivatives are
93 given by:

$$\begin{aligned} D_\mu \psi_L &= \left(\partial_\mu + ig' \frac{Y}{2} B_\mu + ig T_a W_\mu^a \right) \psi_L, \\ D_\mu \psi_R &= \left(\partial_\mu + ig' \frac{Y}{2} B_\mu \right) \psi_R, \end{aligned} \quad (2.5)$$

94 where W_μ^i , $i = 1, 2, 3$ and B_μ are the gauge bosons for the $SU(2)_L$ and $U(1)_Y$ factors,
95 respectively, and g and g' are the corresponding gauge couplings. It can be shown that
96 these fields transform in the adjoint representation of the gauge group.

CHAPTER 2. NEUTRINO PHYSICS

Table 2.2: Neutral current couplings.

	u	d	ν_e	e
$2v_f$	$1 - \frac{8}{3}\sin^2\theta_W$	$-1 + \frac{4}{3}\sin^2\theta_W$	1	$-1 + 4\sin^2\theta_W$
$2a_f$	1	-1	1	-1

97 So far, the theory only contains massless particles, as adding bare mass terms to
 98 the Lagrangian would spoil the gauge symmetry. Therefore, the mass terms need to
 99 be induced by a spontaneous violation of the symmetries. In the SM, the responsible
 100 for this is the Higgs mechanism. The Higgs doublet is coupled to the gauge bosons
 101 through the covariant derivative, and to the fermions through the Yukawa couplings.
 102 Upon spontaneous symmetry breaking, the vacuum expectation value of the Higgs field
 103 generate the mass terms of the particles.

104 In order to obtain the physical intermediate vector boson states, we need to perform
 105 the following redefinitions:

$$\begin{aligned} A_\mu &= \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu, \\ Z_\mu &= \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu, \\ W_\mu^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2), \end{aligned} \tag{2.6}$$

106 where A_μ is the photon field, and Z_μ and W_μ^\pm are the neutral and the charged weak
 107 boson fields, respectively. The Weinberg angle, θ_W , relates the weak coupling constants
 108 and the electric charge:

$$e = g' \cos \theta_W = g \sin \theta_W. \tag{2.7}$$

109 At this point, the interacting part of the electroweak Lagrangian can be re-written
 110 as the sum of three contributions: the electromagnetic (EM), charged-current (CC) and
 111 neutral-current (NC) components:

$$\begin{aligned} \mathcal{L}_{\text{EW}}^{\text{int}} &= \mathcal{L}_{\text{EM}} + \mathcal{L}_{\text{CC}} + \mathcal{L}_{\text{NC}} \\ &= -eA_\mu J_{\text{EM}}^\mu - \frac{g}{2\sqrt{2}} (W_\mu^+ J_{\text{CC}}^\mu + \text{h.c.}) - \frac{g}{2\cos\theta_W} Z_\mu J_{\text{NC}}^\mu, \end{aligned} \tag{2.8}$$

2.2. TROUBLE IN THE NEUTRINO SECTOR

112 with the currents defined as:

$$\begin{aligned} J_{\text{EM}}^\mu &= \sum_f Q_f \bar{f} \gamma^\mu f, \\ J_{\text{CC}}^\mu &= \sum_\ell \bar{\nu}_\ell \gamma^\mu (1 - \gamma_5) \ell + \sum_f \bar{u}_f \gamma^\mu (1 - \gamma_5) d_f, \\ J_{\text{NC}}^\mu &= \sum_f \bar{f} \gamma^\mu (v_f - a_f \gamma_5) f, \end{aligned} \quad (2.9)$$

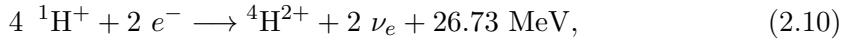
113 where f denotes any SM fermion, ℓ and ν_ℓ a charged lepton and a neutrino of any flavour,
114 and u_f and d_f an up-like and a down-like quarks of any flavour. For the NC case, the
115 values of the v_f and a_f couplings are given in Tab. 2.2.

116 As seen in Eq. (5.8), in the electroweak theory neutrinos are coupled to the Z boson
117 in a universal way. Therefore, by measuring the so-called invisible decay width of the Z
118 boson we have an estimate of the number of light (i.e. lighter than the Z boson) neutrino
119 flavours. This number was measured by LEP in a combined analysis of $e^+ e^- \rightarrow \mu^+ \mu^-$
120 and $e^+ e^- \rightarrow$ hadrons to be $N_\nu = 2.9840 \pm 0.0082$ [8].

121 2.2 Trouble in the neutrino sector

122 2.2.1 The solar neutrino problem

123 Neutrinos are produced everywhere in vast amounts. One of the most prominent sources
124 of neutrinos in our vicinity is our Sun. The Sun is powered mainly by two nuclear fusion
125 reactions, the p – p chain and the CNO cycle. In both cases, the overall reaction is:



126 where part of the released energy is lost to the neutrinos. The electron neutrinos
127 produced are often labelled after the processes that generate them. Figure 2.1 shows the
128 solar neutrino flux as a function of the neutrino energy, broken down by the production
129 process.

CHAPTER 2. NEUTRINO PHYSICS

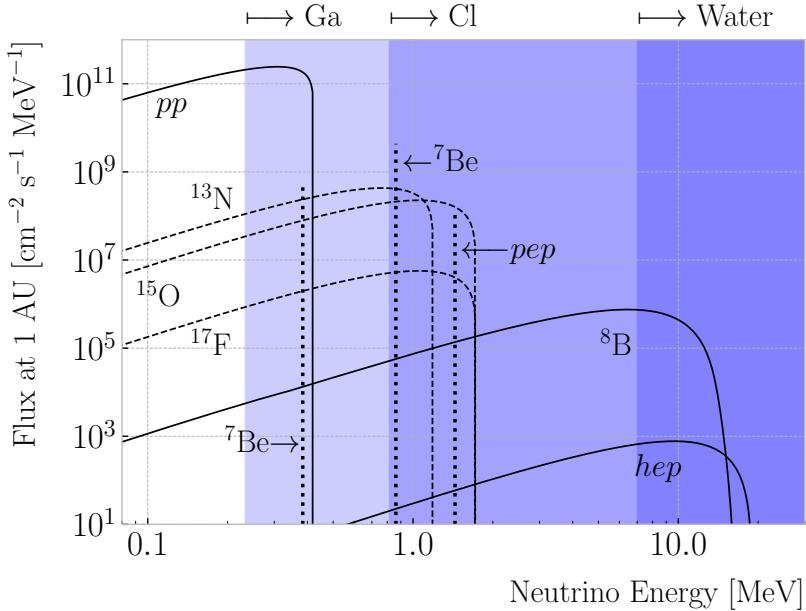


Figure 2.1: Solar neutrino fluxes for the solar model BS05(OP). The detection thresholds for Gallium, Chlorine and water-based experiments are also shown. Figure adapted from Ref. [10].

130 In the late 1960s, the Brookhaven Solar Neutrino Experiment, led by R. Davis, started
 131 data taking with the goal of measuring the solar neutrino flux [9]. The experiment
 132 used a tank containing 380 m^3 of tetrachloroethene (C_2Cl_4), a liquid commonly used
 133 in dry-cleaning, located 1.5 km underground in the Homestake mine, in Lead, South
 134 Dakota. The incoming neutrinos would get captured following the reaction:



135 therefore allowing to measure the neutrino flux by counting the ${}^{37}\text{Ar}$ isotopes. The
 136 threshold for this reaction is 0.814 MeV, just below the 0.862 MeV line from the ${}^7\text{Be}$
 137 ground state transition.

138 The results of the experiment were compared to the theoretical predictions made by
 139 J. Bahcall [11]. During its operation from 1968 to 2002, the experiment observed a solar
 140 ν_e flux that was approximately a third of the total prediction [12].

141 In the early 1990s, the SAGE [13] and GALLEX [14] experiments started operations.

2.2. TROUBLE IN THE NEUTRINO SECTOR

142 The detection principle used for both experiments was similar to that of the Homestake
143 experiment, but using ^{71}Ga instead of C_2Cl_4 . With a detection threshold of 0.233 MeV,
144 the Gallium-based experiments were able to observe the pp neutrino flux. Both
145 experiments measured a solar electron neutrino flux that was a factor of two lower
146 than the predictions, demonstrating that this deficit was energy-dependent.

147 In the early 2000s, the SNO experiment put an end to the solar neutrino puzzle
148 [15, 16]. Thanks to its directionality capabilities, being a Cherenkov light detector, as
149 well as to its heavy water target, SNO measured the total solar neutrino flux through
150 the NC process:



151 where $\alpha = e, \mu, \tau$. This measurement agreed with the solar model predictions. Then,
152 measuring the CC reaction:



153 they were able to establish that the ν_μ and ν_τ solar fluxes are in fact non-zero, revealing
154 that electron neutrinos were transitioning into different flavours.

155 2.2.2 The atmospheric neutrino problem

156 When cosmic-rays interact with the atoms in the upper atmosphere, a plethora of
157 hadrons, mainly π and K mesons, are produced. In particular, for the charged pions,
158 we have the following decay chain dominates:



159 and similar for the antiparticles. For neutrino energies < 1 GeV, the ratio:

$$\frac{N(\nu_\mu + \bar{\nu}_\mu)}{N(\nu_e + \bar{\nu}_e)}, \quad (2.15)$$

160 of produced neutrinos and antineutrinos is, in good approximation, equal to two [17].

CHAPTER 2. NEUTRINO PHYSICS

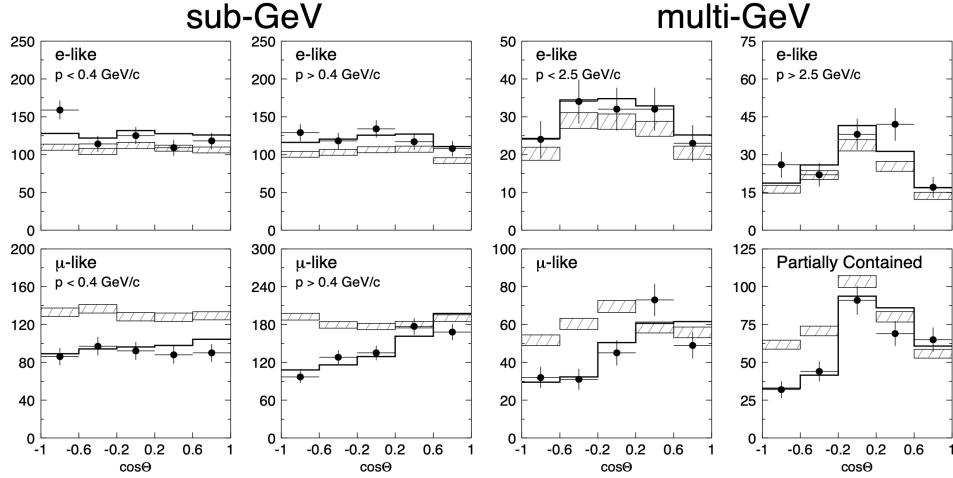


Figure 2.2: Zenith angle distributions for the selected ν_e (top row) and ν_μ (bottom row) events in the SK detector. The hatched region corresponds to the expectation in the case of no oscillations, whereas the solid line indicates the best-fit in the case of $\nu_\mu \rightarrow \nu_\tau$ oscillations. Figure taken from Ref. [22].

161 During the 1980s, several proton decay experiments, like Kamiokande [18], IMB [19],

162 MACRO [20], and Soudan-2 [21], measured the flux of atmospheric neutrinos. This was

163 an important part of their research programme, as the atmospheric neutrinos constitute

164 their main background. All these experiments reported an atmospheric neutrino ratio

165 lower than the predictions.

166 A few years before the SNO discovery, in 1998, Super-Kamiokande (SK) collaboration

167 measured the atmospheric ν_e and ν_μ spectra as a function of the zenith angle [22].

168 Upward-going particles have negative zenith angle, $\cos \Theta < 0$, indicating that they

169 entered from the bottom of the detector. These upward-going neutrinos had to travel

170 through the Earth in order to reach the detector, allowing SK to probe a broad range

171 of baselines. Figure 2.2 shows the reported distributions (black dots), compared to the

172 no oscillations prediction (hatched region). This measurement confirmed that muon

173 neutrinos transition to other flavours, and that this phenomenon depends both on the

174 energy and the path length of the neutrino.

175 The SK and SNO findings provided definitive evidence for the existence of neutrino

176 oscillations, and therefore non-zero neutrino masses. This constitutes one of the

177 groundbreaking discoveries of modern physics and has acted as driving force for beyond

2.3. MASSIVE NEUTRINOS

the Standard Model (BSM) physics. The minimal extension of the SM we can do to address these phenomena is introducing different masses for at least two of the neutrinos. This way, we are left with three neutrino mass eigenstates ν_1 , ν_2 , and ν_3 , with masses m_1 , m_2 , and m_3 respectively, which in general will not coincide with the flavour eigenstates, ν_e , ν_μ , and ν_τ .

2.3 Massive neutrinos

The existence of neutrino oscillations imply that neutrinos are massive particles. However, as we have seen before, within the SM neutrinos are massless, as they do not have a mass term in the Lagrangian. If one wants to give neutrinos a mass, the particle content of the SM needs to be expanded.

A way of generating massive neutrinos while maintaining gauge invariance is by introducing an arbitrary number of sterile neutrinos N_i , $i = 1, \dots, m$. These allow for two different types of neutrino mass terms:

$$-\mathcal{L}_{M_\nu} = \sum_{i=1}^m \sum_{j=1}^3 M_D^{ij} \bar{N}_i \nu_{Lj} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m M_N^{ij} \bar{N}_i N_j^c + \text{h.c.}, \quad (2.16)$$

where M_D is a complex $m \times 3$ matrix and M_N a complex and symmetric $m \times m$ matrix. The first term, often referred to as the Dirac mass term, arises from the corresponding Yukawa interaction after the spontaneous electroweak symmetry breaking, similar to the other fermions. The second term, called the Majorana mass term, is allowed in the Lagrangian, as it is a singlet of the gauge group. However, it violates lepton number conservation by two units.

If one imposes lepton number symmetry conservation, the Majorana term must banish, $M_N = 0$. In this case, if $m = 3$ we can identify the sterile neutrinos as the right-handed component of the neutrino field. The Dirac mass matrix can be diagonalised using two unitary matrices, V_R^ν and V_L^ν , as:

$$M_D = V_R^\nu \text{ diag}(m_1, m_2, m_3) V_L^{\nu\dagger}, \quad (2.17)$$

CHAPTER 2. NEUTRINO PHYSICS

201 where m_i , $i = 1, 2, 3$ are the masses of the three neutrino mass eigenstates.

202 The neutrino mass term can be written in term of the resulting eigenstates as:

$$-\mathcal{L}_{M_\nu} = \sum_{i=1}^3 m_i \bar{\nu}_{Di} \nu_{Di}, \quad (2.18)$$

203 with:

$$\nu_{Di} = \left(V_L^{\nu\dagger} \nu_L \right)_i + \left(V_R^{\nu\dagger} N \right)_i. \quad (2.19)$$

204 In this scenario, both the low energy particle budget and the symmetries of the SM
 205 have to be modified. Moreover, the masses of the neutrinos are generated exclusively
 206 through the Higgs mechanism, which does not explain why they are much smaller than
 207 those of the charged leptons.

208 Going back to the general case, we can re-write Eq. (2.16) in matrix form as:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2} \begin{pmatrix} \bar{\nu}_L^c & \bar{N} \end{pmatrix} \begin{pmatrix} 0 & M_D^T \\ M_D & M_N \end{pmatrix} \begin{pmatrix} \nu_L \\ N^c \end{pmatrix} + \text{h.c.} = \bar{\nu}^c M_\nu \nu + \text{h.c.}, \quad (2.20)$$

209 with $\nu = (\nu_L, N^c)^T$ being a $(3+m)$ -dimensional vector grouping the active and the
 210 sterile neutrinos. The matrix M_ν , which is a complex $(3+m) \times (3+m)$ symmetric
 211 matrix, can be diagonalised by means of a unitary matrix V^ν , yielding:

$$M_\nu = V^\nu \text{ diag}(m_1, m_2, \dots, m_{3+m}) V^{\nu T}. \quad (2.21)$$

212 Using this eigendecomposition, the neutrino mass term can be expressed as:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2} \sum_{i=1}^{3+m} m_i \bar{\nu}_{Mi} \nu_{Mi}, \quad (2.22)$$

213 where the states ν_{Mi} , commonly referred to as Majorana neutrinos, are defined as:

$$\nu_{Mi} = \left(V^{\nu\dagger} \nu \right)_i + \left(V^{\nu\dagger} \nu \right)_i^c, \quad (2.23)$$

214 in such a way that the Majorana condition, $\nu_M^c = \nu_M$, holds true.

2.4. NEUTRINO OSCILLATION FORMALISM

215 As a consequence of the Majorana condition, the neutrino and the antineutrino states
 216 can be described in terms of a single field. As opposed to the charged leptons, which
 217 need to be represented by a four-component or Dirac spinor, the Majorana neutrino is
 218 described by a two-component or Weyl spinor.

219 If the eigenvalues of the Majorana mass matrix, M_N , are much larger than the
 220 electroweak symmetry breaking scale, the diagonalisation of M_ν leads to 3 light and m
 221 heavy neutrino states:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2}\bar{\nu}_l M_l \nu_l + \frac{1}{2}\bar{\nu}_h M_h \nu_h, \quad (2.24)$$

222 where the two mass matrices are given by:

$$\begin{aligned} M_l &\simeq -V_l^T M_D^T M_N^{-1} M_D V_l, \\ M_h &\simeq V_h^T M_N V_h, \end{aligned} \quad (2.25)$$

223 with V_l and V_h two unitary matrices.

224 This scenario represents the so-called see-saw mechanism [23–27]. The name comes
 225 from the fact that the masses of the heavy states are proportional to M_N , whereas for
 226 the light states they are proportional to M_N^{-1} . While both the heavy and the light
 227 neutrinos are Majorana particles, it can be shown that the heavy states are mainly
 228 right-handed, whereas the light ones are mostly left-handed.

2.4 Neutrino oscillation formalism

230 Neutrino oscillations were first proposed in 1958 by B. Pontecorvo [28], inspired by the
 231 neutral kaon oscillation phenomenon [29]. Neutral kaons, K^0 and \bar{K}^0 , have opposite
 232 strangeness (± 1) and are produced in strong processes. It was observed that, when
 233 having a beam initially pure of neutral kaons of one type, these would transition into
 234 their antiparticles while propagating. Because the weak interaction does not conserve
 235 strangeness, neutral kaons can change their identity via the processes shown in Fig. 2.3.

236 The mixing considered initially by Pontecorvo was between the neutrino and the

CHAPTER 2. NEUTRINO PHYSICS

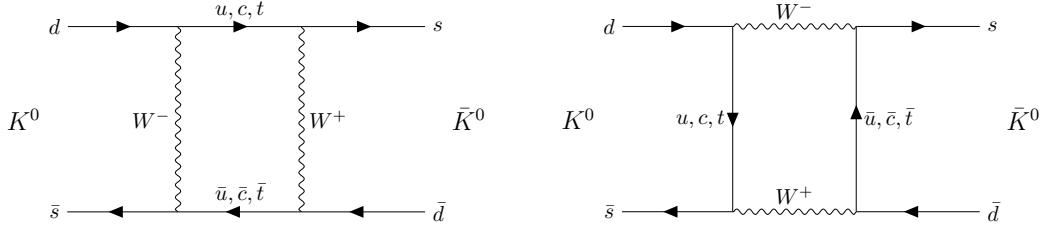


Figure 2.3: $K^0 \rightleftharpoons \bar{K}^0$ mixing through W^\pm exchange.

237 antineutrino states, as only one neutrino flavour was known at the time. After the
238 discovery of the muon neutrino, the mixing between flavours was also explored [30].

239 In the general case, we have 3 active and m sterile neutrinos, resulting in $3 + m$
240 neutrino mass eigenstates. Working in the mass basis, the leptonic charged-current
241 Lagrangian can be written as:

$$-\mathcal{L}_{CC}^{lep} = \frac{g}{\sqrt{2}} (\bar{e}_L, \bar{\mu}_L, \bar{\tau}_L) \gamma^\mu U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_{3+m} \end{pmatrix} W_\mu^+ + \text{h.c.}, \quad (2.26)$$

242 where U is a $3 \times (3 + m)$ matrix which obeys $UU^\dagger = I_{3 \times 3}$, but in general will not be
243 unitary, $U^\dagger U \neq I_{(3+m) \times (3+m)}$.

244 The leptonic mixing matrix, U , establishes how the neutrino mass states couple to
245 the charged leptons. In general, a complex $n \times n$ matrix can be fully specified by $2n^2$ real
246 parameters. If the matrix is unitary, then the number of independent parameters reduces
247 to n^2 , as one has to impose n normalisation and $n(n - 1)$ orthogonality constraints.
248 In our case, we can further reduce the number of parameters by performing a phase
249 redefinition of the charged lepton fields, without affecting the physics. This is not true
250 for the neutrinos. As they may be their own antiparticles, one is not allowed to remove
251 any physically relevant phases. If we consider n generations of leptons, the total number
252 of parameters in the mixing matrix is $n^2 - n$. Out of these, half of them are mixing
253 angles, while the other half are complex phase factors.

254 Considering the extended SM without any additional sterile neutrino states, the
255 resulting 3×3 mixing matrix is unitary. This matrix, often called the Pontecorvo-Maki-

2.4. NEUTRINO OSCILLATION FORMALISM

256 Nakagawa-Sakata (PMNS) matrix [31, 32], relates the set of active neutrinos and the
257 three mass eigenstates as:

$$|\nu_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle, \quad (2.27)$$

258 where the Greek index α denotes the flavour $\{e, \mu, \tau\}$ and the Latin index i the mass state
259 $\{1, 2, 3\}$. This leptonic mixing matrix may be parametrized in terms of 6 parameters, 3
260 of which are mixing angles θ_{12} , θ_{13} and θ_{23} , one CP-violating phase δ_{CP} and 2 Majorana
261 phases α and β :

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha} & 0 \\ 0 & 0 & e^{i\beta} \end{pmatrix}, \quad (2.28)$$

262 where $c_{ij} \equiv \cos \theta_{ij}$ and $s_{ij} \equiv \sin \theta_{ij}$. This matrix is analogous to the Cabibbo-Kobayashi-
263 Maskawa (CKM) matrix in the quark sector. If neutrinos are Dirac fermions, we can
264 drop the Majorana phases in the PMNS matrix, as in this case we can perform the
265 phase redefinitions. However, these phases play no role on the neutrino oscillation
266 phenomenology.

267 In the case that additional sterile neutrinos states are present, the full leptonic mixing
268 matrix would not be unitary in general. For instance, in the see-saw scenario, the 3×3
269 submatrix for the three light Majorana neutrinos is not unitary. However, the deviations
270 from unitarity are of the order $\mathcal{O}(M_D/M_N)$, and therefore expected to be negligible.

271 2.4.1 Oscillations in vacuum

272 Consider the case where a neutrino of flavour α is produced at $t = 0$, and then it
273 propagates through vacuum. Such a state will evolve in time according to the relation:

$$|\nu_\alpha(\vec{x}, t)\rangle = \sum_{i=1}^3 U_{\alpha i}^* e^{-i(E_i t - \vec{p}_i \cdot \vec{x})} |\nu_i(\vec{x} = \vec{0}, t = 0)\rangle, \quad (2.29)$$

274 in the plane wave approximation, as the mass eigenstates are also eigenstates of the free
275 Hamiltonian.

CHAPTER 2. NEUTRINO PHYSICS

276 This way, the probability for the neutrino to transition from flavour α to flavour β
277 will be given by:

$$P(\nu_\alpha \rightarrow \nu_\beta) = |\langle \nu_\beta | \nu_\alpha(\vec{x}, t) \rangle|^2 = \left| \sum_{i=1}^3 \sum_{j=1}^3 U_{\alpha i}^* U_{\beta j} e^{-i(E_i t - \vec{p}_i \cdot \vec{x})} \langle \nu_j | \nu_i \rangle \right|^2 \\ = \left| \sum_{i=1}^3 U_{\alpha i}^* U_{\beta i} e^{-i(E_i t - \vec{p}_i \cdot \vec{x})} \right|^2, \quad (2.30)$$

278 where we have used the orthogonality relation $\langle \nu_i | \nu_j \rangle = \delta_{ij}$. A usual approximation to
279 take at this point is to consider ultra-relativistic neutrinos, i.e. $p_i \simeq E$, so we can write
280 the dispersion relations as:

$$E_i = \sqrt{p_i^2 + m_i^2} \approx E + \frac{m_i^2}{2E}. \quad (2.31)$$

281 In the end, assuming $t \approx L$ where L is the distance between the production and the
282 detection points, the probability for the $\nu_\alpha \rightarrow \nu_\beta$ transition becomes:

$$P(\nu_\alpha \rightarrow \nu_\beta) = \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j} e^{-i \frac{\Delta m_{ij}^2}{2E} L} \\ = \delta_{\alpha\beta} - 4 \sum_{i < j} \Re e [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin^2 \left(\frac{\Delta m_{ij}^2}{4E} L \right) \\ + 2 \sum_{i < j} \Im m [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin \left(\frac{\Delta m_{ij}^2}{2E} L \right), \quad (2.32)$$

283 where Δm_{ij}^2 is the difference of the squared masses of the j th and i th neutrino mass
284 eigenvalues. At this point, it is usual to write the phase responsible for the oscillations
285 as:

$$\Delta_{ij} \equiv \frac{\Delta m_{ij}^2}{4E} L \simeq 1.27 \frac{\Delta m_{ij}^2}{(\text{eV}^2)} \frac{L}{(\text{km})} \frac{(\text{GeV})}{E}. \quad (2.33)$$

286 Notice that, in the case of antineutrinos, the only difference would be the sign of the
287 last term in the oscillation probability. As the process $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ is the CP-mirror image
288 of $\nu_\alpha \rightarrow \nu_\beta$, the differences between their oscillation probabilities would be a measure of

2.4. NEUTRINO OSCILLATION FORMALISM

289 CP symmetry violation:

$$\begin{aligned} A_{CP}^{\alpha\beta} &= P(\nu_\alpha \rightarrow \nu_\beta) - P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) \\ &= 4 \sum_{i<j} \Im \left[U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* \right] \sin 2\Delta_{ij}. \end{aligned} \quad (2.34)$$

290 Assuming that CPT invariance holds, then the following relation must be true:

$$P(\nu_\alpha \rightarrow \nu_\beta) = P(\bar{\nu}_\beta \rightarrow \bar{\nu}_\alpha), \quad (2.35)$$

291 as these two process are related by the CPT symmetry. From the definition of probability,
292 we also must have:

$$\sum_{\beta} P(\nu_\alpha \rightarrow \nu_\beta) = \sum_{\beta} P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = 1, \quad (2.36)$$

293 where the sum includes all flavours, including α . From these two constraints, one can
294 probe that:

$$A_{CP}^{\alpha\beta} = -A_{CP}^{\beta\alpha}, \quad (2.37)$$

295 and in particular:

$$A_{CP}^{\alpha\alpha} = 0. \quad (2.38)$$

296 A direct consequence of this last relation is that there are no observable CP-violating
297 effects in the so-called disappearance experiments. One needs to perform appearance
298 experiments, where the flavour detected is different from the original flavour, in order
299 to measure the CP asymmetry. Neutrino experiments often report the amount of CP-
300 violation through the Jarlskog invariant. In terms of the parametrisation typically used
301 to write the PMNS matrix, it is given by:

$$J = \frac{1}{8} \cos \theta_{13} \sin 2\theta_{12} \sin 2\theta_{13} \sin 2\theta_{23} \sin \delta_{CP}. \quad (2.39)$$

302 The Jarlskog invariant can be used to compare the amount of CP-violation in the lepton
303 and the quark sectors, where $J = 3.12^{+0.13}_{-0.12} \times 10^{-5}$ in the latter [33].

CHAPTER 2. NEUTRINO PHYSICS

304 2.4.2 Oscillations in matter

305 When neutrinos propagate through matter, their oscillation can be affected in mainly
 306 two ways. First, neutrinos can inelastically scatter with nuclei, thus destroying the
 307 coherent propagation of their quantum state. Nevertheless, in most cases this effect is
 308 negligible (even in very dense mediums like the core of the Sun). Second, neutrinos can
 309 also experience coherent or forward scatterings, that can affect their oscillation but not
 310 lose the coherent propagation of the state.

311 The first proposed model to account for neutrino oscillations in matter was proposed
 312 by Mikhaev, Smirnov and Wolfenstein (MSW) [34]. It relies on the fact that, as the
 313 only charged lepton present in ordinary matter is the electron, electron neutrinos can
 314 undergo both charged and neutral-current interactions with matter whereas for muon
 315 and tau neutrinos just neutral currents are possible.

316 An illustrative way to introduce the MSW mechanism is by considering the two
 317 flavours case. It can be shown that the evolution of the two flavour eigenstates in vacuum
 318 is given by the following time-dependent Schrödinger equation:

$$i \frac{d}{dt} \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = H_V \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}, \quad (2.40)$$

319 with a vacuum Hamiltonian given by:

$$H_V = \frac{\Delta m^2}{4E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}, \quad (2.41)$$

320 where Δm^2 is the mass splitting between the two neutrino states and θ the only mixing
 321 angle. For simplicity, I omit the terms of the Hamiltonian that are proportional to the
 322 identity, as they do not affect the oscillation phenomenology.

323 The NC contribution to the matter potential is identical for all the flavours, and has
 324 the form:

$$V_{NC} = -\frac{G_F}{\sqrt{2}} N_n(x), \quad (2.42)$$

2.4. NEUTRINO OSCILLATION FORMALISM

325 where G_F is the Fermi constant and $N_n(x)$ the local neutron density. Because it is
 326 common to all flavours, I do not take it into account in the effective Hamiltonian, as it
 327 would appear as a term proportional to the identity. The CC component only affects
 328 the electron neutrino (and antineutrino). It can be written as:

$$V_{\text{CC}} = \pm \sqrt{2} G_F N_e(x), \quad (2.43)$$

329 with $N_e(x)$ being the local electron density in the material. In the end, the effective
 330 Hamiltonian which describes the propagation of the flavour eigenstates in matter only
 331 contains an extra $\nu_e - \nu_e$ element:

$$H_M = H_V + \begin{pmatrix} V_{\text{CC}} & 0 \\ 0 & 0 \end{pmatrix}. \quad (2.44)$$

332 The solution to the Schrödinger equation greatly simplifies if one considers the case
 333 of a constant matter density. In that case, the effective Hamiltonian can be diagonalised,
 334 obtaining the effective neutrino mass eigenstates in matter. It can be re-written in the
 335 same form as the vacuum Hamiltonian:

$$H_M = \frac{\Delta m_m^2}{4E} \begin{pmatrix} -\cos 2\theta_m & \sin 2\theta_m \\ \sin 2\theta_m & \cos 2\theta_m \end{pmatrix}, \quad (2.45)$$

336 where the effective mass splitting and the effective mixing angle are given by:

$$\begin{aligned} \Delta m_m^2 &= \lambda \Delta m_m^2, \\ \sin 2\theta_m &= \frac{\sin 2\theta_m}{\lambda} \end{aligned} \quad (2.46)$$

337 with:

$$\begin{aligned} \lambda &= \sqrt{(\cos 2\theta - A)^2 + \sin^2 2\theta}, \\ A &= \pm \frac{2\sqrt{2} G_F N_e E}{\Delta m^2}. \end{aligned} \quad (2.47)$$

338 In terms of the effective matter oscillation parameters, the transition probability

CHAPTER 2. NEUTRINO PHYSICS

339 $\nu_e \rightarrow \nu_\mu$ (in the two flavour approximation) reads:

$$P(\nu_e \rightarrow \nu_\mu) = \sin^2 2\theta_m \sin^2 \left(\frac{\Delta m_m^2}{2E} L \right) \quad (2.48)$$

340 From this last equation one can see that, when $\cos 2\theta = A > 0$ the oscillations are
341 greatly enhanced. This effect is known as the MSW resonance. For the neutrinos, this
342 resonant condition is only satisfied if $\Delta m^2 > 0$ (the opposite is true for antineutrinos).
343 This is can be exploited by long baseline experiments, which can gain sensitivity to the
344 neutrino mass hierarchy through matter effects.

345 2.4.3 Current status of neutrino oscillations

346 A wide range of neutrino experiments provide experimental input to the neutrino
347 oscillation framework, both using natural or synthetic neutrino sources. The results
348 from one of the neutrino global fit analyses, shown in Tab. 2.3¹, summarise well our
349 current understanding of the different oscillation parameters.

350 **Solar neutrino experiments** detect neutrinos produced in thermonuclear reactions
351 inside the Sun, mainly from the so-called *pp* chain and the CNO cycle. These neutrinos
352 have a typical energy in the range from 0.1 to 20 MeV. These experiments (Homestake
353 [35], GALLEX [14], SAGE [13], Borexino [36], Super-Kamiokande [37] and SNO [38])
354 provide the best sensitivities to θ_{12} and Δm_{21}^2 .

355 **Atmospheric neutrino experiments** detect the neutrino flux produced when
356 cosmic rays scatter with particles in Earth's atmosphere. These collisions generate particle
357 showers that eventually produce electron and muon neutrinos (and antineutrinos). Their
358 energies range from few MeV to about 10^9 GeV. Experiments, like Super-Kamiokande
359 [39] and IceCube [40] use atmospheric neutrinos to measure oscillations and are specially
360 sensitive to θ_{23} and Δm_{32}^2 .

361 **Reactor neutrino experiments** look for the $\bar{\nu}_e$ spectrum produced by nuclear
362 reactors, with energies in the MeV scale. Depending on the distance to the source,

¹These are the results reported during M. Tórtola's talk at Neutrino 2024 (see this link). I need to keep an eye and see if they publish these or other updated results in the near future.

2.5. OPEN QUESTIONS IN THE NEUTRINO SECTOR

Table 2.3: Summary of neutrino oscillation parameters determined in the Neutrino Global Fit of 2020 [1].

Parameter	Best fit $\pm 1\sigma$	3σ range
Δm_{21}^2 [eV $^2 \times 10^{-5}$]	$7.55^{+0.22}_{-0.20}$	6.98 – 8.19
$ \Delta m_{31}^2 $ [eV $^2 \times 10^{-3}$] (NO)	$2.51^{+0.02}_{-0.03}$	2.43 – 2.58
$ \Delta m_{31}^2 $ [eV $^2 \times 10^{-3}$] (IO)	$2.41^{+0.03}_{-0.02}$	2.34 – 2.49
$\sin^2 \theta_{12}/10^{-1}$	3.04 ± 0.16	2.57 – 3.55
$\sin^2 \theta_{23}/10^{-1}$ (NO)	$5.64^{+0.15}_{-0.21}$	4.23 – 6.04
$\sin^2 \theta_{23}/10^{-1}$ (IO)	$5.64^{+0.15}_{-0.18}$	4.27 – 6.03
$\sin^2 \theta_{13}/10^{-2}$ (NO)	$2.20^{+0.05}_{-0.06}$	2.03 – 2.38
$\sin^2 \theta_{13}/10^{-2}$ (IO)	$2.20^{+0.07}_{-0.04}$	2.04 – 2.38
δ_{CP}/π (NO)	$1.12^{+0.16}_{-0.12}$	0.76 – 2.00
δ_{CP}/π (IO)	$1.50^{+0.13}_{-0.14}$	1.11 – 1.87

363 long-baseline experiments like KamLAND [41] are sensitive to the solar mass splitting
364 Δm_{21}^2 whereas much shorter baseline experiment such as RENO [42] or DayaBay [43]
365 measure θ_{13} and Δm_{31}^2 .

366 **Accelerator experiments** measure neutrino fluxes generated in particle accelerators.
367 Usually mesons are produced in the accelerator to be focused into a beam, then some
368 decay to muon neutrinos and the rest are absorbed by a target. Depending on the
369 configuration one can obtain a beam made of mostly neutrinos or antineutrinos. The
370 typical energies of these neutrinos are in the GeV range. Experiments such as NOvA
371 [44], T2K [45], MINOS [46], OPERA [47] and K2K [48] (and in the future DUNE [49])
372 are primarily sensitive to θ_{13} , θ_{23} and Δm_{32}^2 . Also, in the coming years DUNE [49] and
373 Hyper-Kamiokande [50] will be sensitive to δ_{CP} .

374 2.5 Open questions in the neutrino sector

375 A crucial question that remains open these days, and is of vital importance for oscillation
376 phenomena, is whether the mass eigenvalue ν_3 is the heaviest (what we call normal
377 ordering) or the lightest (referred to as inverted ordering) of the mass eigenstates. In

CHAPTER 2. NEUTRINO PHYSICS

378 other words, this means that we do not know the sign of Δm_{32}^2 , so we can either have
379 $m_1 < m_2 < m_3$ (NO) or $m_3 < m_1 < m_2$ (IO).

380 Another big puzzle is related to the value of δ_{CP} . Nowadays it is poorly constrained,
381 with all values between π and 2π being consistent with data. A prospective measurement
382 different from $\delta_{CP} = 0, \pi$ will predict CP-violation in the leptonic sector, and thus
383 contribute along with the one measured in the quark sector to the total amount of
384 CP-violation. Although it is true that these two contributions by themselves are not
385 enough to explain the matter anti-matter asymmetry in our universe, the amount of
386 CP-violation in the leptonic sector can be key to explain such imbalance.

387 Both of these questions, because of their nature, could be understood thanks to
388 future oscillation experiments.

389 Notwithstanding, there are other mysteries that can not be unveiled just by conducting
390 oscillation experiments, as certain quantities do not influence these phenomena. Among
391 these there is the question of the absolute values of the neutrino masses. Depending
392 on the value of the lightest of the neutrino masses we can have different mass spectra,
393 from hierarchical $m_1 \ll m_2 < m_3$ (NO) or $m_3 \ll m_1 < m_2$ (IO) to quasi-degenerate
394 $m_1 \simeq m_2 \simeq m_3$.

395 Other open question concerns the nature itself of the neutrinos. If neutrinos are Dirac
396 particles then their mass term can be generated through the usual Higgs mechanism
397 by adding right-handed neutrino fields. However, if they are Majorana particles and
398 therefore their own antiparticles, there is no need to add extra fields to have the mass
399 term in the Lagrangian. Experiments like SuperNEMO [51], SNO+ [52] and NEXT
400 [53], which search for neutrino-less double beta decay, will be able to determine whether
401 neutrinos are Dirac or Majorana.

402 2.6 Neutrino interactions

403 The study of neutrino-nucleus interactions is of great importance for long baseline
404 neutrino oscillation experiments. The interaction model provides a mapping between

2.6. NEUTRINO INTERACTIONS

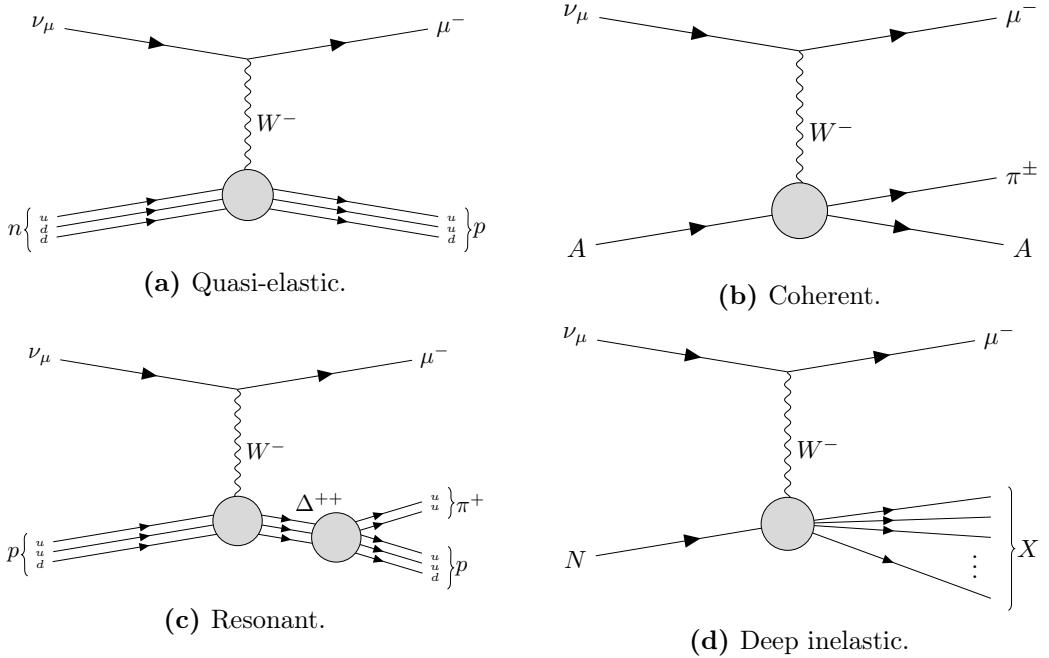


Figure 2.4: Feynman diagrams of the four most relevant CC interactions to long baseline neutrino oscillation experiments.

405 the energy of the incoming neutrino and the final state particles after the interaction.
 406 Because in this kind of experiments neutrinos are obtained as secondary decay products
 407 of mesons, typically charged pions and kaons, their energies are not known a priori. Not
 408 only that, but the kinematics of the interacting nucleon are also unknown. Therefore, we
 409 rely on the neutrino interaction models to provide this relation between the observables
 410 in the detector and the true kinematics of the neutrino. Interaction modelling is expected
 411 to be the one of the leading sources of systematic uncertainties in the next generation of
 412 long baseline experiments [54–56].

413 In the case of neutrino interactions with nuclei, at the energies relevant for long
 414 baseline oscillation experiments, around the GeV-scale, the process is dominated by
 415 the interaction between the neutrino and a single nucleon within the nuclear medium.
 416 Figure 2.4 shows examples of the four most common neutrino CC interactions. In this
 417 diagrams A indicated that the interaction happened with the nucleus as a whole, whereas
 418 N denotes a single nucleon.

419 At low energies, below 1 GeV, quasi-elastic (QE) interactions dominate. In a CCQE

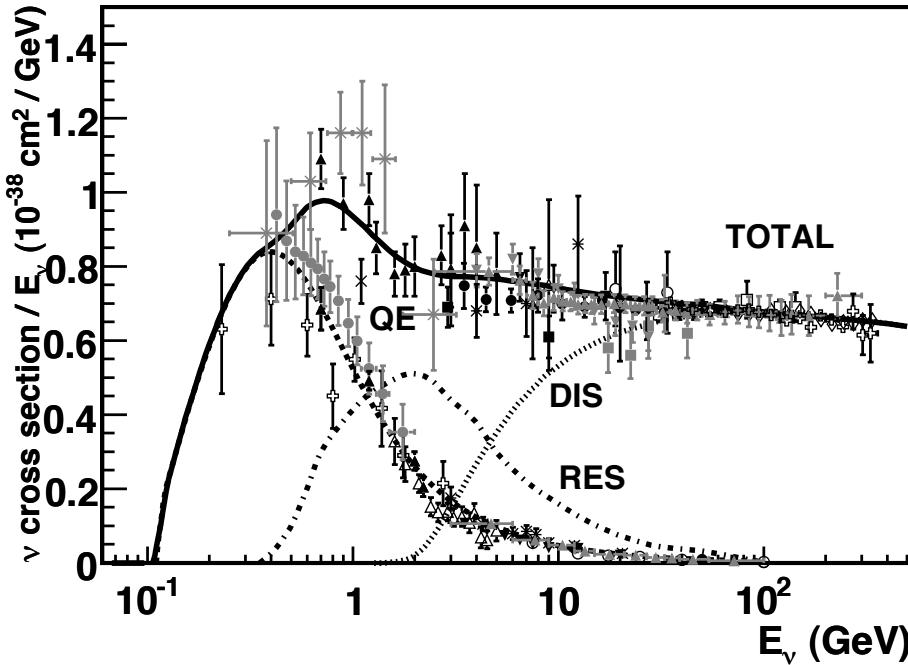


Figure 2.5: Total ν_μ CC cross section per nucleon as a function of the neutrino energy. The contributions from the different channels are shown. Figure taken from Ref. [57].

interaction a neutrino (antineutrino) interacts with a neutron (proton), converting it into a proton (neutron) which is then ejected from the nucleus together with the resulting charged lepton. At energies above 1 GeV the neutrino is able to excite the nucleon into a baryonic resonance, which promptly decays into a nucleon and a pion. These are the so-called resonant (RES) interactions. Neutrinos also interact with entire nucleus coherently, in the process known as coherent (COH) interaction. This kind of reactions also produce a single pion in the final state. At high neutrino energies, above 5 GeV, deep inelastic scattering (DIS) takes place. In these processes, the neutrino interacts with a single quark within the nucleon, breaking the nucleon and producing a hadronic shower.

Figure 2.5 shows a compilation of measurements of the total ν_μ CC cross section (see Ref. [57] for the details of the different experimental results). Also shown are the contributions from the different interaction modes. The contribution of the CCCOH interaction is omitted, as it is negligible compared to the others. This shows how the

2.6. NEUTRINO INTERACTIONS

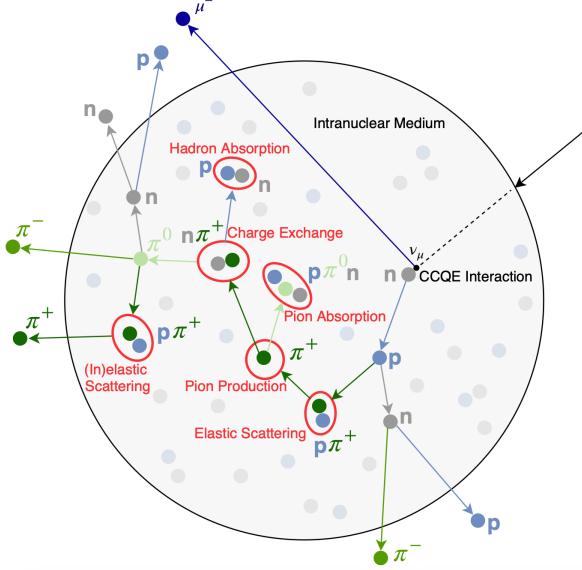


Figure 2.6: Schematic representation of a ν_μ CCQE interaction with a neutron inside a nucleus. The reaction produces a muon and a proton, which travel through the nuclear medium. The outgoing proton undergoes various kinds of hadronic FSIs on its way out. Figure taken from Ref. [58].

434 interaction model needs to accurately predict the neutrino-nucleon cross section for the
 435 different interaction modes across a broad energy range, to obtain the correct relative
 436 contributions.

437 Nuclear effects alter the neutrino cross section, as well as the multiplicities of the
 438 final state particles. Therefore, the interaction models need to account for the effects
 439 introduced by the nuclei. There are several models available to describe the initial state
 440 of the nucleus, like the relativistic Fermi gas model [59], spectral functions [60] or the
 441 random phase approximation [61]. The other main effect that interaction models have to
 442 deal with are the so-called final state interactions (FSI). These are the interactions of the
 443 particles produced in the neutrino-nucleon scattering as they travel through the nuclear
 444 medium. Typically, the lepton exits the nucleus without interacting. However, hadrons
 445 tend to get scattered, absorbed or re-emitted. These effects are usually described by
 446 means of intra-nuclear cascade models [62]. Figure 2.6 illustrates the effects of FSI on
 447 the observable particle content in the detector after a ν_μ CCQE interaction.

448 There exists a rich experimental programme dedicated to the measurement of neutrino

CHAPTER 2. NEUTRINO PHYSICS

449 cross sections. The list of such experiments in the recent years include MiniBooNE
450 [63], MINERvA [64], MicroBooNE [65] and SBND [66]. Additionally, thanks to their
451 near detectors, long baseline experiments can perform cross section measurements.
452 Some recent examples are NOvA [67] or T2K [68]. Future oscillation experiments
453 will greatly benefit from these measurements, as the measurement of the oscillation
454 parameters depends on the cross section modelling. However, there are alternative
455 data-driven approaches to extract the oscillation probabilities without relying on a
456 neutrino interaction model, which are planned to be explored in the next generation of
457 experiments [69, 70].

The Deep Underground Neutrino Experiment

461 *Deep in the human unconscious is a pervasive need for a logical universe that
462 makes sense. But the real universe is always one step beyond logic.*

463 – Frank Herbert, *Dune*

464 The Deep Underground Neutrino Experiment (DUNE) is a next generation long-
465 baseline neutrino experiment [71]. It will aim to address several questions in neutrino
466 physics, study neutrinos from astrophysical sources and search for beyond the standard
467 model physics.

468 This Chapter reviews the main goals of the DUNE experiment, the design of the far
469 detector modules and their data acquisition (DAQ) system, and the role that the near
470 detector plays in the physics program of DUNE.

471 3.1 Overview

472 The main physics goals of DUNE are:

- 473 • measure the neutrino mass hierarchy, the amount of CP violation in the leptonic
474 sector and the θ_{23} octant,
- 475 • detect rare low energy neutrino events, like neutrinos from supernova bursts, and
- 476 • search for proton decay and other beyond the standard model phenomena.

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

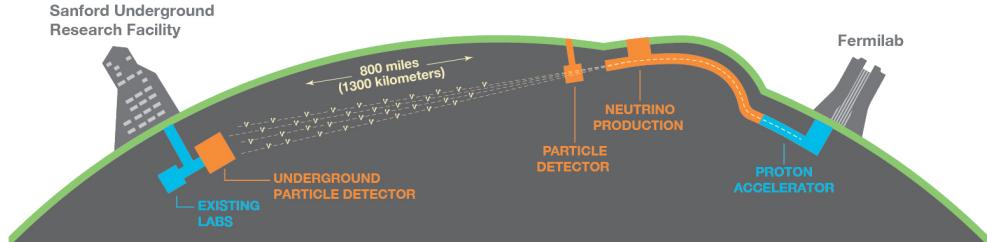


Figure 3.1: Schematic diagram of the DUNE experiment and the LBNF beamline [71].

The design of DUNE has been tailored with these goals in mind. It will consist of two neutrino detectors. A near detector (ND) complex will be placed at Fermilab, 574 m downstream of the neutrino production point, whereas a larger far detector (FD) will be built in the Sandford Underground Research Facility (SURF), South Dakota, approximately 1300 km away. Figure 3.1 shows a simplified view of the various components of DUNE (not to scale).

The beam neutrinos will be provided by the Long-Baseline Neutrino Facility (LBNF) beamline, the multi-megawatt wide-band neutrino beam planned for Fermilab. It will produce neutrinos travelling in the direction of SURF, with the capability to switch between neutrino and antineutrino mode.

Before arriving to the FD, the neutrino beam meets the ND complex, which serves as the experiment's control. The design of the DUNE ND is mainly driven by the needs of the oscillation physics program, as its main role is to measure the unoscillated neutrino energy spectra. From these we can predict the unoscillated spectra at the FD, which can be compared to the spectra measured at the FD to extract the oscillation parameters. Additionally, the ND has a physics programme of its own, including cross section measurements and BSM physics searches.

The technology chosen for the FD modules of DUNE is the liquid Argon time projection chamber (LArTPC). Its four modules will record neutrino interactions from the accelerator-produced beam arriving at predictable times. As it also aims at recording rare and low energy events, like supernovae and solar neutrinos, the FD requires trigger

3.2. PHYSICS GOALS OF DUNE

Table 3.1: Summary of the two-phased plan for DUNE. Adapted from Ref. [72].

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

498 schemes which can deal with both kinds of physics, and also maximum uptime.

499 DUNE is planned to be built using a staged approach consisting on two phases,
500 which are summarised in Tab. 3.1. Phase I consists of a FD with 50% of the total
501 fiducial mass, a reduced version of the ND complex and a 1.2 MW proton beam. It will
502 be sufficient to achieve some early physics goals, like the determination of the neutrino
503 mass ordering. For its Phase II, DUNE will feature the full four FD modules, a more
504 capable ND and a 2.4 MW proton beam. The physics milestones for the two phases are
505 given in Tab. 3.2, in a staging scenario which assumes that Phase II is completed after
506 6 years of operation.

507 A summary of the DUNE science program can be found in the DUNE FD Technical
508 Design Report (TDR) Volume I [71]. For a detailed discussion on the two-phased
509 approach the reader is referred to the DUNE Snowmass 2021 report [72].

510 3.2 Physics goals of DUNE

511 As noted in the literature (see for instance Ref. [1] for a review), the parameter space of
512 the neutrino oscillation phenomena within the three-flavour picture is quite constrained
513 by current experimental data. However, there are still crucial open questions, like the
514 mass ordering, the value of δ_{CP} or the θ_{23} octant. One of the main goals of DUNE is to
515 determine precisely the values of these parameters [73].

516 To address these questions DUNE can look to the subdominant oscillation channel
517 $\nu_\mu \rightarrow \nu_e$ ($\bar{\nu}_\mu \rightarrow \bar{\nu}_e$) and study the energy dependence of the ν_e ($\bar{\nu}_e$) appearance probability.
518 When we focus on the antineutrino channel $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ there is a change in the sign of δ_{CP} ,
519 thus introducing CP-violation. Moreover, due to the fact that there are no positrons in

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Table 3.2: Exposure and time required to achieve the different physics milestones of the two phases. The predictions assume a Phase II staging scenario where FD modules 3 and 4 are deployed in years 4 and 6 and both the beam and ND are upgraded after 6 years. Adapted from Ref. [72].

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

520 the composition of Earth, there is a sign difference for the matter effect contribution
 521 when looking to the antineutrino channel. This asymmetry is proportional to the baseline
 522 length L and is sensitive to the sign of Δm_{31}^2 , and thus to the neutrino mass ordering.

523 Another of the main physics goals of DUNE is the search for baryon-number violating
 524 processes. Specifically, it will try to answer the question of whether protons are stable
 525 or not. There is no symmetry argument that forbids protons from decaying, but its
 526 apparent stability seems to suggest that baryon number is conserved [74]. However,
 527 proton decay is a usual feature of grand-unified theories, where electromagnetic, weak
 528 and strong interactions are unified above a certain energy scale [75].

529 As the energy deposition scale for this kind of searches is nearly the same as the one
 530 for long-baseline neutrino oscillations, DUNE will be able to look for them. It has several
 531 advantages over other experiments, such as excellent imaging and particle identification,
 532 which can be translated to lower backgrounds.

533 The last of the main objectives of DUNE is the detection of neutrinos originated in
 534 supernovae explosions, what is called a supernova neutrino burst (SNB). These neutrinos
 535 carry with them information about the core-collapse process, from the progenitor to the
 536 explosion and the remnant; but also may have information about new exotic physics. So
 537 far, the only neutrino events ever recorded from such a process were a few dozens of $\bar{\nu}_e$

3.3. LBNF BEAMLINE

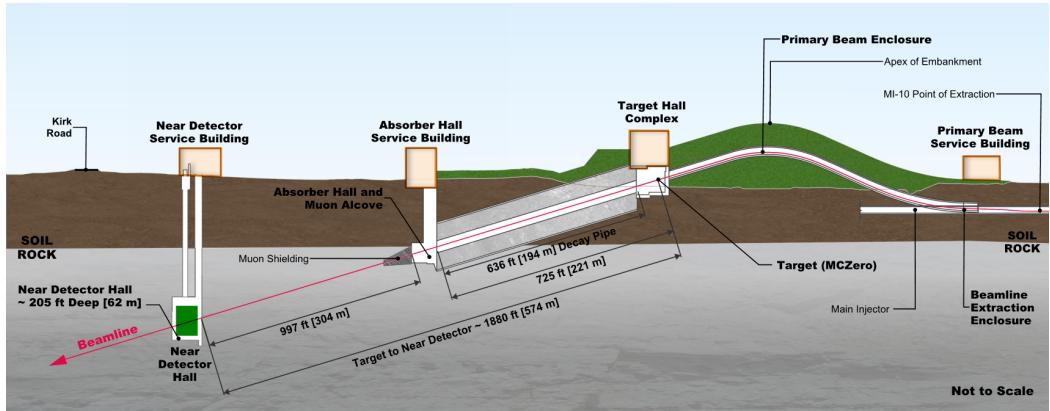


Figure 3.2: Schematic longitudinal section of the LBNF beamline at Fermilab (not to scale). Figure taken from Ref. [78].

538 events from the 1987A supernova located in the Magellanic Cloud, 50 kpc away from
 539 Earth [76,77].

540 DUNE aims to collect SNB events. Although these are quite rare, as the expected
 541 supernovae explosion events are about one every few decades for our galaxy and
 542 Andromeda, the long lifetime of the experiment (around a few decades as well) makes it
 543 reasonable to expect some. Nowadays the main sensitivity to SNB of most experiments
 544 is to the $\bar{\nu}_e$ through inverse beta decay. One of the advantages of DUNE is its expected
 545 sensitivity to ν_e , since the dominant channel will be ν_e CC scattering.

546 Moreover, due to the stringent requirements that the main physics goals set for
 547 DUNE, it will allow also to perform searches for all kind of BSM physics. Among
 548 others, DUNE will be able to look for: active-sterile neutrino mixing, non-unitarity of
 549 the PMNS matrix, non-standard interactions, Lorentz and CPT violations, neutrino
 550 trident production, light-mass DM, boosted DM and heavy neutral leptons. The reader
 551 is referred to the DUNE FD TDR Volume II [73] for a full discussion of the physics
 552 scope of DUNE.

553 3.3 LBNF beamline

554 The LBNF project is responsible for producing the neutrino beam for the DUNE detectors.
 555 A detailed discussion of the LBNF program can be found in the DUNE/LBNF CDR

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

556 Volume III [78].

557 A schematic diagram of the longitudinal section of the LBNF beamline is shown in
 558 Fig. 3.2. First, a beam of 60 – 120 GeV protons is extracted from the Fermilab Main
 559 Injector. This beam is aimed towards the target area, where it collides with a cylindrical
 560 graphite target to produce pions and kaons.

561 The diffuse, secondary beam of particles is focused by a pair of magnetic horns.
 562 These select the positively charged particles when operated in Forward Horn Current
 563 (FHC) mode, or the negatively charged ones when the current is reversed, also known as
 564 Reverse Horn Current (RHC) mode. The focused secondary beam then enters a 194 m
 565 decay pipe where the pions and kaons will predominantly produce $\mu^+\nu_\mu$ pairs when in
 566 FHC mode (or $\mu^-\bar{\nu}_\mu$ in RHC mode).

567 At the end of the decay pipe a hadron absorber removes the undecayed hadrons and
 568 muons from the beam, which reduces the ν_e ($\bar{\nu}_e$) and $\bar{\nu}_\mu$ (ν_μ) contamination coming
 569 from the μ^+ (μ^-) decays. The resulting neutrino flux at the FD is shown in Fig. 3.3,
 570 both for FHC (left) and RHC (right) modes. These predictions show the intrinsic $(\bar{\nu})_e$
 571 contamination and wrong sign component from wrong sign and neutral meson decays,
 572 as well as muons decaying before reaching the absorber.

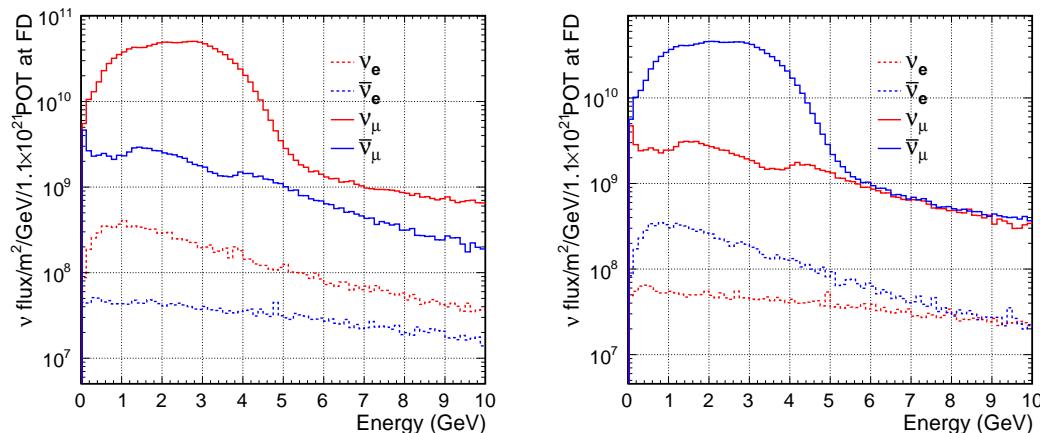


Figure 3.3: Predicted neutrino fluxes at the FD in FHC mode (left panel) and RHC mode (right panel). Figures taken from Ref. [73].

3.4. NEAR DETECTOR

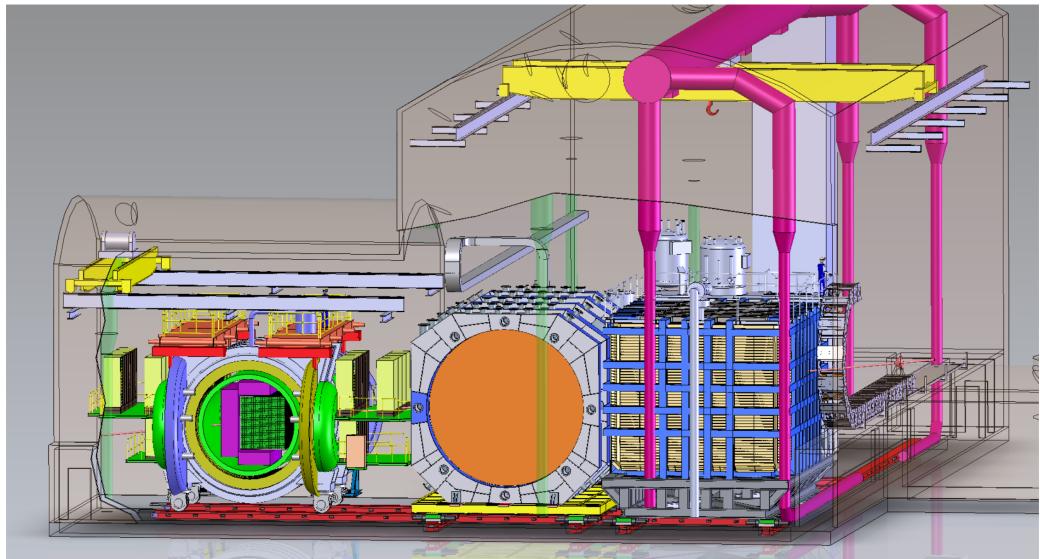


Figure 3.4: Representation of the ND hall in Phase II, showing the different subcomponents. From right to left, in the direction of the beam, we have ND-LAr, ND-GAr and SAND. Figure taken from Ref. [79].

573 3.4 Near Detector

574 To estimate the oscillation parameters we measure the neutrino energy spectra at the FD.
575 This reconstructed energy arises from a convolution of the neutrino flux, cross section,
576 detector response and the oscillation probability. Using theoretical and empirical models
577 to account for the other effects, one can extract the oscillation probability using the
578 measurement. However, these models have associated a number of uncertainties that
579 are then propagated to the oscillation parameters.

580 One of the main roles of the ND is to measure the neutrino interaction rates before
581 the oscillation effects become relevant, i.e. close to the production point. By measuring
582 the ν_μ and ν_e energy spectra, and that of their corresponding antineutrinos, at the ND
583 we can constrain the model uncertainties. A complete cancellation of the uncertainties
584 when taking the ratio between the FD and ND measurements is not possible, as that
585 would require both detectors to have identical designs and the neutrino fluxes to be
586 the same. Because of the distance, the flux probed by the FD will have a different
587 energy and flavour composition than that at the ND, as neutrinos oscillate and the beam

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

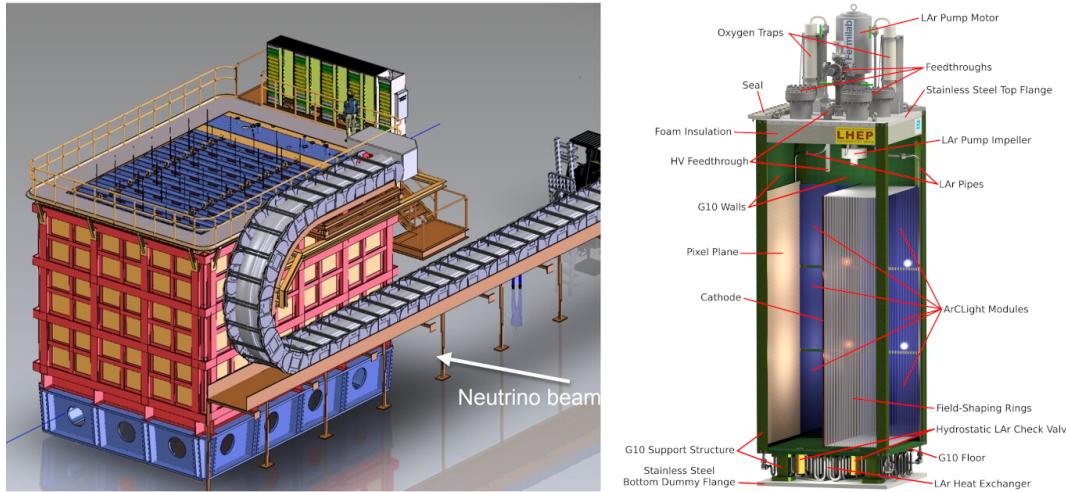


Figure 3.5: Schematic representation of the external components of ND-LAr, including the cryostat and the PRISM movable system (left) and detailed drawing of one ArgonCube module (right). Figure adapted from Ref. [71].

spreads. The differences in the flux also determine the design of the detectors, therefore the ND is limited in its capability to match the FD design.

Nevertheless, having a highly capable ND, DUNE can minimise the systematic uncertainties affecting the observed neutrino energy. The ND data can be used to tune the model parameters by comparison with the prediction. Then, one uses the tuned model to predict the unoscillated FD spectra. Comparing the prediction with the measured spectra it is possible to extract the oscillation parameters.

Additionally, the ND will have a physics program of its own. In particular, it will measure neutrino cross sections that will then be used to constrain the model used in the long-baseline oscillation analysis. It will also be used to search for BSM phenomena such as heavy neutral leptons, dark photons, millicharged particles, etc.

The DUNE ND can be divided in three main components, a LArTPC known as ND-LAr, a magnetised muon spectrometer, which will be the Temporary Muon Spectrometer (TMS) in Phase I and ND-GAr in Phase II, and the System for on-Axis Neutrino Detection (SAND). The layout of the Phase II DUNE ND can be seen in Fig. 3.4. The first two components of the ND will be able to move off-axis, in what is called the Precision Reaction-Independent Spectrum Measurement (PRISM) concept. More details

3.4. NEAR DETECTOR

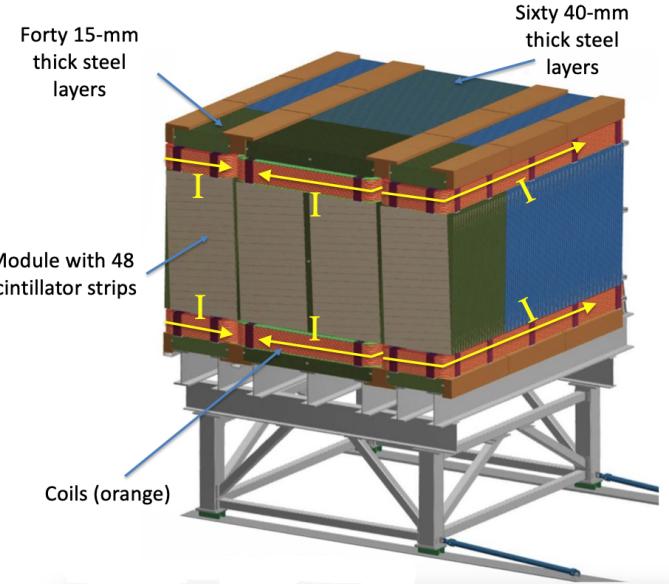


Figure 3.6: Schematic view of the TMS detector, highlighting its main parts. Figure adapted from Ref. [71].

605 on the purpose and design of the ND can be found in the DUNE ND Conceptual Design
 606 Report (CDR) [79].

607 **3.4.1 ND-LAr**

608 ND-LAr is a LArTPC, as the ND needs a LAr component to reduce cross section and
 609 detector systematic uncertainties in the oscillation analysis. However, its design differs
 610 significantly from those proposed for the FD modules. Because of the high event rates
 611 at the ND, approximately 55 neutrino interaction events per $10 \mu\text{s}$ spill, ND-LAr will be
 612 built in a modular way. Each of the modules, based on the ArgonCube technology, is a
 613 fully instrumented, optically isolated TPC with a pixelated readout. The pixelisation
 614 allows for a fully 3D reconstruction and the optical isolation reduces the problems due
 615 to overlapping interactions. Figure 3.5 shows a representation of the external parts of
 616 ND-LAr (left) and a detailed diagram of an ArgonCube module (right).

617 With a fiducial mass of 67 t and dimensions $7 \text{ m} (\text{w}) \times 3 \text{ m} (\text{h}) \times 5 \text{ m} (\text{l})$, ND-LAr
 618 will be able to provide high statistics and contain the hadronic systems from the beam
 619 neutrino interactions, but muons with a momentum higher than 0.7 GeV will exit the

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

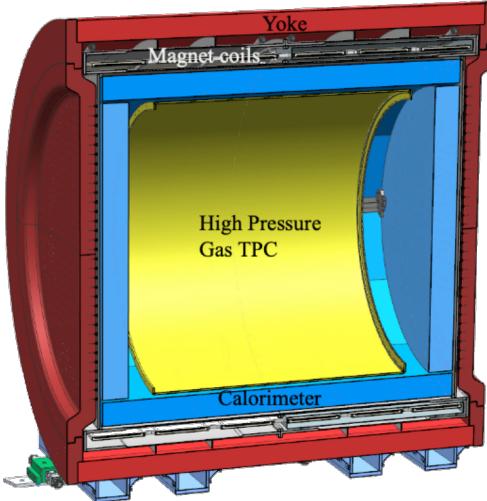


Figure 3.7: Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet. Figure adapted from Ref. [80].

620 detector.

621 3.4.2 TMS/ND-GAr

622 To accurately estimate the neutrino energy, the momentum of the outgoing muons needs
 623 to be determined. That is the reason why a muon spectrometer is needed downstream
 624 of ND-LAr.

625 In Phase I that role will be fulfilled by TMS. It is a magnetised sampling calorimeter,
 626 with alternating steel and plastic scintillator layers. Figure 3.6 shows a schematic view
 627 of the TMS detector. The magnetic field allows a precise measurement of the sign of the
 628 muon, so one can distinguish between neutrino and antineutrino interactions.

629 After the Phase II upgrade, TMS will be replaced with a more capable near detector.
 630 The current technology considered is ND-GAr. This detector is a magnetised, high-
 631 pressure GAr TPC (often denoted as HPgTPC) surrounded by an electromagnetic
 632 calorimeter (ECal) and a muon tagger. A cross section of its geometry can be seen
 633 in Fig. 3.7. ND-GAr will be able to measure the momenta of the outgoing muons
 634 while also detect neutrino interactions inside the GAr volume. This allows ND-GAr
 635 to constrain the systematic uncertainties even further, as it will be able to accurately

3.4. NEAR DETECTOR

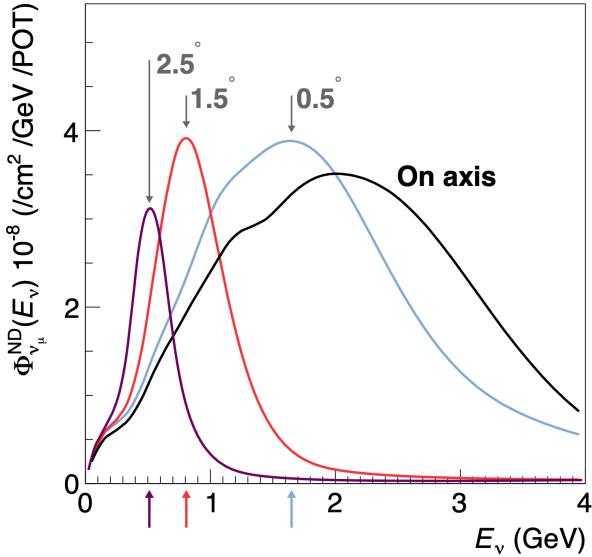


Figure 3.8: Predicted beam muon neutrino flux at the ND location for different off-axis positions. Figure taken from Ref. [79].

636 measure neutrino interactions at low energies thanks to the lower tracking thresholds of
 637 GAr.

638 3.4.3 PRISM

639 In general, the observed peak neutrino energy of a neutrino beam decreases as the
 640 observation angle with respect to the beam direction increases. This feature has been
 641 used in other long-baseline neutrino experiments, like T2K (2.5° off-axis) and NOvA
 642 (0.8° off-axis), to achieve narrower energy distributions. The DUNE PRISM concept
 643 exploits this effect using a movable ND. Within PRISM both ND-LAr and the muon
 644 spectrometer (TMS in Phase I and ND-GAr in Phase II) can be moved up to 3.2°
 645 off-axis, equivalent to move the detectors 30.5 m laterally through the ND hall.

646 This allows to record additional data samples with different energy compositions.
 647 Figure 3.8 compares the on-axis muon neutrino flux at the ND with the fluxes at different
 648 off-axis positions. As the off-axis position increases the neutrino flux becomes closer to
 649 a monoenergetic beam with a lower peak energy. These samples can be used to perform
 650 a data-driven determination of the relation between true and reconstructed neutrino

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

651 energy, to reduce the dependence on the interaction model. The off-axis samples are
652 linearly combined to produce a narrow Gaussian energy distribution centered on a target
653 true energy. From the combination coefficients one can build a sample of reconstructed
654 neutrino events that will determine the energy mapping.

655 The PRISM samples will be used to form a flux at the ND location similar in shape
656 to the oscillated flux measured by the FD. This method can be used to extract the
657 oscillation parameters with minimal input from the neutrino interaction model.

658 3.4.4 SAND

659 The role of SAND is to monitor the beam stability by measuring the on-axis neutrino
660 energy spectra. As the PRISM program requires that ND-LAr and its downstream
661 muon spectrometer spend about half of the time in off-axis positions, it is not possible
662 to monitor the stability with the movable detectors. Moreover, for the success of PRISM
663 it is essential to have a stable beam configuration, or, at least, a quick assessment and
664 modeling of the distortions.

665 The SAND detector is magnetised, and features an inner low density tracker, a LAr
666 target with optical readout and surrounding sampling calorimeter.

667 3.5 A More Capable Near Detector

668 In DUNE Phase II, a more capable near detector is needed to achieve the ultimate physics
669 goals of the experiments. The current leading proposal for this detector is ND-GAr.
670 As mentioned previously, it will fulfill the role of TMS, measuring the momentum and
671 sign of the charged particles exiting ND-LAr. Additionally, it will be able to measure
672 neutrino interactions inside the HPgTPC, achieving lower energy thresholds than those
673 of the ND and FD LArTPCs. By doing so ND-GAr will allow to constrain the relevant
674 systematic uncertainties for the LBL analysis even further. A detailed discussion on the
675 requirements, design, performance and physics of ND-GAr can be found in the DUNE
676 ND CDR [79] and the ND-GAr white paper [81].

3.5. A MORE CAPABLE NEAR DETECTOR

677 3.5.1 Requirements

678 The primary requirement for ND-GAr is to measure the momentum and charge of
679 muons from ν_μ and $\bar{\nu}_\mu$ CC interactions in ND-LAr, in order to measure their energy
680 spectrum. To achieve the sensitivity to the neutrino oscillation parameters described
681 in the DUNE FD TDR Volume II [73], ND-GAr should be able to constrain the muon
682 energy within a 1% uncertainty or better. The main constraint will come from the
683 calibration of the magnetic field, which will be performed using neutral kaon decays in
684 the HPgTPC.

685 Another requirement for ND-GAr is the precise measurement of neutrino interactions
686 on argon for the energies relevant to the neutrino oscillation program. The goal is to
687 constrain the cross section systematic uncertainties in the regions of phase space that
688 are not accessible to ND-LAr. This requires the kinematic acceptance for muons in
689 ND-GAr to exceed that of ND-LAr, being comparable to the one observed in the FD.

690 ND-GAr should also be able to help establishing the relationship between true and
691 reconstructed energy from neutrino interactions on argon with low thresholds, being
692 sensitive to particles that are not observed or may be misidentified in ND-LAr. In
693 particular, ND-GAr needs to have low tracking thresholds in order to measure the
694 spectrum of pions and protons produced in final-state interactions (FSI). It also must
695 be able to accurately measure the pion multiplicity in 1, 2 and 3 pions final states, to
696 inform the pion mass correction in the LArTPCs.

697 3.5.2 Reference design

698 The final design of ND-GAr is still under preparation. However, a preliminary baseline
699 design was in place at the time of the ND CDR. This section summarises the main
700 features of that design, as it is also the one used for the default geometry in our simulation.
701 A DUNE Phase II white paper, discussing the different options under consideration for
702 the ND-GAr design, is in progress.

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

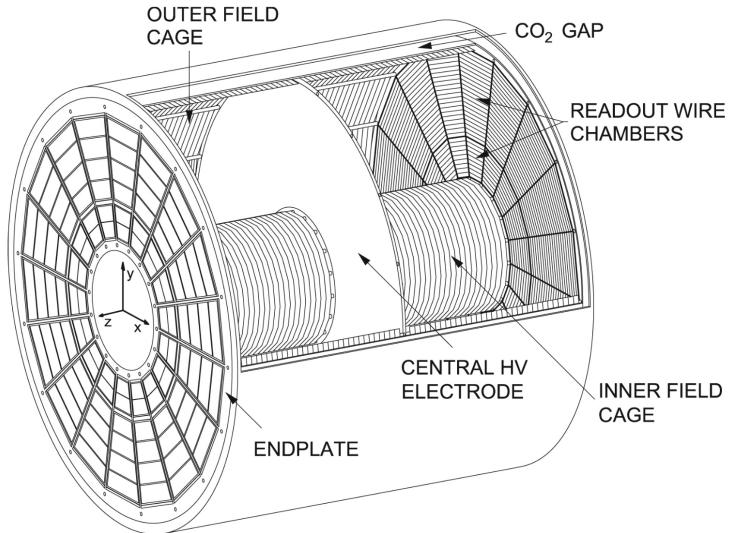


Figure 3.9: Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers. Figure taken from Ref. [79].

703 HPgTPC

704 The reference design for the ND-GAr HPgTPC follow closely that of the ALICE TPC.
 705 It is a cylinder with a central high-voltage cathode, generating the electric field for
 706 the two drift volumes, with a maximum drift distance of 2.5 m each. The anodes will
 707 be instrumented with charge readout chambers. The original design repurposed the
 708 multi-wire proportional readout chambers (MWPCs) of ALICE, however some of the
 709 current R&D efforts focus on a gas electron multiplier (GEM) [82] option instead. Figure
 710 3.9 shows a schematic diagram of the ALICE TPC design. The basic ND-GAr geometry
 711 will resemble this, except for the inner field cage.

712 It will use a 90:10 molar fraction Ar: CH_4 mixture at 10 bar. With this baseline gas
 713 mixture light collection is not possible, as the quenching gas absorbs most of the VUV
 714 photons. Additional R&D efforts are underway, to understand if different mixtures allow
 715 for the light signal to be used to provide a t_0 while maintaining stable charge gain.

3.5. A MORE CAPABLE NEAR DETECTOR

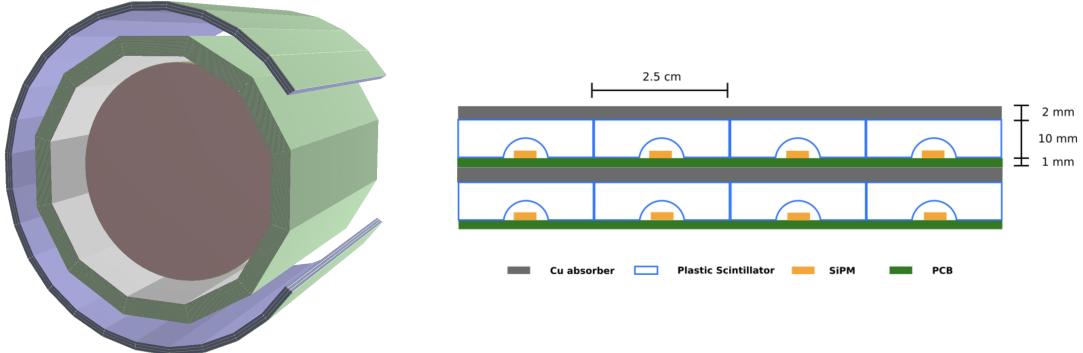


Figure 3.10: View of the 12-sided ECal barrel and outer muon tagger geometries (left) and layout of the ECal tile layers for the 2 mm Cu, 10 mm scintillator option (right). Figure adapted from Ref. [79].

716 ECal

717 The main role of the ND-GAr ECal is the calorimetric measurement of the electron
 718 energies and the reconstruction of photons, in particular those from neutral pion decays.
 719 Also, the ECal is able to provide a t_0 timestamp for neutrino interactions, by associating
 720 its activity to the tracks in the HPgTPC. The ECal will also be able to perform
 721 neutron reconstruction using time of flight and reject external backgrounds, thanks to
 722 its sub-nanosecond time resolution.

723 The ECal design features three independent subdetectors, two end caps at each side
 724 and a barrel surrounding the HPgTPC. Each of the detectors is divided in modules,
 725 which combine alternating layers of plastic scintillator and absorber material readout
 726 by SiPMs. The inner scintillator layers consist of $2.5 \times 2.5 \text{ cm}^2$ high-granularity tiles,
 727 whereas the outer ones are made out of 4 cm wide cross-strips spanning the whole
 728 module length. The current barrel geometry consists of 8 tile layers and 34 strip layers,
 729 while the end caps feature 6 and 36 respectively. The thickness of the scintillator layers
 730 is 7 mm and 5 mm for the Pb absorber layers. The 12-sided geometry of the ECal barrel
 731 (left) and the layout of the tile layers (left)¹ can be seen in Fig. 3.10.

¹The figure shows the layout of the tile layers for a previous design with 2 mm Cu absorber and 10 mm plastic scintillator, as mentioned in the text the current choice is 5 mm Pb absorber and 7 mm scintillator.

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

732 Magnet

733 The ND-GAr magnet design, know as the Solenoid with Partial Yoke (SPY), consists of
734 two coupled solenoids with an iron return yoke. The idea behind the design is to have a
735 solenoid as thin as possible, as well as a return yoke mass distribution that minimises
736 the material budget between ND-LAr and ND-GAr. The magnet needs to provide a
737 0.5 T field in the direction perpendicular to the beam, parallel to the drift electric field.
738 It needs to host the pressure vessel and the surrounding ECal, which points to a inner
739 diameter of ~ 6.4 m.

740 The solenoid is a single layer coil, based on niobium titanium superconducting
741 Rutherford cable. The total length of the coil is 7.5 m. The bobbin will be split in four
742 segments grouped in pairs with two identical cryostats, connected in series. The iron
743 yoke features an aperture in the upstream side to allow the muons coming from ND-LAr.
744 Still, its material will be enough to reduce the magnetic field reaching SAND, and also
745 stop the charged pions produced inside the HPgTPC.

746 Muon system

747 The design of the ND-GAr muon system is still in a preliminary stage. Its role is to
748 distinguish between muons and pions punching through the ECal. This is especially
749 important for wrong-sign determination, to separate these from neutral current events.

750 In its current form, the muon system consists of three layers of longitudinal sampling
751 structures. It alternates 10 cm Fe absorber slabs with 2 cm plastic scintillator strips.
752 The transverse granularity required is still under study.

753 3.5.3 R&D efforts

754 There are several ND-GAr-related prototypes, mostly focused on the TPC charge
755 readout and electronics. The priority is to test the full readout chain, in a high-pressure
756 environment, using a gas mixture with high argon fraction. A detailed summary of these
757 can be found in the DUNE Phase II white paper [80].

3.5. A MORE CAPABLE NEAR DETECTOR

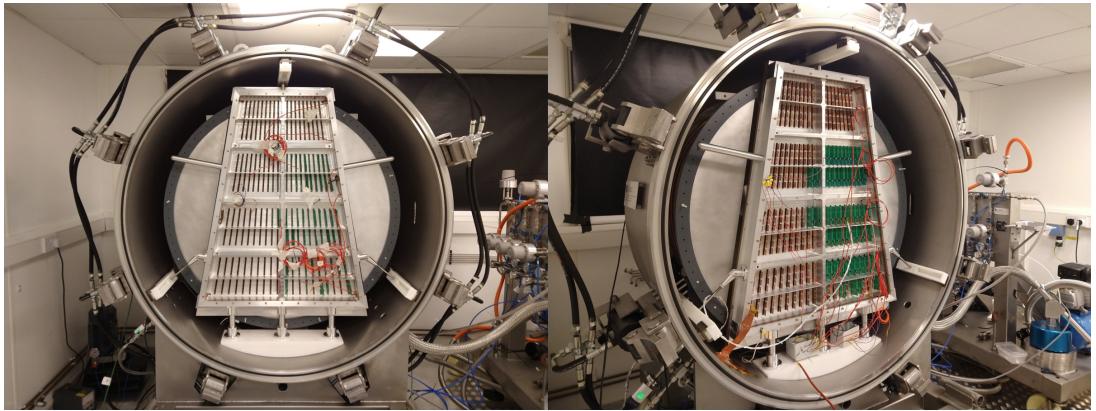


Figure 3.11: Photographs of the TOAD pressure vessel at RHUL. The TPC is mounted inside the vessel, and the OROC is supported by an aluminium frame. Figure taken from Ref. [83].

758 Multi-Wire Proportional Chambers

759 As mentioned before, the original ND-GAr design repurposes the MWPCs of the ALICE
760 TPC, which became available after the recent upgrade [84]. These were operated using
761 a 90:10:5 Ne:CO₂:N₂ gas mixture at 1 atm. Therefore, their performance needed to be
762 studied in an argon gas environment at high pressure.

763 The Gas-argon Operation of ALICE TPC (GOAT) test stand tested the ALICE
764 readout chambers at high pressure. In particular, it used one of the previously operated
765 ALICE inner MWPCs, also known as IROCs, in a pressure vessel rated to 10 atm. It
766 measured the gas gain at various pressure points, voltages and gas mixtures.

767 The Test stand of an Overpressure Argon Detector (TOAD) tested an ALICE outer
768 MWPC, also known as OROC, up to 5 atm. During its time at RHUL, it was used to
769 study the achievable gas gain of the OROC [83]. At the moment, it is being commissioned
770 at Fermilab for a full detector test of the readout electronics and the DAQ.

771 Figure 3.11 shows the interior of the TOAD pressure vessel. The TPC is mounted
772 inside the vessel on three rails. The back of the OROC, supported by an aluminium
773 frame, can be seen at the front.

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

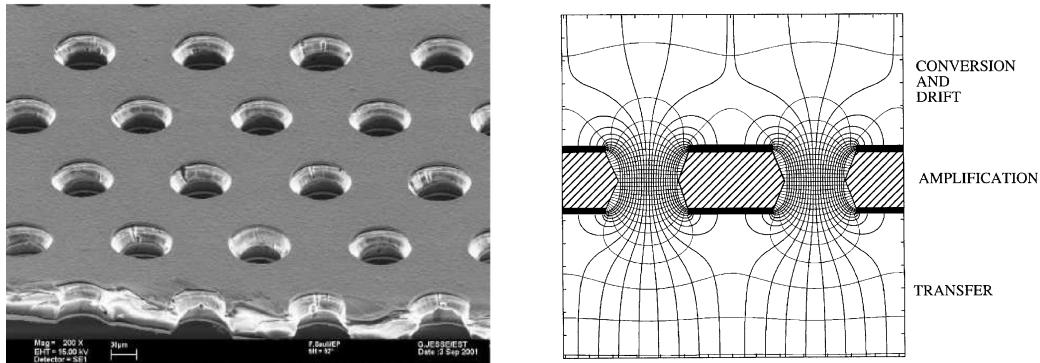


Figure 3.12: Left panel: electron microscope image of a $50\ \mu\text{m}$ thick GEM electrode, with hole pitch and diameter of 140 and $70\ \mu\text{m}$, respectively. Right panel: Schematics of a GEM electrode cross section, showing the electric field lines around the holes. Figures taken from Ref. [85].

774 Gas Electron Multiplier

775 An alternative to the MWPC option is the use of GEMs. These are a type of micro-patter
 776 detector, where the ionisation electrons passing through the holes in the GEM layers
 777 are accelerated by a high intensity electric field. The acceleration causes the electrons
 778 to ionise the medium, resulting in an avalanche which increase the signal exponentially
 779 [82]. GEMs are used in numerous experiments that need a high spatial resolution, like
 780 ALICE [86] and CMS [87] after their upgrades.

781 Figure 3.12 (left panel) shows an electron microscope picture of a $50\ \mu\text{m}$ thick GEM
 782 electrode, with a pitch between neighbouring holes of $140\ \mu\text{m}$ and a hole diameter of
 783 $70\ \mu\text{m}$. A schematic representation of the cross section of a GEM layer is shown in Fig.
 784 3.12 (left panel).

785 The Gaseous Argon T0 (GAT0²) prototype studies the use of thick GEMs made out
 786 of glass to achieve optical imaging of the primary ionisation. Using a 10 atm pressure
 787 vessel, the goal is to study different argon-based mixtures that allow for a precise t_0
 788 determination.

789 The GEM Over-pressurized with Reference Gases (GORG³) test stand is currently
 790 testing a GEM-based charge readout, using a triple-GEM stack.

²Spanish for cat.

³Persian for wolf.

3.6. FAR DETECTOR

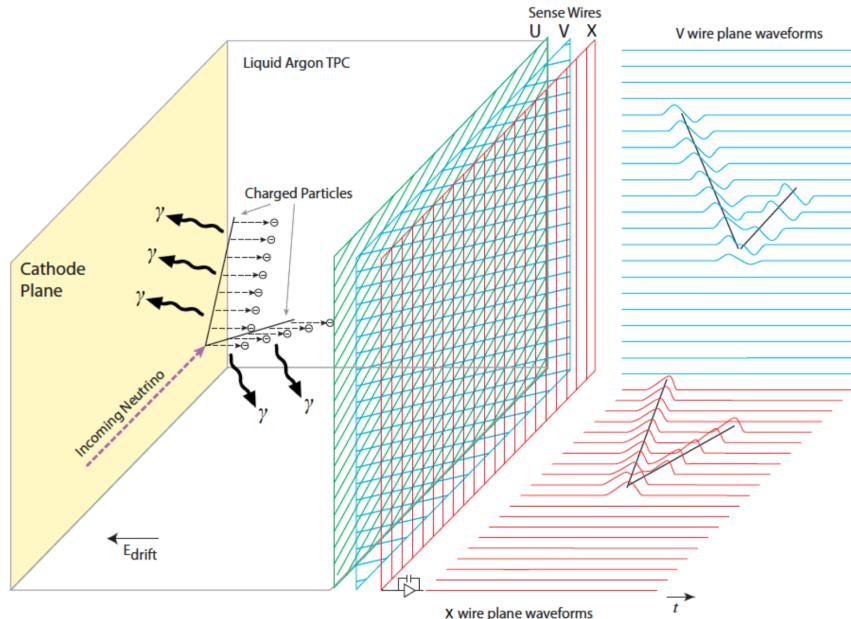


Figure 3.13: Schematic diagram showing the operating principle of a LArTPC with wire readout. Figure taken from Ref. [71].

791 3.6 Far Detector

792 The DUNE FD complex will sit 1300 km away from the beam target and 1.5 km
 793 underground at SURF, South Dakota. Two caverns will host the four FD modules, two
 794 of them per cavern, each embedded in cryostats of dimensions 18.9 m (w) \times 17.8 m (h) \times
 795 65.8 m (l). A central, smaller cavern will host the cryogenic system.

796 Three out of the four modules will be liquid argon (LAr) time projection chamber
 797 detectors, often refer to as LArTPCs, with a LAr fiducial mass of at least 10 kt each.
 798 The first and second FD modules, FD-1 and FD-2, will use a Horizontal Drift (HD)
 799 technology, whereas the third module, FD-3, will have a Vertical Drift (VD) direction.
 800 The technology for the fourth module is still to be decided,

801 For each event, with energies ranging from a few MeV to several GeV, these detectors
 802 collect both the scintillation light and the ionisation electrons created when the charged
 803 particles produced in neutrino-nucleus interactions ionise the argon nuclei. In both HD
 804 and VD designs the characteristic 128 nm scintillation light of argon is collected by a
 805 photon detection system (PDS). This light will indicate the time at which electrons

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

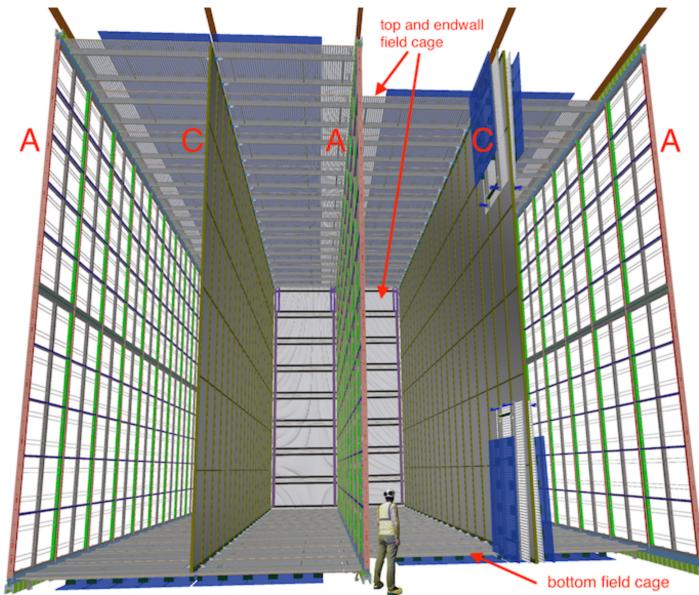


Figure 3.14: Proposed design for the FD-1 and FD-2 modules following the HD principle. Figure taken from Ref. [71].

806 start to drift, thus enabling reconstruction over the drift coordinate when compared
 807 to the time when the first ionisation electron arrives to the anode. Reconstruction of
 808 the topology in the transverse direction is achieved using the charge readout. Fig. 3.13
 809 illustrates the detection principle described, for the case of a HD detector with a wire
 810 readout.

811 3.6.1 Horizontal Drift

812 The HD design the ionisation electrons produced as charged particles traverse the LAr
 813 drift horizontally towards the anode planes, due to the effect of an electric field. These
 814 anode planes are made out of three layers of wire readout. This design, previously
 815 known as single-phase (SP), was tested by the ProtoDUNE-SP detector at CERN. The
 816 prototype collected data from a hadron beam and cosmic rays, providing high-quality
 817 data sets for calibration and performance studies.

818 Each FD HD detector module is divided in four drift regions, with a maximum drift
 819 length of 3.5 m, by alternating anode and cathode walls. The surrounding field cage
 820 ensures the uniformity of the 500 V/cm horizontal electric field across the drift volumes.

3.6. FAR DETECTOR

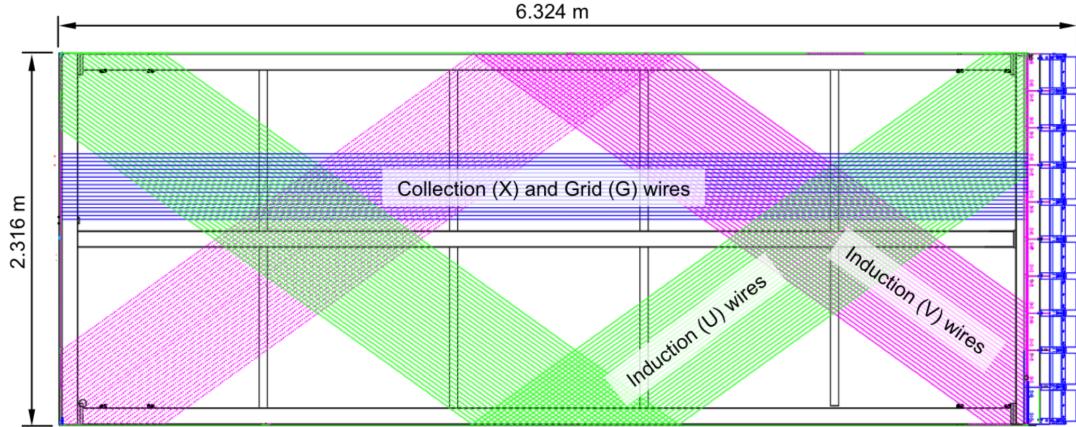


Figure 3.15: Schematic representation of an APA. The black lines represent the APA steel frame. The green and magenta lines correspond to the direction of the U and V induction wires respectively. The blue lines indicate the direction of the X collection wires and the wire shielding G. Figure taken from Ref. [71].

821 The three anode walls, which constitute the charge readout of the detector, are built by
 822 stacking anode plane assemblies (APAs), 2 high times 25 wide. The design of the HD
 823 modules is shown in Fig. 3.14.

824 Each APA is made of 2560 active wires arranged in three layers, plus an extra grid
 825 layer, wrapped around a metal frame. The two induction wire planes, U and V, sit at
 826 $\pm 35.7^\circ$ to the vertical on each side of the APA. The collection and shielding plane wires,
 827 X and G, run parallel to the vertical direction. The ionisation electrons drift past the
 828 induction planes, generating bipolar signals on those wires, and are collected by the
 829 collection plane, producing a monopolar positive signal. The spacing between the wires
 830 is ~ 5 mm, and it defines the spatial resolution of the APA.

831 The front-end readout electronics, or cold electronics as they are immerse in the LAr,
 832 are attached to the top of the up APAs and the bottom of the down APAs. Mounted on
 833 the front-end mother boards we have a series of ASICs that digitize the signals from the
 834 collection and induction planes. Each wire signal goes to a charge-sensitive amplifier,
 835 then there is a pulse-shaping circuit and this is followed by the analogue-to-digital
 836 converter. This part of the process happens inside the LAr to minimise the number of
 837 cables penetrating the cryostat. The digitised signals come out finally via a series of

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

838 high-speed serial links to the warm interface boards (WIBs), from where the data is sent
839 to the back-end DAQ through optical fibers.

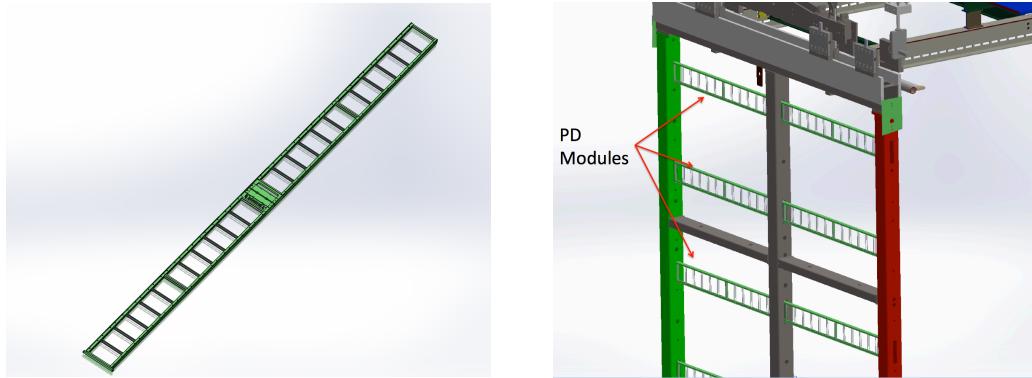


Figure 3.16: A PDS module containing 24 X-ARAPUCAs (left) and the location of the modules on the APAs (right). Figure taken from Ref. [71].

840 The PDS uses modules of X-ARAPUCA devices, mounted on the APA frames
841 between the wire planes. Each X-ARAPUCA consists of layers of dichroic filter and
842 wavelength-shifter. They shift the VUV scintillation light into the visible spectrum,
843 sending then the visible photons to silicon photomultiplier (SiPM) devices. The PDS
844 modules are 209 cm × 12 cm × 2 cm bars, containing 24 X-ARAPUCAs. There are 10
845 of these PDS modules per APA. Fig. 3.16 shows a PDS module (left) and the placement
846 of the modules on the APAs (right).

847 3.6.2 Vertical Drift

848 In the VD case the ionisation electrons will drift vertically until they meet a printed
849 circuit board-based (PCB) readout plane. It is based on the original dual-phase (DP)
850 design deployed at CERN, known as ProtoDUNE-DP, used a vertical drift design with
851 an additional amplification of the ionization electrons using a gaseous argon (GAr) layer
852 above the liquid phase. The VD module incorporates the positive features of the DP
853 design without the complications of having the LAr-GAr interface.

854 The current design of the FD VD module counts with two drift chambers with a
855 maximum drift distance of 6.5 cm. A cathode plane splits the detector volume along the
856 drift direction while the two anode planes are connected to the bottom and top walls

3.6. FAR DETECTOR

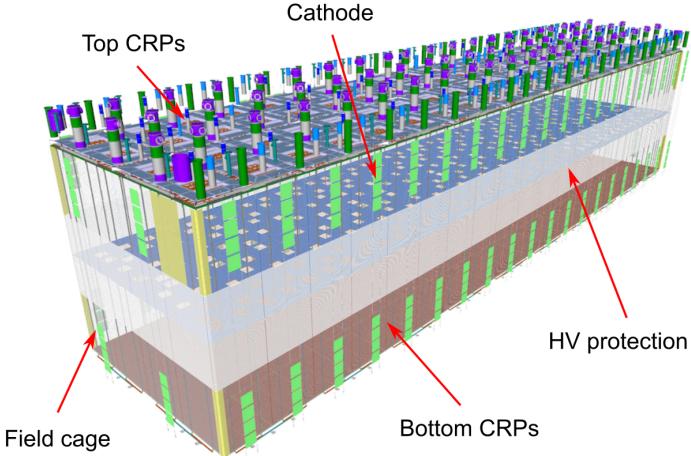


Figure 3.17: Proposed design for the FD-3 module following the VD principle. Figure adapted from Ref. [88].

857 of the detector. The layout of the VD module is shown in Fig. 3.17. Compared with
 858 the HD design, the VD option offers a slightly larger instrumented volume and a more
 859 cost-effective solution for the charge readout.

860 As in the HD design, each drift volume features a 500 V/cm electric field and a
 861 field cage that ensures its uniformity. The anode planes are arrays of 3.4 m × 3 m
 862 charge-readout planes (CRPs). These are formed by a pair of charge-readout units
 863 (CRUs), which are built from two double-sided perforated PCBs, with their perforations
 864 aligned. The perforations allow the drift electrons to pass between the layers.

865 The PCB face opposite to the cathode has a copper guard plane which acts as
 866 shielding, while its reverse face is etched with electrode strips forming the first induction
 867 plane. The outer PCB has electrode strips on both faces, the ones facing the inner PCB
 868 form the second induction plane while the outermost ones form the collection plane. Fig.
 869 3.18 shows the layout of the electrode strips for the top (left) and bottom (right) CRUs.
 870 The magenta and blue lines represent the first and second induction planes respectively,
 871 and the green lines correspond to the collection plane.

872 The PDS in the VD module will use the same X-ARAPUCA technology developed
 873 for the HD design. The plan is to place the PDS modules on the cryostat walls and on

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

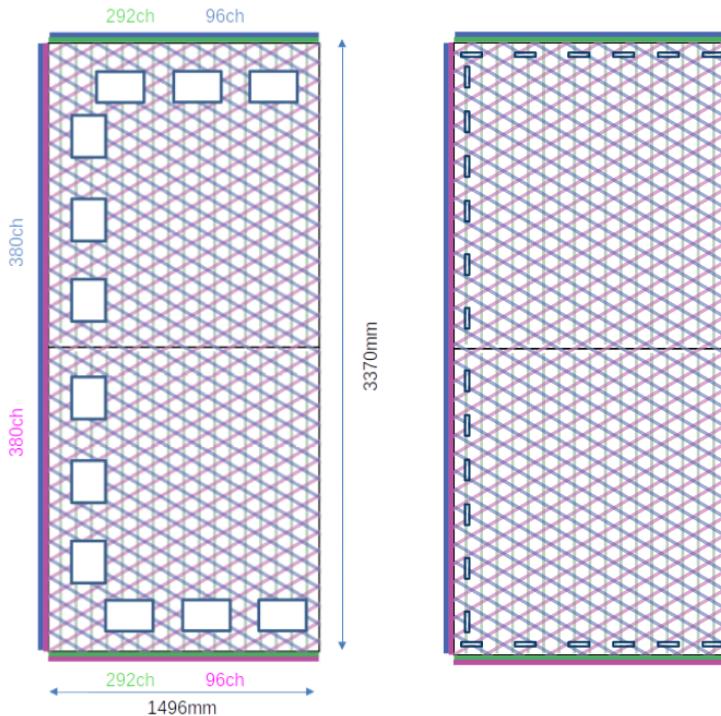


Figure 3.18: Schematic representation of the electrode strip configuration for a top (left) and bottom (right) CRU. Figure taken from Ref. [88].

874 the cathode, in order to maximise the photon yield.

875 **3.6.3 FD Data Acquisition System**

876 The data acquisition (DAQ) system receives, processes and stores data from the detector
877 modules. In the case of DUNE the DAQ architecture is designed to work for all FD
878 modules interchangeably, except some aspects of the upstream part which may depend
879 on the specific module technology.

880 The enormous sample rate and the number of channels in TPC and PD readouts
881 will produce a very large volume of data. These pose really strong requirements and
882 challenges to the DUNE FD DAQ architecture. It will be required to read out data of
883 the order of ten thousand or more channels at rates of a few MHz. To cope with the
884 huge data volume, segmented readouts and compression algorithms are used to reduce
885 the data rate to manageable levels.

886 The DAQ system of the DUNE FD is composed of five different subsystems. The

3.6. FAR DETECTOR

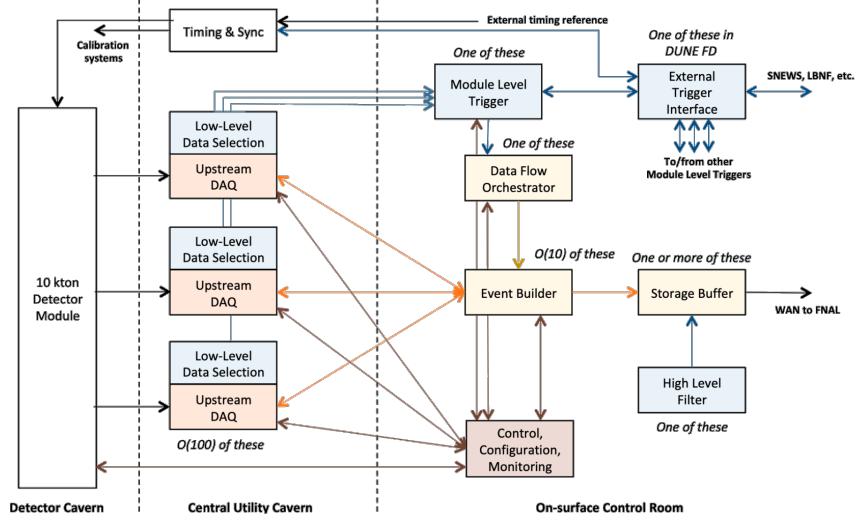


Figure 3.19: Detailed diagram of the DUNE FD DAQ system. Figure taken from Ref. [89].

887 first one is the upstream DAQ, which receives the raw data from the detector, buffers it
 888 and perform some low-level pre-processing. The minimally processed data is then fed
 889 into a hierarchical data selection system, which then performs a module level trigger
 890 decision. In case of a positive decision a trigger command is produced and executed by
 891 the data flow orchestrator, located in the back-end (BE) DAQ subsystem. Subsequently
 892 the DAQ BE retrieves the relevant data from the buffers located in the upstream DAQ,
 893 adds all the data into a cohesive record and saves it to permanent storage. Watching
 894 over all the other subsystems we also have the control, configuration and monitoring
 895 subsystem and the time and synchronization subsystem. Figure 3.19 shows a schematic
 896 diagram of the DAQ system, showing the different subsystems and their relations.

897 A notorious challenge for the DUNE DAQ system comes from its broad physics
 898 goals. We must be prepared to process events spanning a wide range of time windows
 899 (from 5 ms in the case of beam and cosmic neutrinos and nucleon decay to 100 s in the
 900 case of SNBs) and therefore this requires a continuous readout of the detector modules.
 901 Moreover, because of the off-beam measurements we need to ensure the capabilities
 902 of online data processing and self-triggering. Having this into account, together with
 903 the technical constraints, the DUNE FD DAQ faces a series of challenges: it needs to

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

- 904 be fault tolerant and redundant to reduce downtime, accommodate new components
- 905 while it keeps serving the operational modules, have large upstream buffers to handle
- 906 SNB physics, be able to support a wide range of readout windows and last reduce the
- 907 throughput of data to permanent storage to be at most 30 PB/year.

Matched Filter approach to Trigger

Primitives

911 *It is a capital mistake to theorize before one has data. Insensibly one begins to
912 twist facts to suit theories, instead of theories to suit facts.*

913 – Arthur Conan Doyle, *A scandal in Bohemia*

914 The DAQ system is responsible of what data will be collected in the DUNE FD.
915 Therefore, it has the capability of either expanding or limiting our physics reach,
916 depending of its specifications. This is important for the low energy physics program, as
917 more it requires more sensitive and reliable methods to pick the relevant signals.

918 In this Chapter, I present of a possible method to improve the sensitivity of the
919 DUNE FD by enhancing the production of hits in the online processing. This is possible
920 thanks to a more efficient filter strategy, the matched filter, which benefits the induction
921 channels of the detector.

922 4.1 Motivation

923 The lowest-level objects that are formed within the DUNE FD DAQ system are the
924 so-called trigger primitives (TPs). This represent the hits on a channel, and are used as
925 input to the rest of the DAQ trigger chain. The TPs are formed in the hit finder chain.
926 A schematic representation of it is shown in Fig. 4.1. This chain takes the raw ADC
927 data from the detector, removes the constant pedestal of the signal using a dynamical

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES



Figure 4.1: Schematic representation of the Trigger Primitive Generation chain in the DUNE FD.

median estimation method, applies a filter to the signal, and tries to find peaks over a certain threshold. These peaks form the TPs, which contain information such as the start and end time over the threshold, the maximum ADC value and the corresponding ADC integral. Currently, there are two implementations of the hit finder chain, one firmware-based and other software-based.

The filter implemented in the firmware of the upstream DUNE FD DAQ is a 32nd-order low-pass finite impulse-response (FIR) filter. The output of such filter for a discrete system can be written as:

$$y[i] = \sum_{j=0}^N h[j]x[i-j], \quad (4.1)$$

where N is the order of the filter, y is the output sequence, x is the input sequence and h is the set of coefficients of the filter. The current implementation within `dtp-firmware` [90] uses a set of 16 non-zero integer coefficients. For the software case, only a 5th-order filter is used, as the filtering turns out to be a very CPU expensive task.

Filtering is a vital step in the hit finder chain. It helps to suppress the noise and enhance the signal peaks with respect to the noiseless baseline. A good filtering strategy allows us to use lower thresholds when forming the trigger primitives (TPs) and thus increasing the sensitivity of our detector to low energy physics events. In such events, the hits produced by the ionisation electrons tend to have lower amplitudes than those of interest to the LBL physics programme of the DUNE experiment.

This is particularly important for the induction planes. In general, signal peaks in the induction wires have smaller amplitude than the ones in the induction plane. This, together with the fact that the pulse shapes are bipolar, reduces our capacity to

4.2. SIGNAL-TO-NOISE RATIO DEFINITION

949 detect the hits on these channels. The inefficiency of detecting TPs in the induction
950 planes (denoted as U and V planes) lead trigger algorithms to focus mainly on the TPs
951 from the collection plane (so-called X plane). As a result, the possibility of making
952 trigger decisions based on the coincidence of TPs across the three wire planes remains
953 nowadays unexploited in DUNE. This will be beneficial for low energy events, as it
954 adds redundancy to the algorithms, as well as for other physics that requires online
955 directionality information, like the supernova pointing.

956 A possible improvement of the current hit finder chain could require optimising
957 the existing or choosing a new filter implementation. A filter strategy which improves
958 the induction signals may be able to enhance the detection efficiency of TPs from the
959 induction planes and ideally make it comparable to that of the collection plane.

960 The goal is to implement a better finite-impulse response filter design and to evaluate
961 its performance relative to the current filter. To do so, we need to take into account the
962 limitations of the firmware: the FIR filter shall have maximum 32 coefficients (so-called
963 taps) whose values are 12-bit unsigned integers. Although it is technically possible to
964 include non-integer coefficients, it would be a technical challenge as we have 40 FIR
965 instances per APA, as there are 4 FIR per optical link and 10 optical links per APA. With
966 these restrictions, the task is to provide a set of 32 coefficients which yield an optimal
967 filter performance for the induction wires. A solution compatible with the software hit
968 finder implementation is not considered, due to its current limitations concerning the
969 filtering stage.

970 4.2 Signal-to-noise ratio definition

971 I introduce the signal to noise ratio (S/N) as a measure of the FIR filter performance
972 and demonstrate how to extract its value for a set of ProtoDUNE-SP data. The S/N
973 metrics allow us to compare different filter implementations and serve as a basis for more
974 detailed studies presented later in this document. Specifically, I use the ADC capture
975 `felix-2020-07-17-21:31:44` (data capture taken for firmware validation purposes). I

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

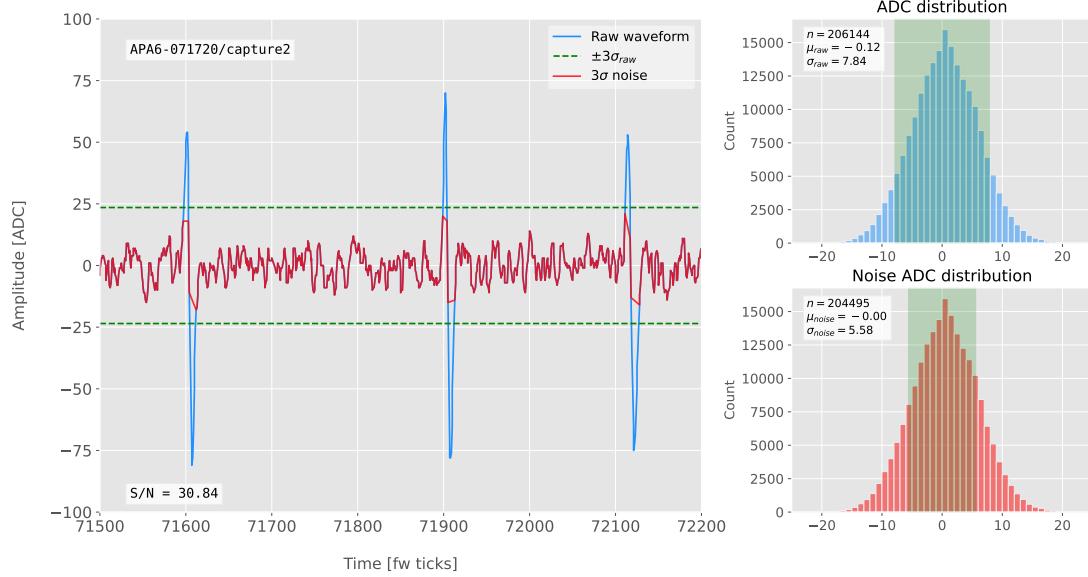


Figure 4.2: Left panel: Zoomed unfiltered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$. The resulting noise waveform is also shown (red line). Top right panel: ADC distribution for channel 7840, where the green shaded region represents $\pm \sigma_{\text{raw}}$. Bottom right panel: noise ADC distribution for channel 7840, where the green shaded region represents $\pm \sigma_{\text{noise}}$.

defined S/N as the height of the signal peaks relative to the size of the noise peaks. To quantify this quantity channel by channel one first need to estimate the standard deviation of the ADC data for each channel, σ_{ADC} . Then, I define the corresponding noise waveform to be the ADC values in the range $\pm 3\sigma_{\text{ADC}}$. From this new noise data one can estimate again the mean and standard deviation, μ_{noise} and σ_{noise} , so I can write the S/N for any given channel as:

$$\text{S/N} = \frac{\max [ADC] - \mu_{\text{noise}}}{\sigma_{\text{noise}}}, \quad (4.2)$$

where $\max [ADC]$ is simply the maximum ADC value found in the corresponding channel.

One can apply this definition of the S/N with a waveform from one of the channels

of the data capture¹. Fig. 4.2 shows a zoomed region of the waveform corresponding to

¹All the original work was done within the `dtp-simulation` package [91], which offers a variety of tools to read raw data and emulate the TPG block (pedestal subtraction, filtering and hit finder). However, the results shown in this report were re-worked later using the C++ based `dtpemulator` package [92]. Its main purpose is the emulation of the TPG block and, in the same way as its predecessor,

4.3. LOW-PASS FIR FILTER DESIGN

channel 7840 (blue line), where one can clearly see three signal peaks and continuous additive noise (we actually see 6 peaks, 3 positive and 3 negative, but, because by design for induction channels the expected signal pulse shapes are bipolar, I treat them as a collection of 3 individual signal peaks). I estimated the standard deviation of this raw waveform to be $\sigma_{raw} = 7.84$ ADC, so I am able to define the noise waveform (red line) as the ADC values in the range ± 23.52 ADC. This way one obtains $\mu_{noise} = 0$ and $\sigma_{noise} = 5.58$ ADC, which gives S/N = 30.84.

We can repeat this calculation now for the corresponding filtered waveform (using the current firmware FIR filter). In Fig. 4.3 I plotted the same time window for the filtered waveform from channel 7840 (blue line). In this case, the standard deviation of the waveform is larger than before, giving $\sigma_{raw} = 10.99$ ADC. The resulting noise waveform (red line) results from selection the ADC values in the range ± 32.91 ADC, giving now $\mu_{noise} = -0.47$ ADC and $\sigma_{noise} = 7.03$ ADC. Finally, one obtains S/N = 24.68. Notice that the value of S/N decreases after the filtering. Clearly, one can see that the noise baseline has increased by a factor of 1.35 when we applied the FIR filter and at the same time the amplitude of the signal peaks has remained almost unchanged, leading to this poorer S/N value.

4.3 Low-pass FIR filter design

In general, when one uses a method to optimize the frequency response of a digital filter, such as the Parks-McClellan algorithm, one finds a set of N real coefficients that give the best response for the specified pass-band and order of the filter [93].

In our case, as the sampling frequency is defined as 1 ticks^{-1} , the Nyquist frequency will simply be $1/2 \text{ ticks}^{-1}$. The current implementation of the filter seems to have as pass-band the range $[0, 0.1] \text{ ticks}^{-1}$. This can be seen in Fig. 4.4, where I show the power spectrum, in decibels, of such filter implementation (blue solid line). For instance, the Park-McClellan algorithm finds the optimal Chebyshev FIR filter taking as input

it has been cross-checked against the current firmware implementation.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

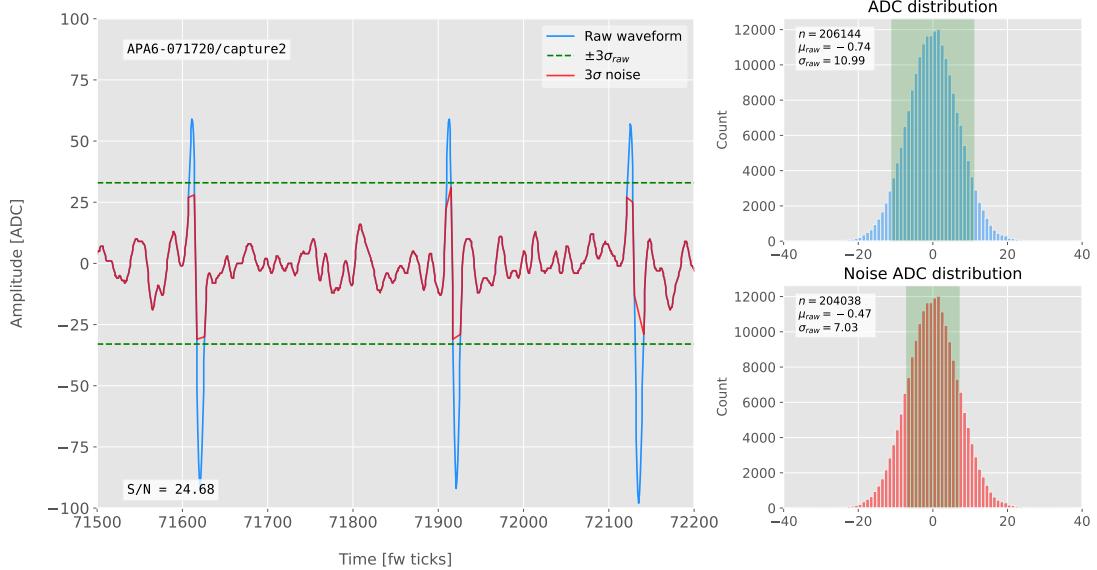


Figure 4.3: Left panel: Zoomed filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The filter used was the current implementation of the low-pass FIR filter in `dtp-firmware`. The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$. The resulting noise waveform is also shown (red line). Top right panel: ADC distribution for channel 7840 after filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$. Bottom right panel: noise ADC distribution for channel 7840 after filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$

1011 the boundaries of the target pass-band and stop-band, which can be written in the form:

$$\left\{ \begin{array}{l} [0, f_c] \\ [f_c + \delta f, f_N] \end{array} \right. , \quad (4.3)$$

1012 where f_c is the cut-off frequency, δf is the transition width and f_N is the aforementioned
 1013 Nyquist frequency. A similar behaviour to the one in the current filter can be obtained
 1014 by setting $f_c = 0$ and $\delta f = 0.1 \text{ ticks}^{-1}$. The response of the resulting filter is also shown
 1015 in Fig. 4.4 (blue solid line). Notice that the suppression of the stop-band is enhanced for
 1016 this optimal filter. For comparison I included the power response of the filter obtained
 1017 by taking the integer part of the coefficients resulting from the Parks-McClellan method
 1018 (red dashed line). One can see that it does not suppress that much the stop-band, in a
 1019 similar way to the current implementation of the filter.

1020 At this point, I tried to improve the performance of the FIR filter using the Park-

4.3. LOW-PASS FIR FILTER DESIGN

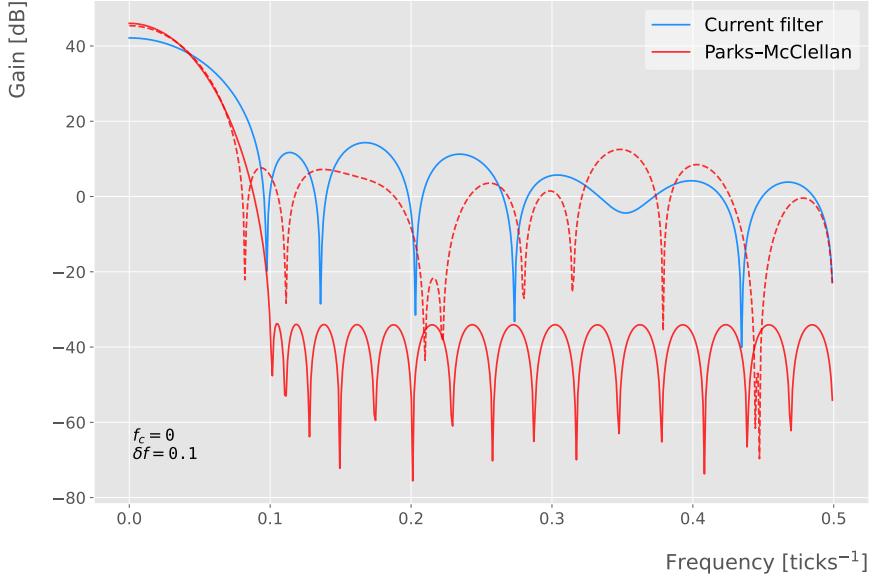


Figure 4.4: Power spectrum in decibels for the current implementation of the low-pass FIR filter in dtp-firmware (blue line), compared to the response of an optimal filter obtained using the Parks-McClellan algorithm for the same pass-band (red line). Also for comparison I include the spectrum of the optimal filter when taking only the integer part of the coefficients (red dashed line).

1021 McClellan method, i.e. maximize the overall S/N, using the available data captures. I
 1022 did so by varying the values of the two quantities that parametrize the pass-band and
 1023 stop-band, the cut-off frequency f_c and the transition width δf .

1024 Fig. 4.5 shows the average relative change in the S/N (i.e. the ratio between the
 1025 value of the S/N after and before the filtering) for capture **felix-2020-07-17-21:31:44**,
 1026 when using filters designed with the Parks-McClellan algorithm for the specified values
 1027 of the cut-off frequency f_c and the transition width δf , restricted to integer values for
 1028 the filter coefficients. One can clearly distinguish different regions where we get an
 1029 improvement of up to a factor of 1.35 for the U plane. For large values of $f_c + \delta f$ the
 1030 ratio tends to 1, as expected (in that limit the width of the stop-band goes to 0, meaning
 1031 that no frequencies are filtered out and thus the waveform remains the same).

1032 Using the configuration which gives the best mean performance for the three
 1033 planes (see bottom right panel of Fig. 4.5), i.e. $f_c = 0.068$ ticks⁻¹ and $\delta f =$
 1034 0.010 ticks⁻¹, we can see how such filter affects the different channels. Fig. 4.6 shows the

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

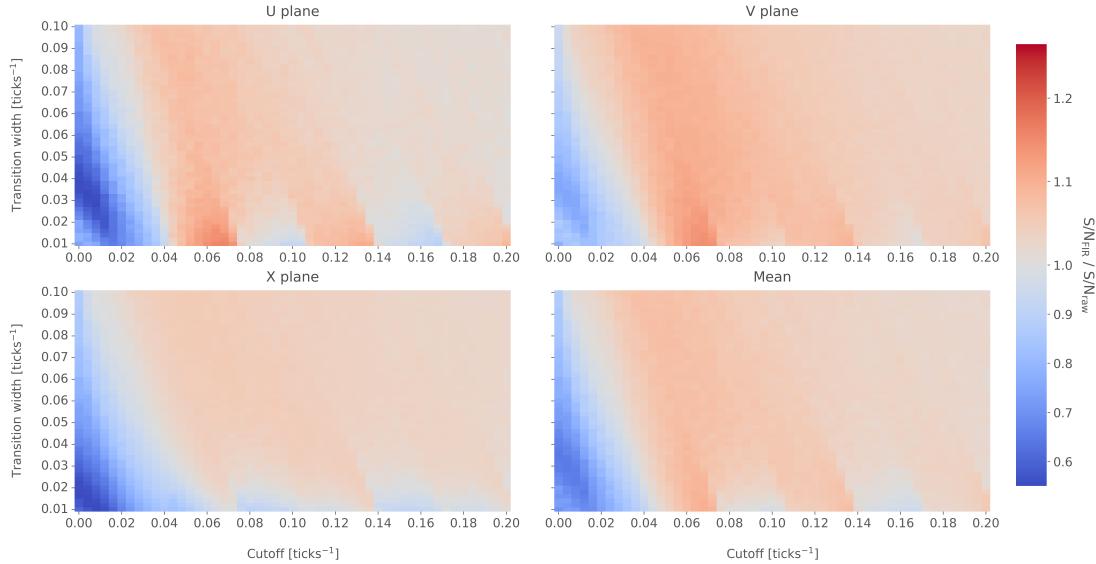


Figure 4.5: Relative change in the S/N for the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44`, using different values of the cutoff frequency f_c and the transition width δf . The optimal Chebyshev filters were applied using just the integer part of the coefficients given by the Parks-McClellan algorithm.

1035 distribution of the S/N improvement values for all the channels in the raw ADC capture
 1036 `felix-2020-07-17-21:31:44`, separated by wire plane, after the optimal Chebyshev
 1037 filter was applied. One can see that there is a clear improvement for both U and V
 1038 induction wire planes, obtaining a mean change of 1.25 and 1.30 for them respectively.
 1039 However, in the case of the collection plane X the mean of this distribution is roughly 1,
 1040 meaning that a good fraction of channels in that plane get a slightly worse S/N after the
 1041 filter is applied. In any case, this is not a big issue as the S/N for collection channels is
 1042 usually much higher than the one for induction channels.

1043 The results I obtained optimising the low pass filter with the Parks-McClellan method
 1044 are promising. Nonetheless, the improvement found is rather marginal so I wondered
 1045 if there could be an alternative approach to the filtering problem which yields better
 1046 outputs. At this point, I found a possible alternative in matched filters. By construction,
 1047 this kind of filters offer the best improvement on the S/N.

4.4. MATCHED FILTERS

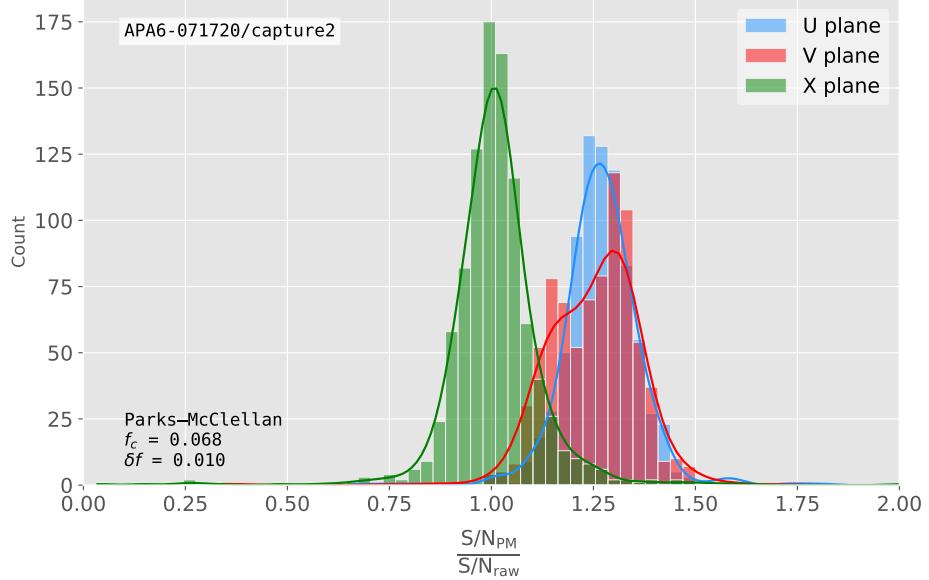


Figure 4.6: Distribution of the relative change of the S/N on the different wire planes from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` after the optimal Chebyshev filter was applied. The filter was computed with the Parks-McClellan algorithm using a cutoff of $f_c = 0.068 \text{ ticks}^{-1}$ and a transition width $\delta f = 0.010 \text{ ticks}^{-1}$.

1048 4.4 Matched filters

1049 In the context of signal processing, a matched filter is the optimal linear filter for
 1050 maximising the signal-to-noise ratio (S/N) in the presence of additive noise, obtained
 1051 by convolving a conjugated time-reversed known template with an unknown signal to
 1052 detect the presence of the template in the signal [94].

1053 Given a known signal sequence $s(t)$ and another (a priori unknown) noise sequence
 1054 $n(t)$, the input signal can be written as:

$$x(t) = s(t) + n(t). \quad (4.4)$$

1055 Now, considering a linear time-invariant filter, whose impulse-response function I

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

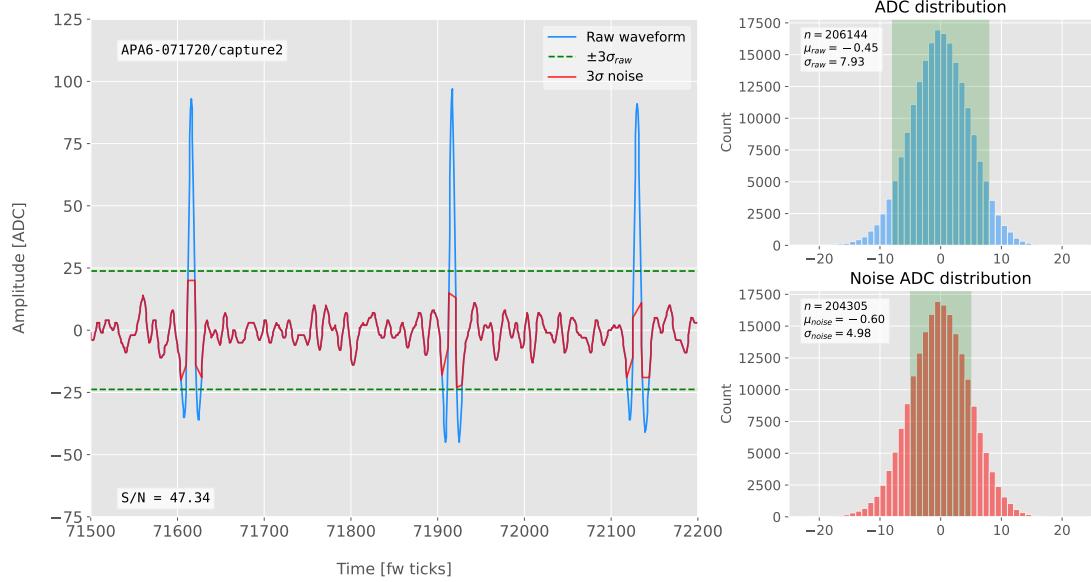


Figure 4.7: Left panel: Zoomed match filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture *felix-2020-07-17-21:31:44* (blue line). The filter used was directly extracted from the data, being the 32 values around the first peak in the original waveform. The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$. The resulting noise waveform is also shown (red line). Top right panel: ADC distribution for channel 7840 after match filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$. Bottom right panel: noise ADC distribution for channel 7840 after match filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$

1056 will refer to as $h(t)$, one can write the output signal as:

$$\begin{aligned} y(t) &= x(t) * h(t) \\ &= (s(t) + n(t)) * h(t) \\ &= y_s(t) + y_n(t), \end{aligned} \tag{4.5}$$

1057 where $y_s(t)$ and $y_n(t)$ are simply the outputs of the filter due to the signal and the noise
1058 components respectively.

1059 The goal of the matched filter is to detect the presence of the signal $s(t)$ in the input
1060 sample $x(t)$ at a certain time t_0 , which effectively means we need to maximise the S/N.
1061 This way, what one wants is to have a filter which gives a much bigger output when the
1062 known signal is present than when it is not. Putting it in other words, the instantaneous
1063 power of the signal output $y_s(t)$ should be much larger than the average power of the

4.4. MATCHED FILTERS

1064 noise output $y_n(t)$ at some time t_0 .

1065 For the case of the filtered signal, one can easily re-write it as an inverse Fourier
1066 transform:

$$y_s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega H(\omega)S(\omega)e^{i\omega t}, \quad (4.6)$$

1067 where $H(\omega)$ and $S(\omega)$ are the Fourier transforms of the impulse-response function (i.e.
1068 the transfer function of the filter) and of the input signal, respectively.

1069 Now focusing on the noise, we can use the Wiener-Khinchin theorem [95] to write
1070 the mean power of the noise after filtering as:

$$E|y_n(t)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega |H(\omega)|^2 S_n(\omega), \quad (4.7)$$

1071 where $S_n(\omega)$ is the power spectral density of the noise.

1072 Having these, one can write the instantaneous S/N at time t_0 as:

$$\begin{aligned} \left(\frac{S}{N} \right)_{t_0} &= \frac{|y_s|^2}{E|y_n(t)|^2} \\ &= \frac{1}{2\pi} \frac{\left| \int_{-\infty}^{\infty} d\omega H(\omega)S(\omega)e^{i\omega t_0} \right|^2}{\int_{-\infty}^{\infty} d\omega |H(\omega)|^2 S_n(\omega)}. \end{aligned} \quad (4.8)$$

1073 Once we have this expression, we need to find the upper limit of it to determine what
1074 would be the optimal choice for the transfer function. One can use the Cauchy-Schwarz
1075 inequality, which in the present case takes the form:

$$\left| \int_{-\infty}^{\infty} dx f(x)g(x) \right|^2 \leq \int_{-\infty}^{\infty} dx |f(x)|^2 + \int_{-\infty}^{\infty} dx |g(x)|^2, \quad (4.9)$$

1076 for any two analytical functions $f(x)$ and $g(x)$. One can prove that making the choice:

$$\begin{aligned} f(x) &= H(\omega) \sqrt{S_n(\omega)} e^{i\omega t_0}, \\ g(x) &= \frac{S(\omega)}{\sqrt{S_n(\omega)}}, \end{aligned} \quad (4.10)$$

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1077 leads to the following upper bound for the S/N:

$$\left(\frac{S}{N}\right)_{t_0} \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{|S(\omega)|^2}{S_n(\omega)}. \quad (4.11)$$

1078 From Eqs. (4.8), (4.9) and (4.10) one can also derive the form of the transfer function
1079 such that the upper bound is exactly reached [96]:

$$H(\omega) \propto \frac{S^*(\omega)e^{-i\omega t_0}}{S_n(\omega)}. \quad (4.12)$$

1080 From this last expression we can clearly see the way the matched filter acts. As the
1081 transfer function is proportional to the Fourier transform of the signal it will try to only
1082 pick the frequencies present in the signal [97].

1083 The matched filter transfer function can be greatly simplified if the input noise is
1084 Gaussian. In that case, the power spectral density of the noise is a constant, so it can be
1085 re-absorbed in the overall normalisation of the transfer function. Moreover, considering
1086 that the input signal is a real function, one can simply set $S^*(\omega) = S(-\omega)$, which gives:

$$H(\omega) \propto S(-\omega)e^{-i\omega t_0}. \quad (4.13)$$

1087 For a discrete signal, one can think of the input and impulse-response sequences as
1088 vectors of \mathbb{R}^N . Then, the matched filter tries to maximise the inner product of the signal
1089 and the filter while minimising the output due to the noise by choosing a filter vector
1090 orthogonal to the later. In the case of additive noise, that leads to the impulse-response
1091 vector:

$$h = \frac{1}{\sqrt{s^\dagger R_n^{-1} s}} R_n^{-1} s, \quad (4.14)$$

1092 where s is a reversed signal template sequence of length N equal to the order of the filter
1093 and R_n is the covariance matrix associated with the noise sequence n . For the Gaussian
1094 noise case, the covariance matrix is simply the unit matrix, so the above expression

4.4. MATCHED FILTERS

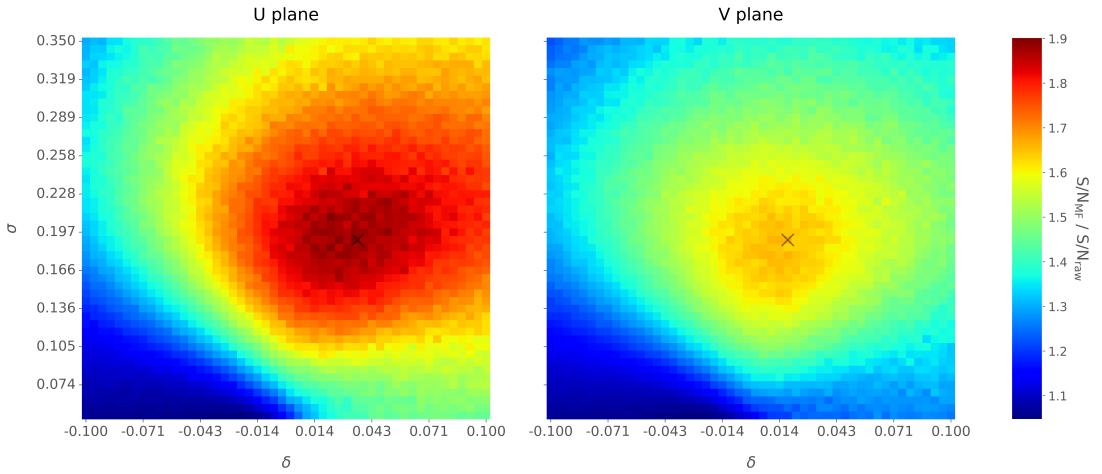


Figure 4.8: Relative improvement in the S/N for the raw data capture `felix-2020-07-17-21:31:44`, using the matched filter following the parametrisation in Eq. (4.16). The black crosses in both panels denote the location of the maximum ratio value.

1095 simplifies again to:

$$h = \frac{s}{|s|}. \quad (4.15)$$

1096 To test whether this choice of filter is appropriate one needs to choose a signal
 1097 template. As an example of how a matched filter would affect our signal, I simply
 1098 took the filter coefficients to be the 32 ADC values around a signal peak present in the
 1099 data. In Fig. 4.7 (left panel) I plotted a zoomed region for channel 7840 in the raw
 1100 data capture `felix-2020-07-17-21:31:44`, after applying the matched filter described
 1101 before (blue line). When compared to the raw and FIR filtered case (see Figs. 4.2 and
 1102 4.3), after applying the match filter the standard deviation of the noise waveform (red
 1103 line) decreases and at the same time the signal peaks are enhanced. This leads to an
 1104 improvement of the S/N by a factor of 1.92 when compared to the raw waveform.

1105 In order to obtain the matched filter that is more suitable for our data, I explored
 1106 different configurations of signal templates. In order to perform this exploration, I
 1107 parametrised the signal using the bipolar function:

$$f(x) = -A(x + \delta) e^{-x^2/\sigma^2}, \quad (4.16)$$

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1108 where the parameter δ controls the asymmetry between the positive and negative peaks
1109 and σ controls their width. The amplitude parameter A is set such that it keeps the
1110 height of the biggest peak to be less than 200 ADC in absolute value.

1111 As this parametrisation is only adequate for bipolar signals I will focus exclusively
1112 on the induction channels. Also, the optimal configurations I found for the U and V
1113 plane will be kept separate, i.e. I will have two sets of coefficients that will be applied to
1114 either the U and V planes of wires. I do so as I found this was the choice giving the
1115 best performance. Even so, as I will discuss, the differences are not very pronounced. In
1116 case it is not technically possible to separate channels in the firmware according to the
1117 wire plane they come from and use different sets of filter coefficients for them, we can
1118 just find a common unique set of coefficients. In such case, I do not expect our results
1119 to change dramatically.

1120 In Fig. 4.8 I present the results of our parameter scan, for channels in the induction
1121 planes U (left panel) and V (right panel). For each configuration of σ and δ the resulting
1122 matched filter was applied to all channels in the corresponding plane within the data
1123 capture `felix-2020-07-17-21:31:44`, the S/N improvement was computed with respect
1124 to the raw waveforms and then the S/N mean value was kept as a score for such filter.
1125 One can see that the improvement obtained for the U plane is in general higher than the
1126 one for the V plane. In any case, I got substantially higher ratios than the ones obtained
1127 for the low-pass FIR filters. For the optimal configurations I attained improvements up
1128 to a factor of 1.85 for the U plane and 1.65 for the V plane.

1129 The sets of optimal matched filter coefficients were obtained for the parameters
1130 $\delta = 0.035$, $\sigma = 0.191$ for the U plane and $\delta = 0.018$, $\sigma = 0.191$ for the V plane. I
1131 show these two sets of coefficients in Fig. 4.9 (left panel). Also in Fig. 4.9 (right
1132 panel) I plot the distribution of the S/N improvement after the optimal match filters
1133 for the U and V were applied to the corresponding channels in the raw data capture
1134 `felix-2020-07-17-21:31:44`. As mentioned before, the mean improvement achieved
1135 for the U plane channels is slightly bigger than the one for the V channels. Note, however,
1136 that the spread of the distribution for the V plane is also smaller than the one for the U

4.5. MONTE CARLO STUDIES

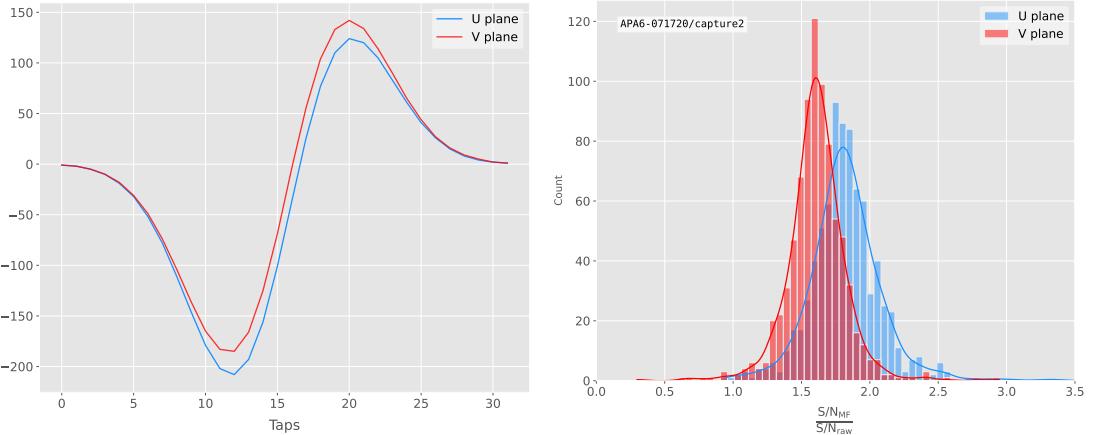


Figure 4.9: Left panel: Optimal matched filter coefficients for the U (blue line) and V (red line) planes. The filters were computed with our parametrisation in Eq. (4.16) for the parameter values $\delta = 0.035$, $\sigma = 0.191$ and $\delta = 0.018$, $\sigma = 0.191$ respectively. Right panel: Distribution of the relative change of the S/N on the two induction wire planes from the ProtoDUNE-SP raw data capture *felix-2020-07-17-21:31:44* after their respective optimal matched filters were applied.

1137 plane.

1138 Overall, one can see that the improvements on the S/N are much more significant in
 1139 the case of the matched filter than it is for the low-pass FIR filters. The analysis of this
 1140 and other raw data captures from ProtoDUNE-SP suggest that matched filters increase
 1141 the S/N of induction channels by a factor of 1.5 more than the optimal low-pass FIR
 1142 filters.

1143 Although these results are by themselves great points in favour of the matched
 1144 filter, more studies are needed to completely assess the robustness of this approach. I
 1145 proceeded then to test the matched filter with simulated data samples.

1146 4.5 Monte Carlo studies

1147 In order to further test the matched filter, the next step was to generate and process
 1148 data samples using *LArSoft* [98]. In this way, one can control the particle content of
 1149 the samples, the orientation of the tracks and their energy, and therefore see how the
 1150 matched filter behaves in various situations.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

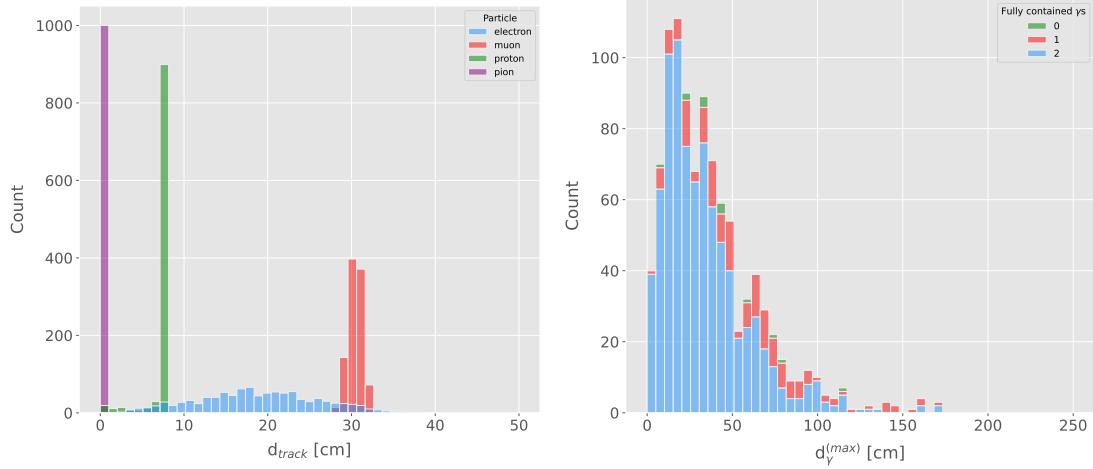


Figure 4.10: Left panel: distributions of the particles track length in the liquid argon for the generated $E_k = 100$ MeV monoenergetic samples, electrons (blue), muons (red), protons (green) and neutral pions (purple). Right panel: distribution of the length of the longest photon in the neutral pion sample after the decay process $\pi^0 \rightarrow \gamma\gamma$.

1151 To begin with, I prepared different monoenergetic and isotropic samples containing
 1152 a single particle per event. Each sample contains a different particle species, namely
 1153 electrons, muons, protons and neutral pions all with a kinetic energy of $E_k = 100$ MeV.
 1154 I chose these because of the fairly different topologies they generate in the liquid argon,
 1155 ranging from shower-like to track-like.

1156 These were generated with the single particle gun and the Geant4 stage of the
 1157 *LArSoft* simulation [98] was performed with the standard configuration for the DUNE
 1158 FD 10kt module.

1159 For simplicity, I restricted the particles to start drifting in a single TPC volume
 1160 (in this case TPC 0), so I can focus exclusively on the signals coming from one APA.
 1161 The chosen kinetic energy for all the particles in my first trial is $E_k = 100$ MeV, so a
 1162 necessary check is to see if all our tracks will be typically contained in one TPC volume.
 1163 Fig. 4.10 (left panel) shows the distributions of the track lengths in the liquid argon
 1164 of all generated particles with $E_k = 100$ MeV. One can see that, in the case of the
 1165 track-like particles (i.e. muons and protons), their length distributions are quite sharp
 1166 and centered at relatively low distances (30 and 8 cm, respectively). For electrons, the
 1167 distribution is quite broad but it does not extend past ~ 30 cm. The case of neutral

4.5. MONTE CARLO STUDIES

1168 pions can be misleading, as they decay promptly the track length associated with the
1169 true Monte Carlo particle is always < 1 cm. In Fig. 4.10 (right panel) I show the
1170 effective length distribution of the longest photon after the pion decays as $\pi^0 \rightarrow \gamma\gamma$,
1171 highlighting the number of fully contained photons in the TPC volume per event (either
1172 zero, one or both). One can see that the vast majority of events has both photons
1173 contained and that just a negligible number of them has none of them contained in the
1174 TPC volume. In any case, for the sake of caution, I will only keep the pion events with
1175 both photons contained.

1176 Once I have prepared a sample at the Geant4 level, I need to process it through the
1177 detector simulation. In order to make adequate estimations of the noise levels and run
1178 the filtering and hit finder as I did with the ProtoDUNE data, one needs to turn off
1179 the default zero-suppression of the waveforms produced by the simulation. At this first
1180 stage I am only concerned with the waveforms with the noise added, so I keep the noise
1181 addition option as true in the configuration. However, for studies related to the hit finder
1182 performance one will also need to store the noiseless waveforms in order to retrieve the
1183 truth information of the hits. I will discuss this approach next.

1184 After the detector simulation stage, one needs to extract the no zero-suppressed noisy
1185 waveforms, along with their offline channel numbers, and store them in a certain format
1186 to be analysed later. To reduce the amount of data that will go for processing, I used the
1187 information from the Geant4 step of the simulation to select only the active channels,
1188 i.e. the channels where some ionisation electrons arrive. Moreover, as said previously, I
1189 only extract the waveforms from APA 0 and exclusively the ones coming from induction
1190 channels. The resulting ROOT file contains a tree with two branches, one containing
1191 the waveforms for each event and channel and the other with the corresponding offline
1192 channel numbers.

1193 Finally, to extract the truth values for the orientation of the tracks and the energies
1194 of the particles I used a modified analysis module. This gives a ROOT file with a single
1195 tree, containing several branches with different information such as the components of
1196 the initial momentum of the particles, initial and final xyz location, track length, etc.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

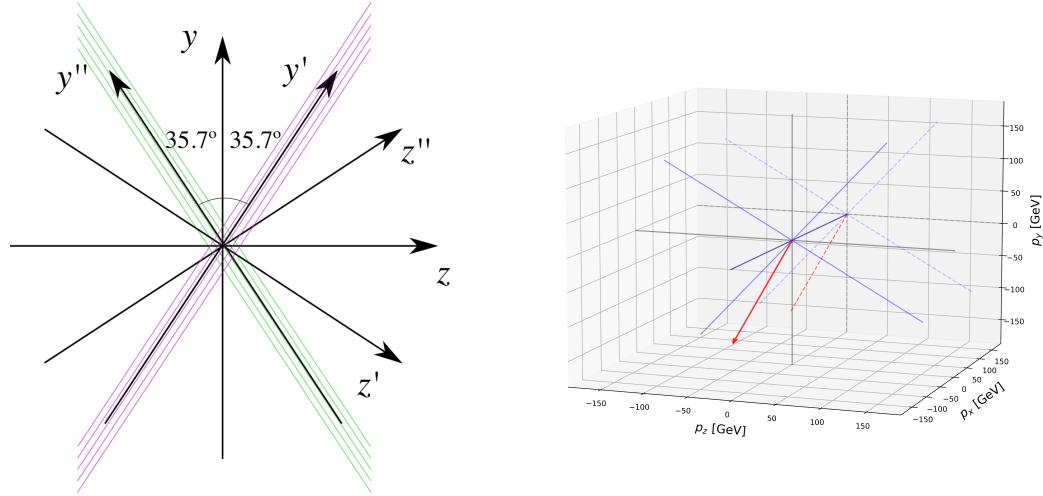


Figure 4.11: Left panel: schematic representation of the two new rotated reference frames used in this analysis (denoted as prime and double prime), viewed from the yz plane. The magenta stack of lines represent the wires in the U plane, whereas the green lines correspond to the wires in the V plane. Right panel: 3D representation of the momentum of one of the generated monoenergetic muons (red arrow) in the original reference frame (black lines), along with the new reference frame used for the U plane waveforms (blue lines). In the yz plane I added the projection of these three.

1197 For the analysis of the resulting waveforms and truth values I used a custom set of
 1198 Python libraries (available at [???]). Among other functionalities, these enable the user
 1199 to read the ROOT files, export the raw data as pandas objects, apply the filters and
 1200 compute the S/N of both the raw and filtered signals. So far, the default configuration
 1201 for the filtering uses the set of optimal matched filter coefficients that I found using the
 1202 ProtoDUNE data samples.

1203 Additionally, for the analysis of the samples it was necessary to use two different
 1204 reference frames, to study separately the signals coming from the U and V induction wire
 1205 planes. As I am focussing on a single APA, the U and V wires have a different orientation
 1206 in the yz plane. In the case of U wires, these are tilted 35.7° clockwise from the vertical
 1207 (y direction), whereas the V wires are at the same angle but in the counter clockwise
 1208 direction. Because of this, the best option is to deal with two new coordinate systems
 1209 rotated by $\pm 35.7^\circ$ along the x axis, so the new y' and y'' directions are aligned with the
 1210 U and V induction wires. Fig. 4.11 (left panel) shows a schematic representation of

4.5. MONTE CARLO STUDIES

the original reference frame together with the two rotated ones (denoted by primed and double primed). This way, one can easily understand how parallel was a track to the wires in the two induction planes. Fig. 4.11 (right panel) shows a 3D representation of the momentum of a track (red arrow) in the original reference frame (black lines), along with the new reference frame for U wires (blue lines). I added the projection in the yz plane of this three, to show the usefulness of the new reference frame to tell whether a track is parallel or normal to the wires in the induction plane.

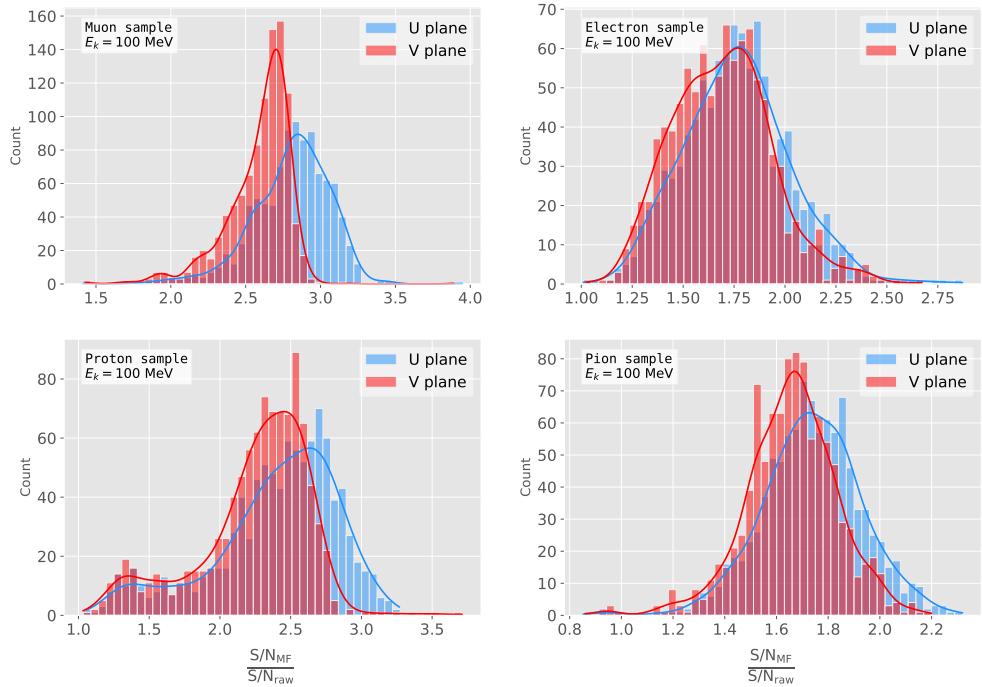


Figure 4.12: Distributions of the mean S/N improvement per event for the corresponding sample after applying the matched filters. Here I separated the change in the U plane (blue) and the V plane (red) channels. From top left to the right: muon, electron, proton and neutral pion. All the events have a fixed kinetic energy of $E_k = 100$ MeV.

Fig. 4.12 shows the distribution of the average S/N improvement per event when one applies the optimal matched filters. I produced separate distributions for the channels in the U (red) and V (blue) induction wire planes. Notice that the S/N distributions for the track-like particles, i.e. muons (top left panel) and protons (bottom left panel), have significantly larger mean values than the distributions of the shower like particles, i.e. electrons (top right panel) and neutral pions (bottom right panel). An important

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

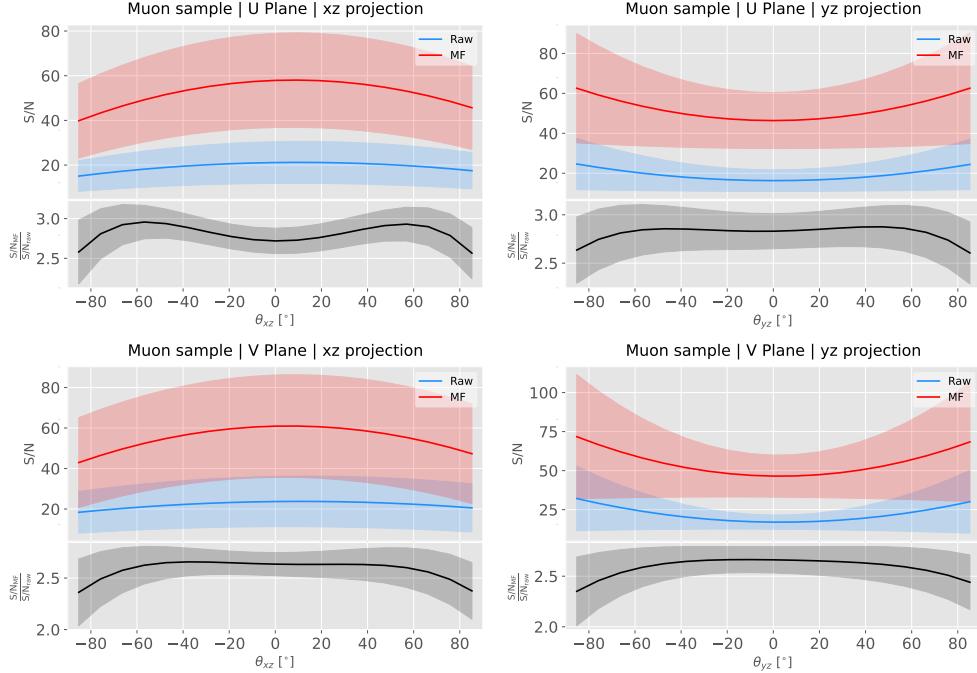


Figure 4.13: Angular dependence of the mean S/N and the S/N improvement, for the different monoenergetic samples considered (from top to bottom: electrons, muons, protons and neutral pions). The two columns on the left represent the values for the U plane waveforms. The top subplots show the mean S/N for raw (green) and filtered (red) waveforms whereas the bottom subplots depict the averaged S/N improvement (black).

difference between these results and the ones seen before for ProtoDUNE data is that, overall, the improvements that I get for simulated data are bigger. This could be due either to the default noise model used in the *LArSoft* simulation or to the simulated hits having higher energy than the ones in the recorded data. Nonetheless, the concluding message is that the previously optimised matched filters give an overall significant improvement of the S/N for the different samples.

About the convention I followed for the plots and results, in the case of the raw and filtered S/N of each event in the sample I simply took the average of the quantities over all the active channels in the event. That is, if a certain event has N_{chan} active channels

4.5. MONTE CARLO STUDIES

1233 these two quantities are computed as:

$$\begin{aligned}(S/N_{fir})_{event} &= \frac{\sum_{i=0}^{N_{chan}} (S/N_{fir})_i}{N_{chan}}, \\ (S/N_{raw})_{event} &= \frac{\sum_{i=0}^{N_{chan}} (S/N_{raw})_i}{N_{chan}}.\end{aligned}\quad (4.17)$$

1234 However, for the ratio of the raw and filtered S/N (what I called the S/N improvement)
1235 per event I am not just taking the ratio of the previous two quantities but computing
1236 the average of the individual ratios per channel in the event:

$$\left(\frac{S/N_{fir}}{S/N_{raw}}\right)_{event} = \frac{\sum_{i=0}^{N_{chan}} \left(\frac{S/N_{fir}}{S/N_{raw}}\right)_i}{N_{chan}}, \quad (4.18)$$

1237 and so:

$$\left(\frac{S/N_{fir}}{S/N_{raw}}\right)_{event} \neq \frac{(S/N_{fir})_{event}}{(S/N_{raw})_{event}}. \quad (4.19)$$

1238 4.5.1 Angular dependence

1239 Having these monoenergetic samples, one can also study the angular dependence of the
1240 performance of the matched filter. This is an important point, as it is a well established
1241 fact that for certain configurations (an extreme case configuration being signals normal
1242 to the wire plane and perpendicular to the induction wires at the same time) the S/N is
1243 much lower than average as the corresponding waveforms are severely distorted. In this
1244 sense, I am interested to see how the matched filter behaves for these cases and how the
1245 S/N improvement on those compare to the average.

1246 Fig. 4.13 shows the angular dependence of the S/N for the monoenergetic $E_k =$
1247 100 MeV isotropic muons, for the different induction wire planes and projections. The
1248 angles for each event are given by the components of the initial value of the momentum
1249 of the particles, taking the angles of the projections on the xz and yz planes with respect
1250 to the z axis (more accurately, one needs to compute these angles twice for each event, a
1251 pair for the $xy'z'$ coordinate system and the other for the $xy''z''$). The top row shows the
1252 dependence on the angles corresponding to the U plane, i.e. $\theta_{xz'}$ and $\theta_{y'z'}$, whereas the

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

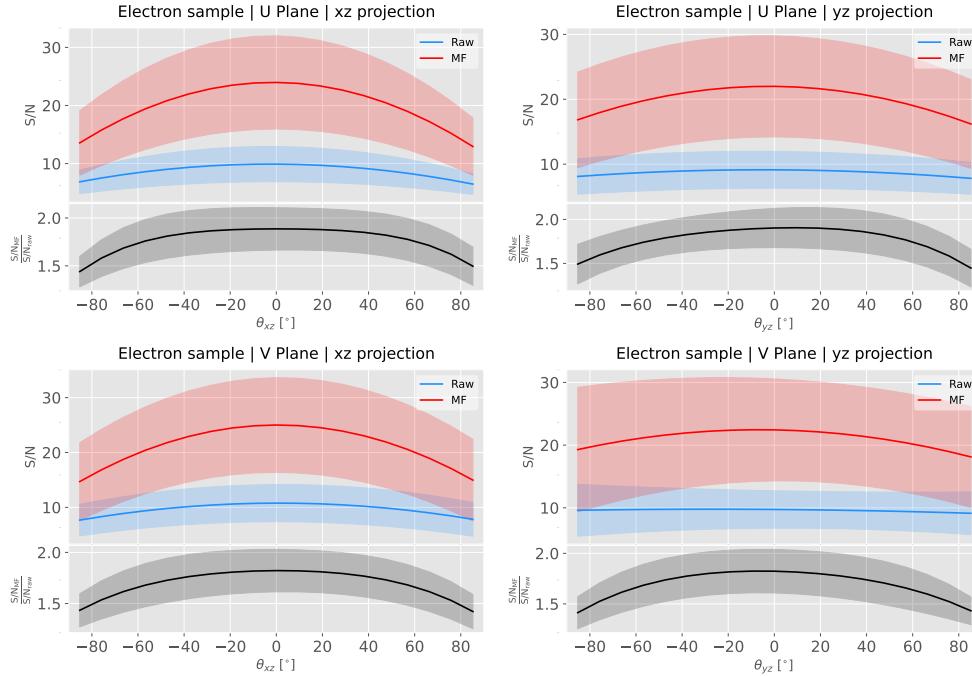


Figure 4.14: Angular dependence of the mean S/N and the S/N improvement, for the different monoenergetic samples considered (from top to bottom: electrons, muons, protons and neutral pions). The two columns on the left represent the values for the U plane waveforms. The top subplots show the mean S/N for raw (green) and filtered (red) waveforms whereas the bottom subplots depict the averaged S/N improvement (black).

bottom row shows the angular dependence viewed from the V plane, $\theta_{xz''}$ and $\theta_{y''z''}$. In each plot, the top subplot represents the mean values of the S/N for the raw (blue) and matched filtered (red) signals, and the bottom subplot the averaged S/N improvement (black). The solid lines represent the mean value obtained for the corresponding angular value, whereas the semitransparent bands represent one standard deviation around the mean at each point.

As expected, the S/N is in general higher when tracks are parallel to the APA (i.e. $\theta_{xz} \sim 0$) and lower when it is normal to the plane ($\theta_{xz} \sim \pm 90^\circ$). In the same way, tracks parallel to the wires ($\theta_{yz} \sim \pm 90^\circ$) tend to have higher S/N than those perpendicular to these ($\theta_{yz} \sim \pm 0$).

Fig. 4.14 shows the corresponding angular dependence information for the $E_k = 100$ MeV electrons sample. Notice that, in this case, the S/N behaviour discussed above does not hold. A possible explanation can be that, because most hits in these events

4.5. MONTE CARLO STUDIES

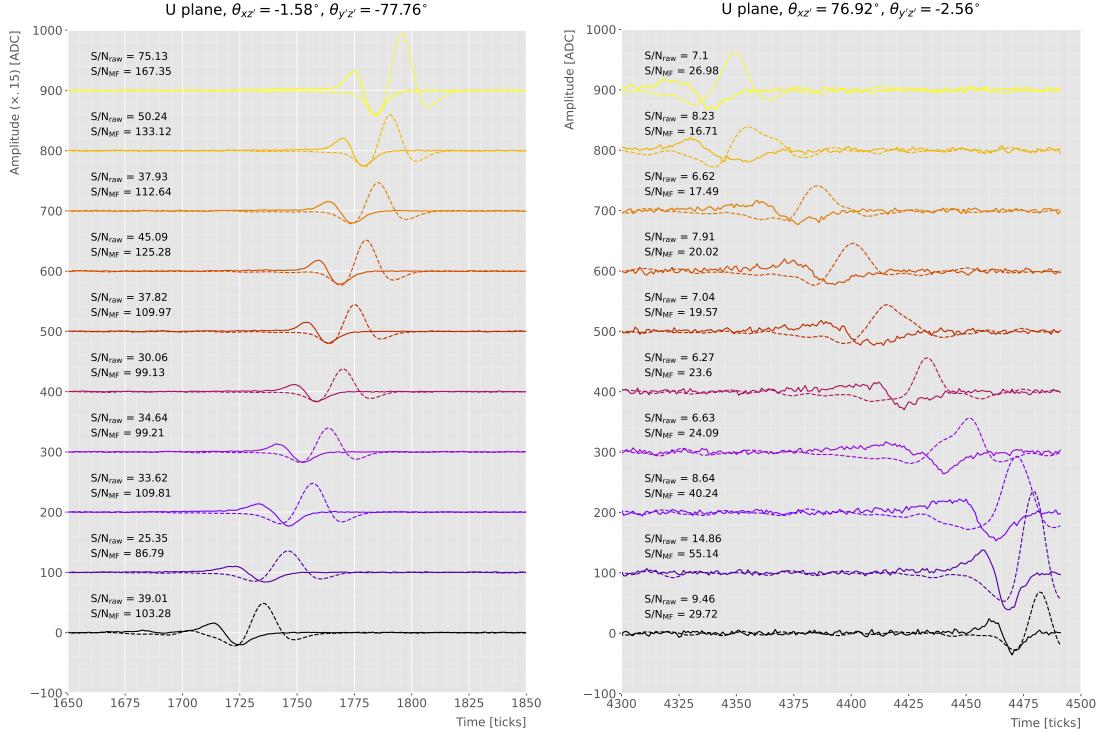


Figure 4.15: Selected consecutive waveforms corresponding to two monoenergetic $E_k = 100$ MeV muon events, one is parallel to the APA and to the wires in the U plane (left panel) and the other is normal to the APA plane and perpendicular to the U plane wires (right panel). The solid lines represent the raw waveforms whereas the dashed lines correspond to the waveforms after the matched filter was applied. The waveforms on the left panel have been scaled by a factor of 0.15 to have similar amplitude to the ones on the right panel.

1266 are produced by the secondary particles generated in the EM shower, the signal peaks
 1267 whose S/N ratios were computed do not correspond to the directional information of
 1268 the primary electron.

1269 4.5.2 Distortion and peak asymmetry

1270 As a little case of study, I selected two of the simulated $E_k = 100$ MeV monoenergetic
 1271 muon events. With respect to the U induction plane, one is parallel to the APA (low
 1272 $\theta_{xz'}$) and to the wires (high $\theta_{yz'}$) and the other is normal to the APA plane (high $\theta_{xz'}$)
 1273 and perpendicular to the wires (low $\theta_{yz'}$). As expected from the results on the angular
 1274 dependence discussed above, the former has a higher S/N (before and after the filtering)

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

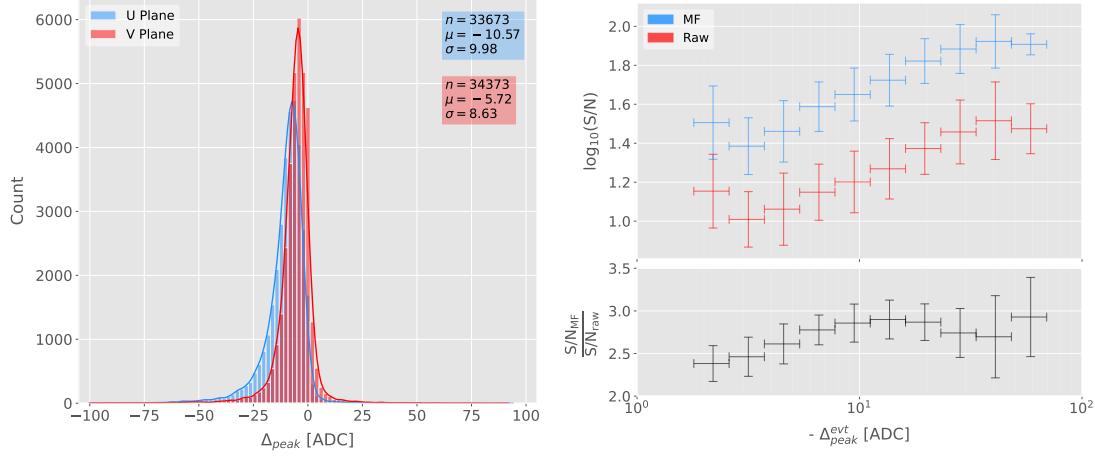


Figure 4.16: Left panel: peak asymmetry distribution for the case of the monoenergetic $E_k = 100$ MeV muon sample. Each value corresponds to a single bipolar signal peak from a channel in any event. The blue distribution represents the peaks on U plane channels, whereas the red corresponds to signal peaks in V wires. Right panel: relation between the mean peak asymmetry per event with the S/N for U channel waveforms from the $E_k = 100$ MeV muon sample. The top subplot shows the decimal logarithm of the mean S/N for the raw (red) and the matched filtered (blue) waveforms. The bottom subplot contains the mean S/N improvement ratio after the matched filter was applied.

when compared to the latter. An interesting thing to notice about these two samples is that, even though one has a much bigger S/N than the other, it is the one with the smallest S/N the one that got the biggest averaged S/N improvement. In Table 4.1 I included all the relevant parameters of these two $E_k = 100$ MeV muon events I am considering, namely, the angles with respect to the $xy'z'$ reference frame, the values of the S/N, the S/N improvement and also the so-called peak asymmetry Δ_{peak} that I will discuss next.

Table 4.1: Characteristic parameters of the two monoenergetic muon events selected, relative to the U plane: projected angles in the xz' and $y'z'$ planes, S/N values for the raw and filtered waveforms, mean improvement of the S/N and peak asymmetry.

	$\theta_{xz'}$ ($^\circ$)	$\theta_{y'z'}$ ($^\circ$)	S/N _{raw}	S/N _{MF}	$S/N_{MF}/S/N_{raw}$	Δ_{peak} (ADC)
High ("parallel")	-1.58	-77.76	41.65	112.44	2.83	-35.73
Low ("normal")	76.92	-2.56	8.07	25.46	3.12	-10.38

One can try to understand better what is going on with these two events by looking at the raw and filtered data from some of their active channels. Fig. 4.15 shows a

4.5. MONTE CARLO STUDIES

selection of consecutive raw and filtered U plane waveforms from the event with high S/N (left panel) and the one with low S/N (right panel). Notice that to show both collections of waveforms at a similar scale I had to apply a factor of 0.15 to the waveforms with high S/N. Additionally, next to each waveform I included the values of the raw and matched filtered S/N for the corresponding channel. The first thing to notice in this plot is that the amplitude of the signal peaks from the normal track have a much smaller amplitude, and also appear quite distorted when compared to the others. On the other hand, although the matched filtered S/N is still smaller, the relative improvement is bigger than in the parallel case.

A way I found to quantify the difference between the shapes within these two events is their different peak asymmetry. One can define the peak asymmetry as the (signed) difference between the positive and the negative peaks of the bipolar shape, i.e.:

$$\Delta_{peak} \equiv h_+ - h_-, \quad (4.20)$$

where both heights h_+ and h_- are positive defined. Fig. 4.16 (left panel) shows the distribution of this peak asymmetry for all the waveforms corresponding to channels in the U (blue) and V (red) planes for the monoenergetic muon sample. One can see that these distributions are clearly shifted to negative values (with mean values $\mu_{\Delta}^U = -10.57$ ADC and $\mu_{\Delta}^V = -5.72$ ADC respectively). It is interesting to notice that the peak asymmetry value of the sample with high S/N sits at the left tail of the distribution whereas the corresponding value of the sample with low S/N lies around the mean.

Now, one can try to correlate the peak asymmetry with the S/N and the S/N change per event. Fig. 4.16 (right panel) shows the result of comparing (minus) the mean peak asymmetry per event to the averaged raw (red) and matched filtered (blue) S/N per event (top subplot). The horizontal lines sit at the mean value obtained in the fit and represent the width of the $-\Delta_{peak}$ bins used, while the vertical lines indicate one standard deviation around that mean value. Notice that, when taking decimal logarithm

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

on both, there is an approximate linear relation between these quantities, except for peak asymmetry values bigger than -5 ADC where the S/N remains constant.

Also, in the bottom subplot of Fig. 4.16 (right panel) I show the relation between the peak asymmetry and the mean S/N improvement. In this case, one see that there is a maximum at $\Delta_{peak} \sim -10$ ADC. As mentioned previously, this is also the value of the mean of the peak asymmetry distribution. In fact, it is expected that our filter favours the signal peaks with the most common values of the peak asymmetry, as this was one of the features I target in our filter coefficient optimisation through the parameter δ .

These results suggest that events with poorer values of the mean S/N, usually associated to non-favourable track orientations, tend to have smaller values of the mean peak asymmetry (in absolute value). Nonetheless, because our matched filters have been optimised to account for these asymmetries, the improvement on the S/N for these events is sizeable if not better than the one for events which already had a high S/N.

4.5.3 Hit sensitivity

One of the advantages of the matched filter, directly related to increasing the S/N, is the capability of picking hits that before fell below the threshold. For instance, Fig. 4.17 shows the raw ADC data from an example event (electron, $E_k = 100$ MeV) with the produced true hits superimposed (black boxes), together with the hits produced by the standard hit finder chain (blue circles), i.e. using the current FIR filter, and the hits obtained using the matched filters (green triangles). Both the standard and the matched filter hit finders were run with a threshold of 10 ADC. Notice that the standard hits match well the true ones at the initial part of the event (where we have a track-like object), but they miss most of the hits produced by the EM shower at later times. On the other hand, the hits produced with the matched filter have a better agreement with the true hits even for the more diffuse shower activity.

Even though now I get more hits with this combination of matched filter and low threshold as a results of the enhancement of the signal peaks relative to the noise level, it is also true that I pick some spurious hits not related to any real activity if one lowers the

4.5. MONTE CARLO STUDIES

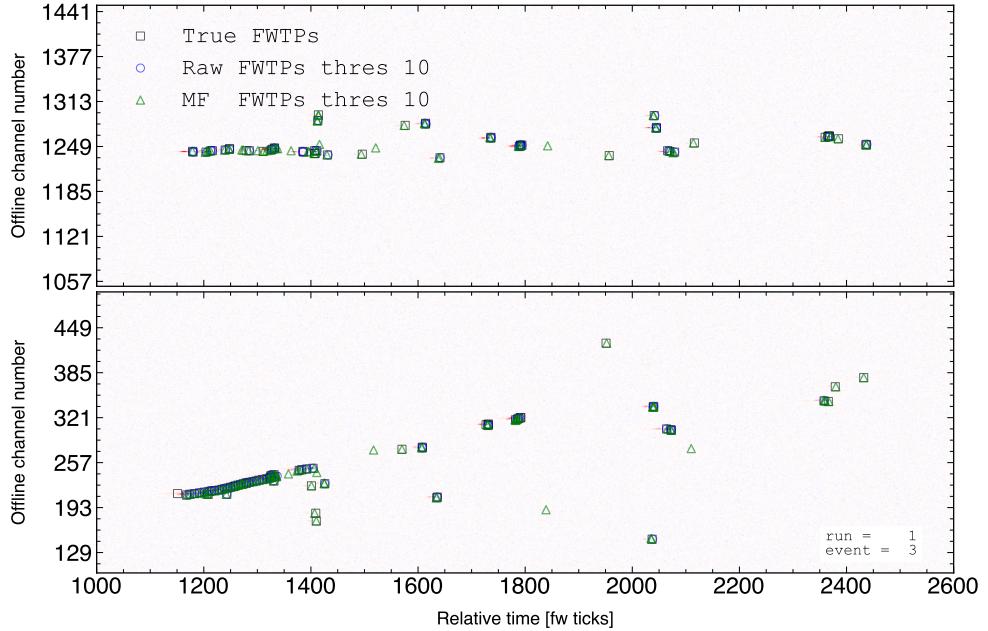


Figure 4.17: Raw data display in the plane time (in firmware ticks) vs. offline channel number for an $E_k = 100$ MeV electron event. The produced true hits are superimposed (black boxes) as well as the hits coming from the standard hit finder chain (blue circles) and the hit finder using the matched filter (green triangles).

thresholds too much. Therefore, some optimisation of the threshold is needed. Basically one will need to make a trade-off between precision and sensitivity.

Having this in mind, I tried to compare the produced hits one gets from the standard hit finder and the ones resulting from applying the matched filter with the true hits. By running the hit finders on our samples with different values of the threshold one can understand, for instance, how low one can set the threshold without getting mostly spurious hits and then evaluate the gains obtained from this.

Because now I am also interested in seeing how the hit sensitivity changes with the energy, I prepared new isotropic samples with the same types of particles as before (muons, electrons, protons and neutral pions) but with a flat kinetic energy distribution ranging from 5 to 100 MeV.

In order to estimate the hit sensitivity, given a certain sample, one needs to recover the set of true hits to be able to compare these with the ones produced. To do so, a modification in the procedure I was using to extract the raw waveforms is needed.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1352 For this kind of study I run the detector simulation in two steps, first I produce the
1353 waveforms without noise and extract them in the same format I used for the raw data,
1354 then the noise is added and the noisy waveforms are then written to a file as well.

1355 To have a better comparison between the true hits and the ones produced from
1356 the raw waveforms after applying the two filters, I applied also the FIR filter and the
1357 matched filters to the noiseless waveforms and then I run the hit finder with a minimal
1358 threshold (in this case I used 1 ADC) on these noiseless filtered waveforms. In this way
1359 I generated two sets of true hits, I will refer to them as standard true hits (with the
1360 current/default FIR filter) and matched filter true hits respectively. This allows a more
1361 precise matching between the different groups of hits produced, as it will account for
1362 any delays and distortions introduced by the FIR and the matched filters.

1363 In the case of the raw waveforms (with noise), I run the hit finder on them, with
1364 different values of the threshold, after applying either the FIR or the matched filters. I
1365 will name them simply standard hits and matched filter hits respectively. Then, I match
1366 the generated hits to the true hits (the standard hits with the standard true hits and
1367 the matched filter hits with the matched filter true hits). The matching is performed by
1368 comparing the channel number and the timestamp of the hits. To count as a match,
1369 I require that all hits with the same channel number and timestamp have overlapping
1370 hit windows, i.e. the time windows between their hit end and hit start times need to
1371 overlap. If more than one hit in one of the groups have hit overlap with the same hit in
1372 the other group I only count the hit with closer hit peak time value.

1373 To quantify the performance of the two hit finder approaches, I use a classical method
1374 from statistical classification known as confusion matrix [99]. This is basically a way of
1375 sorting the outputs of a binary classifier, considering the true values of the classification
1376 and the predicted values. It divides the outputs in four categories: true positive (TP,
1377 both true and predicted values are 1), false negative (FN, true value is 1 but predicted
1378 is 0), false positive (FP, true value is 0 but predicted is 1) and true negative (TN, both
1379 true and predicted values are 0)).

1380 The contents of the confusion matrix allow us to compute other derived scores to

4.5. MONTE CARLO STUDIES

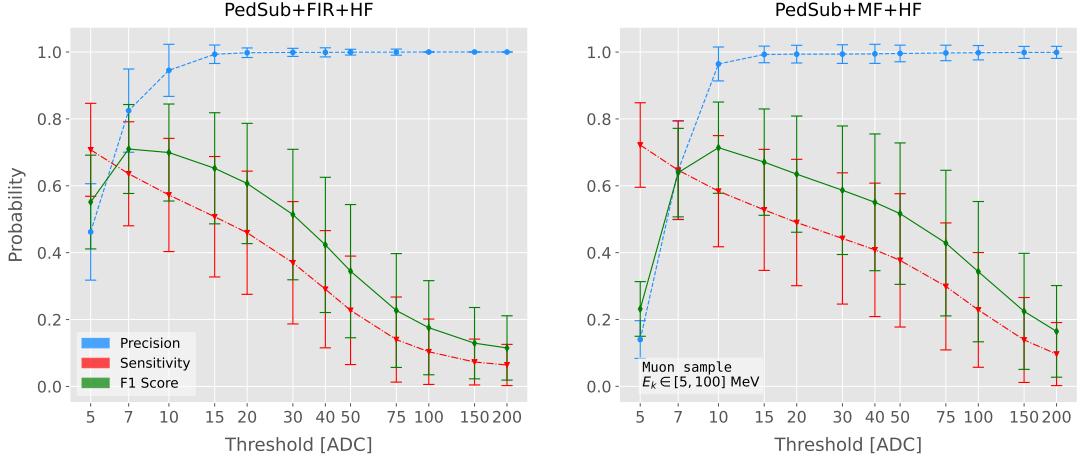


Figure 4.18: Dependence of the precision (blue), sensitivity (red) and F_1 (green) scores on the threshold values used in the hit finder, for the FIR (left panel) and matched filter (right panel) cases. The results were obtained after matching the hits to the true hits in the case of the isotropic muon sample with kinetic energy in the range 5 to 100 MeV, taking only into account the induction plane channels. The points represent the mean value while the error bars indicate one standard deviation around that mean value.

1381 judge the performance of our classifiers. In this study, I will make use of three of these
 1382 metrics, namely the precision or positive predictive value:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4.21)$$

1383 the sensitivity or true positive rate:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (4.22)$$

1384 and the F_1 score [100]:

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (4.23)$$

1385 which is the harmonic mean of the precision and the sensitivity.

1386 In our specific case I am not going to make use of the true negative value, as its
 1387 definition in this context can be ambiguous because one does not have clear instances in
 1388 the classification process. This way, I will only count the number of true positives as the
 1389 total amount of hits I can match between true and raw populations, the number of false

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1390 negatives will be the number of missing true hits and the false positive the number of
1391 hits which do not match any true hit.

1392 In Fig. 4.18 I show the precision (blue), sensitivity (red) and F_1 (green) scores I
1393 obtained for different values of the threshold used in the hit finder for the case of the
1394 muon sample. Because the matched filters are only applied to induction channels, I only
1395 consider here hits coming from the U and V planes. The panel on the left corresponds
1396 to the scores I got when I ran the hit finder on the FIR filtered waveforms, whereas the
1397 right panel contains the scores for the matched filter case. The points are centered at
1398 the threshold value used and represent the mean value obtained for each score using all
1399 the generated events, while the error bars indicate one standard deviation around the
1400 mean value.

1401 One can see that the precision for the matched filter case is lower when the thresholds
1402 are very low, as the noise baseline is slightly amplified, but then rises to high values
1403 quicker than for the FIR case. The other difference one can spot is that the sensitivity
1404 in the FIR case starts dropping faster at around the same threshold values where the
1405 precision stabilizes around 1, while in contrast for the matched filter this rapid decrease
1406 starts at higher threshold values. A similar scan for the same thresholds was performed
1407 for the electron sample in the same energy range, yielding similar results.

1408 In Fig. 4.19 I show the averaged hit sensitivity versus the kinetic energy of the
1409 events, both for the matched filter hits (blue) and the standard hits (red). The left
1410 panel corresponds to the muon sample, whereas the one on the right corresponds to the
1411 electron sample, both with kinetic energies between 5 and 100 MeV. In each panel the
1412 top plot corresponds to hits in the U plane, while the bottom plot contains the same
1413 information for the V plane. Each plot contains two subplots, the one on the top shows
1414 the hit sensitivity values for the matched filter and standard hits separate, while the
1415 bottom subplot depicts the ratio between the matched filter and standard sensitivities.
1416 The horizontal lines are placed at the mean value obtained in the fit and represent the
1417 width of the E_k bins used, while the vertical error bars indicate one standard deviation
1418 around that mean value. In both cases the threshold used was 30 ADC, as I required

4.5. MONTE CARLO STUDIES

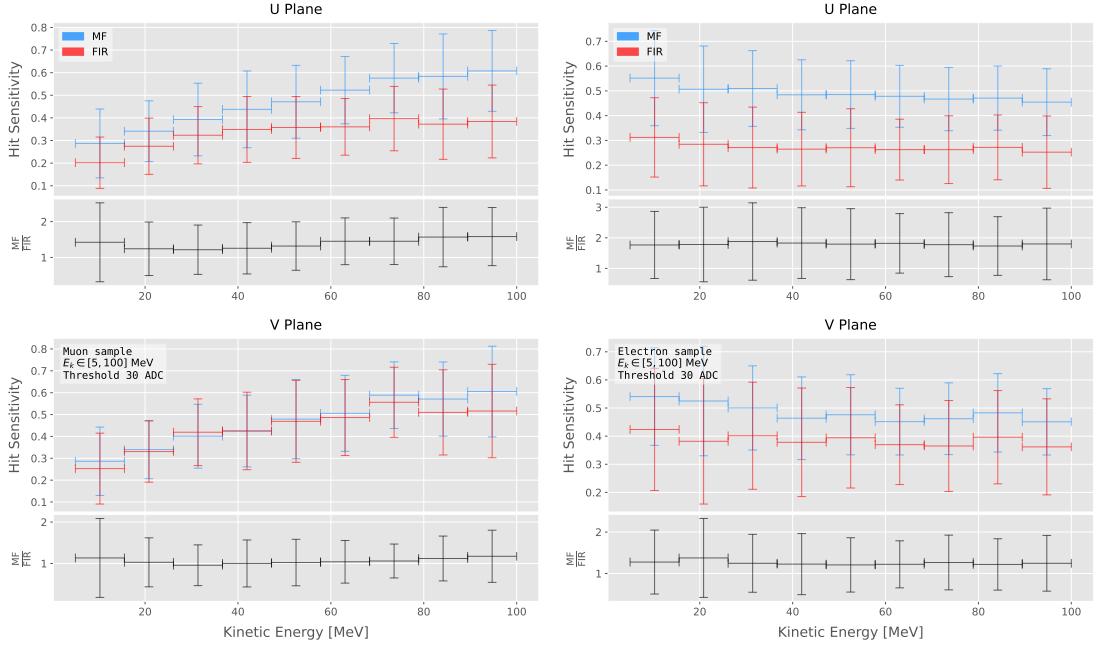


Figure 4.19: Dependence of the averaged hit sensitivity on the kinetic energy of the events for the matched filter (blue) and standard (red) hits, for the case of the muon (left panel) and electron (right panel) samples, separated between U (top plots) and V (bottom plots) induction wire planes. The top subplots contain the hit sensitivities for the two hit finder alternatives, while the bottom subplots show the ratio between the two. The horizontal lines sit at the mean value and represent the size of the energy bins, while the vertical error bars indicate one standard deviation around that mean value.

1419 the precision to be higher than 0.99 for both matched filter and standard cases.

1420 One can see that, in general, the improvements are better for the U than for the V
 1421 plane. While for the U channels I achieved a mean improvement of 50% and 80% for
 1422 muons and electrons respectively, the improvement in the V plane is stalled at 10% and
 1423 25%. Nevertheless, if I look at the sensitivities for the matched filter hits in both planes
 1424 one can see these have similar mean values for each energy bin, while on the contrary
 1425 for the standard hits the sensitivity remains relatively high for the V plane. This way, it
 1426 looks there was a less significant gain because the hit sensitivity was already high.

1427 Another interesting observation is the different behaviors for muons and electrons.
 1428 While hit sensitivity for muons grows significantly with energy, in the case of electrons
 1429 this slightly decreases the higher the kinetic energy of the event is. In any case, when it
 1430 comes to the improvement on the sensitivities, this remains almost constant in all cases.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

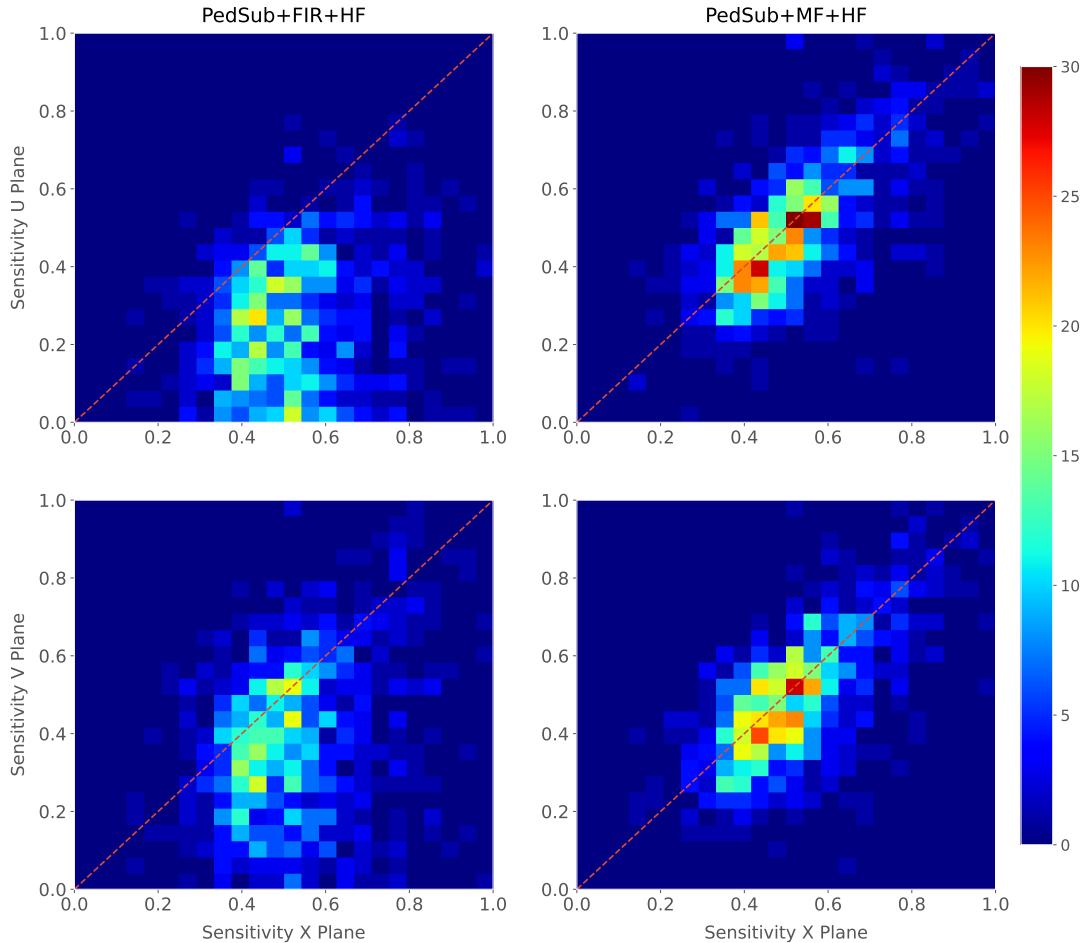


Figure 4.20: Distributions of the hit sensitivity in the U (top panels) and V (bottom panels) planes versus the hit sensitivity in the X plane, both for the standard hits (left panels) and the matched filter hits (right panels), in the case of the electron sample and a threshold of 30 ADC.

Furthermore, we can look at how the concurrence of hits between the different wire planes has changed. For any given event, I expect to have a similar number of hits in the three planes. As the ionisation electrons need to cross the U and V planes prior to reach the collection plane X they will induce current in those wire planes. A way to check the concurrence of hits across planes is looking at the relation between the hit sensitivities for each individual event. One cannot expect the sensitivities to be exactly equal across planes, but ideally they should be normally distributed around the diagonal.

Fig. 4.20 shows the hit sensitivity in the U (top panels) and V (bottom panels) planes versus the hit sensitivity in the X plane, for the case of the standard hits (left

4.5. MONTE CARLO STUDIES

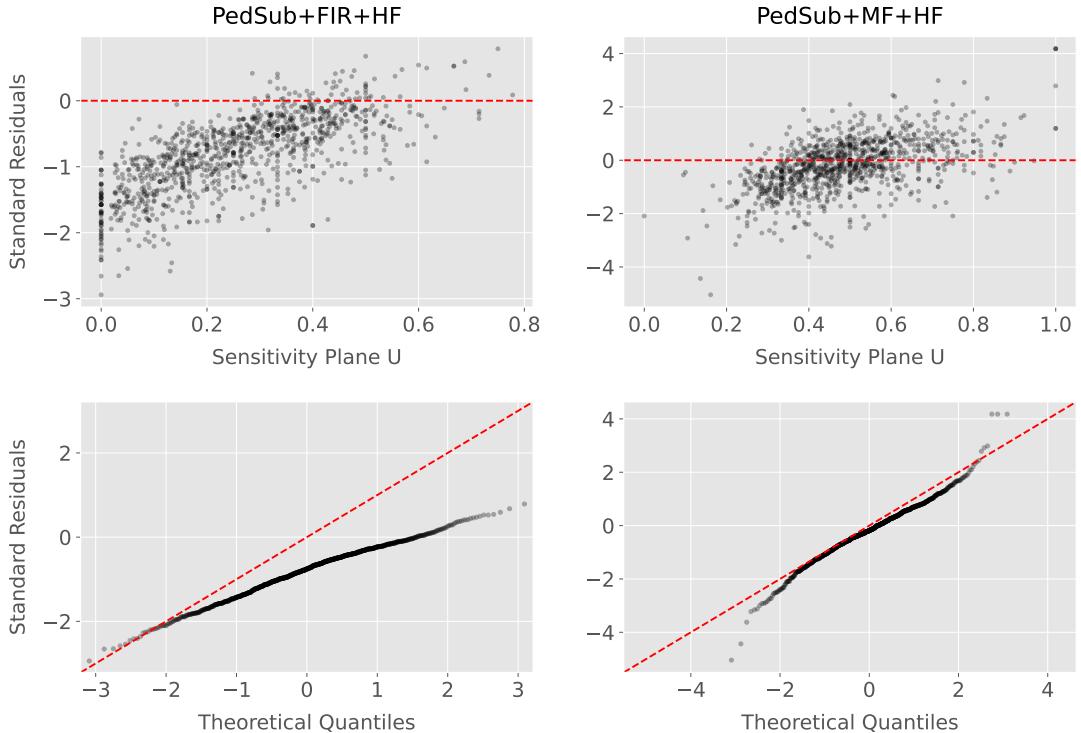


Figure 4.21: Top panels: standard residual plots of the hit sensitivities between the X and U planes. Bottom panels: quantile-quantile plots of the hit sensitivity standard residuals between the X and U planes. In all cases, the left panel corresponds to the standard hits while the right panel represents the matched filter case, all from the electron sample with a 30 ADC threshold.

1440 panels) and the matched filter hits (right panels). All plots were generated for the
 1441 electron sample and a threshold of 30 ADC. From these one can see a clear trend,
 1442 when I use the standard hit finder chain the sensitivities in the induction planes are
 1443 systematically lower than the hit sensitivity in the X plane, i.e. most of the points sit
 1444 below the diagonal (red dashed line). In contrast, when the matched filters are applied,
 1445 the majority of the events are distributed around the diagonal. This points out that the
 1446 concurrence of hits across planes has improved.

1447 To exemplify the improvement I obtained, one can consider the residuals of the hit
 1448 sensitivities for the X and U planes. Assuming the diagonal hypothesis, i.e. given a
 1449 dataset of the form (x, y) for any x I take the predicted y value to be equal to the value
 1450 of x , I can compute the standard residuals for the hit sensitivities in U given the ones for
 1451 X . In Fig. 4.21 (top panels) I show these standard residuals against the corresponding

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

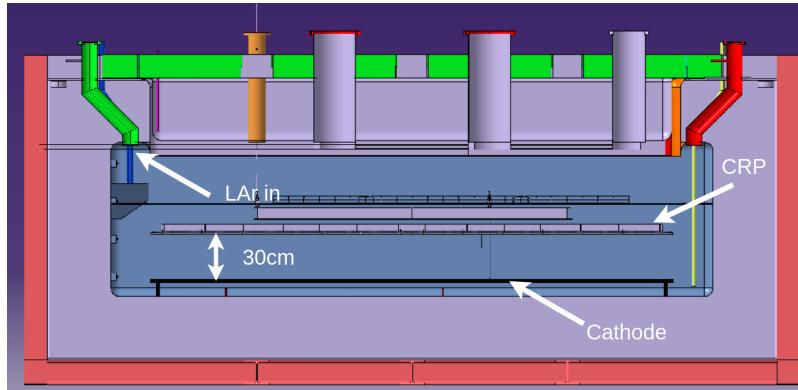


Figure 4.22: Schematic diagram of the vertical drift ColdBox setup at CERN.

values of the hit sensitivity in the U plane, for our electron sample with kinetic energy between 5 and 100 MeV. If I compare the scatter points in the case of the standard hits (left panel) and the matched filter hits (right panel), I see that the residuals of the standard hit finder case follow a certain pattern and their mean deviates from 0.

To see clearly if the residuals are normally distributed, in Fig. 4.21 (bottom panels) I plot the corresponding quantile-quantile plot for both the standard (left panel) and matched filter (right panel) standard residuals. One can clearly see that the points for the standard case follow a strongly non-linear pattern, suggesting that the residuals do not follow a normal distribution. In contrast, for the matched filter hits the points conform to a roughly linear path, implying that in this case the normality condition is fulfilled.

All these results hint at the fact that the concurrence of hits across the wire planes can be strengthened by applying the matched filters.

4.6 VD ColdBox data taking

Between February and April 2023 the vertical drift (VD) ColdBox setup at CERN, shown in Fig. 4.22, was recommissioned for cold electronics testing with CRP5. That provided another opportunity for testing the firmware trigger primitive generation in a real LArTPC. However, during the two run periods new software-related complications that were not observed in previous running conditions arose.

4.6. VD COLDBOX DATA TAKING

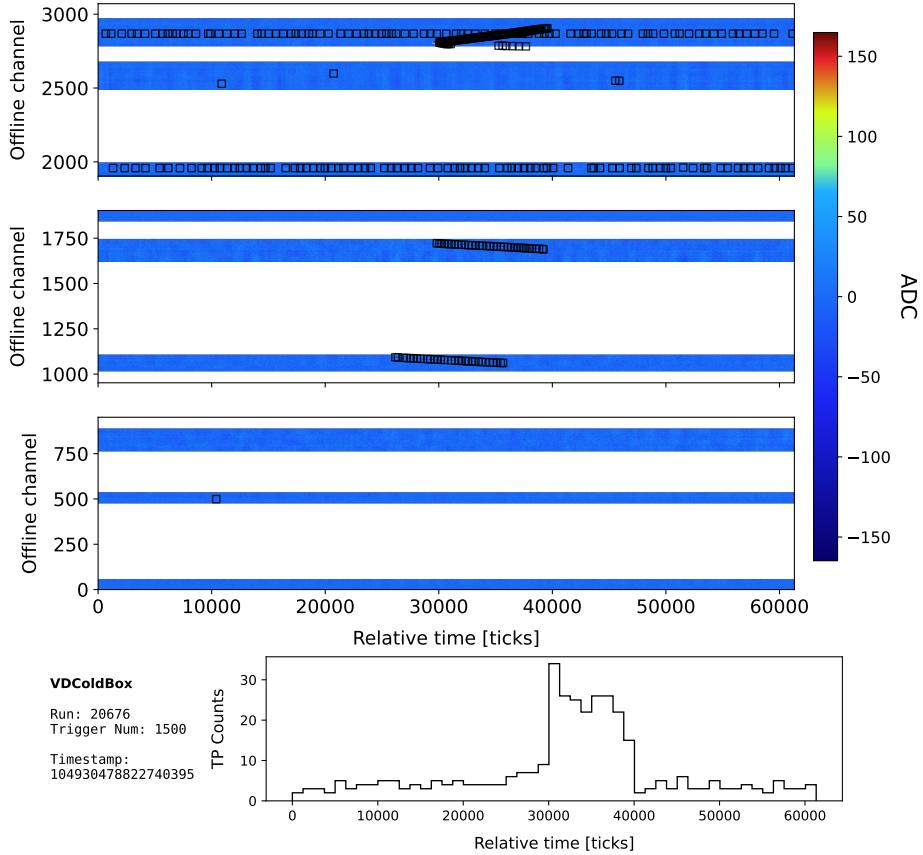


Figure 4.23: Event display of the data taken with the matched filter and HMA trigger at the VD ColdBox. The display shows the data from 3 ADC links for the full trigger window, with the black squares representing the produced TPs. The bottom panel represents the TP counts as a function of time in the trigger window.

1471 These prevented us from taking data with the whole system. As a palliative measure,
 1472 new configurations were made that allowed to run with TP generation enabled just in a
 1473 subset of the ADC links. With these workarounds, we managed to run with up to three
 1474 out of twelve ADC links and the horizontal muon trigger algorithm (HMA).

1475 Additionally, an alternative firmware version was prepared featuring the matched
 1476 filter coefficients optimised for the induction plane hit finding. The version of the filter
 1477 we used for the data taking is slightly different from the one of the previous studies, as
 1478 in this case we needed to apply the same filter coefficients to all channels irrespective of
 1479 the readout plane they come from. With this, we also managed to run with three ADC
 1480 links and the HMA trigger. Fig. 4.23 shows an example event display from the longest

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

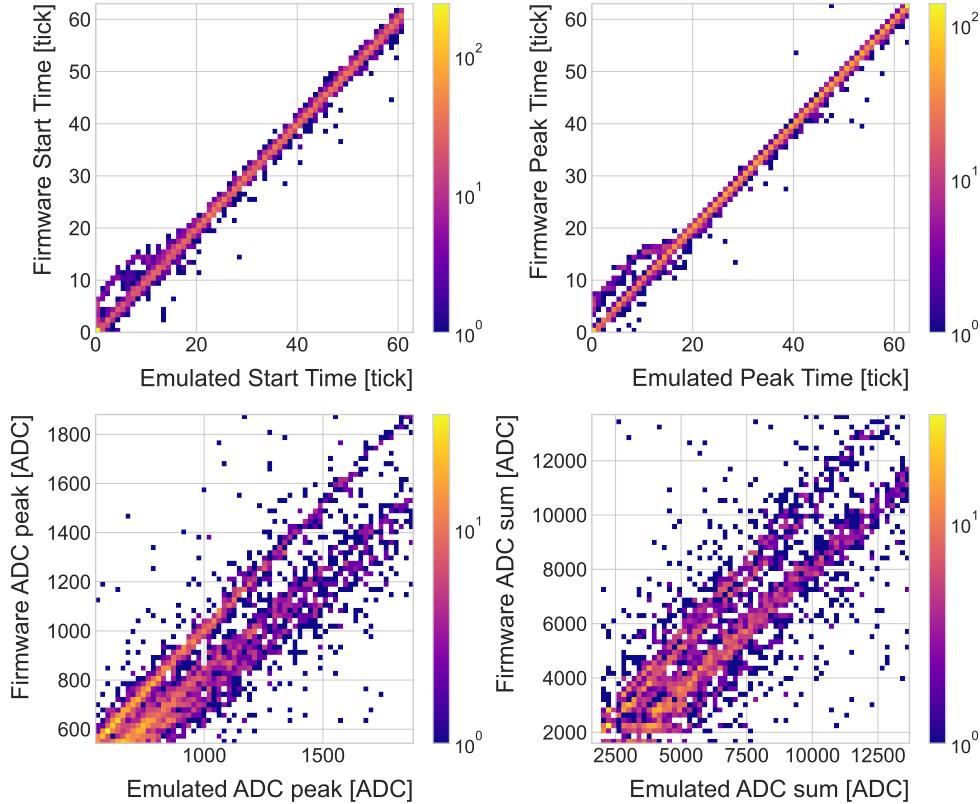


Figure 4.24: Comparison between firmware-produced and simulated TP quantities for a matched filter run at the VD ColdBox.

1481 run we recorded with the matched filter firmware.

1482 We used the recorded data, together with our standalone TPG simulation tool, to
 1483 perform comparisons between the firmware and simulated TPs. One such comparison
 1484 for a matched filter run can be seen in Fig. 4.24. The agreement achieved is within the
 1485 expectation, from what we have seen in previous samples.

1486 All the studies presented demonstrate the robustness of the matched filter approach
 1487 to form TP. I have used both ProtoDUNE-SP data and MC samples to assess its impact
 1488 on the S/N and TP production of the induction channels. Additionally, I have shown
 1489 that it is possible to run with it in a real detector environment, after the tests at the
 1490 VD ColdBox setup.

1492 DM searches with neutrinos from the Sun

1493 *He stepped down, trying not to look long at her, as if she were the Sun, yet he
1494 saw her, like the Sun, even without looking.*

1495 – Leo Tolstoy, *Anna Karenina*

1496 The idea of detecting neutrino signals coming from the Sun’s core to probe DM
1497 is not new. The main focus of these searches has usually been high-energy neutrinos
1498 originated from DM annihilations into heavy particles [101–104], although recent studies
1499 have proposed to look at the low-energy neutrino flux arising from the decay of light
1500 mesons at rest in the Sun [105–108] previously thought undetectable.

1501 In this Chapter, I try to demonstrate the capability of DUNE to constrain different
1502 DM scenarios. I used the neutrino fluxes arising from DM annihilations in the core
1503 of the Sun to compute the projected limits that DUNE would be able to set on the
1504 annihilation rates in the Sun and the DM scattering cross sections.

1505 **5.1 Gravitational capture of DM by the Sun**

1506 The Sun and the centre of the Earth are possible sources of DM annihilations, specially
1507 interesting because of their proximity. Their gravitational attraction ensured the capture
1508 of DM from the local halo through repeated scatterings of DM particles crossing them.
1509 Only neutrinos produced from DM annihilations can escape the dense interior of these
1510 objects. Therefore, neutrino telescopes are the most useful experimental layouts to
1511 pursue DM searches from their cores.

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

1512 The neutrino flux from DM annihilations inside the Sun depends on the DM capture
 1513 rate, which is proportional to the DM scattering cross section, and the annihilation rate,
 1514 which is proportional to the velocity-averaged DM annihilation cross-section. The total
 1515 number of DM particles inside the Sun follows the Boltzmann equation [105]:

$$\frac{dN_{DM}}{dt} = C_{\odot} - A_{\odot}N_{DM}^2, \quad (5.1)$$

1516 where C_{\odot} and A_{\odot} are the total Sun DM capture and annihilation rates respectively.
 1517 In this expression I neglected the evaporation term, proportional to N_{DM} , which only
 1518 contribute for $m_{DM} \lesssim 4$ GeV [109]. As the current threshold of neutrino telescopes
 1519 is a few GeV, this region falls below the probed range but can be important in future
 1520 low-energy projects.

1521 This equation has an equilibrium solution:

$$N_{DM}^{eq} = \sqrt{\frac{C_{\odot}}{A_{\odot}}}, \quad (5.2)$$

1522 which represents the amount of DM inside the Sun if the capture and annihilation have
 1523 reached equilibrium. As the Sun is approximately 4.6 Gyr old, it is usually assumed that
 1524 equilibrium has been achieved. Therefore, the anomalous neutrino flux from the Sun
 1525 would only depend on the DM scattering cross section, enabling us to set limits on this
 1526 quantity. If one does not assume equilibrium, some assumptions on the DM annihilation
 1527 cross section are necessary to extract predictions from neutrino signals.

1528 Here, I am going to consider three possible scenarios for the DM interactions: DM
 1529 scattering off electrons, spin-dependent (SD) and spin-independent interactions off nuclei.
 1530 For the case of these last two, the cross sections will be given in terms of the SD and
 1531 SI elastic scattering DM cross section off protons (assuming that DM interactions off
 1532 protons and neutrons are identical), σ_p^{SD} and σ_p^{SI} , as [105, 110]:

$$\sigma_i^{SD} = \left(\frac{\tilde{\mu}_{A_i}}{\tilde{\mu}_p} \right)^2 \frac{4(J_i + 1)}{3J_i} |\langle S_{p,i} \rangle + \langle S_{n,i} \rangle|^2 \sigma_p^{SD}, \quad (5.3)$$

5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

$$\sigma_i^{\text{SI}} = \left(\frac{\tilde{\mu}_{A_i}}{\tilde{\mu}_p} \right)^2 A_i^2 \sigma_p^{\text{SI}}, \quad (5.4)$$

where $\tilde{\mu}_{A_i}$ is the reduced mass of the DM-nucleus i system, $\tilde{\mu}_p$ is the reduced mass of the DM-proton system, A_i and J_i the mass number and total angular momentum of nucleus i and $\langle S_{p,i} \rangle$ and $\langle S_{n,i} \rangle$ the expectation value of the spins of protons and neutrons averaged over all nucleons, respectively (see Ref. [111] for a review on spin expectation values).

Since the Sun is mainly composed of Hydrogen, the capture of DM from the halo is expected to occur mainly through spin-dependent scattering. However, since the spin-independent cross section is proportional to the square of the atomic mass, heavy elements can contribute to the capture rate (even though they constitute less than 2% of the mass of the Sun). Heavy elements can also contribute to the spin-dependent cross section if the DM has also momentum-dependent interactions.

DM particles can get captured by the Sun if after repeated scatterings off solar targets their final velocity is lower than the escape velocity of the Sun. In the limit of weak cross sections, this capture rate can be approximately written as [110]:

$$C_{\odot}^{\text{weak}} = \sum_i \int_0^{R_{\odot}} dr \ 4\pi r^2 \int_0^{\infty} du_{\chi} \ \frac{\rho_{\chi}}{m_{\chi}} \frac{f_{v_{\odot}}(u_{\chi})}{u_{\chi}} \omega(r) \int_0^{v_e(r)} dv \ R_i^-(\omega \rightarrow v) |F_i(q)|^2, \quad (5.5)$$

where the summation extends over all possible nuclear targets. In this expression, R_{\odot} is the radius of the Sun, ρ_{χ} is the local DM density, m_{χ} the mass of the DM particle, $f_{v_{\odot}}(u_{\chi})$ the DM velocity distribution seen from the Sun's reference frame, $R_i^-(\omega \rightarrow v)$ is the differential rate at which a DM particle with velocity v scatters a solar target of mass m_i to end up with a velocity ω and $|F_i(q)|$ is the nuclear form factor of target i .

The differential scattering rate takes a rather simple form when considering velocity-independent and isotropic cross sections. In that case, this quantity is given by [110, 112]:

$$R_i^-(\omega \rightarrow v) = \frac{2}{\sqrt{\pi}} \frac{\mu_{i,+}^2}{\mu_i} \frac{v}{\omega} n_i(r) \sigma_i \left[\chi(-\alpha_-, \alpha_+) + \chi(-\beta_-, \beta_+) e^{\mu_i(\omega^2 - v^2)/u_i^2(r)} \right], \quad (5.6)$$

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

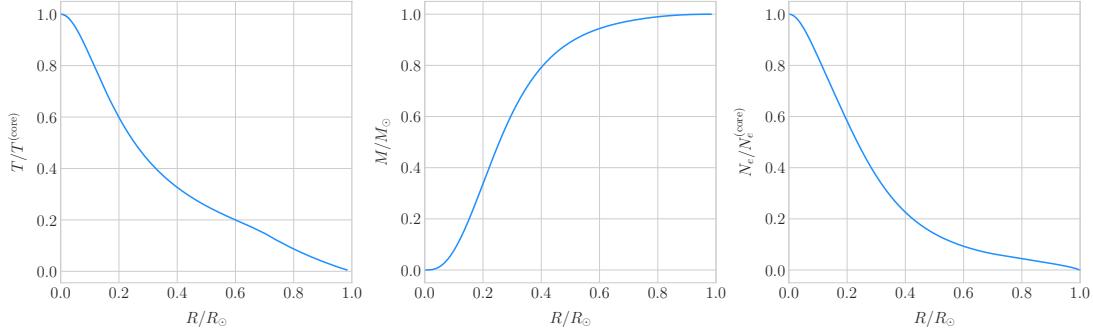


Figure 5.1: Input solar parameters used in the capture rate computation as a function of the solar radius, from left to right: temperature (with respect to the temperature at the core), mass (in solar masses) and electron number density (with respect to the electron density at the core). All quantities shown correspond to the standard solar model BS2005-OP [10].

1554 where μ_i is the ratio between the DM mass and the mass of target i , $\mu_{i,\pm}$ is defined as:

$$\mu_{i,\pm} \equiv \frac{\mu_i \pm 1}{2}, \quad (5.7)$$

1555 $n_i(r)$ is the density profile of target i in the solar medium, $u_i(r)$ is the most probable
1556 velocity of target i given by:

$$u_i(r) = \sqrt{\frac{2T_\odot(r)}{m_i}}, \quad (5.8)$$

1557 where $T_\odot(r)$ is the temperature of the Sun, the quantities α_\pm and β_\pm are defined as:

$$\alpha_\pm \equiv \frac{\mu_{i,+}v \pm \mu_{i,-}\omega}{u_i(r)}, \quad (5.9)$$

$$\beta_\pm \equiv \frac{\mu_{i,-}v \pm \mu_{i,+}\omega}{u_i(r)}, \quad (5.10)$$

1558 and the function $\chi(a, b)$ is a Gaussian integral of the form:

$$\chi(a, b) \equiv \int_a^b dx e^{-x^2}. \quad (5.11)$$

1559 Finally, if one assumes the DM halo velocity distribution in the galactic rest frame
1560 to be a Maxwell-Boltzmann distribution, one can write the halo velocity distribution for

5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

1561 an observer moving at the speed of the Sun with respect to the DM rest frame as:

$$f_{v_\odot}(u_\chi) = \sqrt{\frac{3}{2\pi}} \frac{u_\chi}{v_\odot v_d} \left(e^{-\frac{3(u_\chi - v_\odot)^2}{2v_d^2}} - e^{-\frac{3(u_\chi + v_\odot)^2}{2v_d^2}} \right), \quad (5.12)$$

1562 where:

$$\omega^2 = u_\chi^2 + v_e(r)^2, \quad (5.13)$$

1563 is the DM velocity squared, v_\odot the relative velocity of the Sun from the DM rest frame
 1564 and $v_d \simeq \sqrt{3/2}v_\odot$ the velocity dispersion.

1565 For the case of strong scattering cross section, Eq. (5.5) ceases to be valid, as it
 1566 escalates indefinitely with the cross section. In that limit, the capture rate saturates to
 1567 the case where the probability of interaction is equal to one, which can be written as:

$$C_\odot^{\text{geom}} = \pi R_\odot^2 \left(\frac{\rho_\chi}{m_\chi} \right) \langle v \rangle \left(1 + \frac{3}{2} \frac{v_e^2(R_\odot)}{v_d^2} \right) \xi(v_\odot, v_d), \quad (5.14)$$

1568 where $v_d = \sqrt{8/3\pi}v_d$ is the mean velocity in the DM rest frame and the factor $\xi(v_\odot, v_d)$
 1569 accounts for the suppression due to the motion of the Sun:

$$\xi(v_\odot, v_d) = \frac{v_d^2 e^{-\frac{3v_\odot^2}{2v_d^2}} + \sqrt{\frac{\pi}{6}} \frac{v_d}{v_\odot} (v_d^2 + 3v_e^2(R_\odot) + 3v_\odot^2) \operatorname{Erf} \left(\sqrt{\frac{3}{2}} \frac{v_\odot}{v_d} \right)}{2v_d^2 + 3v_e^2(R_\odot)}. \quad (5.15)$$

1570 Having these into account, one can write the total capture rate as a combination of
 1571 both contributions, allowing a smooth transition between the two, as:

$$C_\odot = C_\odot^{\text{weak}} \left(1 - e^{C_\odot^{\text{geom}} / C_\odot^{\text{weak}}} \right). \quad (5.16)$$

1572 I computed the capture rate from Eq. (5.16) in the case of interactions with
 1573 electrons. To do so, I used the standard solar model BS2005-OP [10]. Fig. 5.1 shows the
 1574 three parameters from the solar model that are needed for the computation, the solar
 1575 temperature (left panel), mass (central panel) and electron density (right panel) profiles.

1576 For the case of the interactions off nuclei, the computations are more convoluted

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

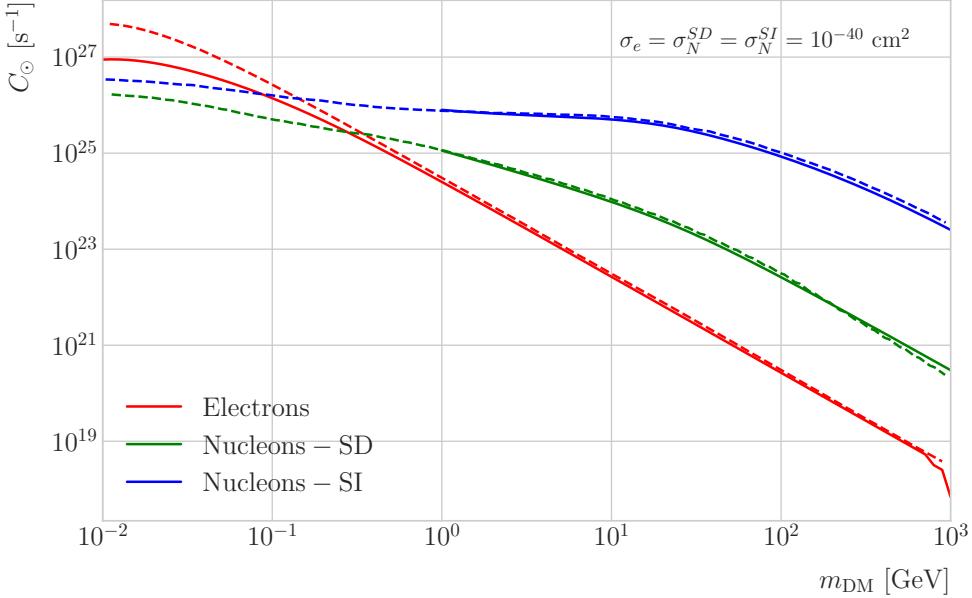


Figure 5.2: Capture rates as a function of the DM mass for the DM-electron interactions (red lines), SD DM-nucleons interactions (green lines), and SI DM-nucleons interactions (blue lines). Solid lines represent the values computed in this work while the dashed lines are the one given in Ref. [110]. All the rates are shown for a choice of scattering cross section of $\sigma_i = 10^{-40} \text{ cm}^2$.

as one needs to add up the contributions of the different most abundant nuclei in the Sun. Also, in contrast to the electron scenario where the form factor is trivially $|F_e(q)|^2 = 1$, for any nucleus i one would need to consider some appropriate nuclear density distribution (either a Gaussian approximation, a Woods-Saxon distribution, etc) which would complicate the calculations even further.

That is the reason why, at this stage of our study, I decided to take an alternative approach to the computation of the DM-nucleus capture rates. I used the **DarkSUSY** software, that allows us to compute these quantities performing a full numerical integration over the momentum transfer of the form factors. The default standard solar model used by **DarkSUSY** is BP2000¹ [113].

In Fig. 5.2 I show the results I obtained for the capture rates, for the case of interactions off electrons (red solid line), SD (green solid line) and SI (blue solid line)

¹This is what they say in their manual, but I fear it is somewhat outdated. It appears to me this model is relatively old and do not see why they are not using others like [10]. Maybe one can double-check in the code to make sure.

5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

interactions of nucleons. In all cases I used a value of the scattering cross sections of $\sigma_i = 10^{-40} \text{ cm}^2$. Note here one of the limitations of the DarkSUSY approach, one can not extend the computation below $m_{\text{DM}} = 1 \text{ GeV}$. Nevertheless, this is not something to worry about in this case, as I will discuss next. As a comparison, I added also the values computed in Ref. [110] (same color scheme, dashed lines). One can see there is good agreement between these and the DarkSUSY computation of the SD and SI interactions for $m_{\text{DM}} \geq 1 \text{ GeV}$. In this regime their computations also matches quite well our result for the electron capture rate. However, these start to differ significantly below $m_{\text{DM}} = 1 \text{ GeV}$, being their estimate up to a factor of 5 bigger than ours for low masses.

Let us comment briefly about the assumption I made before about not including an evaporation term in the Boltzmann equation. If I include this term in the equation (which will be proportional to the number of DM particles) the equilibrium solution takes the form:

$$N_{\text{DM}}^{eq} = \sqrt{\frac{C_{\odot}}{A_{\odot}}} \frac{1}{\kappa + \frac{1}{2} E_{\odot} \tau_{eq}}, \quad (5.17)$$

where E_{\odot} is the total evaporation rate, τ_{eq} is the equilibrium time in the absence of evaporation:

$$\tau_{eq} = \frac{1}{\sqrt{C_{\odot} A_{\odot}}}, \quad (5.18)$$

and κ is defined as:

$$\kappa \equiv \sqrt{1 + \left(\frac{E_{\odot} \tau_{eq}}{2} \right)^2}. \quad (5.19)$$

Now, it is easy to proof that in case evaporation dominates $\kappa \gg 1$ and therefore:

$$N_{\text{DM}}^{eq} \simeq \frac{C_{\odot}}{E_{\odot}}. \quad (5.20)$$

In contrast, if evaporation is irrelevant $\kappa \simeq 1$ and one recovers Eq. (5.2).

In this way, one can define the evaporation mass as the mass for which the number

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

1609 of DM particles in equilibrium approaches Eq. (5.20) at 10% level:

$$\left| N_{DM}^{eq}(m_{\text{evap}}) - \frac{C_{\odot}(m_{\text{evap}})}{E_{\odot}(m_{\text{evap}})} \right| = 0.1 N_{DM}^{eq}(m_{\text{evap}}). \quad (5.21)$$

1610 This can be regarded as the minimum testable mass one can reach using the annihilation
1611 products of the DM in the Sun.

1612 It was reported in Ref. [110] that, in the case of both SD and SI DM interactions
1613 off nuclei, this value ranges from 2 to 4 GeV depending on the specific scattering
1614 cross section value, compatible with the usual assumptions in the literature. What is
1615 interesting is the case of the electron capture. It was found that, when one applies a
1616 cutoff in the velocity distribution of the DM trapped in the Sun slightly below the escape
1617 velocity, the evaporation mass for the DM-electron interaction decreases remarkably. For
1618 a moderate choice of $v_c(r) = 0.9v_e(r)$ one gets an evaporation mass of around 200 to
1619 600 MeV. This possibility opens a region of the parameter space that could be tested
1620 with neutrino detectors.

1621 5.2 Neutrino flux from DM annihilations

1622 When WIMPs annihilate inside the Sun a flux of high-energy neutrinos is expected from
1623 heavy quarks, gauge bosons and $\tau^+\tau^-$ final states, which decay before losing energy in
1624 the dense solar medium, as they will produce a continuum spectra up to $E_{\nu} \sim m_{\chi}$ (in
1625 the case of direct annihilation to neutrinos one would have a line at $E_{\nu} = m_{\chi}$) [106].
1626 This kind of signal has been extensively studied in the literature, allowing to put strong
1627 limits on the SD WIMP-proton cross section for large m_{χ} . However, the number of
1628 high-energy neutrinos per WIMP annihilation is small and the spectrum depends on the
1629 unknown final state. Moreover, background rejection is easier for large m_{χ} but neutrinos
1630 with $E_{\nu} \gtrsim 100$ GeV are significantly attenuated by interactions in the Sun.

1631 Nevertheless, most WIMP annihilation final states eventually produce a low-energy
1632 neutrino spectrum. In this case one does not just consider the more massive final states
1633 but also annihilations into e^+e^- , $\mu^+\mu^-$ and light quarks [105]. In particular, light

5.3. COMPUTING LIMITS FROM SOLAR NEUTRINO FLUXES

mesons would be produced and stopped in the dense medium, thus decaying at rest and producing a monoenergetic neutrino signal. The decay-at-rest of kaons will produce a $E_\nu = 236$ MeV ν_μ while in the case of pions one would have a $E_\nu = 29.8$ MeV ν_μ . In practice only K^+ and π^+ contribute to these signals, as K^- and π^- are usually Coulomb-captured in an atomic orbit and get absorbed by the nucleus. There is also a low-energy neutrino signal coming from muon decays, which are produced in kaon or pion decays, leptonic decays of other hadrons and heavy leptons or even directly from WIMP annihilations, which can decay at rest and contribute to the previous low-energy neutrino flux with a well known spectrum below 52.8 MeV.

These monoenergetic MeV neutrinos were previously considered undetectable but, due to the large yield, the known spectra and the modern advances in the detector technology, these low-energy neutrino flux can be a good probe of the SD WIMP-proton cross-section in standard solar WIMP capture scenario, as it is sensitive to low WIMP masses and insensitive to the particular final state. A good place to look for these signals are next-generation neutrino experiments such as DUNE.

5.3 Computing limits from solar neutrino fluxes

In order to use the neutrino fluxes from DM annihilations in the Sun, the first thing I need to do is to determine the expected number of atmospheric background events, for a given exposure, after directionality selection has been applied. I can write this number as:

$$N_B = \eta_B \int d\Omega \int_{E_{min}}^{E_{max}} dE_\nu \frac{d^2\Phi_{atm}^\mu}{dE_\nu d\Omega} \times \left(A_{eff}^{(\mu)}(E_\nu) T \right), \quad (5.22)$$

where η_B is the background efficiency, E_{min} and E_{max} the minimum and maximum energies to integrate over, $d^2\Phi_{atm}^\mu/dE_\nu d\Omega$ the differential flux of atmospheric muon neutrinos, $A_{eff}^{(\mu)}$ is the effective area of DUNE to muon neutrinos and T is the exposure time. The effective area can be expressed as the product of the neutrino-nucleus scattering cross section and the number of nuclei in the fiducial volume of the detector. This way

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

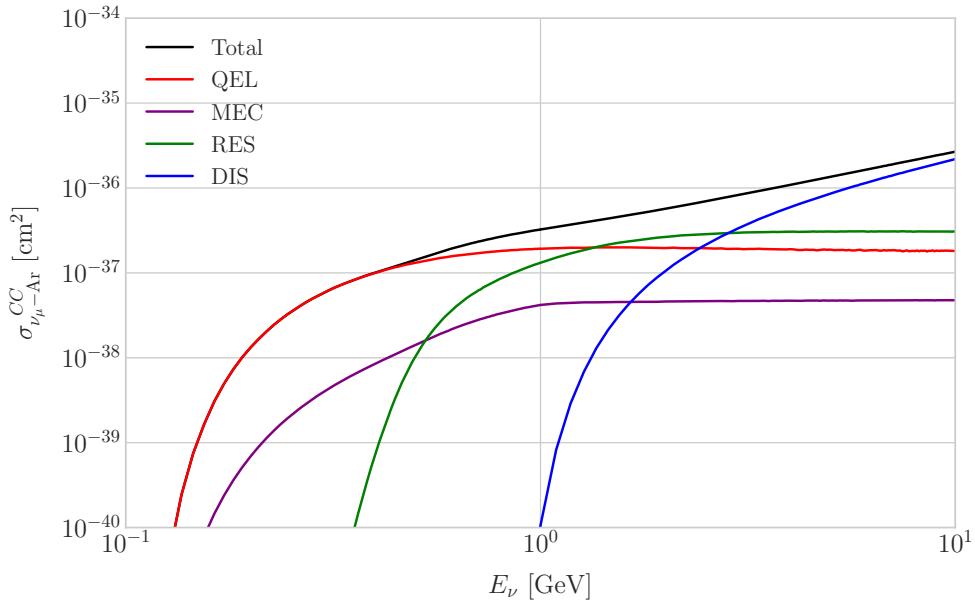


Figure 5.3: NuWro computed $\nu_\mu - {}^{40}\text{Ar}$ charged-current scattering cross section as a function of the neutrino energy. The black line shows to the total cross section, whereas the others correspond to the different contributions (in red quasi-elastic scattering, in green resonant pion exchange, in blue deep inelastic scattering and in purple meson exchange current).

1659 for DUNE I can write:

$$A_{eff}^{(\mu)}(E_\nu) = (6.0 \times 10^{-10} \text{ m}^2) \left(\frac{\sigma_{\nu-\text{Ar}}^{(\mu)}(E_\nu)}{10^{-38} \text{ cm}^2} \right) \left(\frac{M_{target}}{40 \text{ kT}} \right), \quad (5.23)$$

1660 where $\sigma_{\nu-\text{Ar}}^{(\mu)}$ is the $\nu_\mu - {}^{40}\text{Ar}$ charged-current scattering cross section. In Fig. 5.3 I
 1661 show the computed value of this cross section as a function of the neutrino energy E_ν ,
 1662 in the range of interest both for the atmospheric background and signal events. It was
 1663 computed using the NuWro Monte Carlo neutrino event generator [114], including the
 1664 charged-current contributions of the quasi-elastic scattering (red line), resonant pion
 1665 exchange (green line), deep inelastic scattering (blue line) and meson exchange current
 1666 (purple line).

1667 The background rejection will depend on the resolution of the detector and the
 1668 selection one applies on the events. A geometry argument can be used to estimate
 1669 the maximum background rejection one can achieve in this case, considering one can

5.3. COMPUTING LIMITS FROM SOLAR NEUTRINO FLUXES

1670 efficiently discriminate all events coming from a direction different from that of the
 1671 Sun. In that case, the optimal background efficiency will simply be the relative angular
 1672 coverage of the Sun. Taking the angular diameter of the Sun as seen from the Earth to
 1673 be 0.5° , I have:

$$\eta_B^{(opt)} \approx \frac{\pi \left(\frac{0.5}{2}\right)^2}{360 \times 180} \simeq 3.03 \times 10^{-6}. \quad (5.24)$$

1674 This value will give a very optimistic estimate of the number of background events.
 1675 However, it can be regarded as an lower limit, as it represents the best case scenario.

1676 In Fig. 5.4 I show the fluxes of atmospheric neutrinos at the Homestake mine during
 1677 solar minimum, taken from Ref. [115]. The values are averaged over the two angular
 1678 directions. In blue I have the flux of muon neutrinos while in red I indicate the flux
 1679 of electron neutrinos. Additionally, the dashed lines correspond to both antineutrino
 1680 species.

1681 Using these values for the muon neutrino and the corresponding total CC cross
 1682 section, one can compute the number of expected background events by integrating over
 1683 the given energy range (as in this case the angular integral is trivial). As for the energy
 1684 range to integrate over, I choose the range for DUNE specified in [73], $E_{min} = 10^{-1}$ GeV
 1685 and $E_{max} = 10$ GeV. Taking all these into account, I found the number of background
 1686 events to be:

$$N_B \simeq \eta_B \times (3.827 \times 10^4) \times \left(\frac{\text{exposure}}{400 \text{ kT yr}} \right). \quad (5.25)$$

1687 In order to estimate the sensitivity of DUNE to this kind signal, one can consider a
 1688 hypothetical data set where the number of observed neutrinos is taken to be the expected
 1689 number of background events rounded to the nearest integer, $N_{obs} = \text{round}(N_B)$ [116].
 1690 Now, if I assume that the number of signal and background events seen by DUNE are
 1691 given by Poisson distributions with means equal to the expected number of signal and
 1692 background events, N_S and N_B , one can denote by N_S^{90} to the number of expected
 1693 signal events such that the probability of having an experimental run with a number of
 1694 events greater than N_{obs} is 90%. This number can be obtained as the numerical solution

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

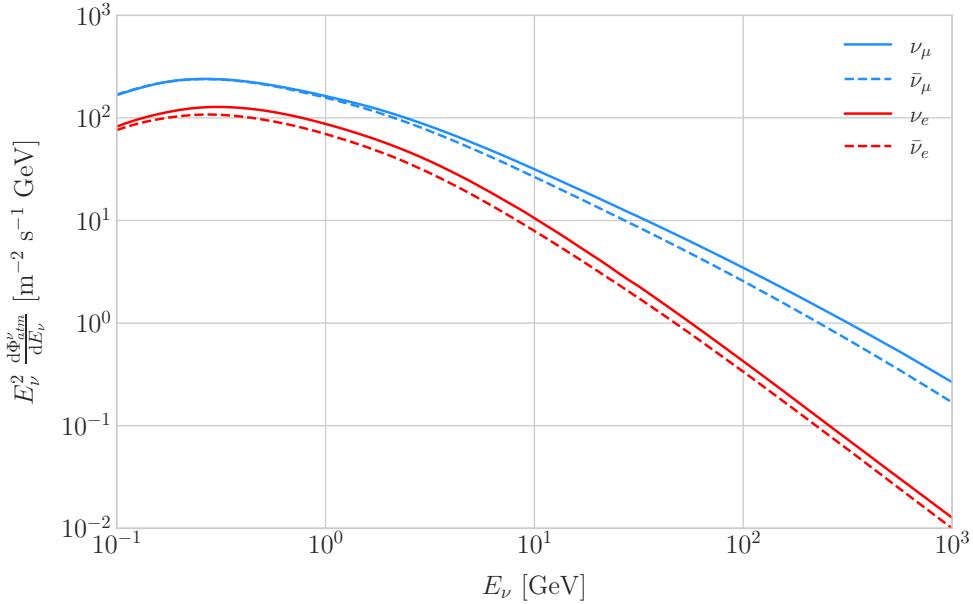


Figure 5.4: Expected atmospheric neutrino flux as a function of the neutrino energy E_ν at Homestake at solar minimum, taken from Ref. [115]. The blue solid (dashed) line correspond to muon neutrinos (antineutrinos) and the red solid (dashed) line correspond to electron neutrinos (antineutrinos).

1695 to the equation:

$$1 - \frac{\Gamma(N_{obs} + 1, N_S^{90} + N_B)}{N_{obs}!} = 0.9, \quad (5.26)$$

1696 where $\Gamma(x, y)$ is the upper incomplete gamma function.

1697 The number of signal events is related to the neutrino flux from DM annihilations in
1698 a similar way as the background events to the atmospheric neutrino flux. In this case I
1699 have:

$$N_S = \eta_S \Gamma_A^{eq} \int_{z_{min}}^{z_{max}} dz \frac{dN_\nu}{dAdN_A dz} \times (A_{eff}^\mu(z)T), \quad (5.27)$$

1700 where η_S is the signal efficiency, Γ_A^{eq} is the total annihilation rate of DM particles at
1701 equilibrium, $\Gamma_A^{eq} = A_\odot (N_{DM}^{eq})^2$, z_{min} and z_{max} the minimum and maximum relative
1702 energies to integrate over (in such a way that $z_{min,max} \leq E_{min,max}/m_{DM}$ for each m_{DM})
1703 and $dN_\nu/dAdN_A dz$ the muon neutrino flux per DM annihilation in the Sun.

1704 Knowing N_S^{90} one can use the relation in Eq. (5.27) to obtain $\Gamma_A^{eq,90}$ for different
1705 values of the DM mass. From there I can directly translate those values into the

5.4. EXAMPLE: KALUZA-KLEIN DARK MATTER

1706 upper limits for DUNE on the DM scattering cross sections, for a given exposure. The
1707 relation between the annihilation rate and the DM-nucleon cross section comes from the
1708 equilibrium condition through the solar DM capture rate. The details of the evolution
1709 of the number of DM particles inside the Sun and the computation of the capture rates
1710 are discussed in App. 5.1.

1711 5.4 Example: Kaluza-Klein Dark Matter

1712 Even though there are plenty of BSM theories which provide viable dark matter
1713 candidates, Kaluza-Klein type of models [117, 118] within the universal extra dimensions
1714 (UED) paradigm naturally predict the existence of a massive, stable particle that can
1715 play the role of the dark matter. In the UED scenario all the SM fields can propagate
1716 in one or more compact extra dimensions [119], as opposed to the idea of brane worlds
1717 [120, 121], where just gravity can propagate in the bulk while SM particles live at fixed
1718 points.

1719 Furthermore, in UED there is no violation of the translational invariance along the
1720 extra dimensions, thus leading to degenerate KK modes masses and also the conservation
1721 of the KK number in the effective four dimensional theory. At loop level, radiative
1722 corrections and boundary terms shift the masses of the KK modes and break KK
1723 number conservation into a KK parity. As a result, this theory only contains interactions
1724 between an even number of odd KK modes and therefore the lightest among the first KK
1725 excitations will be stable. This particle is usually denoted as the lightest Kaluza-Klein
1726 particle (LKP) and its mass is proportional to $1/R$, being R the size of the extra
1727 dimension.

1728 A viable DM candidate needs to be electrically neutral and non-baryonic, therefore
1729 good candidates among the first Kaluza-Klein excitations would be the KK neutral
1730 gauge bosons and the KK neutrinos [122]. Another possible candidate is the first KK
1731 excitation of the graviton, which receives negligible radiate contributions and therefore
1732 has a mass almost equal to $1/R$, but it has been shown that the lightest eigenstate from

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

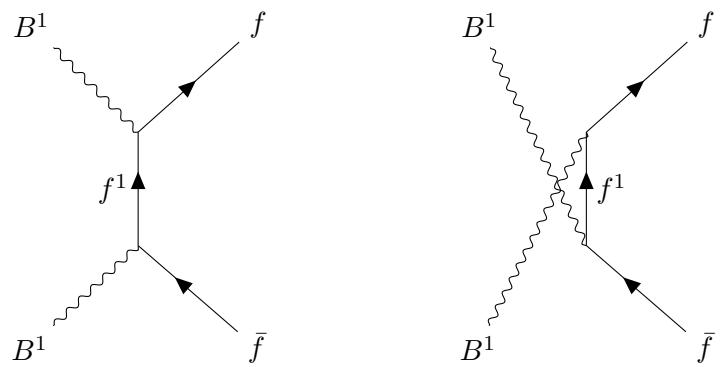


Figure 5.5: Feynman diagrams for $B^1 B^1$ annihilation into SM fermions.

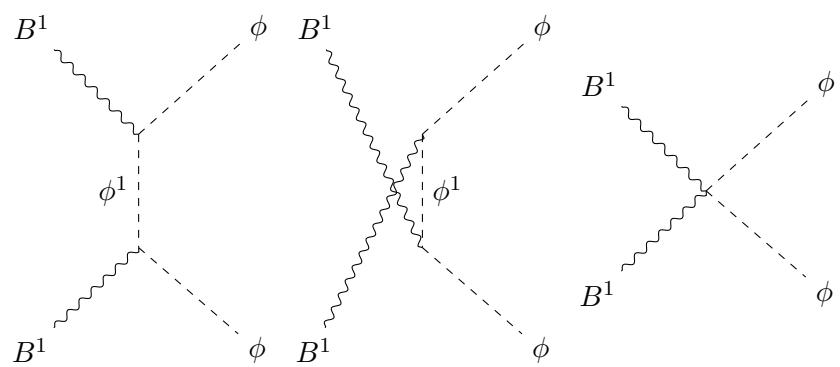


Figure 5.6: Feynman diagrams for $B^1 B^1$ annihilation into a Higgs boson pair.

5.4. EXAMPLE: KALUZA-KLEIN DARK MATTER

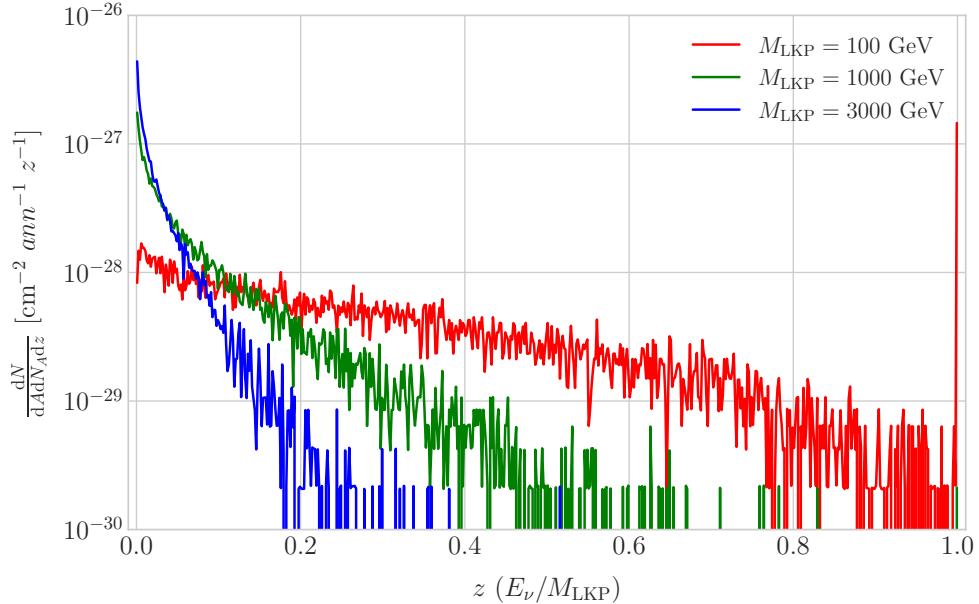


Figure 5.7: Computed spectra of muon neutrinos at the DUNE FD site from B^1 annihilations in the Sun for three different values of M_{LKP} , plotted in relative energy units for legibility.

the mixing of the gauge mass states (B^1, W_3^1) would be lighter, as B^1 and W_3^1 receive negative radiative corrections [123]. It is also understood that, when these corrections become sizeable, the eigenstates become approximately pure B^1 and W_3^1 states as the Weinberg mixing angle grows small with the KK number [123]. In that case, the LKP can be well-approximated as being entirely B^1 .

I need to compute the neutrino flux produced by the annihilations of the LKP in the core of the Sun, taking into account their propagation in the solar medium, as well as neutrino oscillations. To this end I used `WimpSim` [124, 125] to generate one million annihilation events in the Sun over a time span of four years and propagate them to the DUNE FD location ($44^\circ 20' \text{ N}, 103^\circ 45' \text{ W}$), for different values of M_{LKP} . In Fig. 5.7 I show the obtained muon neutrino spectra arriving to the detector from LKP annihilations in the Sun, per unit area and per annihilation, plotted in relative energy units for different values of the mass. As one could expect the spectra get steeper the higher is the mass, due to the absorption of high-energy neutrinos in the solar medium. Also, one can see the peak at $z = 1$ due to the direct annihilation into

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

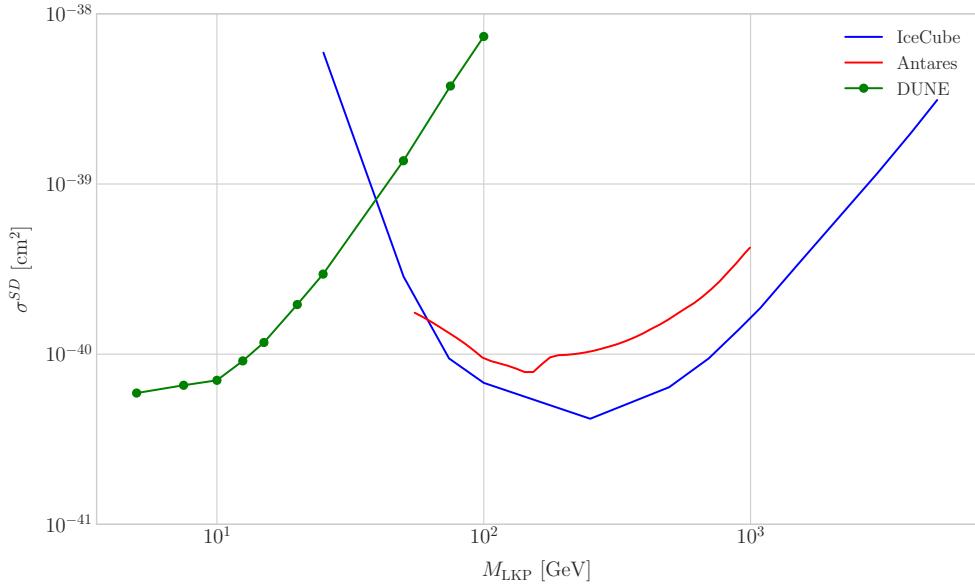


Figure 5.8: Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent B^1 -proton scattering cross section as a function of M_{LKP} (green dots). I also show the previous limits from IceCube [126] (blue line) and Antares [127] (red line) on the LKP cross section. The shaded area represents the disfavoured region (at 95% confidence level) on the mass of the LKP from LHC data [128].

1748 neutrinos $\chi\chi \rightarrow \nu\bar{\nu}$.

1749 Now, one can estimate the sensitivity of DUNE to this particular model by using
 1750 the methods I previously discussed. To begin with, I will use the optimistic estimation
 1751 of the background efficiency in Eq. (5.24) to get our upper bound. Using it, one can
 1752 directly compute the number of expected background events to be $N_B = 0.1101$ for an
 1753 exposure of 400 kT yr. Then, Eq. (5.26) give us a value of $N_S^{90} = 2.20$ for the 90%
 1754 exclusion number of expected signal events. By using the NuWro generated cross sections
 1755 and the computed neutrino fluxes from B^1 annihilations in the Sun I can estimate the
 1756 limits on the SD and SI DM-nucleus cross section using the relation in Eq. (5.2) and
 1757 the capture rates I computed with DarkSUSY.

1758 In Fig. 5.8 I show the projected sensitive for DUNE on the spin-dependent B^1 -
 1759 proton scattering cross section versus the mass of the DM particle, for a exposure of
 1760 400 kT yr (green dots). I also include the previous results from IceCube [126] (blue line)
 1761 and Antares [127] (red line). The shaded area represents the disfavoured region from

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

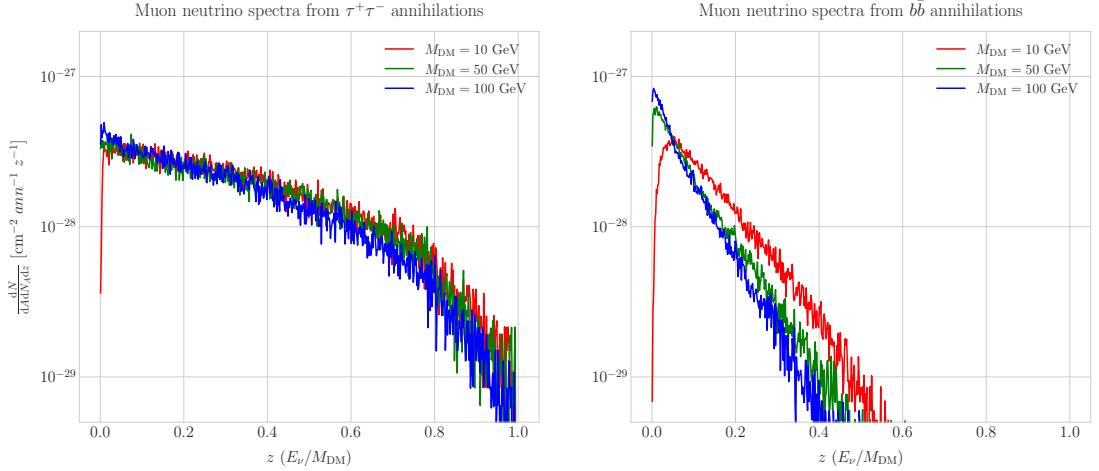


Figure 5.9: Computed spectra of muon neutrinos at the DUNE FD site from $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel) annihilations in the Sun for the DM masses $m_{\text{DM}} = 10 \text{ GeV}$ (red line), 50 GeV (green line) and 100 GeV (blue line), plotted in relative energy units.

1762 combined searches for UED by ATLAS and CMS [128].

1763 From the experimental point of view, this estimation lacked a detailed simulation of
 1764 the detector response and thus this must be consider as a mere optimistic sensitivity
 1765 computation. However, it shows the potential of DUNE to constrain this kind of exotic
 1766 scenarios, showing the region where it will be in a position to compete with other neutrino
 1767 telescopes. A more detailed analysis is needed if I am to make a realistic estimation.
 1768 Even though the region of the parameter space where DUNE would be sensitive to this
 1769 particular model is quite constrained by collider searches [128] and other rare decay
 1770 measurements [129, 130], it still constitutes an alternative indirect probe.

1771 5.5 High energy DM neutrino signals

1772 To have better estimates on the capability of the DUNE FD to constrain the parameter
 1773 space of DM using solar neutrino fluxes, I need to start accounting for the detector
 1774 resolution effects and the topologies of the different signatures. As a starting point, I
 1775 will focus on specific annihilation channels. For the case of DUNE, the relevant ones
 1776 are mainly the hard channels $\tau^+\tau^-$ and $\nu\bar{\nu}$ and the soft channel $b\bar{b}$. These are the open
 1777 annihilation channels for relatively low mass WIMPs that will actually give neutrino

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

1778 fluxes. Other channels, like W^+W^- and ZZ , are open for more massive WIMPs, but
1779 those will produce usually a higher energy neutrino flux that will be out of reach for
1780 DUNE (usually the maximum neutrino energy is taken to be $E_{max} = 10$ GeV).

1781 In Fig. 5.9 I show the `WimpSim` [124, 125] generated muon neutrino spectra at the
1782 DUNE FD location ($44^\circ 20' N$, $103^\circ 45' W$) from $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel)
1783 annihilations in the core of the Sun, for different DM masses. Here, one can clearly see
1784 the meaning of the previous distinction between hard and soft channels. For the same
1785 DM mass value, the muon neutrino spectrum from the $\tau^+\tau^-$ channel is more flat and
1786 reaches higher energies than the one from the $b\bar{b}$ channel, which drops faster.

1787 In this case, I prepared two sets of files, one for $\tau^+\tau^-$ and the other for $b\bar{b}$, for DM
1788 masses in the range from 5 to 100 GeV (actually for $b\bar{b}$ the first mass point I took is
1789 7.5 GeV, as a WIMP with $m_{DM} = 5$ GeV can not kinematically self annihilate into $b\bar{b}$).
1790 Then, I prepared the `WimpSim` output fluxes in a specific way to use them as inputs to
1791 `NuWro`, which simulates the neutrino interaction with the argon.

1792 Because `WimpSim` outputs an event list together with the fluxes, I can use the former
1793 to generate the events. The direction of these is given in terms of the azimuth and
1794 altitude angles viewed from the specified location, so first I need to convert these into the
1795 DUNE FD coordinates. Once I have done it, each event can be processed with `NuWro`.
1796 To increase the number of samples and optimise the computation time, I generate 100
1797 interactions (i.e. `NuWro` events) for each `WimpSim` event². I restrict the event generation
1798 to charged current interactions, but I allow all the different contributions to the CC
1799 cross section, i.e. quasielastic scattering (QEL), meson exchange current process (MEC),
1800 resonant pion production (RES) and deep inelastic scattering (DIS). I just take into
1801 account the CC contribution because I am only interested in final states with charged
1802 leptons, as we have better chances of reconstructing the kinematics of CC events.

1803 For the atmospheric fluxes I follow a similar procedure, only that this time I do not
1804 have a set of events but the fluxes binned in azimuth and altitude angles. This way, I

²This also solves a problem related with the generation of the neutrino interactions in `NuWro`, as if you only produce one event each time you launch `NuWro` it will always produce an interaction of the dominant interaction type for that particular energy.

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

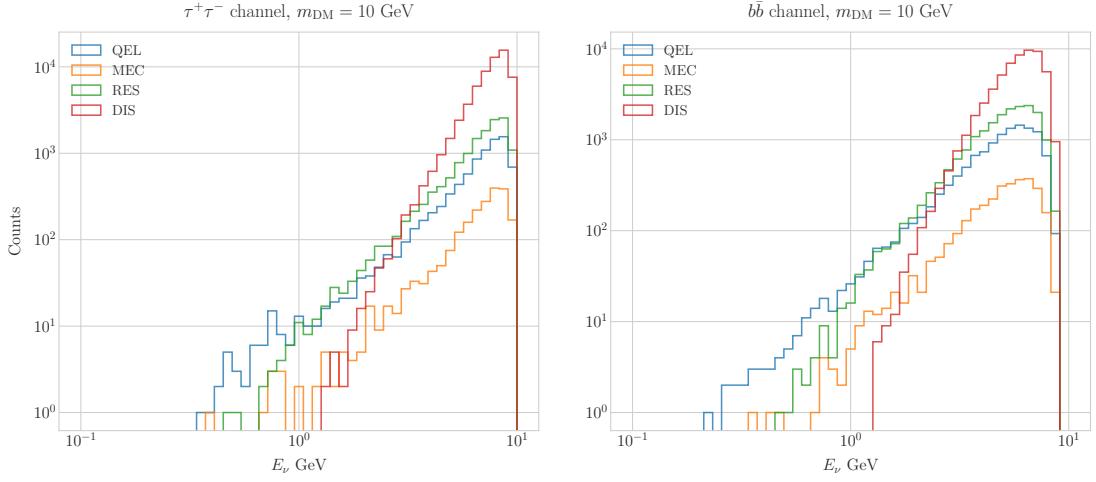


Figure 5.10: Distribution of the muon neutrino energies from the $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel) annihilation channels, for $m_{\text{DM}} = 10 \text{ GeV}$, separated by CC interaction type: QEL (blue), MEC (orange), RES (green) and DIS (red).

1805 transform these to DUNE coordinates and process the fluxes for each bin separated with
 1806 `NuWro`.

1807 At this point, I have two sets of events with different energies and final states.
 1808 In Fig. 5.10 one can see the distribution of the muon neutrino energies for the case
 1809 $m_{\text{DM}} = 10 \text{ GeV}$, both for the $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel) channels, separated
 1810 by interaction. One can clearly see that there are different energy regimes where the
 1811 primary interaction type is different. This leads to a plurality of event topologies,
 1812 therefore making it difficult to implement a general approach to the selection of events
 1813 in detriment of the background. As a way to proceed, I decided to split our samples,
 1814 based on the different interaction modes and contents of the final state, into a CC DIS
 1815 sample and a single proton CC QEL sample.

1816 **5.5.1 DIS events**

1817 To begin with, I consider the high energy part of the spectrum. In this region DIS events
 1818 dominate, i.e. interactions of the form $\nu_\mu + q_d(\bar{q}_u) \rightarrow \mu^- + q_u(\bar{q}_d)$. Therefore, our final
 1819 estates will contain a muon and a hadronic jet from the fragmentation of the outgoing
 1820 quark. As all these events have $E_\nu \gtrsim 1 \text{ GeV}$ the momentum transfer to the remnant

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

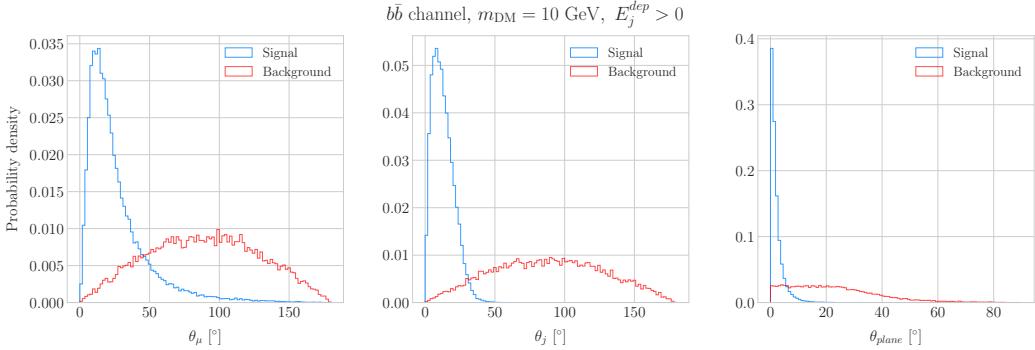


Figure 5.11: Distributions of θ_μ (left panel), θ_j (central panel) and θ_{plane} (right panel) for the $b\bar{b}$ sample with $m_{\text{DM}} = 10 \text{ GeV}$ (blue) and the atmospheric background (red).

nucleus is negligible, for this reason the neutrino energy can be effectively reconstructed just taking into account the momenta of the muon and the jet. This technique was successfully used in Ref. [131] to select monoenergetic DM solar neutrino events from $\nu\bar{\nu}$ annihilation channels.

Using momentum conservation one sees that the plane generated by the momenta of the muon and the jet needs to also contain the momentum of the neutrino. As we are interested in neutrinos coming from the Sun, the momentum of the neutrino can be regarded as known beforehand. This will allow us to define the angle of the outgoing muon and jet with respect to the incoming neutrino. Moreover, one can also use that information to reject poorly reconstructed jets, checking for deviations of these from the momentum conservation plane.

To account for the limited angular resolution of the detector, I smeared the momenta of the muons and hadrons. In a liquid argon TPC muons are expected to be tracked with high precision, therefore I take the associated angular resolution to be 1° . In the case of jets, it is expected that for the hadrons dominating the cascade a detector like DUNE has an angular resolution between 1° to 5° [73], so I take the latter, more conservative, estimate.

As a first selection step, I will just take into account particles with kinetic energies above the detection threshold of DUNE. For muons and photons the specified threshold energy is 30 MeV, for charged pions 100 MeV and for other hadrons 50 MeV [73]. This

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

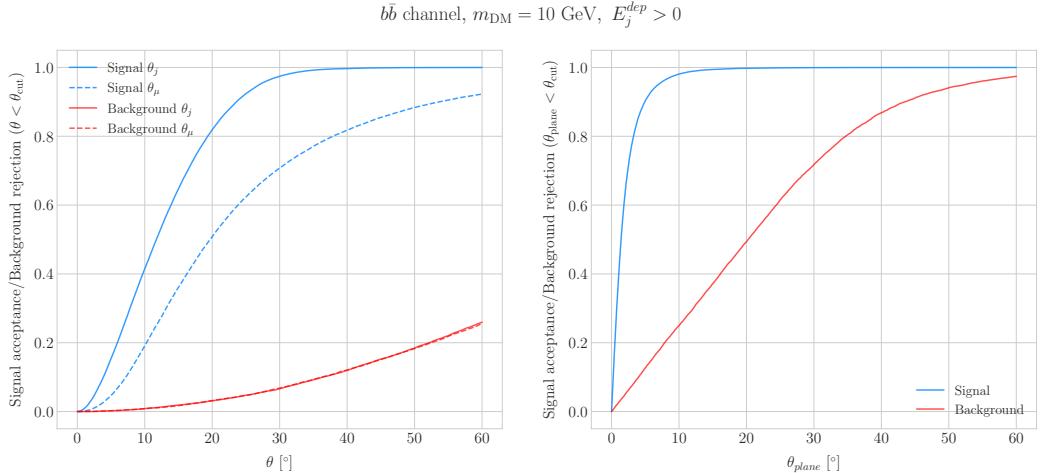


Figure 5.12: Left panel: signal efficiencies (blue lines) and background rejections (red lines) for events passing the cuts $\theta < \theta_{cut}$ for the jet (solid lines) and muon (dashed lines) angles. Right panel: signal efficiency (blue line) and background rejection (red line) for events passing the cut $\theta_{plane} < \theta_{cut}$ for the momentum conservation plane deviation.

way, if the outgoing muon in a certain event has an energy lower than the required threshold I will drop such event. For the case of hadrons and photons, I will only require to have at least one particle above the energy threshold, so then one can compute the jet momentum using the (smeared) momenta of the N particles above threshold as:

$$\vec{p}_j = \sum_{i=1}^N \vec{p}_i. \quad (5.28)$$

Additionally, I will also define an estimation of the deposited hadronic energy as:

$$E_j^{dep} = m_{^{39}\text{Ar}} - m_{^{40}\text{Ar}} + \sum_{i=1}^N \sqrt{|\vec{p}_i|^2 + m_i^2}. \quad (5.29)$$

This quantity is useful to select events with enough hadronic visible energy in the detector. For events where most of the hadronic energy is scattered across plenty of hadrons with individual energies below the detection threshold, this estimation will give $E_j^{dep} \leq 0$. In these cases it could be expected that the jet momentum is poorly reconstructed, and therefore I require events to pass the cut $E_j^{dep} > 0$.

For the events I can compute the angles for the muon and jet with respect to the

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

1852 incoming neutrino as:

$$\cos \theta_\mu = \hat{p}_\nu \cdot \hat{p}_\mu, \quad (5.30)$$

$$\cos \theta_j = \hat{p}_\nu \cdot \hat{p}_j, \quad (5.31)$$

1853 and the deviation from the momentum conservation plane as:

$$\sin \theta_{plane} = \left| \frac{\hat{p}_\mu \times \hat{p}_\nu}{|\hat{p}_\mu \times \hat{p}_\nu|} \cdot \hat{p}_j \right|. \quad (5.32)$$

1854 In Fig. 5.11 I show some distributions of these quantities for the case of the $b\bar{b}$ sample
1855 with $m_{DM} = 10$ GeV (blue histograms) and for the atmospheric backgrounds (red).
1856 In order to select the atmospheric events I followed the same criteria as for the signal
1857 events. However, because in the signal case I used the true direction of the neutrino
1858 as input, as it should be that of the Sun at that time and therefore known, in the
1859 atmospheric case I used a set of solar positions as our ansatz for the neutrino direction.
1860 From the distributions, one can see that the muon and the jet for the signal events are
1861 predominantly forward and also that the deviations from the momentum conservation
1862 plane are peaked at zero, as one should expect.

1863 Now, I can start applying cuts to maximise our signal selection efficiency while at
1864 the same time I try to minimise the amount of atmospheric background events passing
1865 the selection. To this end, I will need to find some lower and upper cuts for θ_j and
1866 θ_μ and an upper bound for θ_{plane} . In Fig. 5.12 I show how upper bound cuts in the
1867 different angular variables affect the signal efficiency (blue lines) and the background
1868 rejection (red lines). Notice that the signal efficiency behaves in a quite different way
1869 when I apply cuts in the jet and the muon angles. On the contrary, the cuts on both
1870 variables have a similar effect on the background rejection.

1871 In order to obtain the optimal set of cuts, I perform a multidimensional scan. I
1872 do this separately for the $\tau^+\tau^-$ and the $b\bar{b}$ samples. For each case, I scan the possible
1873 cuts for each mass point and then I take the mean value of the signal efficiency for

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

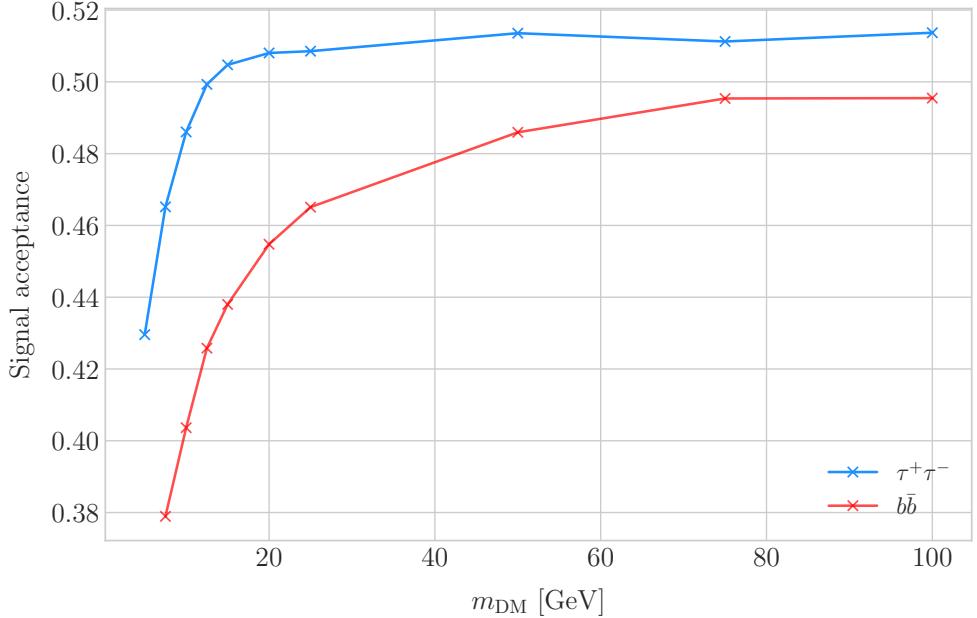


Figure 5.13: Signal efficiencies for the $\tau^+\tau^-$ (blue line) and $b\bar{b}$ (red line) DIS samples as functions of the DM mass, m_{DM} , obtained by applying the optimal angular cuts $\theta_\mu < 27^\circ$, $4^\circ < \theta_j < 26^\circ$ and $\theta_{\text{plane}} < 3.5^\circ$.

1874 each configuration, to get the mean efficiency for each set of cuts. I do a similar scan
 1875 for the atmospheric sample independently. Then, I take the sets of cuts such that
 1876 the background rejection achieved is greater than 99.8% and search for the one which
 1877 maximises the $\tau^+\tau^-$ and $b\bar{b}$ sample mean efficiencies. I found that with the cuts $\theta_\mu < 27^\circ$,
 1878 $4^\circ < \theta_j < 26^\circ$ and $\theta_{\text{plane}} < 3.5^\circ$ I get a background rejection of 99.80% while achieving
 1879 a 49.40% and 44.92% mean signal efficiencies for the $\tau^+\tau^-$ and $b\bar{b}$ signals respectively.

1880 In Fig. 5.13 I show the signal efficiencies as a function of the DM mass for the $\tau^+\tau^-$
 1881 (blue line) and the $b\bar{b}$ (red line) DIS events, after applying the cuts discussed above, as
 1882 well as the energy threshold and hadronic visible energy selections. One can see that
 1883 the efficiency grows with the mass, as annihilations of more massive DM particles will
 1884 produce a neutrino spectrum centered at higher energies, where DIS events dominate.
 1885 Notice also that the efficiency is higher for the $\tau^+\tau^-$ case at every mass point, as in
 1886 general this channel produces neutrinos at higher energies than the corresponding $b\bar{b}$
 1887 channel.

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

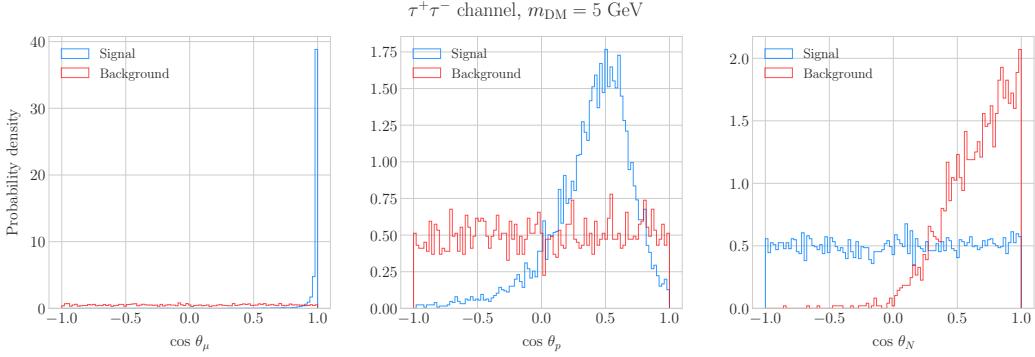


Figure 5.14: Distributions of $\cos \theta_\mu$ (left panel), $\cos \theta_p$ (central panel) and $\cos \theta_N$ (right panel) for the $\tau^+\tau^-$ QEL sample with $m_{\text{DM}} = 5 \text{ GeV}$ (blue) and the atmospheric background (red).

1888 5.5.2 Single proton QEL events

1889 Now, one can try to explore the low energy tail of the neutrino energy distributions. This
 1890 regime is dominated by the QEL interactions, i.e. events of the type $\nu_\mu + n \rightarrow \mu^- + p$.
 1891 In this case, as the typical energies are $E_\nu \lesssim 1 \text{ GeV}$, the momentum transfer to the
 1892 remnant nucleus is sizeable. Therefore, I can not make the approximation I did before
 1893 and assume that the momentum of the muon and the proton will give an adequate
 1894 estimation of the reconstructed neutrino energy.

1895 In any case, as before, I can take the direction of the incoming neutrino as known.
 1896 That way, one can estimate the energy of the neutrino as:

$$E_\nu^{reco} = E_\mu + E_p + m_{39\text{Ar}} - m_{40\text{Ar}}, \quad (5.33)$$

1897 and using momentum conservation I can write the momentum of the remnant nucleus
 1898 as:

$$\vec{p}_N = \hat{p}_\nu (E_\mu + E_p + m_{39\text{Ar}} - m_{40\text{Ar}}) - \vec{p}_\mu - \vec{p}_p. \quad (5.34)$$

1899 As in the previous case, I need to drop the events where the muon or the proton fall
 1900 below the kinetic energy detection threshold [73]. Also, I again apply a smearing to the
 1901 momenta of the particles, a 1% for muons and 5% for protons.

1902 Having done that, one can compute the following angular variables for our selected

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

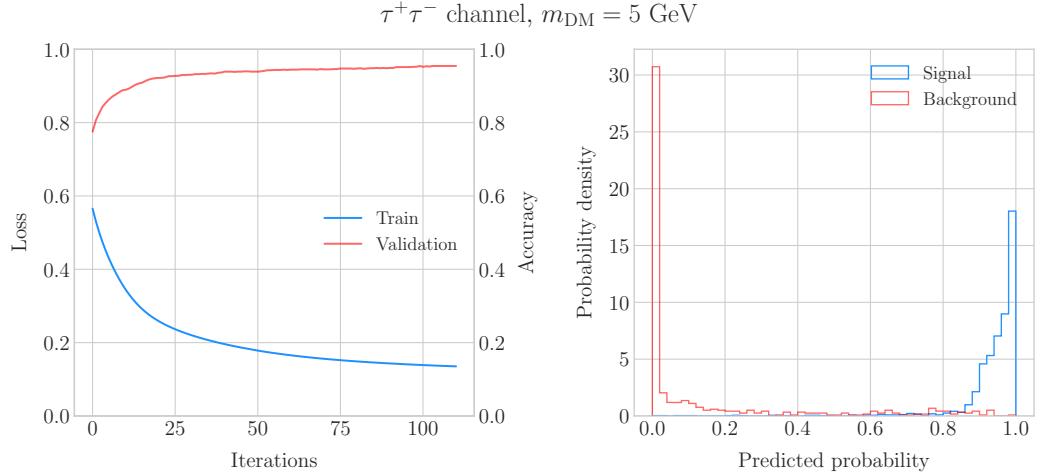


Figure 5.15: Left panel: value of the loss function for the training sample (blue line) and accuracy for the validation sample (red line) versus the number of iterations for the MLP classifier training. Right panel: distributions of the predicted probabilities assigned by the MLP classifier to the test sample for the $\tau^+\tau^-$ QEL signal with $m_{\text{DM}} = 5 \text{ GeV}$ (blue) and the atmospheric background (red).

1903 events:

$$\cos \theta_\mu = \hat{p}_\nu \cdot \hat{p}_\mu, \quad (5.35)$$

$$\cos \theta_p = \hat{p}_\nu \cdot \hat{p}_p, \quad (5.36)$$

$$\cos \theta_N = \hat{p}_\nu \cdot \hat{p}_N. \quad (5.37)$$

1904 Fig. 5.14 shows the distributions of these angular variables for the $\tau^+\tau^-$ QEL
 1905 sample with $m_{\text{DM}} = 5 \text{ GeV}$ (blue) and the atmospheric background (red). Again, for
 1906 the atmospheric events I used a random solar position as the ansatz for the incoming
 1907 neutrino direction. Notice that now, opposed to the DIS case where the signal had very
 1908 sharp distributions for the variables considered, the shapes of the angular distributions
 1909 for signal and background are not that much different.

1910 This effectively means that the usual approach of applying simple angular cuts would
 1911 not work as well as in the previous situation. Therefore, as a possible solution, I tried to
 1912 use a multilayer perceptron (MLP) classifier to separate between signal and background
 1913 events. Thus, the power of the hypothesis test will serve as an estimate of the signal

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

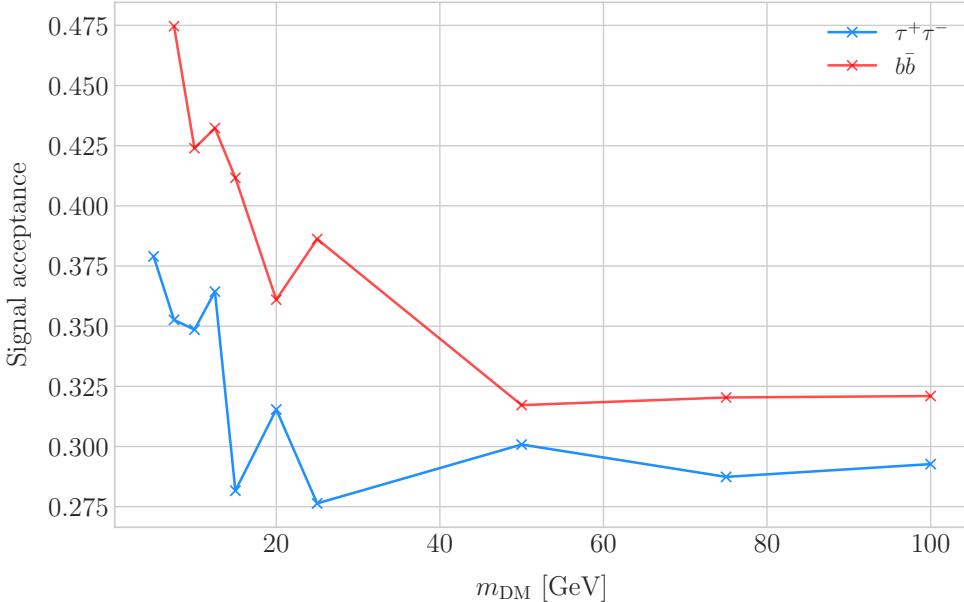


Figure 5.16: Signal efficiencies for the $\tau^+\tau^-$ (blue line) and $b\bar{b}$ (red line) single proton QEL samples as functions of the DM mass, m_{DM} , obtained by requiring a minimum predicted probability from the MLP classifier of 0.97 in order to achieve a background rejection greater than 99.8%.

efficiency, and in the same way one can take the size of the test to be our background rejection.

For each DM mass value and channel, as well as for the background sample, I divide our events into training, validation and test samples. The input variables for the classifier were the reconstructed neutrino energy from Eq. (5.33) and the angular variables defined in Eqs. (5.35 - 5.37). I used the MLP classifier implemented in `scikit-learn` [132], with a total of five hidden layers, the rectified linear unit activation function and adaptive learning rate. In order to account for fluctuations due to artifacts in the training process I repeated the training a thousand times for each sample, redefining each time the training, validation and test subsets, so one can take as our signal efficiency and background rejection the mean values of the powers and sizes of the tests.

The results of one of these training processes for the $\tau^+\tau^-$ QEL signal with $m_{\text{DM}} = 5$ GeV is shown in Fig. 5.15. On the left panel I show the loss function values (blue) and accuracy (red) at each iteration for the training and the validation samples respectively.

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

1928 The training stops either when the maximum number of iterations is reached (1000 in
 1929 this case) or when the accuracy for the validation sample reaches a certain tolerance
 1930 (I chose 10^{-4} as our tolerance). On the right panel I have the distributions for the
 1931 predicted probability by the model, separated in true signal (blue) and background
 1932 (red) events, for the test sample. One can see that both populations are well separated,
 1933 obtaining a power of 44.97% and a size of 0.17% when I require a predicted probability
 1934 greater than 0.97.

1935 Applying this criteria for each sample, I obtain the mean signal efficiencies shown in
 1936 Fig. 5.16. Notice that the efficiencies for the channel $\tau^+\tau^-$ (blue line) are consistently
 1937 lower than the ones for the $b\bar{b}$ channel (red line). This can be due to the fact that, for
 1938 each DM mass point, the neutrino spectrum coming from the $b\bar{b}$ annihilation channel is
 1939 centered at lower energies when compared to the $\tau^+\tau^-$ spectrum. This directly translates
 1940 into more low energy neutrinos undergoing QEL interactions, which give signals that
 1941 can be easily separated from the atmospheric background. This explanation also help us
 1942 understand why in both cases the signal acceptance drops when the DM mass increases.
 1943 In all cases, the background rejection took values between 99.8% to 99.9%. I will assume
 1944 a 99.8% background rejection value in all cases to keep our estimation conservative.

1945 5.5.3 Results

1946 In order to estimate the DM-nucleon cross section sensitivities in the present case I need
 1947 again to compute the expected number of background events. As I am now separating
 1948 events by interaction type Eq. (5.25) does not hold anymore, as in that case I integrated
 1949 over the total neutrino-argon cross section. In this instance, the expected background
 1950 events for DIS events is approximately given by:

$$N_B^{DIS} \simeq \eta_B^{DIS} \times (4.655 \times 10^3) \times \left(\frac{\text{exposure}}{400 \text{ kT yr}} \right), \quad (5.38)$$

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

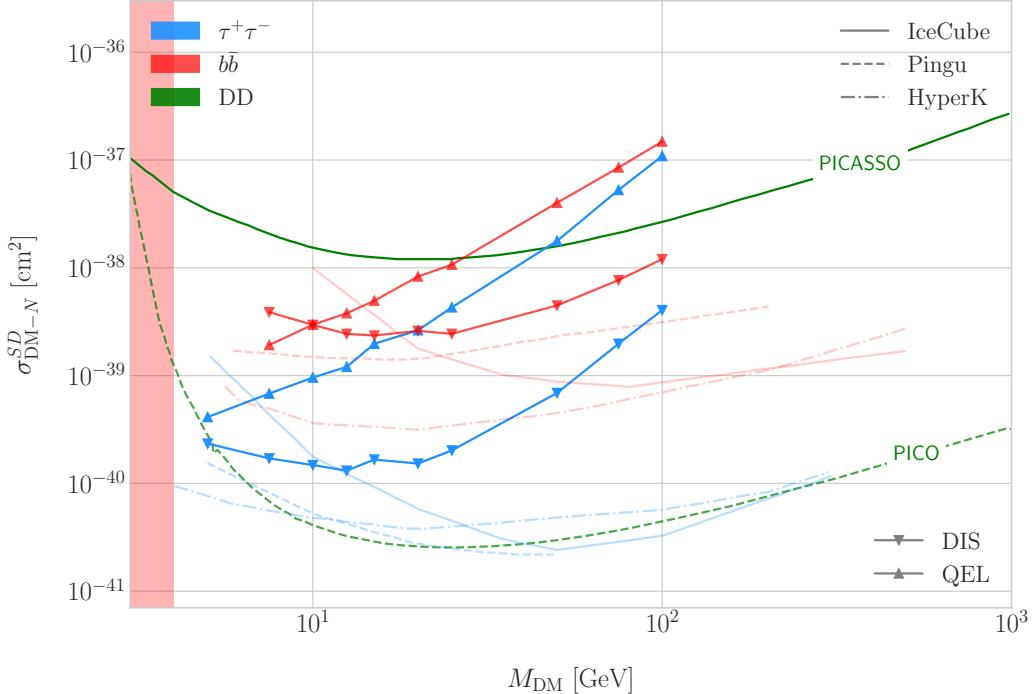


Figure 5.17: Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent DM-nucleon scattering cross section as a function of m_{DM} , for the annihilation channels $\tau^+\tau^-$ (blue) and $b\bar{b}$ (red) separated by interaction mode (up triangles denote DIS interactions whereas down triangles represent QEL interactions). I also show the previous limits from IceCube [133] (solid lines) and the projected sensitivities for Pingu [134] (dashed lines) and Hyper-Kamiokande [135] (dash-dotted lines), as well as the direct detection limits from PICASSO [136] (solid green line) and PICO-60 C₃F₈ [137] (dashed green line).

1951 whereas for QEL events we have:

$$N_B^{QEL} \simeq \eta_B^{QEL} \times (2.248 \times 10^4) \times \left(\frac{\text{exposure}}{400 \text{ kT yr}} \right). \quad (5.39)$$

1952 Now, using these together with Eqs. (5.26) and (5.27) one can obtain the 90% C.L.
1953 upper limit on the total annihilation rate at equilibrium for both kind of events. Then,
1954 applying the computed DM-nucleons capture rates I can translate these into limits on
1955 the DM-nucleon cross section by means of Eqs. (5.2), (5.5) and (5.6).

1956 Fig. 5.17 shows the obtained limits on the SD DM-nucleon cross section for DUNE,
1957 using the DIS (upward triangles) and QEL (downward triangles) events both for the $\tau^+\tau^-$ (blue)
1958 and the $b\bar{b}$ (red) samples, for an exposure of 400 kT yr. I also include the corresponding

5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

1959 current limits from IceCube [133] (solid lines), as well as the projected sensitivities
1960 of Pingu [134] (dashed lines) and Hyper-Kamiokande [135] (dash-dotted lines). For
1961 comparison, I also show the reported direct detection limits from PICASSO [136] (solid
1962 green line) and PICO-60 C₃F₈ [137] (dashed green line).

1963 Notice that, for most of the mass range, the limits one can set by using the DIS
1964 events are stronger than those of the QEL interactions, except for the low mass part
1965 of both the $\tau^+\tau^-$ and the $b\bar{b}$ curves where the QEL events dominate. In general, the
1966 expected sensitivity of DUNE for DM masses $\lesssim 25$ GeV surpasses the stronger current
1967 indirect limits. However, experiments like Hyper-Kamiokande are foreseen to have an
1968 overall better sensitivity in this kind of searches, as they have a bigger active volume
1969 and accept a broader energy range.

1970 A pending question is what happens when we add the RES and MEC charged-current
1971 interaction contributions. In that case it would probably be more convenient to split
1972 the samples by final state interaction topologies. Also, another necessary improvement
1973 would be adding a full detector simulation and reconstructions. This will also require
1974 considering the effect of poorly reconstructed events or final states containing neutral
1975 particles such that they mimic the desired topology at the reconstruction level.

1976 5.6 Example: Leptophilic Dark Matter

1977 In general, the capture rate of DM particles by the Sun via interactions with electrons is
1978 several orders of magnitude smaller than the capture via DM-nucleus scattering. Thus,
1979 it would be sub-leading even when nucleon capture is loop suppressed. As I showed in
1980 Fig. 5.2, the capture rate via scattering off electrons only surpasses the capture rates
1981 via DM-nucleons interactions for DM masses $\lesssim 100 - 500$ MeV.

1982 However, if one considers a model where DM-nucleon interactions are forbidden even
1983 at loop level, then electron interactions will be the sole contributor to DM capture in
1984 the Sun. One can describe such scenario where the DM particles couple to leptons but
1985 not to the quark sector using effective operators.

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

1986 In general, assuming that the DM particle is a Dirac fermion, the dimension six
 1987 operators describing the interaction between two DM particles and two leptons can be
 1988 written as:

$$\mathcal{L}_{eff} = G \sum_i (\bar{\chi} \Gamma_\chi^i \chi) (\bar{\ell} \Gamma_\ell^i \ell), \quad (5.40)$$

1989 where $G = 1/\Lambda^2$ is the effective coupling strength, Λ the cut-off of the effective field
 1990 theory and ℓ denotes any lepton. In principle, one should consider all the possible
 1991 Lorentz structures Γ_f^i in order to have a complete set of effective operators.

1992 However, some combinations will induce interactions with nucleons at loop level.
 1993 As we are specifically interested in interactions which forbid any communication with
 1994 the quark sector, I will not consider those [138]. In addition, some of the effective
 1995 operators give rise to velocity-suppressed scattering cross sections between DM particles
 1996 and leptons. I will also neglect those, as the suppression goes with the square of the DM
 1997 halo velocity which in units of the speed of light is $\sim 10^{-6}$.

1998 This way, the only Lorentz tensor structure that do not induce interactions with
 1999 quarks at loop level and gives a contribution to the scattering cross section that is not
 2000 velocity suppress is the axial-axial interaction. The effective Lagrangian is then given
 2001 by:

$$\mathcal{L}_{eff} = \frac{c_A^\chi c_A^\ell}{\Lambda^2} (\bar{\chi} \gamma^\mu \gamma^5 \chi) (\bar{\ell} \gamma_\mu \gamma^5 \ell), \quad (5.41)$$

2002 where c_A^χ and c_A^ℓ are the couplings for the different species. As the DM coupling appears
 2003 as a common factor for any lepton choice, I will redefine the corresponding coupling c_A^ℓ
 2004 to absorb c_A^χ . Also, for simplicity, I will assume that the couplings between the DM
 2005 particles and the leptons are flavour independent, i.e. I have just two couplings, c_A^e for
 2006 charged leptons and c_A^v for neutrinos.

2007 In the case of a scalar DM particle, the lowest order effective interaction with
 2008 leptons happens through a dimension five operator, generating scalar and pseudoscalar
 2009 interactions. However, the former induces interactions with quarks at two loop level
 2010 whereas the latter gives a velocity suppressed scattering cross section.

2011 From the effective Lagrangian in Eq. (5.41) it can be shown that the axial-axial

5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

contribution to the scattering cross section for the fermionic DM and a charged lepton is given by:

$$\sigma_{\text{DM}-e}^{AA} = 3(c_A^e)^2 \frac{m_e^2}{\pi \Lambda^4}. \quad (5.42)$$

If the DM interacts exclusively with fermions, then the only annihilation channels that will give us a measurable neutrino flux coming out of the Sun are $\tau^+\tau^-$ and $\nu\bar{\nu}$. The former channel, already explored previously in the more mainstream scenario of the DM capture via scattering off nucleons, is open only for $m_{\text{DM}} > m_\tau \simeq 1776.86 \pm 0.12$ MeV [139], a mass region where the solar DM capture by electrons is at least one order of magnitude smaller than the capture via interactions with nucleons. On the contrary, the latter allows us to explore a region where the capture rate via scattering off electrons dominates over the rest.

One downside of focusing in such low mass range is that it falls below the usual limit of $m_{\text{evap}} \sim 4$ GeV usually explored in the literature. The pretext to explore this region is the result discussed previously reported in Ref. [110], where DM evaporation in the Sun for the case of capture via electron scattering could be negligible for masses as low as $m_{\text{evap}} \sim 200$ MeV. This result is quite sensitive to the high velocity tail of the DM velocity distribution in equilibrium inside the Sun, and therefore full numerical simulations would be needed to assess the impact of this effect. However, this falls out of the scope of our work.

In this case, as I have a specific realisation of the interaction between the DM and leptons, one can estimate the relic density of our DM for different values of the couplings and the effective field theory scale Λ . The first step to do so is compute the self-annihilation cross section. Because I consider cold relics, at the freeze-out time our DM particles were non-relativistic and so one can expand the annihilation cross section in terms of the relative velocity v between two annihilating DM particles as [140]:

$$\sigma_{\text{ann}}^{AA}|v| \approx \frac{1}{2\pi\Lambda^4} \sum_\ell \left(c_A^\ell\right)^2 m_\chi^2 \sqrt{1 - \frac{m_\ell^2}{m_\chi^2} \left[\frac{m_\ell^2}{m_\chi^2} + \frac{1}{12} \left(2 - \frac{m_\ell^2}{m_\chi^2}\right) v^2 \right]}, \quad (5.43)$$

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

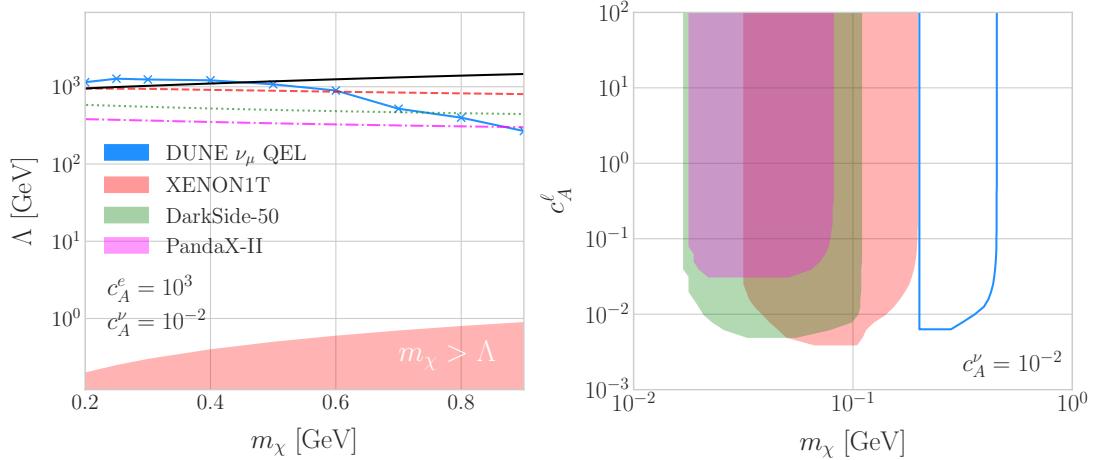


Figure 5.18: Left panel: Projected 90% confidence level sensitivity of DUNE (400 kT yr) to the scale Λ of an EFT containing only leptophilic DM axial-axial interactions (blue line), for the coupling values $c_A^e = 10^3$ and $c_A^\nu = 10^{-2}$. The black line represents the values for which the correct relic density is achieved. Right panel: Excluded values of c_A^ℓ as a function of the DM mass, for a fixed value $c_A^\nu = 10^{-2}$. In both cases the corresponding limits from XENON1T [142] (red), DarkSide-50 [143] (green) and PandaX-II [144] (magenta) are also shown.

2036 where the sum includes all the possible lepton final states with mass m_ℓ .
 2037 Solving the Boltzmann equation for the evolution of the DM density gives as a
 2038 solution a relic density of:

$$\Omega_\chi h^2 \approx \frac{(1.04 \times 10^9) x_F}{M_{Pl} \sqrt{g_*} (a + 3b/x_F)}, \quad (5.44)$$

2039 where $x_F = m_\chi/T_F$ being T_F the freeze-out temperature, g_* the number of relativistic
 2040 degrees of freedom at freeze-out and a and b the terms in the annihilation cross section
 2041 expansion $\sigma_{ann}|v| \approx a + bv^2 + \mathcal{O}(v^4)$. Using the current best fit for the relic DM density
 2042 $\Omega_\chi h^2 = 0.1198 \pm 0.0012$ [141] one can use these relations to compute the required
 2043 effective theory scale Λ at which the correct density is achieved for any combinations of
 2044 m_χ and c_A^ℓ .

2045 As discussed before, in the low DM mass region QEL interactions dominate. Moreover,
 2046 if I focus on direct annihilation to neutrinos, the energy of the muon neutrino flux is
 2047 known as it must be equal to the mass of the DM particle, $E_\nu = m_\chi$. That way, now

5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

2048 I do not need to use Eq. (5.33) in order to estimate the momentum transfer to the
 2049 remnant nucleus, I can simply take:

$$\vec{p}_N = \hat{p}_\nu m_\chi - \vec{p}_\mu - \vec{p}_p. \quad (5.45)$$

2050 To estimate the signal efficiency and background rejection for this case I used again
 2051 the MLP classifier from `scikit-learn`, using the same specifications as before. The
 2052 only difference now is that I add also the reconstructed neutrino energy as one of the
 2053 features to train the classifier with, because the characteristic monoenergetic flux for
 2054 each m_χ value will help to distinguish between signal and background events.

2055 In this case, for masses below ~ 500 MeV I obtain a signal efficiency close to unity
 2056 while keeping a background rejection of 99.9%. For bigger values of the mass, the signal
 2057 efficiency drops significantly if I require to keep the background acceptance under 0.01%.
 2058 However, because this kind of search is dominated by the background, sacrificing the
 2059 signal acceptance to keep the background rejection to a minimum enhances the reach
 2060 of the analysis. This way, for DM masses of the order of $m_\chi \sim 1$ GeV I end up with
 2061 efficiencies as low as 1%.

2062 Now, estimating the number of background events using Eq. (5.39) one can go on
 2063 and apply Eqs. (5.26) and (5.27) together with Eq. (5.42) to derive the sensitivity of
 2064 DUNE to this kind of model. Fig. 5.18 (left panel) shows the potential reach of DUNE
 2065 to constrain the EFT scale Λ this model containing only leptophilic DM axial-axial
 2066 interactions (blue line), for a choice of couplings $c_A^e = 10^3$ and $c_A^\nu = 10^{-2}$. I also included
 2067 the current limits on the DM-electron scattering cross section from XENON1T [142]
 2068 (dashed red line), DarkSide-50 [143] (dotted green line) and PandaX-II [144] (dash-dotted
 2069 magenta line), reworked with Eq. (5.42) to show their implications for the EFT scale.
 2070 The values of Λ for which the correct DM relic density value is achieved for each mass are
 2071 also shown (black line). This tells us that, for that specific choice of couplings, DUNE
 2072 would be sensitive to DM configurations allowed by the relic density constraint up to a
 2073 mass of $m_\chi \sim 400$ MeV.

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2074 In Fig. 5.18 (right panel) I show similar limits for the excluded values of c_A^ℓ as a
2075 function of the DM mass, for a fixed $c_A^\nu = 10^{-2}$. I do not show the limits for other values
2076 of c_A^ν , as this parameter has little effect on the phenomenology at hand. From this view
2077 one can see that DUNE would be able to offer complementary information to the low
2078 energy DM-electron interaction searches performed by direct detection experiments, in a
2079 slightly higher mass range.

2080 With the present example, although it focuses on a very specific realisation of the DM
2081 interactions, I show the potential of DUNE to constrain exotic DM scenarios. Thanks
2082 to its low backgrounds and superb angular resolution DUNE will be able to help with
2083 the systematic searches for dark sectors physics.

2084 5.7 Systematic uncertainties

2085 The estimation of the DM cross sections using neutrinos from WIMP annihilations
2086 inside the Sun is affected by systematic uncertainties from different sources. Surely, the
2087 atmospheric background estimation is also affected by systematic uncertainties. There
2088 are uncertainties common to both types of events, as well as others specific to each. In
2089 this section, I try to provide a comprehensive summary of the main sources of uncertainty
2090 for this analysis, which should be taken into account in any future extensions of the
2091 same.

2092 5.7.1 Systematic uncertainties in the solar WIMP signal

2093 The systematic uncertainties affecting the solar WIMP neutrino signal can be divided in
2094 two categories. On the one hand, we have those affecting the solar WIMP annihilation
2095 rate. On the other hand, there are the ones which modify the neutrino flux resulting
2096 from the annihilations reaching our detector.

- 2097 • **Uncertainties on the annihilation rate.** These include the astrophysical effects
2098 that affect the normalisation of the solar DM neutrino flux. The main contributions
2099 are the solar model choice, the form factor uncertainties (only for SI searches), the

5.7. SYSTEMATIC UNCERTAINTIES

Table 5.1: Systematic uncertainties for the solar WIMP signal events. Table adapted from Ref. [145].

Systematic	Value
Form factor	Does not apply to SD [146]
Solar model	3% [146]
Local DM density	Not relevant for relative interpretations [146, 147]
Dynamics of solar system	Negligible [148]
Velocity distributions	20% at 20 GeV [146, 147]
Oscillation parameters	8% for $\tau^+\tau^-$, 5% for $b\bar{b}$ [149]
Neutrino interactions in the Sun	10%
Matter effects in the Earth	10%

Table 5.2: Systematic uncertainties for the solar WIMP atmospheric background events. Table adapted from Ref. [39].

Systematic	Value
Flux normalisation	25 – 7% for $0.1 < E_\nu \leq 1$ GeV (linear in $\log E_\nu$) 7% up to 10 GeV
$(\nu_\mu + \bar{\nu}_\mu)/(\nu_e + \bar{\nu}_e)$	2% for $E_\nu \leq 1$ GeV 3% for $1 < E_\nu \leq 10$ GeV
$\bar{\nu}_\mu/\nu_\mu$	2% for $E_\nu \leq 1$ GeV 6% for $1 < E_\nu \leq 10$ GeV
K/ π ratio	5% $E_\nu \leq 100$ GeV

2100 gravitational effect of other planets, the local DM density (not relevant for relative
 2101 comparisons, as it affects direct detection experiments in the same way), and the
 2102 DM halo and dispersion velocities.

2103 • **Uncertainties on the neutrino flux.** These are related to the oscillation effects,
 2104 as well as the absorption and regeneration of neutrinos in the Sun. Matter effects
 2105 inside the Earth also affect the neutrino flux the measured at the detectors.

2106 Table 5.1 summarises the contributions of the different sources of uncertainty for the
 2107 signal events. These are the signal systematic uncertainties that have been taken into
 2108 account in previous solar DM searches with neutrinos [145, 147, 149].

CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2109 5.7.2 Systematic uncertainties in the atmospheric background

2110 For the atmospheric background events, one needs to take into account the systematic
2111 uncertainties affecting the atmospheric ν_μ flux. These have been extensively studied
2112 in the context of atmospheric neutrino oscillation measurements. Among these, the
2113 energy-dependent flux normalisation uncertainty is the in the low energy regime. Other
2114 important contributions to the uncertainty come from the ratios between the muon to
2115 electron neutrino and the muon to anti-muon neutrino components of the flux. Additional
2116 uncertainty is introduced by the errors in the pion and kaon production rates calculated
2117 for the hadronic interactions of cosmic rays in the atmosphere [150].

2118 Table 5.2 shows a summary of the leading contributions to the uncertainty on the
2119 atmospheric muon neutrino flux, in the energy range relevant for this analysis.

2120 5.7.3 Common systematic uncertainties

2121 Finally, there are sources of uncertainty common to both signal and backgrounds. These
2122 have two different origins:

- 2123 • **Uncertainties on the neutrino cross section.** These are introduced by the
2124 modelling of the neutrino-nucleus interactions. In the context of the solar WIMP
2125 analysis, these have been estimated to be 10% for DM masses around 10 GeV
2126 [149].

- 2127 • **Uncertainties related to the detector.** They affect the measurement of the
2128 neutrino interaction and the final state particles produced. The main detector
2129 uncertainties relevant to this analysis are those of the energy and angular resolutions
2130 of the DUNE FD. Other effects, like the timing and triggering efficiencies, will
2131 also contribute to the uncertainties. The particular values these will take for this
2132 analysis need to be worked out in the context of DUNE.

6

2133

2134

Particle identification in ND-GAr

2135 *I am no bird; and no net ensnares me; I am a free human being with an
2136 independent will.*

2137

– Charlotte Brontë, *Jane Eyre*

2138 In DUNE Phase II, ND-GAr will fulfill the role of TMS measuring the momentum
2139 and sign of the charged particles exiting ND-LAr. Additionally, it will measure neutrino
2140 interactions inside the HPgTPC. This way, ND-GAr will allow to constrain certain cross
2141 section systematic uncertainties and study the effect of FSI in CC interactions. To
2142 do so, it needs to measure the spectrum of protons and charged pions at low energies,
2143 as well as measure the pion multiplicity. This puts strong requirements to the the
2144 particle identification (PID) capabilities of the detector, as well as stimulates the relevant
2145 developments in the reconstruction.

2146 The goal of the present Chapter is to review the status and design of the GArSoft
2147 package, the simulation and reconstruction software of ND-GAr, and present the different
2148 additions and upgrades that I have added to the reconstruction with the PID in mind.

2149 6.1 GArSoft

2150 GArSoft is a software package developed for the simulation and reconstruction of events
2151 in ND-GAr. It is inspired by the LArSoft toolkit used for the simulation of LArTPC
2152 experiments, like the DUNE FD modules. It is based on `art`, the framework for event
2153 processing in particle physics experiments [151]. Other of its main dependencies are

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2154 ROOT, NuTools, GENIE and Geant4. It allows the user to run all the steps of a generation-
2155 simulation-reconstruction workflow using FHiCL configuration files.

2156 6.1.1 Event generation

2157 The standard generator FHiCLs in GArSoft run the event generation and particle
2158 propagation simulation (i.e. Geant4) in the same job by default. However, it is possible
2159 to split them up if needed. The current version of GArSoft provides five different event
2160 generators, each of them producing `simb::MCTruth` products defined in NuTools. The
2161 available modules are:

2162 • **SingleGen**: particle gun generator. It produces the specified particles with a given
2163 distribution of momenta, initial positions and angles.

2164 • **TextGen**: text file generator. The input file must follow the `hepevt` format¹, the
2165 module simply copies this to `simb::MCTruth` data products.

2166 • **GENIEGen**: GENIE neutrino event generator. The module runs the neutrino-nucleus
2167 interaction generator using the options specified in the driver FHiCL file (flux file,
2168 flavour composition, number of interactions per event, t_0 distribution, ...). Current
2169 default version is v3_04_00.

2170 • **RadioGen**: radiological generator. It produces a set list of particles to model
2171 radiological decays. Not tested.

2172 • **CRYGen**: cosmic ray generator. The module runs the CRY event generator with a
2173 configuration specified in the FHiCL file (latitude and altitude of detector, energy
2174 threshold, ...). Not tested.

¹In brief, each event contains at least two lines. The first line contains two entries, the event number and the number of particles in the event. Each following line contains 15 entries to describe each particle. The entries are: status code, pdg code for the particle, entry of the first mother for this particle, entry of the second mother for this particle, entry of the first daughter for this particle, entry of the second daughter for this particle, x component of the particle momentum, y component of the particle momentum, z component of the particle momentum, energy of the particle, mass of the particle, x component of the particle initial position, y component of the particle initial position, z component of the particle initial position and time of the particle production.

6.1. GArSOFT

2175 The module **GArG4** searches for all the generated **simb::MCTruth** data products, using
2176 them as inputs to the Geant4 simulation with the specified detector geometry. A constant
2177 0.5 T magnetic field along the drift coordinate is assumed. The main outputs of this step
2178 are **simb::MCParticle** objects for the generated Geant4 particles, **gar::EnergyDeposit**
2179 data products for the energy deposits in the HPgTPC and **gar::CaloDeposit** data
2180 products for the energy deposits in the ECal and muon system.

2181 6.1.2 Detector simulation

2182 The standard detector simulation step in GArSoft is all run with a single FHiCL, but
2183 the different modules can be run independently as well. First the **IonizationReadout**
2184 module simulates the charge readout of the HPgTPC, and later the **SiPMReadout** module
2185 runs twice, once for the ECal and then for the muon system, with different configurations.

2186 The **IonizationAndScintillation** module collects all the **gar::EnergyDeposit**
2187 data products, to compute the equivalent number of ionization electrons for each energy
2188 deposit. The **ElectronDriftAlg** module simulates the electron diffusion numerically
2189 both in the longitudinal and transverse directions and applies an electron lifetime
2190 correction factor. The induced charge on the nearest and neighbouring readout pads
2191 is modeled using the provided pad response functions. The digitisation of the data is
2192 then simulated with the **TPCReadoutSimAlg** module. By default, the ADC sampling
2193 rate used is 50.505 MHz. The resulting raw waveforms for each channel are stored with
2194 zero-suppression, in order to save memory and CPU time. The algorithms keep blocks
2195 of ADC values above a certain threshold, plus some adjustable additional early and late
2196 tick counts. The results of these three steps are **gar::raw::RawDigit** data products.

2197 For the ECal and the muon system the **SiPMReadout** module calls either the
2198 **ECALReadoutSimStandardAlg** or **MuIDReadoutSimStandardAlg** modules. These take
2199 all the **gar::CaloDeposit** data products in the corresponding detector and do the
2200 digitisation depending on whether the hit was in a tile or strip layer. They include single
2201 photon statistics, electronic noise, SiPM saturation and time smearing. The resulting
2202 objects are **gar::raw::CaloRawDigit** data products.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

6.1.3 Reconstruction

The reconstruction in GArSoft is also run as a single job by default. It first runs the hit finding, clustering, track fitting and vertex identification in the HPgTPC, followed by the hit finding and clustering in the ECal and muon system. After those it produces the associations between the associations between the tracks and the ECal clusters.

Focusing first on the HPgTPC reconstruction, the `CompressedHitFinder` module takes the zero-suppressed ADCs from the `gar::raw::RawDigit` data products. The reconstructed hits largely correspond to the above threshold blocks, however the hit finder identifies waveforms with more than one maximum, diving them in multiple hits if they dip below a certain threshold. The data products produced are of the form `gar::rec::Hit`. These are the inputs to the clustering of hits in the `TPCHitCluster` module. Hits close in space and time are merged, and the resulting centroids are found. This module outputs `gar::rec::TPCClusters` objects and associations to the input hits.

The following step prior to the track fitting is pattern recognition. The module called `tpcvechitfinder2` uses the `gar::rec::TPCClusters` data products to find track segments, typically called vector hits. They are identified by performing linear 2D fits to the positions of the clusters in a 10 cm radius, one fit for each coordinate pair. A 3D fit defines the line segment of the vector hit, using as independent variable the one whose sum of (absolute value) slopes in the 2D fits is the smallest. The clusters are merged to a given vector hit if they are less than 2 cm away from the line segment. The outputs are `gar::rec::VecHit` data products, as well as associations to the clusters. The `tpcpatrec2` module takes the `gar::rec::VecHit` objects to form the track candidates. The vector hits are merged together if their direction matches, their centers are within 60 cm and their direction vectors point roughly to their respective centers. Once the clusters of vector hits are formed they are used to make a first estimation of the track parameters, simply taking three clusters along the track. The module produces `gar::rec::Track` data products and associations between these tracks and the clusters

6.1. GArSOFT

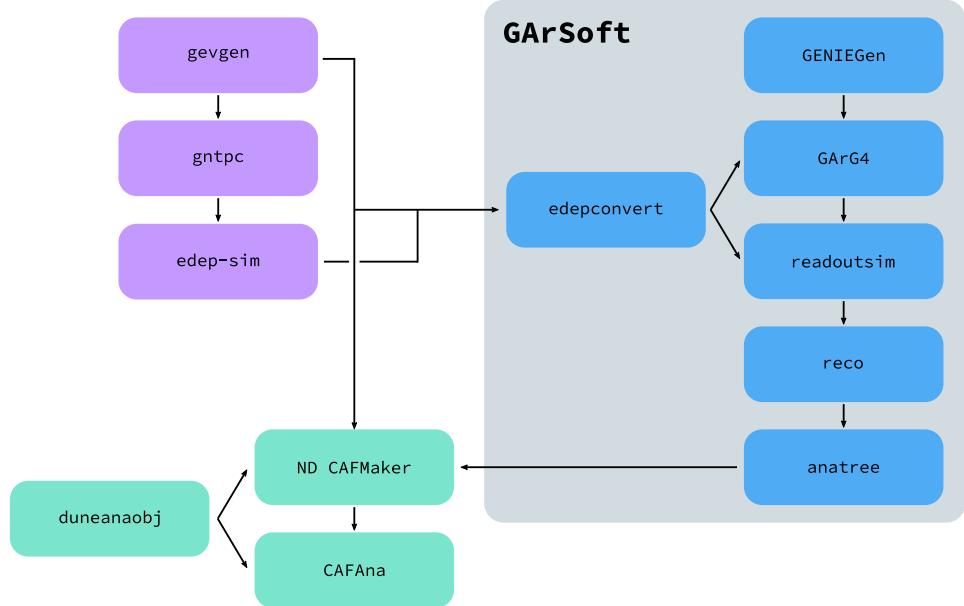


Figure 6.1: Schematic diagram showing the different modules involved in the ND-GAr production.

and vector hits.

The track is fitted by means of a Kalman filter in the `tpctrackfit2` module, using the position along the drift direction as the independent variable. Two different fits are performed per track, a forward and a backwards fit, each starting from one of the track ends. The Kalman filter state vector ($y, z, R, \phi, \tan\lambda$) is estimated at each point along the track using a Bayesian update. The track parameters reported in the forward and backwards fits are the ones computed at the opposite end where the fit started. The main outputs of the track fit are the `gar::rec::Track` objects. Additionally, the module stores the fitted 3D positions along the track in the `gar::rec::TrackTrajectory` data products and the total charge and step sizes for each point also get stored in the form of `gar::rec::TrackIonization` objects.

After the tracking step, the `vertexfinder1` module looks at the reconstructed `gar::rec::Track` products, creating vertex candidates with the track ends that are within 12 cm of each other. The vertices are then fitted using linear extrapolations from the different track ends associated. The results are `gar::rec::Vertex` data products,

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2246 and associations to the tracks and corresponding track ends.

2247 For the ECal and muon tagger, the `SiPMHitFinder` module runs twice with different
2248 configurations, adapted to the particular capabilities of both. The module simply takes
2249 the `gar::raw::CaloRawDigit` products, applies a calibration factor to convert the ADC
2250 counts to MeV and for the strip layer hits it calculates the position along the strip using
2251 the times recorded of both SiPMs. This module produces `gar::rec::CaloHit` data
2252 products. Next, these objects are used as inputs to the `CaloClustering` module. It
2253 merges the hits based on a simple nearest neighbours (NN) algorithm. For the resulting
2254 clusters it also computes the total energy and position of the centroid. The results are
2255 stored as `gar::rec::Cluster` data products, with associations to the hits.

2256 The last step in the reconstruction is associating the reconstructed tracks in the
2257 HPgTPC to the clusters formed in the ECal and muon system. The `TPCECALAssociation`
2258 module checks first the position of the track end points, considering only the points
2259 that are at least 215 cm away from the cathode or have a radial distance to the center
2260 greater than 230 cm. The candidates are propagated up to the radial position, in the
2261 case of clusters in the barrel, or the drift coordinate position, for the end cap cluster, of
2262 the different clusters in the collection using the track parameters computed at the end
2263 point. The end point is associated to the cluster if certain proximity criteria are met.
2264 This module creates associations between the tracks, the end points and the clusters.
2265 The criteria for the associations are slightly different for the ECal and the muon tagger.

2266 6.2 dE/dx measurement in the TPC

2267 Among the parameters extracted from the track fitting, ionisation is particularly useful
2268 for particle identification, as it is a function of the particle velocity. Although for the
2269 case of relativistic particles this dependence is not very strong, measuring the track on
2270 a large number of points may allow us to estimate the amount of ionisation accuratel.
2271 This, paired with a measurement of the momentum, may allow us to identify the particle
2272 type.

6.2. dE/dx MEASUREMENT IN THE TPC

2273 The first calculation of the energy loss per unit length of relativistic particles using a
 2274 quantum-mechanical treatment is due to Bethe [152]. Using this approach, the mean
 2275 ionisation rate of a charged particle traveling through a material medium is (using
 2276 natural units $G = \hbar = c = 1$):

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi Ne^4}{m_e \beta^2} z^2 \left(\log \frac{2m_e \beta^2 \gamma^2}{I} - \beta^2 \right), \quad (6.1)$$

2277 where N is the number density of electrons in the medium, e the elementary charge, m_e
 2278 is the electron mass, z the charge of the particle in units of e , β is the velocity of the
 2279 particle, $\gamma = (1 - \beta^2)^{-1}$ and I denotes the effective ionisation potential averaged over
 2280 all electrons. This relation is known as the Bethe-Bloch formula.

2281 From Eq. (6.1) one can see that the ionisation loss does not depend explicitly on
 2282 the mass of the charged particle, that for non-relativistic velocities it falls as β^{-2} , then
 2283 goes through a minimum and increases as the logarithm of γ . This behaviour at high
 2284 velocities is commonly known as the relativistic rise. The physical origin of this effect
 2285 is partly due to the fact that the transverse electromagnetic field of the particle is
 2286 proportional to γ , therefore as it increases so does the cross section.

2287 It was later understood that the relativistic rise could not grow indefinitely with γ .
 2288 A way to add this feature in the Bethe-Bloch formula is by introducing the so-called
 2289 density effect term. It accounts for the polarisation effect of the atoms in the medium,
 2290 which effectively shield the electromagnetic field of the charged particle halting any
 2291 further increase of the energy loss [153]. Denoting the correction as $\delta(\beta)$, one can rewrite
 2292 Eq. (6.1) as:

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi Ne^4}{m_e \beta^2} z^2 \left(\log \frac{2m_e \beta^2 \gamma^2}{I} - \beta^2 - \frac{\delta(\beta)}{2} \right). \quad (6.2)$$

2293 In general, the form of $\delta(\beta)$ depends on the medium and its state of aggregation,
 2294 involving the usage of tabulated parameters and implicit relations [154].

2295 Another standard method to compute the amount of ionisation a charged particle
 2296 produces is the so-called photo-absorption ionisation (PAI) model proposed by Allison
 2297 and Cobb [155]. Within their approach, the mean ionisation is evaluated using a

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

2298 semiclassical calculation in which one characterises the continuum material medium by
2299 means of a complex dielectric constant $\varepsilon(k, \omega)$. However, in order to model the dielectric
2300 constant they rely on the quantum-mechanical picture of photon absorption and collision.
2301 Therefore, in the PAI model the computation of the ionisation loss involves a numerical
2302 integration of the measured photo-absorption cross-section for the relevant material.

2303 In a particle physics experiment, the typical way of determining the energy loss
2304 per unit length as a function of the particle velocity is studying identified particles
2305 over a range of momenta. Once we have established this relation we can use it for
2306 other, unknown particles. In this sense, it makes sense to have a regular mathematical
2307 expression for this relation that one can use.

2308 It happens that neither the Bethe-Bloch theory nor the PAI model from Allison and
2309 Cobb offer a close mathematical form for the ionisation curve. This is the reason why a
2310 full parametrisation of the ionisation curves can be useful. A parametrisation originally
2311 proposed for the ALEPH TPC [156] and later used by the ALICE TPC [157] group that
2312 manages to capture the features of the ionisation energy loss is:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \left(P_2 - \beta^{P_4} - \log \left[P_3 + \frac{1}{(\beta\gamma)^{P_5}} \right] \right), \quad (6.3)$$

2313 where P_i are five free parameters. Hereafter, we will refer to Eq. (6.3) as the ALEPH
2314 dE/dx parametrisation.

2315 6.2.1 Energy calibration

2316 In order to obtain the amount of energy loss by a charged particle due to ionisation
2317 in our TPC we need to determine the conversion between the charge deposited in our
2318 readout planes and the actual energy depositions. This procedure is known as energy
2319 calibration.

2320 In a general, the first step of the calibration involves a non-uniformity correction,
2321 to make sure that the detector response is uniform throughout the TPC. These are
2322 typically divided into three categories, non-uniformities in the transverse YZ plane,

6.2. dE/dx MEASUREMENT IN THE TPC

2323 non-uniformities along the drift direction X and variations of the detector response
2324 over time (would not apply to us as the detector is not built yet). These would correct
2325 for effects such as electron diffusion and attenuation, space charge effects or channel
2326 misconfiguration. However, because at the moment I am only interested in making sure
2327 we recover a sensible result from our simulation, I will not apply uniformity corrections
2328 to our charge deposits.

2329 Other effects, like electron-ion recombination or ADC saturation, lead to a non-linear
2330 relation between the observed charge and the deposited energy in the detector, with the
2331 observed readout charge saturating at high ionisation energies. In this case, because we
2332 are dealing with gaseous argon and therefore recombination is not as important as in
2333 liquid, we do not simulate recombination effects in the TPC. Even so, the simulation of
2334 the electronic response will still introduce charge saturation, and one needs to correct
2335 for it in order to obtain the exact amount of energy loss due to ionisation.

2336 By default, the track fitting algorithm in GArSoft provides a `TrackIonization`
2337 object associated to each reconstructed track. It contains two collections of charge
2338 deposits, one for each fitting direction, consisting on pairs of charge values (dQ , in ADC)
2339 and step sizes (dx , in cm).

2340 In order to estimate the ionisation loss in the ND-GAr TPC, I have used an MC
2341 sample consisting of single, isotropic protons propagating in the TPC. The starting points
2342 of the protons were sampled inside a $50 \times 50 \times 25$ cm box centered at $(100, -150, 1250)$,
2343 and their momenta are uniformly distributed in the range $0.25 - 1.75$ GeV. I ran the
2344 simulated sample through GArSoft's default detector simulation and reconstruction, and
2345 then a custom analyser module that extracts the ionisation data together with other
2346 reconstructed track information from the Kalman fit.

2347 For studying the energy loss of the protons I select the reconstructed tracks that
2348 range out (i.e. slow down to rest) inside the TPC. A characteristic feature of the energy
2349 loss profile of any stopping ionising particle is the so-called Bragg peak, a pronounced
2350 peak that occurs immediately before the particle comes to rest. From Eq. (6.1) we can
2351 see that this behaviour is expected, as the energy loss for non-relativistic particles is

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

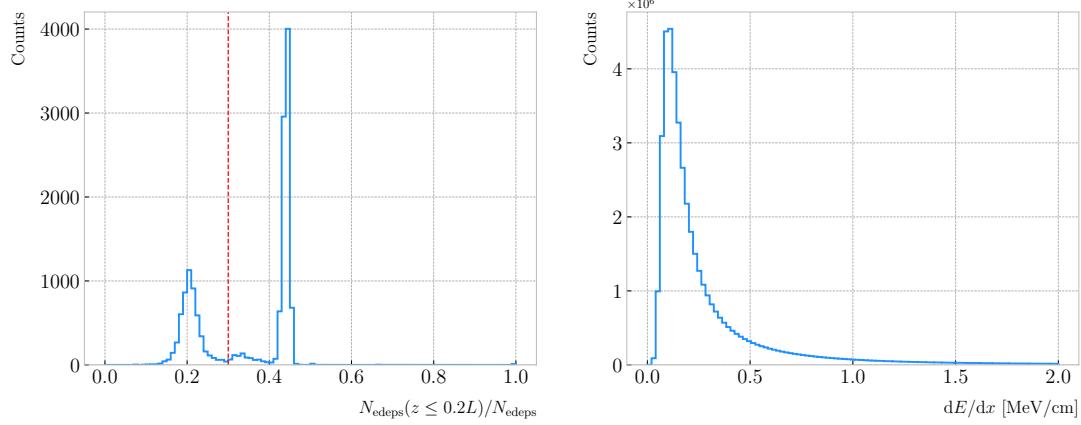


Figure 6.2: Left panel: distribution of the fraction of Geant4-level energy deposits per track with residual range less than 20% of the total track length, for the isotropic proton sample. Right panel: distribution of the ionisation per unit length of the energy deposits in the proton sample after removing the tracks with less than 30% of their energy deposits in the last 20% of the track.

2352 inversely proportional to β^2 . In data, a way of identifying the Bragg peak, and thus
 2353 select the stopping particles, is checking the number of energy deposits towards the
 2354 end of the track. In this case, I count the fraction of the Geant4 simulated energy
 2355 deposits with a residual range value (the distance from a given energy deposit to the
 2356 last deposit in the track trajectory) less than a 20% of the corresponding track length².
 2357 The distribution of this fraction of energy deposits for our proton sample is shown in
 2358 Fig. 6.2 (left panel). We can clearly see two well separated peaks in this distribution,
 2359 one centered at 0.2 and another, narrower, one centered at a higher value. The first
 2360 one corresponds to non-stopping protons, as in that case the number of energy deposits
 2361 towards the end of the track is uniformly distributed due to the absence of the Bragg
 2362 peak. In that way, I apply a cut in this distribution, requiring that at least 30% of the
 2363 simulated energy deposits sit in the last 20% of the tracks, to ensure that the Bragg
 2364 peak is present.

2365 Figure 6.2 (right panel) shows the distribution of the energy loss per unit length for
 2366 the Geant4 simulated energy deposits of the selected stopping protons. We can see that

²As we are applying this selection at the Geant4 level we could have simply selected the stopping protons using the `EndProcess` labels from the simulation. However, the Bragg peak identification method displayed here could serve as a starting point for a selection of stopping protons in real data.

6.2. dE/dx MEASUREMENT IN THE TPC

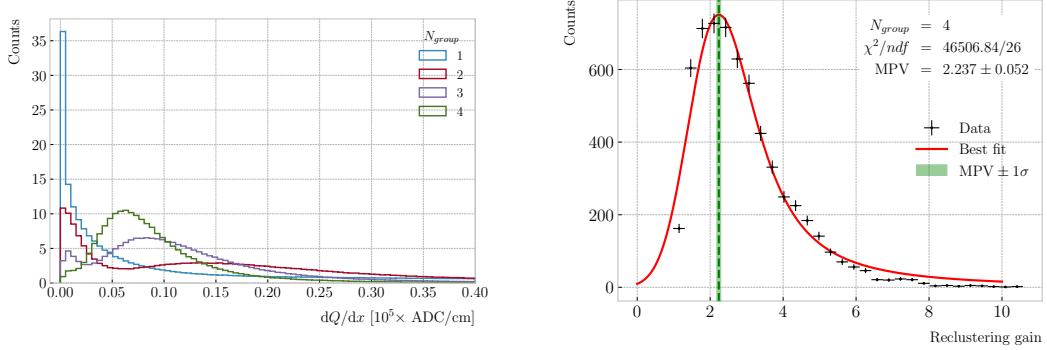


Figure 6.3: Left panel: distribution of the reconstructed ionisation charge per unit length for our MC stopping proton sample. The different colors indicate how many consecutive dQ/dx pairs were grouped together. Right panel: distribution of the median change in dQ/dx per track after $N_{group} = 4$ clusters were reclustered together.

it follows the expected shape of a Landau distribution, which describes the fluctuations of the ionisation energy losses [158]. This distribution has a characteristic asymmetric PDF, with a long right tail that translates into a high probability for high-energy ionisation losses. The origin of these fluctuations is mainly the possibility of transferring a high enough energy to an electron, so it becomes a ionising particle itself.

Now, from the point of view of the reconstruction, the objects that we have available to extract the ionisation information for the different reconstructed tracks are the collections of dQ and dx pairs, as stated before. The dQ values come from adding up the amplitude of all the reconstructed hits in a cluster, which is the input object to the Kalman fit.

Figure 6.3 (left panel) shows the distribution of the ionisation charge deposits per unit length for the track in the stopping proton sample (blue line). As one can notice, this distribution does not resemble the expected shape of the Landau PDF. This distribution peaks sharply at 0 and has a heavy tailed behaviour. Notice, however, how the distribution changes its shape as we group together N_{group} consecutive charge deposit pairs (red, purple and green lines). The distribution in the $N_{group} = 4$ case already has a shape which resembles that of the Geant4-level ionisation per unit length, so I will proceed using this amount of reclustering for the reconstruction-level depositions.

An extra factor I need to account for, when reclustering is applied, is how the overall

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

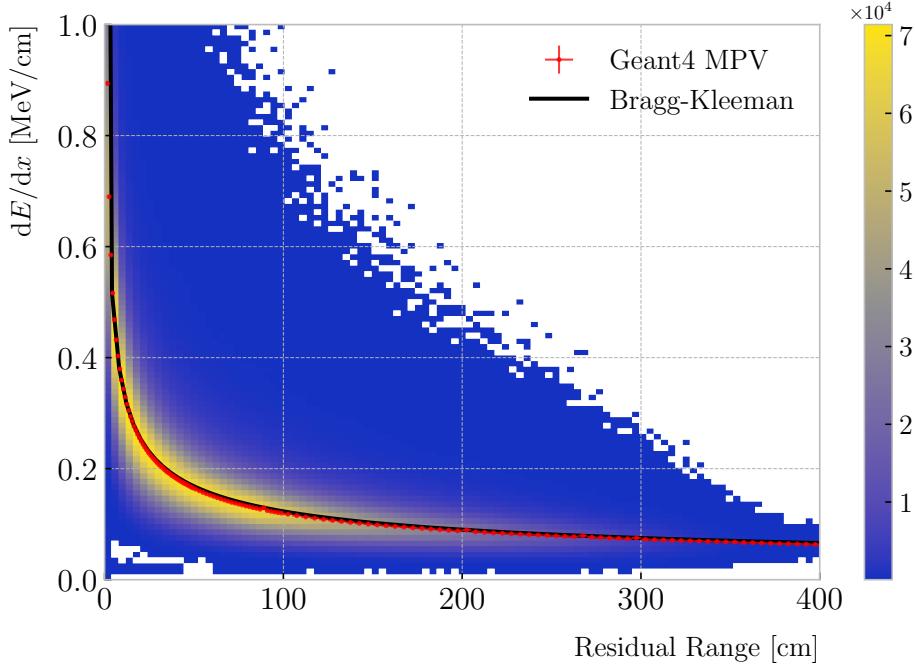


Figure 6.4: Distribution of the Geant4-simulated energy losses per unit length versus residual range for the stopping proton sample. The overlaid points represent the fitted most probable value of the dE/dx distribution in each residual range bin, whereas the curve is their best fit to the Bragg-Kleeman formula from Eq. (6.4).

2386 dQ/dx per track changes. To do so, we can look at the ratio between the median dQ/dx
 2387 after and before the reclustering. Figure 6.3 (right panel) shows the median enhancement
 2388 in dQ/dx per track for the stopping proton sample in the case $N_{group} = 4$. Fitting a
 2389 Landau distribution convolved with a Gaussian³, I estimate the most probable value of
 2390 this ratio to be $G_{group} = 2.24 \pm 0.05$.

2391 At this point, I am left with determining the conversion between the charge deposits
 2392 per unit length dQ/dx and the energy deposits per unit length dE/dx . To this end, we
 2393 need a way of comparing the two. I can use the residual range z to get a prediction of
 2394 the most probable dE/dx by using the following empirical parametrisation [159]:

$$\frac{dE}{dx}(z) = \frac{z^{\frac{1}{p}-1}}{p\Lambda^{\frac{1}{p}}}, \quad (6.4)$$

³In the literature, this distribution is often referred to as Landau+Gaussian or langau. In the following, I will use LanGauss to refer to such PDF.

6.2. dE/dx MEASUREMENT IN THE TPC

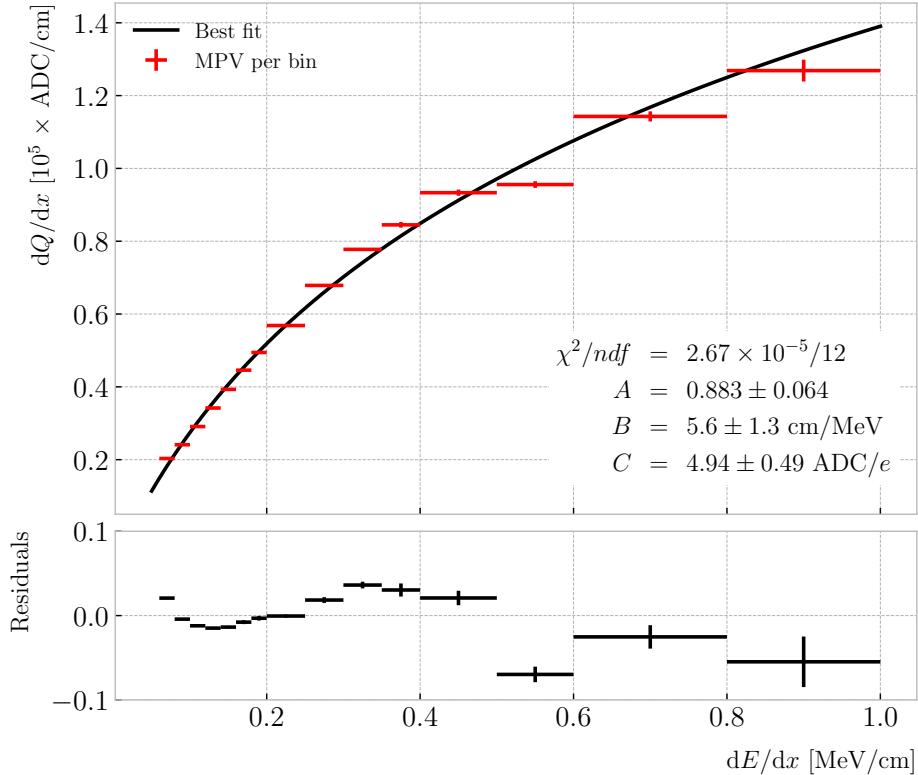


Figure 6.5: Fitted most probable dQ/dx values for each dE/dx bin (red points), obtained from the stopping proton sample. The overlaid curve (black line) represents the best fit to the logarithmic calibration function from Eq. (6.5).

which is quoted in the literature as the Bragg-Kleeman formula. In order to obtain the p and Λ parameters I perform a fit using the energy losses and the residual ranges given by the Geant4 stage of our proton sample.

Within our simulation, the residual range is sampled with a maximum size of 5 mm. Therefore, to perform the fit to the Bragg-Kleeman formula, we can use a fine-grained residual range binning. For each of the residual range bins I extract the dE/dx distribution and fit it to a LanGauss distribution, to obtain the value of the most probable dE/dx in the bin together with a statistical uncertainty. I then fit Eq. (6.4) to these most probable values and the centres of the residual range bins. This procedure is depicted in Fig. 6.4, where I show the distribution of the energy loss per unit length versus the residual range, together with the most probable dE/dx values and their uncertainty in each bin (red points) and the curve with the best fit of the

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2407 Bragg-Kleeman relation to those values (black line). The best fit is obtained for the
 2408 parameter values $p = 1.8192 \pm 0.0005$ and $\Lambda = 0.3497 \pm 0.0008 \text{ cm}/\text{MeV}^p$ ⁴.

2409 Having an analytical expression that relates the residual range to dE/dx , I can take
 2410 our reconstruction-level residual ranges from the stopping proton sample and compute
 2411 the most probable energy loss associated.

2412 In order to parametrise the charge saturation, we can use the following logarithmic
 2413 function inspired by the modified box model for recombination:

$$\frac{dE}{dx} = \frac{e^{\frac{dQ}{dx}B\frac{W_{ion}}{G_{group}C}} - A}{B}, \quad (6.5)$$

2414 where A and B are the calibration parameters we need to determine, W_{ion} is the average
 2415 energy to produce an electron-ion pair, G_{group} is the gain from the reclustering discussed
 2416 above and C is the calibration constant to convert number of electrons to ADC counts,
 2417 commonly refer to as gain (also to be obtained in the fit). In this case, I use a value
 2418 for the electron-ion production energy of $W_{ion} = 26.4 \text{ eV}$ [160]. This value, used in our
 2419 simulation as well, was measured for gaseous argon in normal conditions, and therefore
 2420 should be checked in the future to describe correctly the high-pressure argon-CH₄ mixture
 2421 of ND-GAr.

2422 For the calibration fit I follow a procedure similar to the previous one for Eq. (6.4).
 2423 Binning the dE/dx range, I fit a LanGauss distribution to the corresponding dQ/dx
 2424 distribution to obtain the most probable value. The resulting data points (red bars) are
 2425 shown in Fig. 6.5 (top panel), the horizontal error bars depict the width of the dE/dx
 2426 bin whereas the vertical bars represent the error associated to the most probable value
 2427 estimation. A fit to the logarithmic function in Eq. (6.5) is also shown (black line).
 2428 For this I weighted the data points using the inverse of their relative error, obtaining
 2429 a reduced chi-square value of $\chi^2/ndf = 2.22 \times 10^{-6}$. The best fit parameters I found
 2430 from this fit are $A = 0.883 \pm 0.064$, $B = 5.6 \pm 1.3 \text{ cm}/\text{MeV}$ and $C = 4.94 \pm 0.49 \text{ ADC}/e$.
 2431 Figure 6.5 (bottom panel) shows the residuals between the data points and the fit.

⁴These strange units for Λ come from dimensional analysis, just to keep the Bragg-Kleeman formula (6.4) consistent.

6.2. dE/dx MEASUREMENT IN THE TPC

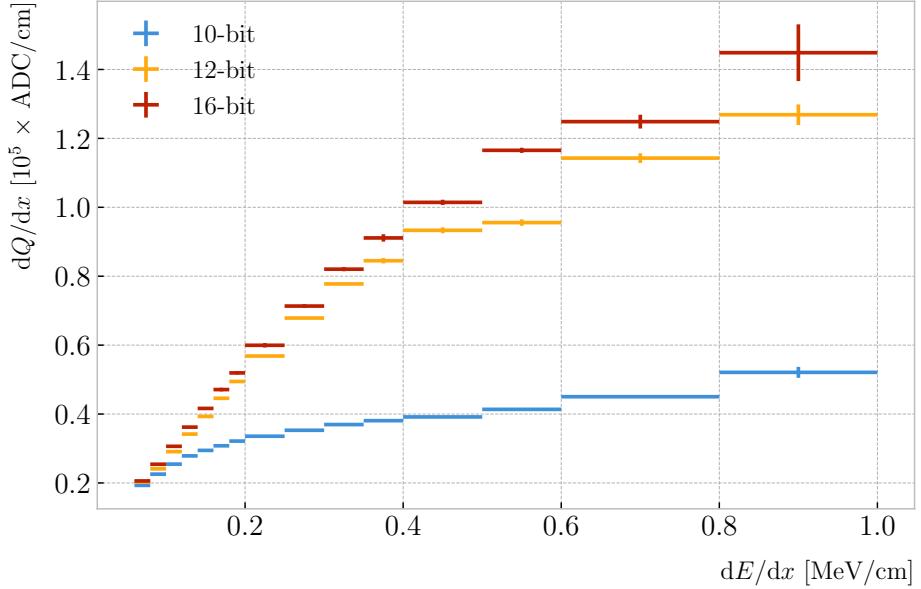


Figure 6.6: Fitted most probable dQ/dx values for each dE/dx bin for three different ADC bit limits, 10 (blue points), 12 (default, yellow points) and 16-bit (red points).

Table 6.1: Calibration parameters obtained from the fit of the ND-GAr simulated stopping proton sample to the calibration function from Eq. (6.5). The fits were performed for the 10, 12, and 16-bit ADC limits.

	χ^2/ndf	Best fit $\pm 1\sigma$		
		A	B (cm/MeV)	C (ADC/e)
10-bit	$1.83 \times 10^{-6}/12$	-9.3 ± 3.9	270 ± 69	27.1 ± 5.4
12-bit	$2.67 \times 10^{-5}/12$	0.883 ± 0.064	5.6 ± 1.3	4.94 ± 0.49
16-bit	$1.44 \times 10^{-5}/12$	0.949 ± 0.024	3.53 ± 0.58	4.52 ± 0.29

2432 The value for the gain I obtained from the fit is in reasonable agreement with our
 2433 expectation. This value is set in GArSoft to 5 ADC/e by default.

2434 One interesting thing to check is what induces this non-linear relation between charge
 2435 and energy. The only effects that modify the amount of electrons reaching the readout
 2436 planes in the simulation are the transverse diffusion and the finite electron lifetime.
 2437 Once the electrons reach the readout chambers, the pad response functions are applied,
 2438 together with an electrons-to-ADC conversion and the ADC saturation limit.

2439 By default, GArSoft applies a 12-bit ADC limit, which can be changed in the
 2440 simulation configuration. However, it can only be increased up to 16-bit, as we represent

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

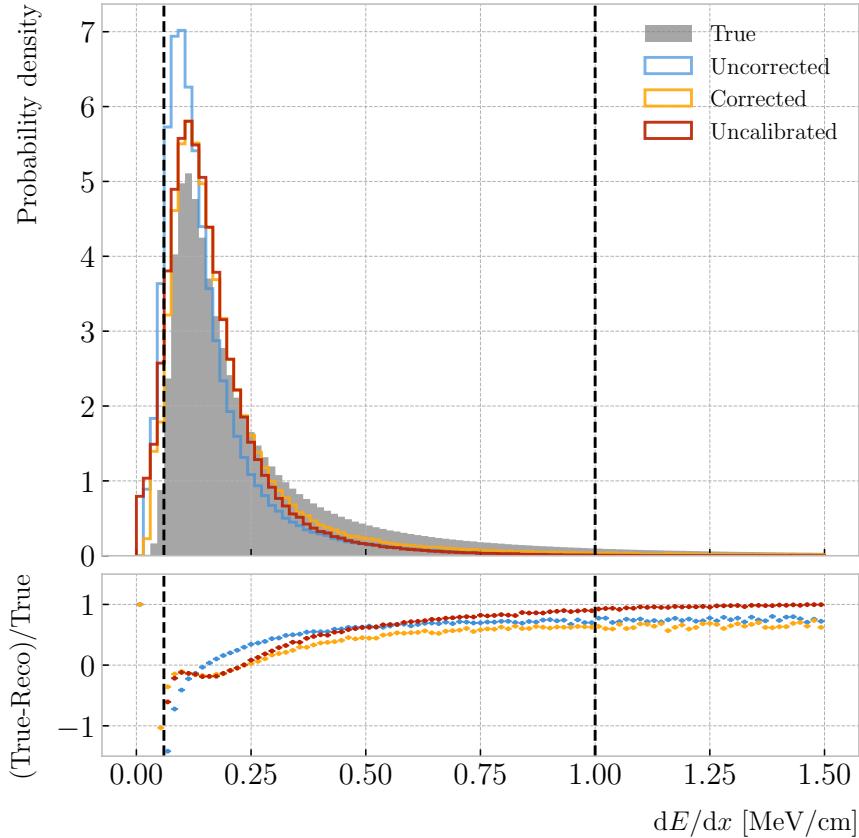


Figure 6.7: Top panel: area normalised dE/dx distributions for the true (solid grey) and the reconstructed energy deposits in the stopping proton sample, both after applying the calibration (blue) and the calibration and the normalisation correction (yellow). Also shown is the distribution obtained by applying a correction factor to the dQ/dx values but not the calibration (red). Bottom panel: fractional residuals for the uncorrected (blue), corrected (yellow) and uncalibrated (red) samples.

2441 the ADC collection as a `std::vector<short>`. This way, I tried to change the saturation
 2442 parameter to see how it affects the relation between reconstructed charge and energy.
 2443 Figure 6.6 shows a comparison between the most probable dQ/dx for 10, 12 and 16-
 2444 bit ADC limits. As expected, the lower the limit is the sooner the charge saturates.
 2445 For higher ADC limits the relation between energy and charge remains linear up to
 2446 higher dE/dx values, but even for the 16-bit limit the saturation is noticeable for values
 2447 $\gtrsim 0.5$ MeV/cm.

2448 Table 6.1 shows the results of fitting the samples with 10 and 16-bits ADC limits to
 2449 the calibration function from Eq. (6.5), using the weights based on their relative error

6.2. dE/dx MEASUREMENT IN THE TPC

as described previously. One interesting feature to notice is how different the best fit points look for the 10-bit ADC saturation when compared to the other two, which are consistent with each other.

At this point we can compare the dE/dx distribution one gets from Geant4, i.e. the true energy loss distribution, and the distribution I found by applying the calibration function to our collection of reconstructed dQ/dx values. Figure 6.7 (top panel) shows the true (solid grey) and reconstructed (blue, labeled as uncorrected) distributions together. The dashed vertical lines indicate the region of validity of the calibration fit, i.e. the left and right edges of the first and last dE/dx bin respectively. Notice that these histograms are area-normalised, as the total number of true energy deposits is much higher than the number of reconstructed charge deposits. This is due to a combination of effects, like the finite spatial resolution of the detector, the hit clustering used in the track fitting and the reclustering we have applied here.

The two distributions are significantly different. That can be seen clearly when looking at the fractional residuals, shown in Fig. 6.7 (bottom panel). In particular, the position of the peak is off, which could bias the mean energy loss predictions. It seems like the difference between these may be due to an overall scaling factor. One possibility is to scale the most probable value of the reconstructed distribution to the most probable value predicted by Geant4. I do this by fitting both distributions using a LanGauss function, obtaining $dE/dx_{MPV, true} = 0.1145 \pm 0.0005$ MeV/cm and $dE/dx_{MPV, reco} = 0.0928 \pm 0.0005$ MeV/cm for the true and reconstructed most probable values respectively. These can be translated into a scaling factor $S = 0.579 \pm 0.006$.

The result of applying the scaling correction can be seen in Fig. 6.7 (top panel). The corrected dE/dx distribution (yellow, labeled as corrected) peaks around the same value the true distribution does, as expected. Moreover, the high energy region is also slightly better described. For low ionisations, below the lower limit of the calibration fit, the differences between true and reconstructed are still significant. This low energy excess may be migration of some events from the peak region. The overall effect of the correction can be seen in the fractional residual plot in Fig. 6.7 (bottom panel).

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

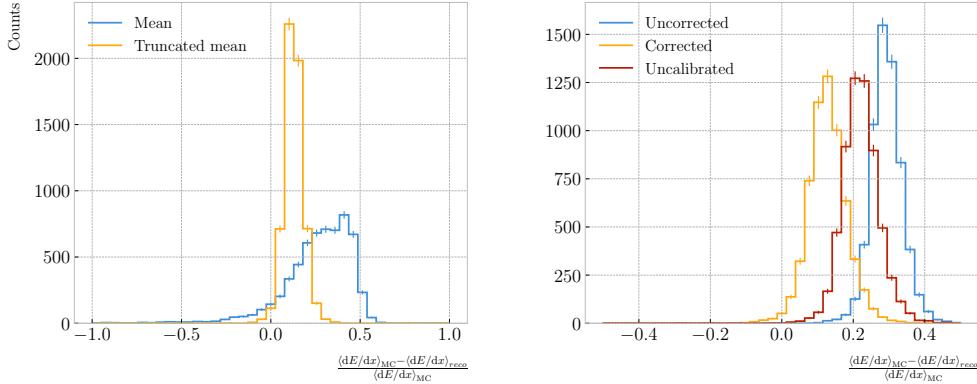


Figure 6.8: Left panel: fractional residuals between the true and the corrected dE/dx means (blue) and the 60% truncated means (yellow), for each event in the stopping proton sample. Right panel: fractional residuals between the true and the uncorrected (blue), corrected (yellow) and uncalibrated (red) dE/dx 60% truncated means, for each event in the stopping proton sample.

2479 One can also check what happens if instead of applying the logarithmic calibration we
 2480 simply scale the dQ/dx distribution (post reclustering) to have the same most probable
 2481 value as the true dE/dx distribution. In this case, following an analogous procedure to the
 2482 one described earlier, I found the scaling factor $S_{uncalibrated} = 0.414 \pm 0.002$ MeV/ADC⁵.
 2483 The resulting distribution (red, labeled as uncalibrated) is also shown in Fig. 6.7 (top
 2484 panel). The behaviour of the new distribution is similar to the corrected case at low
 2485 energy losses, around the peak of the true distribution, but it is worse at describing the
 2486 high energy tail. This is expected, it is in the high ionisation regime where saturation
 2487 effects apply and therefore calibration is needed.

2488 6.2.2 Truncated dE/dx mean

2489 Once we have a collection of dE/dx values for each reconstructed track, we can compute
 2490 the corresponding most probable ionisation loss per unit length of the particle. This
 2491 is the value predicted by the Bethe-Bloch or the PAI models, and together with a
 2492 measurement of the momentum it allows for particle identification.

2493 However, estimating the most probable dE/dx value for each reconstructed track
 2494 is not a trivial task. As mentioned before, the dE/dx distributions follow Landau-like

⁵Notice that now the scaling factor is not dimensionless, as it acts more like a conversion factor here.

6.2. dE/dx MEASUREMENT IN THE TPC

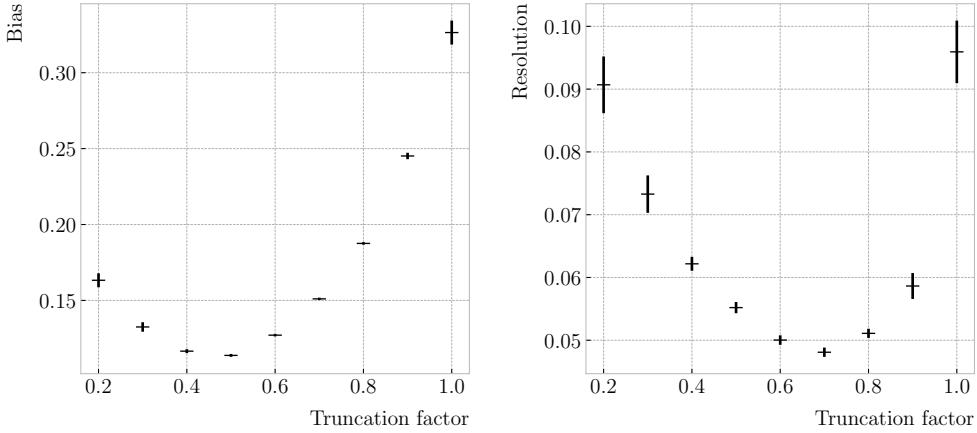


Figure 6.9: Estimated values of the mean dE/dx bias (left panel) and resolution (right panel) obtained using the corrected data from the stopping proton sample, for different values of the truncation factor.

2495 distributions. Therefore, one should perform e.g. a LanGauss fit to correctly estimate
 2496 the most probable values. Automating this kind of fits is often problematic, as they
 2497 usually incur in convergence problems. Moreover, the reconstructed dE/dx distributions
 2498 we obtain tend to have relatively small statistics, which may also produce poor fits. In
 2499 practice, doing these unsupervised fits may degrade our performance, and a more robust
 2500 method is preferred.

2501 A possibility could be taking the mean of the reconstructed dE/dx distribution for
 2502 each particle. The problem with this approach is that the high energy Landau tail,
 2503 combined with our limited statistics, can induce large fluctuations in the computation
 2504 of the mean. Imagine you have two protons with the same kinetic energy, but due to
 2505 reconstruction problems in one case you did not get as many charge deposits reconstructed
 2506 in its high ionisation loss region. If you do not remove the tails the computed dE/dx
 2507 means will be significantly different.

2508 In order to avoid those fluctuations, one can compute the mean of a truncated dE/dx
 2509 distribution instead. By keeping only a given fraction of the lowest energy deposits
 2510 we obtain an estimate of the mean energy loss that is more resilient to reconstruction
 2511 inefficiencies and statistical effects. Figure 6.8 (left panel) shows a comparison between
 2512 the $\langle dE/dx \rangle$ computed by taking the mean of the full distribution (blue line) and the

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

2513 60% lowest energy clusters (yellow line), for the stopping proton sample. The fractional
 2514 residuals are computed for each proton, taking the corresponding means using their
 2515 collections of true and reconstructed energy deposits. One can see that using the simple
 2516 mean translates into a high bias and uncertainty in the $\langle dE/dx \rangle$ estimation, whereas
 2517 applying the truncation reduces both significantly.

2518 Additionally, I performed a comparison between the 60% truncated mean dE/dx
 2519 obtained using the different calibration methods discussed earlier, namely the uncorrected
 2520 (blue), corrected (yellow) and uncalibrated (red) distributions. The results are shown
 2521 in Fig. 6.8 (right panel). While the widths of these distributions are similar, the bias
 2522 obtained for the corrected sample, i.e. calibration function and correction factor applied,
 2523 is a factor of ~ 2 lower than in the uncalibrated case and almost three times smaller
 2524 than for the uncorrected sample.

2525 The next step is to optimise the level of truncation we are going to apply to our
 2526 data. To do so, I used different truncation factors, i.e. the percentage of energy-ordered
 2527 reconstructed energy deposits we keep to compute the mean, on the corrected dE/dx
 2528 sample of the stopping protons. Then, following the same procedure of computing the
 2529 fractional residuals as before, I fitted the resulting histograms using a double Gaussian
 2530 function. This is simply the sum of two Gaussian functions of the type:

$$g(x; \mu, \sigma, A) = A e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (6.6)$$

2531 I do not add the classical normalisation factor of the Gaussian, $1/\sqrt{2\pi}\sigma$, therefore
 2532 the amplitude A simply represents the maximum of the function. One of the two
 2533 Gaussian functions describes the core part of the distribution, while the other captures
 2534 the behaviour of the tails.

2535 For each truncation factor, I look at the bias and the resolution I obtain. I define
 2536 these as the weighted means of the corresponding parameters in the fits:

$$\bar{x} = \frac{A_{core} x_{core} + A_{tail} x_{tail}}{A_{core} + A_{tail}}, \quad (6.7)$$

6.2. dE/dx MEASUREMENT IN THE TPC

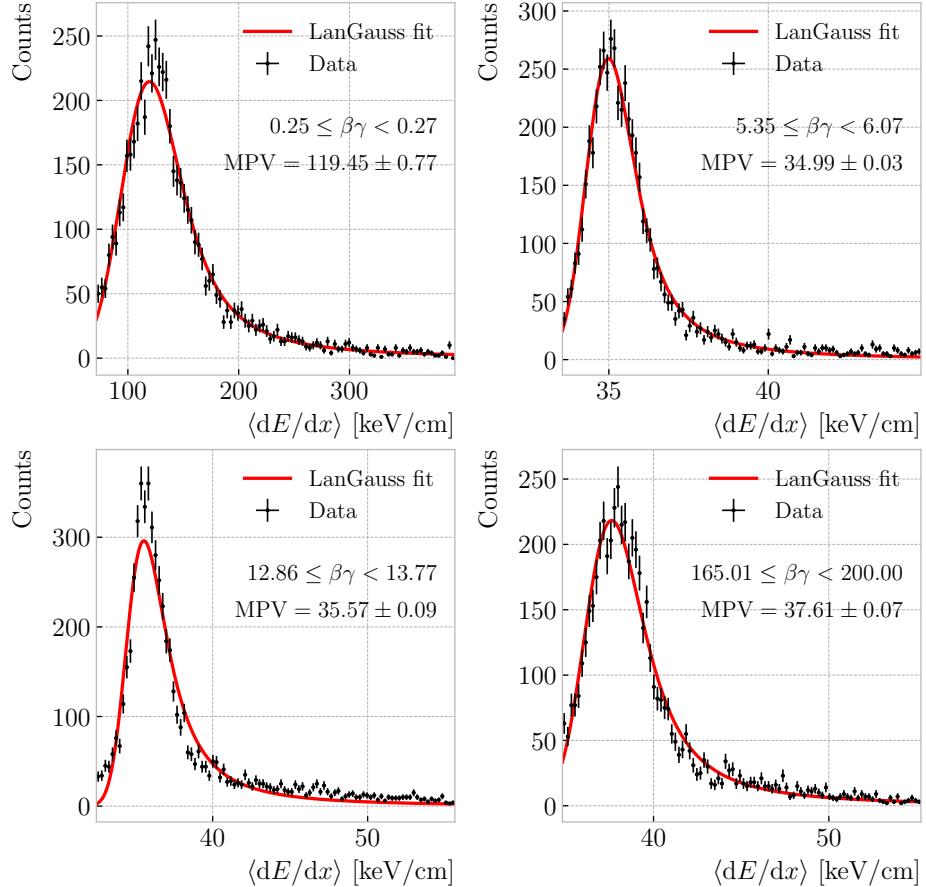


Figure 6.10: Examples of the truncated mean dE/dx LanGauss fits for various $\beta\gamma$ bins, from a simulated FHC neutrino sample.

2537 where A_{core} and A_{tail} are the amplitudes of the core and tail distributions respectively
 2538 and x is either the mean μ or the width σ of said distributions.

2539 Figure 6.9 shows the bias (left panel) and the resolution (right panel) I obtained
 2540 for the stopping proton sample, using different values of the truncation. From these, it
 2541 can be seen that a truncation factor of 50% minimises the bias in the estimation, while
 2542 70% gives the best resolution. That way, I settled on the intermediate value of 60%
 2543 truncation, which yields a $\langle dE/dx \rangle$ resolution of 5.00 ± 0.08 % for stopping protons.

2544 **6.2.3 Mean dE/dx parametrisation**

2545 Now that we have a way to estimate the mean energy loss of a particle in the HPgTPC,
 2546 we can determine the value of the free parameters in the ALEPH formula, Eq. (6.3).

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2547 For this, I used a sample of 10^5 reconstructed FHC neutrino events inside ND-GAr. In
2548 this case I cannot use the stopping proton sample, as we need to cover the full kinematic
2549 range of interest for the neutrino interactions in our detector.

2550 The original data does not contain an estimation of the velocity of the tracks, instead
2551 the tracks have a value for the reconstructed momentum and the associated PDG code
2552 of the Geant4-level particle that created the track. Therefore, one can select some of the
2553 particles in the data, in this case I selected electrons, muons, pions and protons, and
2554 compute β and γ using the reconstructed momentum and their mass. In terms of $\beta\gamma$
2555 the mean dE/dx does not depend on the particle species, so one can consider all the
2556 dataset as a whole. For this fit, I will express β in terms of the $\beta\gamma$ product as:

$$\beta = \frac{\beta\gamma}{\sqrt{1 + (\beta\gamma)^2}}, \quad (6.8)$$

2557 which can be easily proven from the definition of γ .

2558 Next, I bin the data in $\beta\gamma$. I chose a fine binning so as to capture the different
2559 features of the ionisation curve. Instead of fixing the bin width, I select them so each one
2560 has approximately the same statistics. Then, for each $\beta\gamma$ slice, I compute the median
2561 and the interquartile range (IQR) of the $\langle dE/dx \rangle$ distribution. Using these, I make a
2562 histogram in the range [median - IQR, median + 5 IQR], which I fit to a LanGauss
2563 function in order to extract the MPV. Using this range accounts for the asymmetric
2564 nature of the distributions, while also helps avoiding a second, lower maximum present
2565 at low $\beta\gamma$, probably a result of reconstruction failures.

2566 A few examples of these fits are shown in Fig. 6.10. The chosen values of $\beta\gamma$ sit in
2567 very distinct points along the $\langle dE/dx \rangle$ curve, going from the high ionisation region at
2568 low velocities (top left panel), to the minimum point (top right panel), the beginning of
2569 the relativistic rise (bottom left panel), and the plateau produced by the density effect
2570 (bottom right panel).

2571 I used the resulting most probable $\langle dE/dx \rangle$ values and the centres of the $\beta\gamma$ bins as
2572 the points to fit to the ALEPH formula. For this particular fit I used the least-squares

6.2. dE/dx MEASUREMENT IN THE TPC

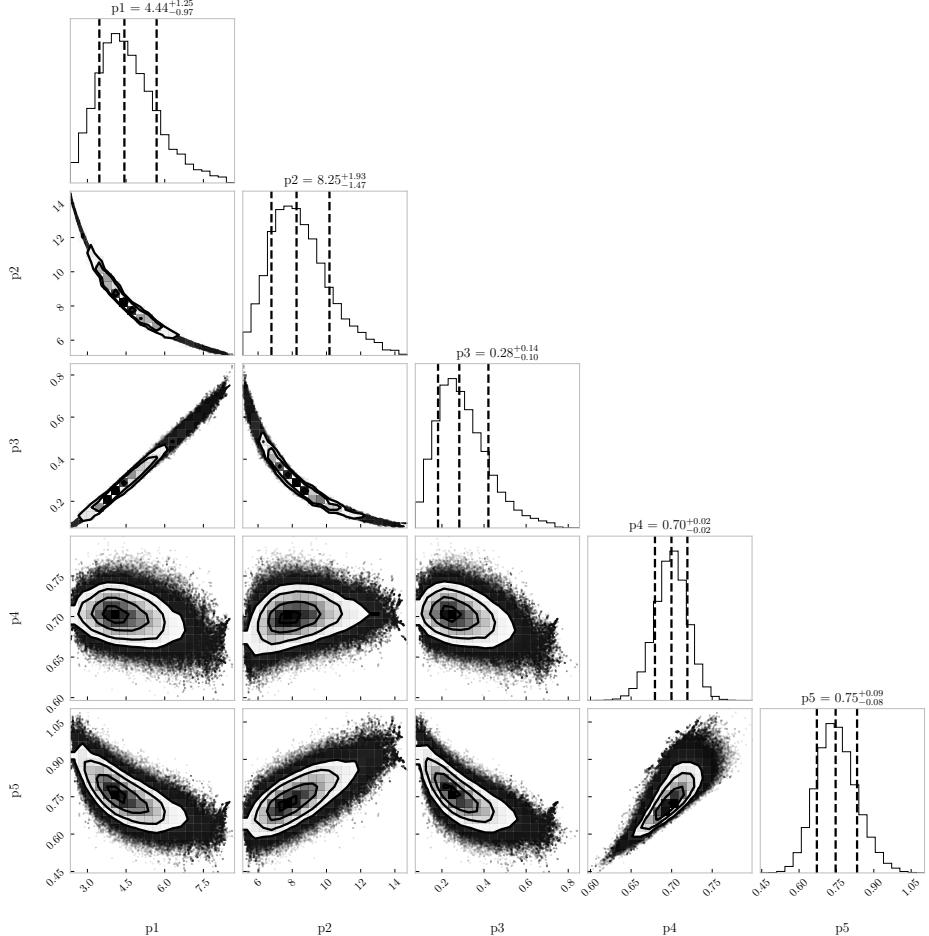


Figure 6.11: Resulting one and two dimensional projections of the posterior probability distributions of the ALEPH $\langle dE/dx \rangle$ parameters obtained by fitting the 60% truncated mean dE/dx values from a FHC neutrino sample in ND-GAr. The vertical dashed lines in the 1D distributions represent the 16th, 50th and 84th percentiles.

method to get a first estimation of the ALEPH parameters. Applying some uniform priors, I then used these values as the starting point of a 100000 steps MCMC. Figure 6.11 shows the posterior probability distributions I obtain for each parameter. The reported best fit points are based on the 16th, 50th, and 84th percentiles in the marginalised distributions.

The resulting fit (black line), compared to the data points (red points) and the underlying distribution is shown in Fig. 6.12 (top panel). The overall fit is good, with a reduced chi-squared of $\chi^2/ndf = 1.02$. However, there are some regions where the fit does not describe the data correctly, like the very low $\beta\gamma$ regime, where the fit severely

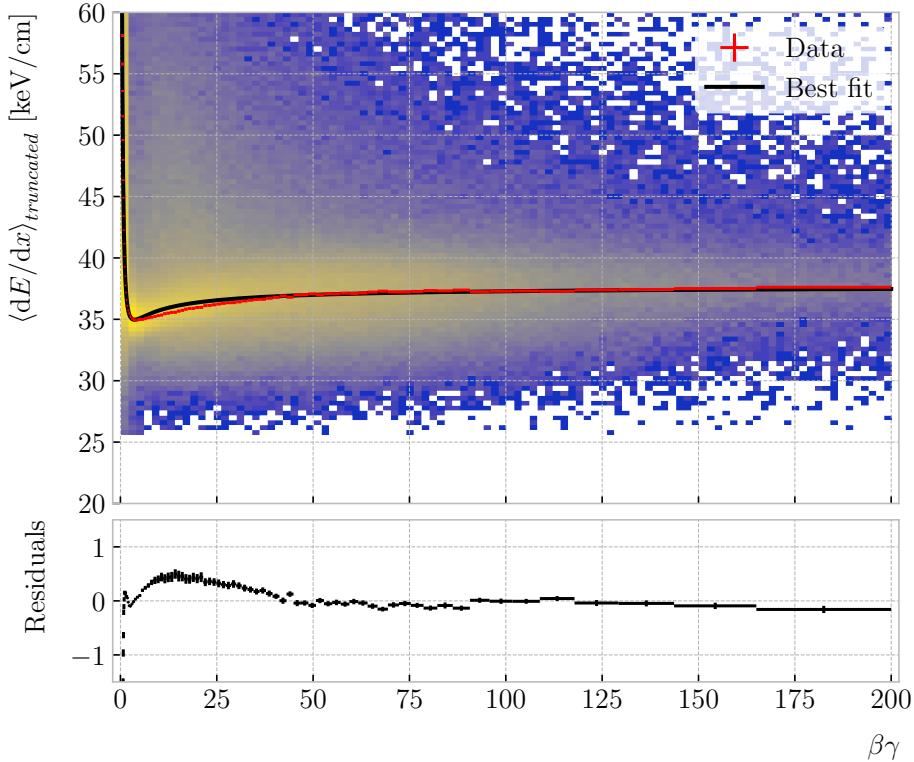


Figure 6.12: Truncated mean dE/dx obtained for the FHC neutrino sample as a function of the $\beta\gamma$ product (upper panel). Also shown are the fitted most probable values for each $\beta\gamma$ bin (red points) and the best fit obtained using the ALEPH parametrisation (black line). The residuals resulting from the fit are shown in the lower panel.

underestimates for energy losses $\gtrsim 50$ keV/cm, and the start of the relativistic raise, where we have a slight overestimation. This is a result of those points having a larger uncertainty when compared to the ones around the dip or the plateau areas. These differences can be better seen in the residual plot, Fig. 6.12 (bottom panel).

6.2.4 Particle identification

6.3 Muon and pion separation in the ECal and MuID

As it could be seen from Fig. 6.13, it is not possible to separate muons and charged pions in the HPgTPC using dE/dx for momenta $\gtrsim 300$ MeV/ c . In ND-GAr, approximately 70% of the interactions in FHC mode will be ν_μ CC (compared to the 47% of $\bar{\nu}_\mu$ CC interactions when operating in RHC mode), while 24% are neutral currents. Out of

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

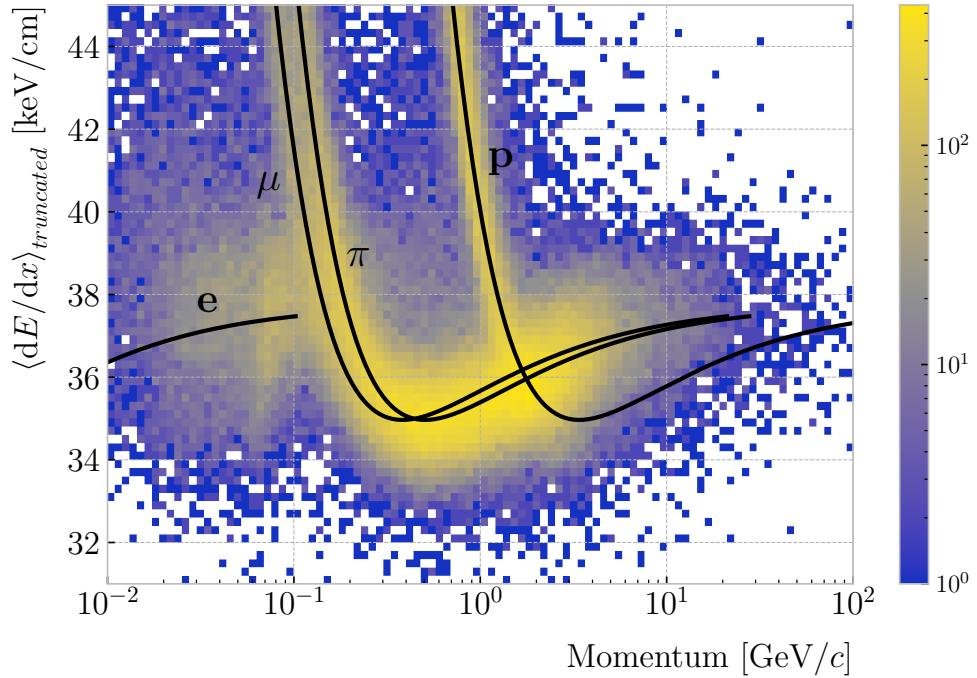


Figure 6.13: Distribution of the 60% truncated mean dE/dx versus reconstructed momentum for the FHC neutrino sample. The black lines indicate the predictions of the ALEPH parametrisation for electrons, muons, charged pions and protons.

these, around 53% and 47% of them will produce at least one charged pion in the final state, respectively. Figure 6.15 shows a comparison between the spectra of the primary muons and the charged pions for ν_μ CC interactions in ND-GAr producing one or more charged pions. From this, one can see that (i) the majority of muons and charged pions are not going to be distinguishable with a $\langle dE/dx \rangle$ measurement, and that (ii) particle identification is necessary both to classify correctly the ν_μ CC events and identify the primary muon within them.

ND-GAr features two other subdetectors which can provide additional information for this task, namely the ECal and MuID. The current ECal design, described in (ref section), consists of 42 layers, made of 5 mm of Pb, 7 mm of plastic scintillator and a 1 mm PCB board. The total thickness of this calorimeter is 1.66 nuclear interaction lengths or 1.39 pion interaction lengths. The MuID design is in a more conceptual stage, however it is envisioned to feature layers with 10 cm of Fe and 2 cm of plastic

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

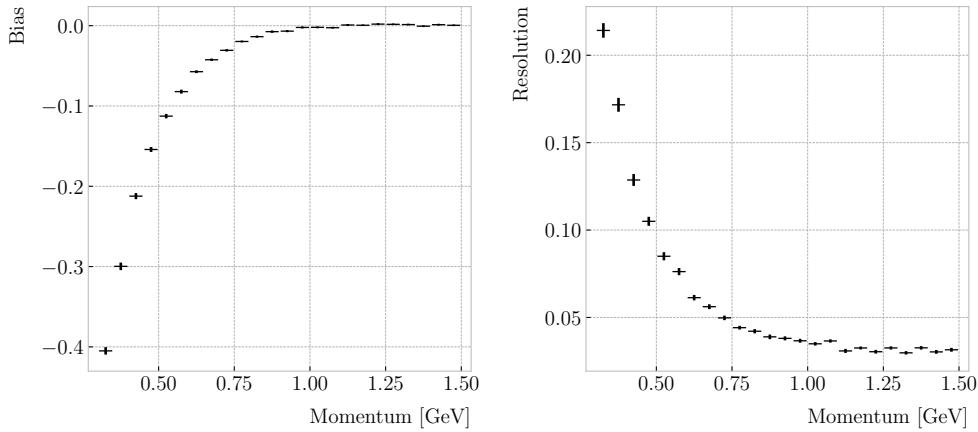


Figure 6.14: Estimated values of the mean dE/dx bias (left panel) and resolution (right panel) obtained for the true protons in a FHC neutrino sample.

scintillator⁶. With its three layers, it will have a thickness of 1.87 or 1.53 nuclear or pion interaction lengths, respectively.

Because pion showers are dominated by inelastic nuclear interactions, the signatures of these particles in the calorimeter will look significantly different from those of muons. Although our ECal is not thick enough to fully contain the hadronic showers of the charged pions at their typical energies in FHC neutrino interactions, they can still be used to understand whether the original particle was more hadron-like or MIP-like. In Fig. 6.16 I show two examples of energy distributions created by a muon (left panel) and a charged pion (right panel) of similar momenta interacting in the ECal. These figures represent the transverse development of the interactions. For each of them, I computed the principal component and centre of mass of the interaction, projecting the position of the hits onto the plane perpendicular to that direction, and taking the distances relative to the centre. It can be seen that the muon follows an almost MIP-like behaviour, being the central bin in the histogram the one with the highest deposited energy. On the other hand, the pion not only deposits more energy overall, but also this energy is more spread-out among the different hits. It is this kind of information that would allow us to tell apart muons from pions.

⁶It is not mentioned anywhere, but I assume that there should also be another layer of PCB board of 1 mm. However, in this case its contribution to the total thickness of the sampling calorimeter would be negligible.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

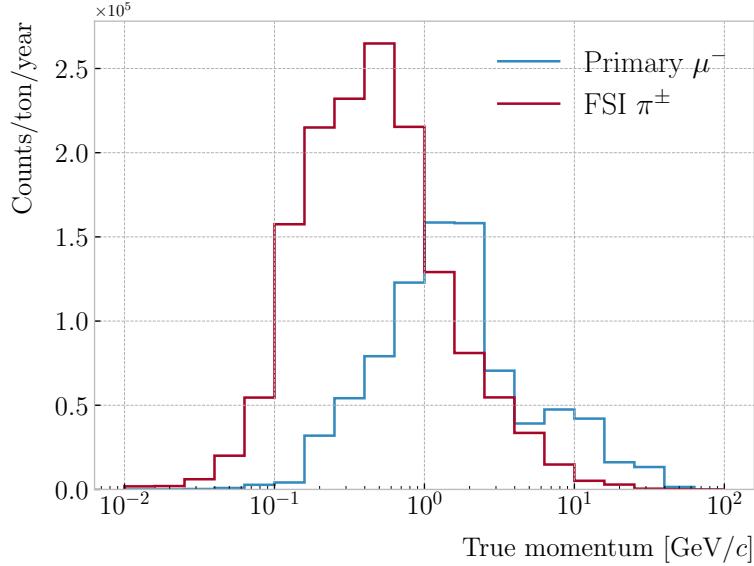


Figure 6.15: True momentum distribution for the primary muon in ν_μ CC $N\pi^\pm$ interactions inside the fiducial volume of ND-GAr (blue line), compared to the post FSI charged pion spectrum (red line).

- 2622 This way, I identify three main action points that need to be addressed if one wants
 2623 to use these detectors to distinguish between muons and charged pions. These are:
 2624 1. the way we make the associations between tracks in the HPgTPC to the activities
 2625 (what in GArSoft we call clusters) in the ECal and the MuID,
 2626 2. what variables or features one can extract from the calorimeters that encapsulate
 2627 the information we are interested about,
 2628 3. and how to carry out the classification problem.

2629 6.3.1 Track-ECal matching

- 2630 One of the main players in the muon and pion separation is the way we associate clusters
 2631 in the ECal to reconstructed tracks in the TPC. Missing some associations or making
 2632 wrong ones can bias the ECal quantities that we can use for classifying particles. The
 2633 current algorithm in GArSoft provides precise associations, i.e. most of the associations
 2634 that it produces are correct, but it appears to miss an important number of associations
 2635 (at least when using the default configuration).

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

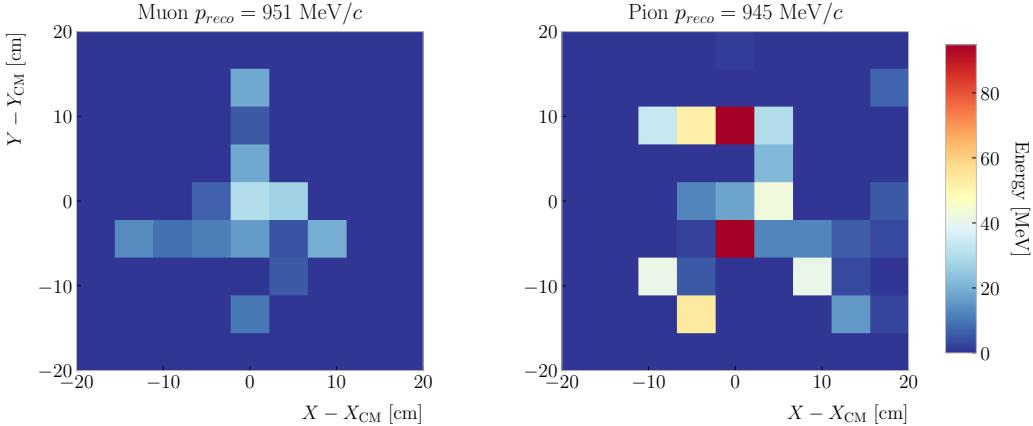


Figure 6.16: Distributions of energy deposits in the ECal for a muon (left panel) and a charged pion (right panel) with similar momenta. The energy is projected onto the plane perpendicular to the principal component of the hits, and the positions are relative to the center of the interaction.

2636 The current TPC track-ECal cluster association algorithm is divided in four parts.
 2637 It first checks whether the track end point fulfils certain conditions to be extrapolated.
 2638 There are two cut values in this step, one for the drift direction and other radial.

2639 If the point can be extrapolated, the code computes the coordinates of the centre
 2640 of curvature using the Kalman fit estimates at the track end (y , z , $1/R$, ϕ , $\tan\lambda$). It
 2641 then compares the distance between this and the cluster in the (z, y) plane with R . This
 2642 introduces another cut in the perpendicular direction.

2643 The next step is different for clusters in the barrel or in one of the end caps. If it
 2644 is a barrel cluster the algorithm extrapolates the track up to the radial distance of the
 2645 cluster. There are three possible outcomes, the extrapolated helix can cut the cylinder
 2646 of radius r_{clus} two, one or zero times. I get the cut point that is closer to the cluster and
 2647 check that it is either in the barrel or the end caps. Computing the difference between
 2648 the x coordinates of the cluster and the extrapolated point, the module checks that this
 2649 is not greater than a certain cut. If the cluster is in an end cap, I propagate the track
 2650 up to the x position of the cluster. Then, the algorithm computes the angle in the (z, y)
 2651 plane between the centre of curvature and the cluster, α , and the centre of curvature
 2652 and the propagated point, α' . A cut is applied to the quantity $(\alpha - \alpha')R$.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

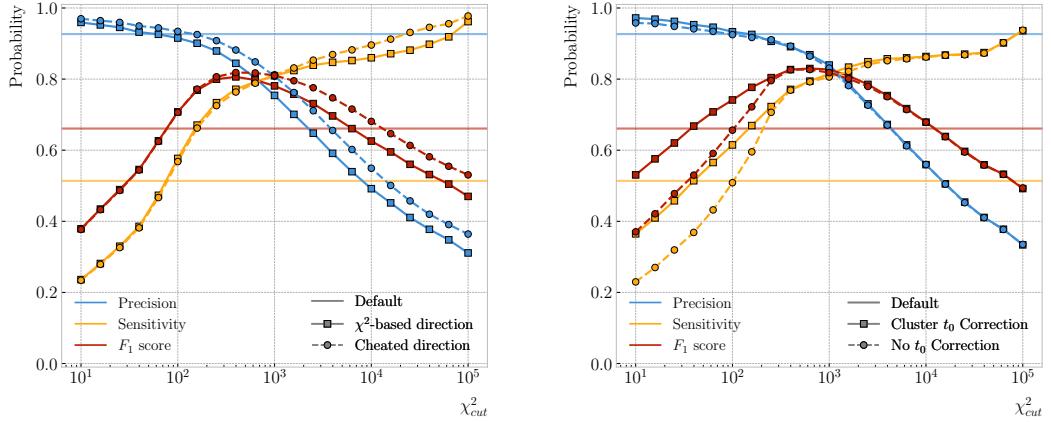


Figure 6.17: Left panel: comparison between the precision (blue), sensitivity (yellow) and F_1 score (red) obtained for the default (horizontal lines) and new algorithms, both with the χ^2 -based direction estimator (squares) and cheating the directions (circles), for different values of the *chi*² cut. Right panel: comparison of the performance of the new algorithm when applying the cluster t_0 correction (squares) and when (circles).

2653 If the cluster contains more than a certain number N of hits, I apply an extra cut to
2654 the dot product of the direction of the track at the propagated x value and the cluster
2655 direction.

2656 The code makes sure to only associate one end of the track (if any) to a cluster.
2657 However, it can associate more than one track to the same cluster. This makes sense,
2658 as different particles can contribute to the same cluster in the ECal, but it makes it
2659 difficult to quantify the relative contributions of the tracks to a certain cluster.

2660 As a way of comparing the performance of this algorithm, a new, simpler association
2661 module was written. The goal was to have a simple and robust algorithm, which depends
2662 on as few parameters as possible and that can produce a one-to-one matching between
2663 tracks and ECal clusters.

2664 For each reconstructed track, the new algorithms applies the same procedure to the
2665 forward and the backward fits irrespective of their end point positions. It first gets the
2666 Kalman fit parameters at the corresponding end point together with the X position, x_0 ,
2667 (y_0 , z_0 , $1/R$, ϕ_0 , $\tan\lambda$).

2668 For each ECal cluster, I compute the radial distance to the centre of the TPC and
2669 find the ϕ value in the range $[\phi_0, \phi_0 + \text{sign}(R)\phi_{max}]$ that makes the propagated helix

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

2670 intersect with the circle defined with such radius. The (x, y, z) position of the helix for
2671 the ϕ value found (if any) is then computed. In case there are two intersections, I keep
2672 the one that minimises the distance between (y, z) and (y_c, z_c) .

2673 I then calculate χ^2 value based on the Euclidean distance between the propagated
2674 point and the cluster:

$$\chi^2/ndf = \frac{\sum_{n=0}^2 (x^{(n)} - x_c^{(n)})^2}{3}. \quad (6.9)$$

2675 If there was no intersection I store a -1 instead. In the end, for each reconstructed track
2676 in the event one ends up with two collections of χ^2 values, one for each ECal cluster
2677 and fit directions.

2678 The current code only supports having ECal clusters associated to one end of each
2679 track. We have two options to decide what track end to keep. The first one tries to
2680 cheat the selection, looking at the distance between the two track ends and the true
2681 start position of the associated MC particle. The second one keeps the track end with
2682 more χ^2 entries below the cut.

2683 This feature of only considering one track end limits the algorithm, making it not
2684 suitable for reconstructing events with particles originating outside the TPC. However,
2685 as for the moment the main concern of the group is the study of neutrino interactions
2686 off the gaseous argon, this is an acceptable assumption.

2687 In order to associate a cluster to a track, I take all clusters with a χ^2 value in the
2688 range $[0, \chi_{cut}^2]$. If a cluster has been assigned to more than one track we leave it with
2689 the one with the lowest χ^2 .

2690 This default behaviour of the algorithm can be modified to associate more than one
2691 track to each cluster. Not only that, but the χ^2 values can be used to assign relative
2692 weights to the different contributions.

2693 To evaluate the performance of the association method, I use a binary classification
2694 approach. In this case, I check the leading MC Track IDs associated to the reconstructed
2695 tracks and ECal clusters. I count an association as true positive (TP) if both Track
2696 IDs coincide. An association is considered false positive (FP) when the Track IDs are

6.3. MUON AND PION SEPARATION IN THE ECal AND MUID

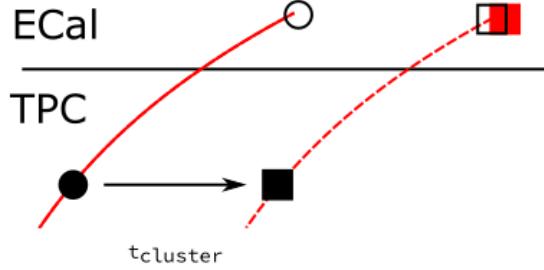


Figure 6.18: Schematics of a possible option to deal with track-ECal associations in non-zero t_0 neutrino interaction events, trying to correct for the drift direction uncertainty in a cluster-by-cluster basis using the cluster time, $t_{cluster}$.

2697 different. If a cluster has not been associated to any track but it shares the Track ID
 2698 with a reconstructed track it is counted as a false negative (FN).

2699 For the testing, I used a sample of 10000 FHC neutrino events inside the HPgTPC.
 2700 Figure 6.17 (left panel) shows the precision (blue line), sensitivity (yellow line), and F_1
 2701 score (red line) I obtained for different values of χ^2_{cut} . For comparison, the same metrics
 2702 computed for the default algorithm with the current configuration are also shown (dashed
 2703 lines). In the case of the new algorithm, I used both the χ^2 -based method to estimate
 2704 the track direction described earlier (square markers) and the cheated direction from the
 2705 Geant-level information (circle markers). For either of these we achieve similar values of
 2706 the precision compared to the old code, while having a considerably higher sensitivity.
 2707 It can be seen that cheating the direction of the tracks only makes a difference at high
 2708 χ^2_{cut} , past the optimal value of the cut around the F_1 score maximum. Therefore, I set
 2709 the χ^2 method as the default.

2710 One of the possible weak points of this approach is that it relies on the position along
 2711 the drift direction to make the decisions. Within the current ND-GAr design implemented
 2712 in GArSoft, the timing information is provided by the ECal. That effectively means
 2713 that prior to make the track-ECal associations the reconstructed x positions of the track
 2714 trajectories differ from the simulated ones by an amount:

$$x_{reco}^{(n)} - x_{sim}^{(n)} = v_{drift} t_0, \quad (6.10)$$

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

where v_{drift} is the mean drift velocity in our medium and the initial time is in the range $t_0 \in [0, t_{spill}]$ where t_{spill} is the spill length. For a $10 \mu\text{s}$ spill this translates into a maximum 30 cm uncertainty on the drift direction position.

The current default in GArSoft sets $t_0 = 0$, but the functionality to randomly sample this within the spill time is in place. Therefore, we need to understand what is the impact of a non-zero t_0 on the associations algorithm and foresee possible ways of minimising a loss in performance.

Figure 6.18 represents a possible option to tackle the association problem when having events with a non-zero initial time t_0 . The black and white circles represent the original points, whereas the squares indicate the corrected positions. The end points of the track and the propagated points up to the cluster radius are indicated using filled and unfilled markers respectively. The red square represents the position of the cluster.

Here I try to correct for the drift coordinate position using the time associated to the cluster. Assuming that the drift time is much larger than the propagation time, $t_{cluster}$ could be used as a good estimation of the t_0 . An alternative can be using the earliest time associated to a hit in said cluster. Doing this for each cluster before computing the χ^2 value could be used as an alternative to knowing the specific value of the t_0 , as when the association is correct this will provide the right correction but its impact is small enough to not change the position significantly in the case the cluster does not correspond to a given track.

I tested the effect of this correction again using a sample of 10000 FHC neutrino events. Figure 6.17 (right panel) shows the precision (blue line), sensitivity (yellow line), and F_1 score (red line) for the case the cluster t_0 correction is applied (square markers) and for the no correction case (circle markers), as a function of χ^2_{cut} . In this case, the differences are particularly notorious at low values of the cut. It makes sense, as the t_0 effect becomes subdominant when the distance we consider grows large. Overall, the correction increases the sensitivity while keeping the precision almost unchanged. As a result, I apply the t_0 correction to the generated samples as the default.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

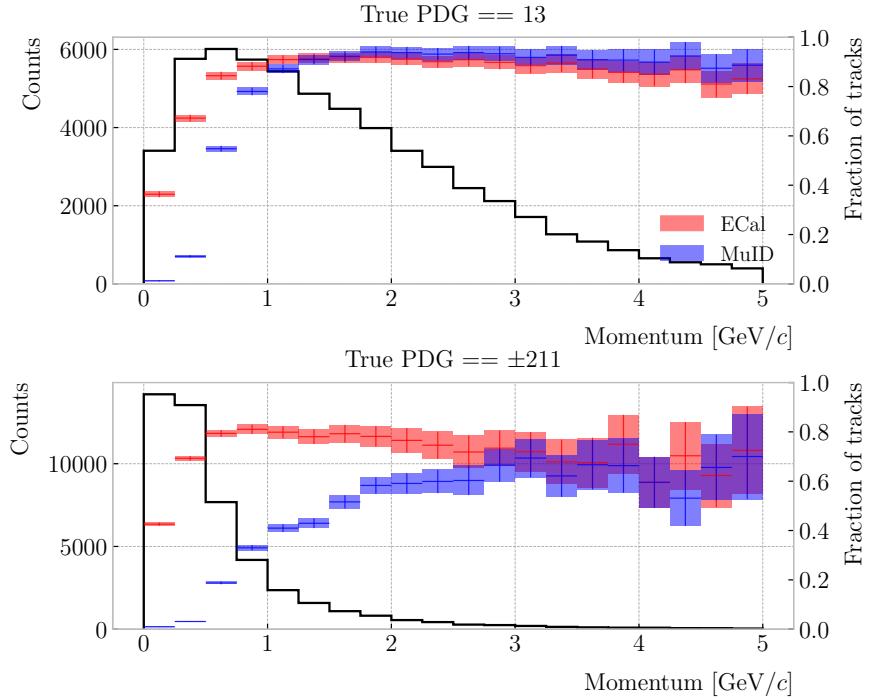


Figure 6.19: Momentum distribution for the reconstructed muons (top panel) and charged pions (bottom panel) in a FHC neutrino sample, together with the fraction of them reaching the ECal (red) and MuID (blue). Each entry corresponds to a reconstructed track, backtracked to a true muon or pion which has not produced any other reconstructed track.

2743 6.3.2 Classification strategy

2744 The problem of the muon and charged pion separation has to be viewed in the broader
 2745 context of the particle identification in our detector. Focusing on the beam neutrino
 2746 interactions, it is clear that we are going to have muons and pions spanning a broad
 2747 momentum range. Not only that, but we will also have other particles with similar
 2748 characteristics that will make the classification even more challenging. Therefore, we are
 2749 presented with a task that will depend heavily on the kinematic range we are looking at
 2750 each time, as both the available information and the possible impurities of other particle
 2751 species vary.

2752 For instance, distinguishing muons from pions could be difficult at low momenta, as
 2753 a great number of them do not reach the ECal. Therefore, we could think of tailoring a
 2754 version of the classification for that particular case, which could be complemented with

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

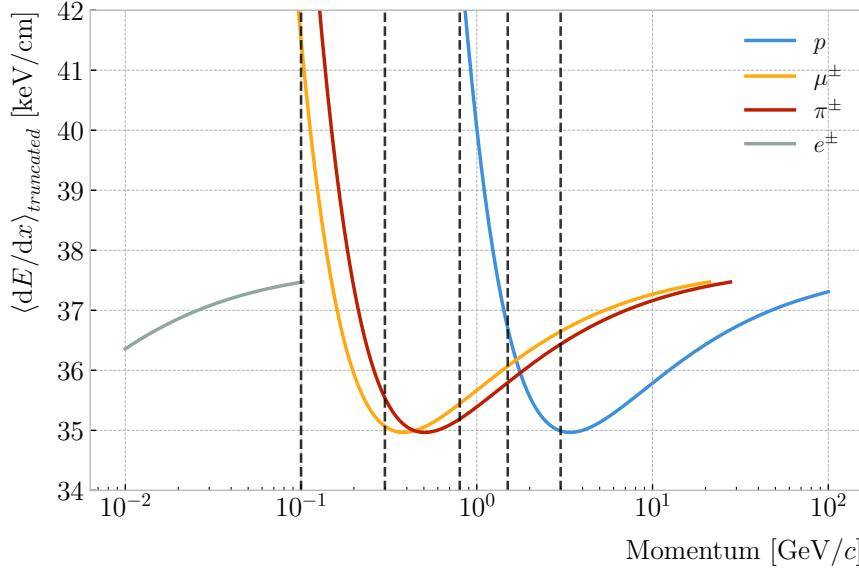


Figure 6.20: Predicted truncated mean dE/dx versus momentum, for electrons, muons, charged pions and protons, obtained using the ALEPH parametrisation. The vertical dashed lines represent the boundaries of the six regions used for the muon and pion classification training.

2755 a dE/dx measurement. Likewise, for momenta $\gtrsim 1$ GeV muons and pions reach the
 2756 calorimeters efficiently, but so do protons. Because of this, one can try to train another
 2757 classifier for this energy range, and rely on other methods to remove as many of the
 2758 protons as possible.

2759 Figure 6.19 shows the momentum distribution of the reconstructed muons (top) and
 2760 pions (bottom) in a FHC sample. It also contains the fraction of particles reaching the
 2761 ECal (red) and MuID (blue), for the different momentum bins. In Fig. 6.20 I show the
 2762 mean dE/dx of different particles as a function of the momentum, computed using the
 2763 ALEPH parametrisation with the best fit parameters found in Subsec. 6.2.2.

2764 Using these two figures as references, I decided to approach the classification by
 2765 dividing the problem into six different momentum regions. A summary of these can be
 2766 found in Tab. 6.2. The basic idea is to exploit all the information that is available in
 2767 each region and . For the problem at hand, I prepared separated samples of isotropic
 2768 single muons and pions, with momenta uniformly distributed along the corresponding
 2769 momentum range. Each sample contains 50000 events of the corresponding particle

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

Table 6.2: Momentum ranges and description of the PID approach assumed for the muon and pion classification task.

Momentum range	Description
$< 0.1 \text{ GeV}/c$	All tracks can be separated with dE/dx
$[0.1, 0.3) \text{ GeV}/c$	Use ECal for reaching muons and pions, dE/dx for the rest
$[0.3, 0.8) \text{ GeV}/c$	Use ECal for muons and pions, dE/dx for protons
$[0.8, 1.5) \text{ GeV}/c$	Use ECal and MuID for muons and pions, dE/dx for protons
$[1.5, 3.0) \text{ GeV}/c$	Use ECal and MuID for muons and pions, ToF for protons
$\geq 3.0 \text{ GeV}/c$	Use ECal and MuID for muons and pions, dE/dx and ToF for protons

2770 species. I did not generate samples for the first region, as it is assumed that the separation
 2771 can be achieved using dE/dx only. For the last region, I generated particles up to a
 2772 momentum of $10 \text{ GeV}/c$, as that is well above the typical energies of muons and pions
 2773 from FHC neutrino interactions in ND-GAr.

2774 Additionally, I prepared another sample of 100000 FHC neutrino events. For each
 2775 interaction, I select the reconstructed particles which were backtracked to true muons or
 2776 charged pions. I use this dataset to perform validation checks, to see how the models
 2777 trained with the single particle data generalise to a more realistic scenario.

2778 To tackle this classification problem, I make use of Boosted Decision Trees (BDT). A
 2779 decision tree uses a flowchart-like structure to make decisions based on some input data.
 2780 It starts from a root node, which represents the complete dataset, and then it splits
 2781 this based on the variable or feature which gives the best separation between classes,
 2782 creating two new nodes. The process repeats for each node until it reaches a certain
 2783 limit, like a maximum number of splits or some tolerance criteria. The last set of nodes
 2784 are often called leave nodes, and represent the final prediction of the classifier.

2785 Boosting refers to a family of methods to combine the predictions from multiple
 2786 classifiers, following a sequential approach where each new model learns from the errors
 2787 of the previous one. The process starts with a simple decision tree, which is used to
 2788 make predictions on the training data. Then, the data points misclassified by the first
 2789 model are assigned higher weights, and another decision tree is trained on the data with

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

adjusted weights. The predictions of the two trees are then combined, and the cycle repeats for a predefined number of iterations. Gradient boosting uses the direction of the steepest error descent to guide the learning process and improve the accuracy with each iteration.

6.3.3 Feature selection and importance

Using the reconstructed tracks as a starting point, I compute a number of ECal and MuID variables for each of them. As there can be more than one cluster associated to a track, what I do is collect all associated clusters and compute these variables from the complete collection of associated hits. For the MuID, because it only features three layers and typically there will be less hits, I also allow single hits to be associated with tracks⁷. I can roughly divide the variables in three types: energy-related, geometry-related and statistical. In the following, I briefly describe the variables related exclusively to the ECal:

• Energy-related ECal

- ECal total energy (ClusterTotalEnergy): sum of the energy of all the ECal hits.
- Mean ECal hit energy (HitMeanEnergy): mean of the hit energy distribution.
- Standard deviation ECal hit energy (HitStdEnergy): standard deviation of the hit energy distribution.
- Maximum ECal hit energy (HitMaxEnergy): maximum of the hit energy distribution.

• Geometry-related ECal

- Mean distance hit-to-cluster (DistHitClusterMean): mean of the distance distribution between the hits and the corresponding cluster's main axis.

⁷At the reconstruction level what happens is that non-clustered hits are put into single hit clusters, instead of being thrown away. This is necessary to keep the consistency of the track-cluster association code.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

- 2814 – RMS distance hit-to-cluster (DistHitClusterRMS): root mean square of the
2815 distance distribution between the hits and the corresponding cluster's main
2816 axis.
- 2817 – Maximum distance hit-to-centre (DistHitCenterMax): maximum of the
2818 distance distribution between the hits and the centre of the TPC.
- 2819 – Time-of-Flight velocity (TOFVelocity): slope obtained when fitting a straight
2820 line to the hit time versus hit distance to the centre (i.e. $d = v \times t$).

2821 • Energy and geometry ECal

- 2822 – Radius 90% energy (Radius90E): distance in the hit-to-cluster distribution
2823 for which 90% of the total energy is contained in the hits that are closer to
2824 the axis (i.e. radius that contains 90% of the energy).

2825 • Statistical ECal

- 2826 – Number of hits (NHits): total number of hits associated to the track.
- 2827 – Number of layers with hits (NLayers): not really a count of all layers with
2828 hits but the difference between the last and the first layer with hits.

2829 Figure 6.21 shows the distributions of three different ECal variables, separating true
2830 muons (blue) and charged pions (red), for the five momentum ranges considered. I chose
2831 to show one feature from each category, namely the mean energy per hit (left column),
2832 the mean distance between the hits and the centre of the cluster (middle column), and
2833 the number of ECal layers with hits (right column). These give an idea of the separating
2834 power of the different features, and how it changes considerably with the energy. In
2835 the number of layers with hits distributions, the peak at 6 is due to the fact that the
2836 first six ECal layers sit inside the pressure vessel⁸. Therefore, some of the particles get
2837 stopped crossing it, never making it to the seventh layer.

⁸Note to self: check this. I thought the ECal barrel had 8 layers of tiles, and that all of them were inside the pressure vessel.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

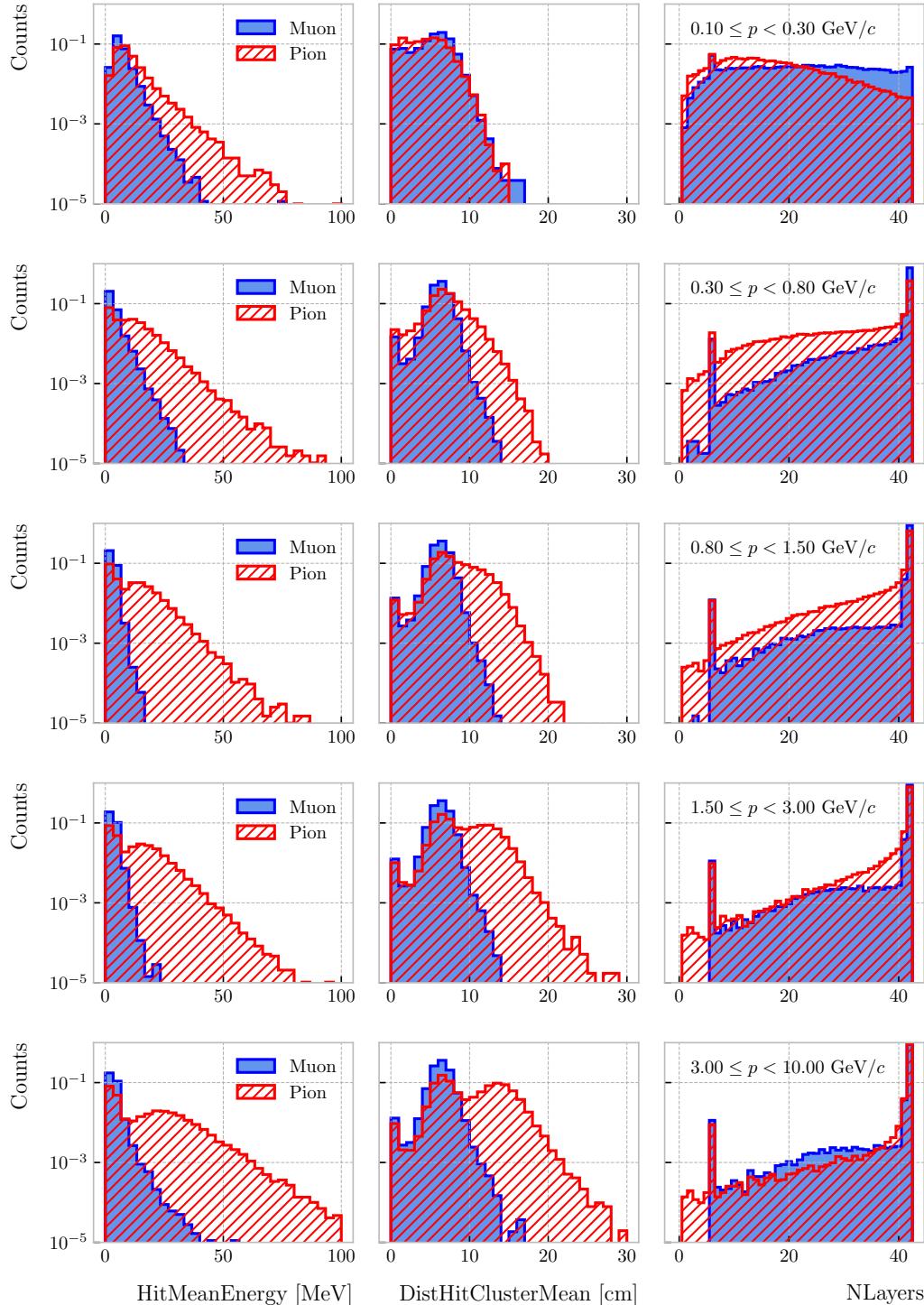


Figure 6.21: Example ECal feature distributions for muons (blue) and charged pions (red) in the five different momentum ranges considered (from top to bottom, in ascending momentum order). From left to right: mean hit energy, mean distance hit-to-cluster, and number of layers with hits.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

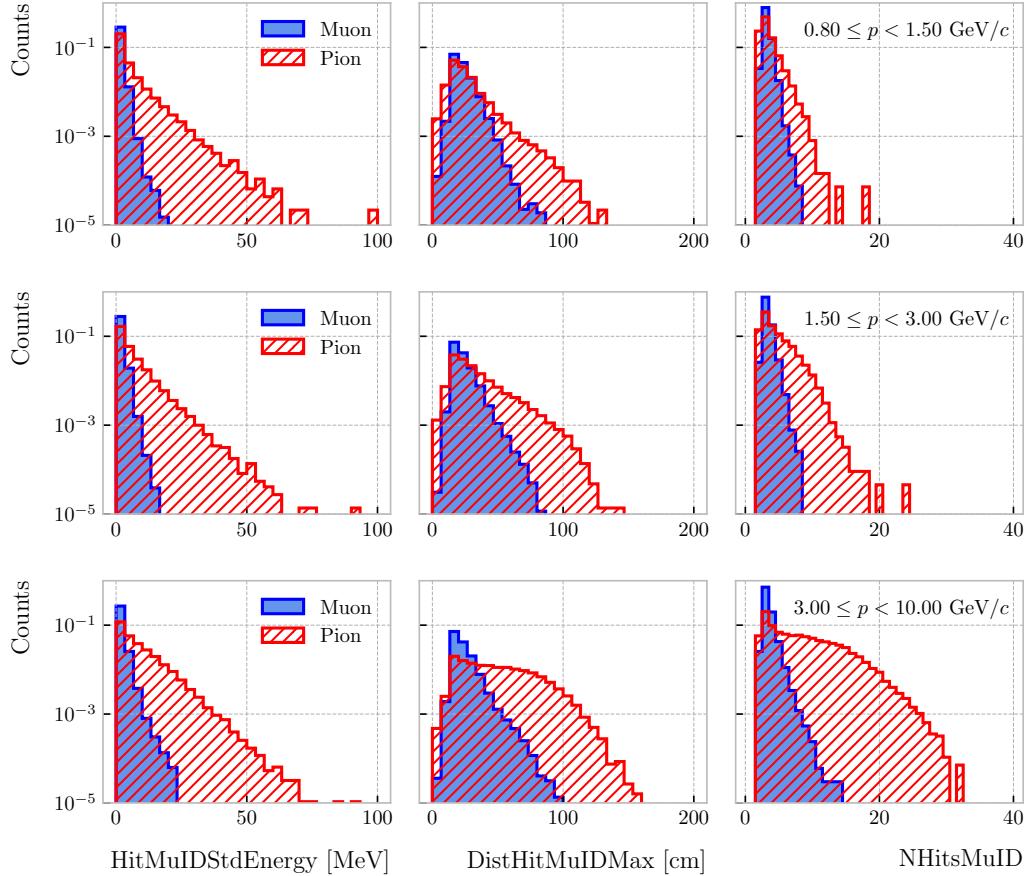


Figure 6.22: Example MuID feature distributions for muons (blue) and charged pions (red) in the three different momentum ranges considered (from top to bottom, in ascending momentum order). From left to right: standard deviation hit energy, maximum distance hit-to-hit, and number of hits.

2838 In the case of the MuID, because at low momenta a significant fraction of the particles
 2839 do not make it past the ECal, I only consider the information coming from this detector
 2840 for momenta ≥ 0.8 GeV/ c , i.e. for the last three momentum regions. The variables I
 2841 extract from it are the following:

2842 • **Energy-related MuID**

- 2843 – MuID total energy (ClusterMuIDTotalEnergy): sum of the energy of all the
 2844 MuID hits.
- 2845 – Mean MuID hit energy (HitMuIDMeanEnergy): mean of the MuID hit energy
 2846 distribution.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

- 2847 – Standard deviation MuID hit energy (HitMuIDStdEnergy): standard deviation
2848 of the MuID hit energy distribution.
- 2849 – Maximum MuID hit energy (HitMuIDMaxEnergy): maximum of the MuID
2850 hit energy distribution.

2851 • Geometry-related MuID

- 2852 – Maximum distance MuID hit-to-hit (DistHitMuIDMax): maximum distance
2853 between pairs of MuID hits (not sure this is a good variable, distribution
2854 looks nuts).
- 2855 – Maximum distance MuID hit-to-centre (DistHitCenterMuIDMax): maximum
2856 of the distance distribution between the MuID hits and the centre of the
2857 TPC.

2858 • Statistical MuID

- 2859 – Number of hits (NHitsMuID): total number of MuID hits associated to the
2860 track.
- 2861 – Number of layers with hits (NLayersMuID): not really a count of all layers
2862 with MuID hits but the difference between the last and the first layer with
2863 MuIDhits.

2864 Figure 6.22 shows the distributions of three different MuID variables, separating true
2865 muons (blue) and charged pions (red), for the three momentum ranges which use the
2866 muon tagger information. In this case I decided to standard deviation of the MuID hit
2867 energy distribution (left column), the maximum distance between the MuID hit pairs
2868 (middle column), and the number of MuID hits (right column). These variables are used
2869 together with the ECal features at high momenta, providing additional disambiguation
2870 power.

2871 Once our features have been defined, one can do some exploratory analysis to
2872 understand how well the variables describe the target class, and avoid the black-box

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

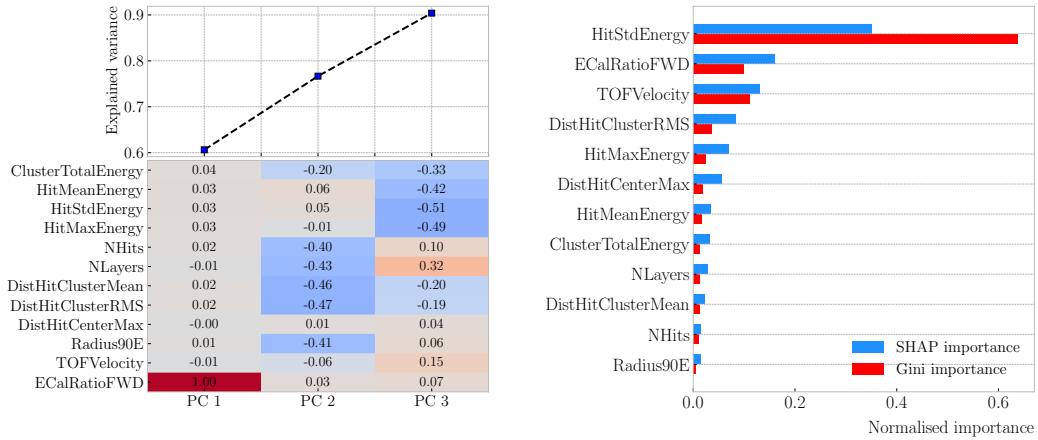


Figure 6.23: Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range $0.3 \leq p < 0.8$ GeV/c.

approach by what features are most relevant for the learning process. This way, I performed a feature analysis for each of the momentum ranges I divided this classification problem into. It follows three steps: first a principal component analysis (PCA), followed by a feature importance study using Gini and Shapley values, and finally a feature permutation importance analysis.

The PCA is useful to understand the variance of the feature space. It is an unsupervised machine learning technique that allows the user to perform a dimensionality reduction. It uses a singular value decomposition of the input features to project them into a lower dimensional space. The idea is to find the matrix \mathbf{C}_m , whose columns are the first m orthonormal eigenvectors of the input covariance matrix. Consider the $n \times p$ real matrix of input data \mathbf{X} , where n is the number of samples and p the number of features. If \mathbf{X} is centred, i.e. the means of its columns are equal to zero, we can write the covariance matrix of \mathbf{X} as $\mathbf{C} = \mathbf{X}^\top \mathbf{X} / (n - 1)$. This matrix can be diagonalised, yielding:

$$\mathbf{C} = \mathbf{V} \mathbf{L} \mathbf{V}^\top, \quad (6.11)$$

where \mathbf{V} is a matrix of eigenvectors and \mathbf{L} a diagonal matrix with eigenvalues λ_i . Then,

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

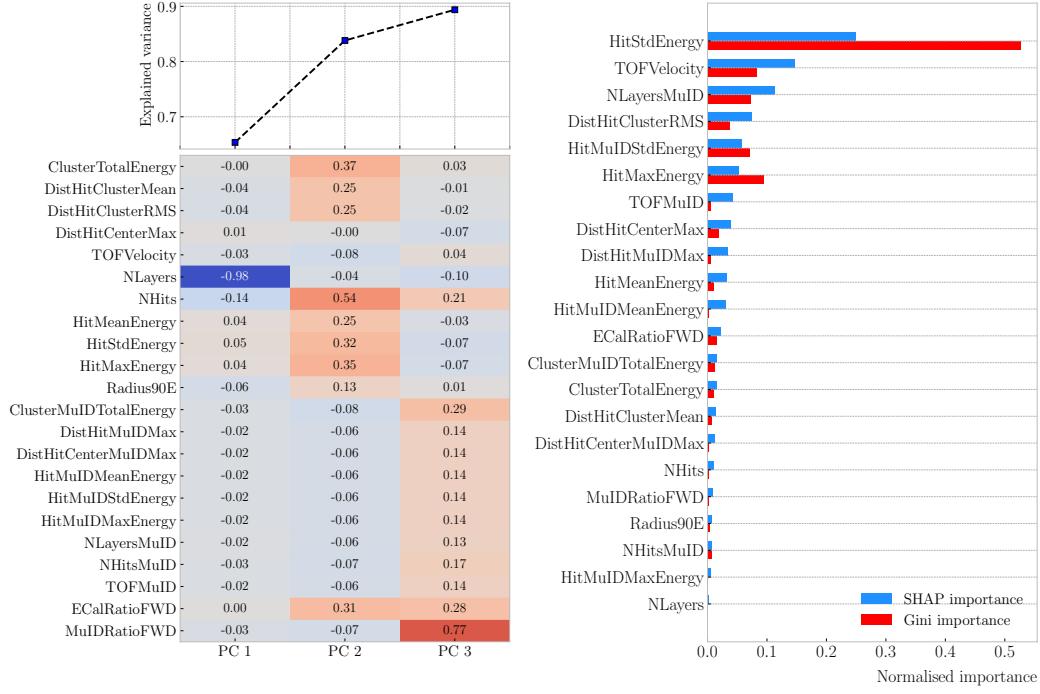


Figure 6.24: Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range $0.8 \leq p < 1.5$ GeV/c.

2887 performing SVD on \mathbf{X} gives us:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{W}^\top, \quad (6.12)$$

2888 where \mathbf{U} is a unitary matrix, whose columns are called left singular vectors, \mathbf{S} is a
2889 diagonal matrix of single values s_i , and \mathbf{W} is another unitary matrix, its columns known
2890 as right singular vectors. This way, we can write:

$$\mathbf{C} = \mathbf{W}\mathbf{S}\mathbf{U}^\top\mathbf{U}\mathbf{S}\mathbf{W}^\top/(n-1) = \mathbf{W}\frac{\mathbf{S}^2}{n-1}\mathbf{W}^\top. \quad (6.13)$$

2891 meaning that the right singular vectors are also the eigenvectors of the covariance matrix.
2892 The SVD can be computed numerically following an iterative approach.

2893 This way, taking an input data vector $X \in \mathbb{R}^n$, the resulting feature vector $Y \in \mathbb{R}^m$

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

2894 is given by:

$$Y = \mathbf{C}_m^\top X. \quad (6.14)$$

2895 The new features capture most of the variance of the original sample, while being lower
2896 dimensional, as $m < n$.

2897 Before applying the PCA reduction one needs to centre and scale the input data.
2898 Centring is necessary when using SVD to obtain the eigenvectors of the covariance
2899 matrix, as only in that case we can do the identification with the right singular vectors
2900 from the input data. Scaling is needed when variables are on different scales, as some
2901 can then dominate the PCA procedure.

2902 I used the PCA module of `scikit-learn`, together with the `RobustScaler`, which
2903 centres the data and scales it based on the interquartile range. In Fig. 6.23 (left panel)
2904 and Fig. 6.24 (left panel) I show the results I obtained from the PCA for the momentum
2905 ranges $0.3 \leq p < 0.8 \text{ GeV}/c$ and $0.8 \leq p < 1.5 \text{ GeV}/c$, respectively. Notice that in
2906 the second case the number of features increases considerably, as this is the first region
2907 which uses the MuID variables. I found that, in all the cases, adding a fourth PC does
2908 not add additional information. As it can be seen in the top panels of the figures, the
2909 cumulative explained variance is already over 80% with three PCs.

2910 The bottom panels show the contribution of the variables to the principal axes. For
2911 the two first momentum regions, I observe a tendency of the energy-related and the
2912 geometry-related ECal variables to be clustered together. For the other ranges, when
2913 I include the MuID variables, there seems to be a division between ECal and MuID
2914 variables. For these, it seems like the number of ECal layers with hits also plays an
2915 important role.

2916 The next step in the analysis is to quantify the importance of the features based on
2917 two additional metrics, namely the Gini and the Shapley values. The Gini importance,
2918 often called mean decrease impurity, is based on how much a feature contributes to the
2919 purity improvement at the splits in each decision tree. The purity is measured in terms

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

2920 of the Gini impurity index, defined as:

$$I_G = 1 - \sum_i f_i, \quad (6.15)$$

2921 where f_i is the fractional abundance of the i -th class. Then, for each split one can
 2922 compute the weighted decrease in impurity as:

$$\Delta_G = \frac{N_t}{N} \left(I_G - \frac{N_t^R}{N_t} I_G^R - \frac{N_t^L}{N_t} I_G^L \right), \quad (6.16)$$

2923 where N represents the total number of samples, N_t the number of samples at the current
 2924 node, N_t^R and N_t^L the number of samples in the right and left children respectively,
 2925 I_G is the Gini impurity at the current node, and I_G^R and I_G^L the Gini impurities of the
 2926 resulting right and left children.

2927 For each decision tree, one will have a normalised vector with the accumulated
 2928 decrease in Gini impurity for each feature. In the case of a BDT, the feature importances
 2929 are simply the mean for all the estimators in the ensemble⁹.

2930 The concept of Shapley values originated in the context of game theory, and it
 2931 measures the marginal contribution of a feature in enhancing the accuracy of a classifier.
 2932 Take F to be the set of all features in a problem, and $S \subseteq F$ the subset of features. To
 2933 compute the Shapley value of the i -th feature, one has to train a model with that feature
 2934 present, $f_{S \cup \{i\}}$, and another model trained without it, f_S . This has to be repeated for
 2935 all possible combinations of subsets $S \subset F \setminus \{i\}$, and evaluating the models predictions
 2936 on the appropriate sets of data x_S . This way, the Shapley value results:

$$\varphi_i = \sum_{S \subset F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]. \quad (6.17)$$

2937 I trained the `GradientBoostingClassifier` from `scikit-learn` with the default
 2938 configuration in order to evaluate both the Gini and Shapley importances. The Gini

⁹Note to self: this appears not to be the case. If you get the `feature_importance` for each tree in the BDT and take the average, the result is not the same to be one reported in the `feature_importance` attribute of the BDT.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

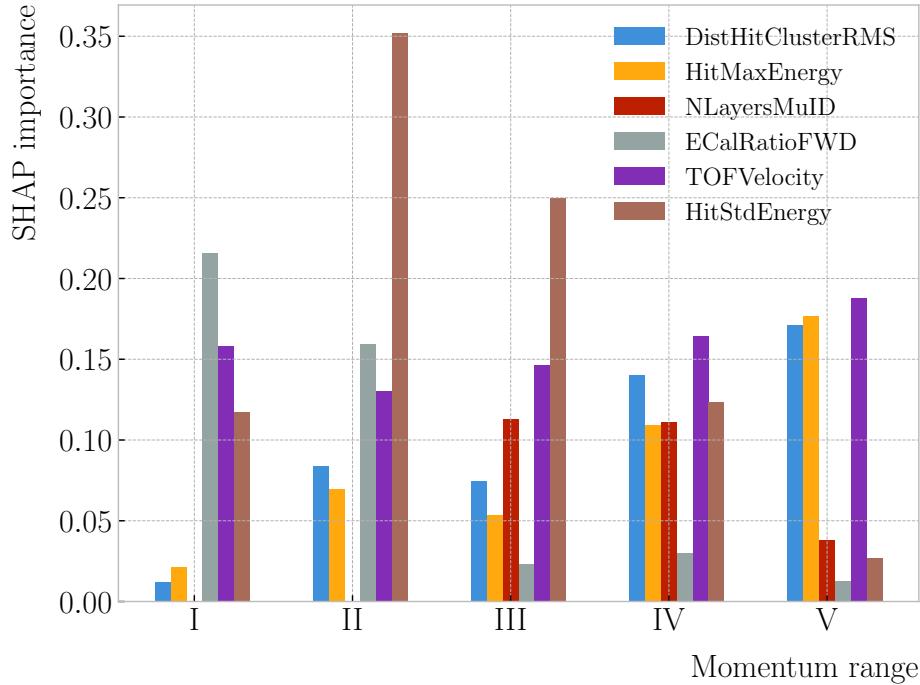


Figure 6.25: Evolution of the SHAP importance for the top six most important features across all five momentum ranges.

2939 scores are automatically computed by `scikit-learn`, using the training data. For the
 2940 Shapley importance, I used the implementation from the `SHAP` package, computing
 2941 it using the test sample. The results can be seen in Fig. 6.23 (right panel) and
 2942 Fig. 6.24 (right panel), again for the momentum ranges $0.3 \leq p < 0.8 \text{ GeV}/c$ and
 2943 $0.8 \leq p < 1.5 \text{ GeV}/c$. The length of the bars denote either the SHAP (blue) or the Gini
 2944 (red) importance of the feature. One interesting thing to notice is that, when looking at
 2945 the Gini importance, there is always one feature that dominates over the rest. This is
 2946 not the case for the SHAP importance, where importances tend to be more balanced.

2947 Across all momentum ranges, I observe that the most important features are. For
 2948 the five momentum ranges considered, only six variables sit in the top five at least once.
 2949 Figure 6.25 shows the evolution of the SHAP importance of these six features. It is
 2950 interesting to see that the time-of-flight variable keeps its importance almost unchanged
 2951 for all momenta. Also, it looks like the ECal energy ratio gets less relevant the higher the
 2952 momentum is, but the RMS of the hit-to-cluster distance distribution and the maximum

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

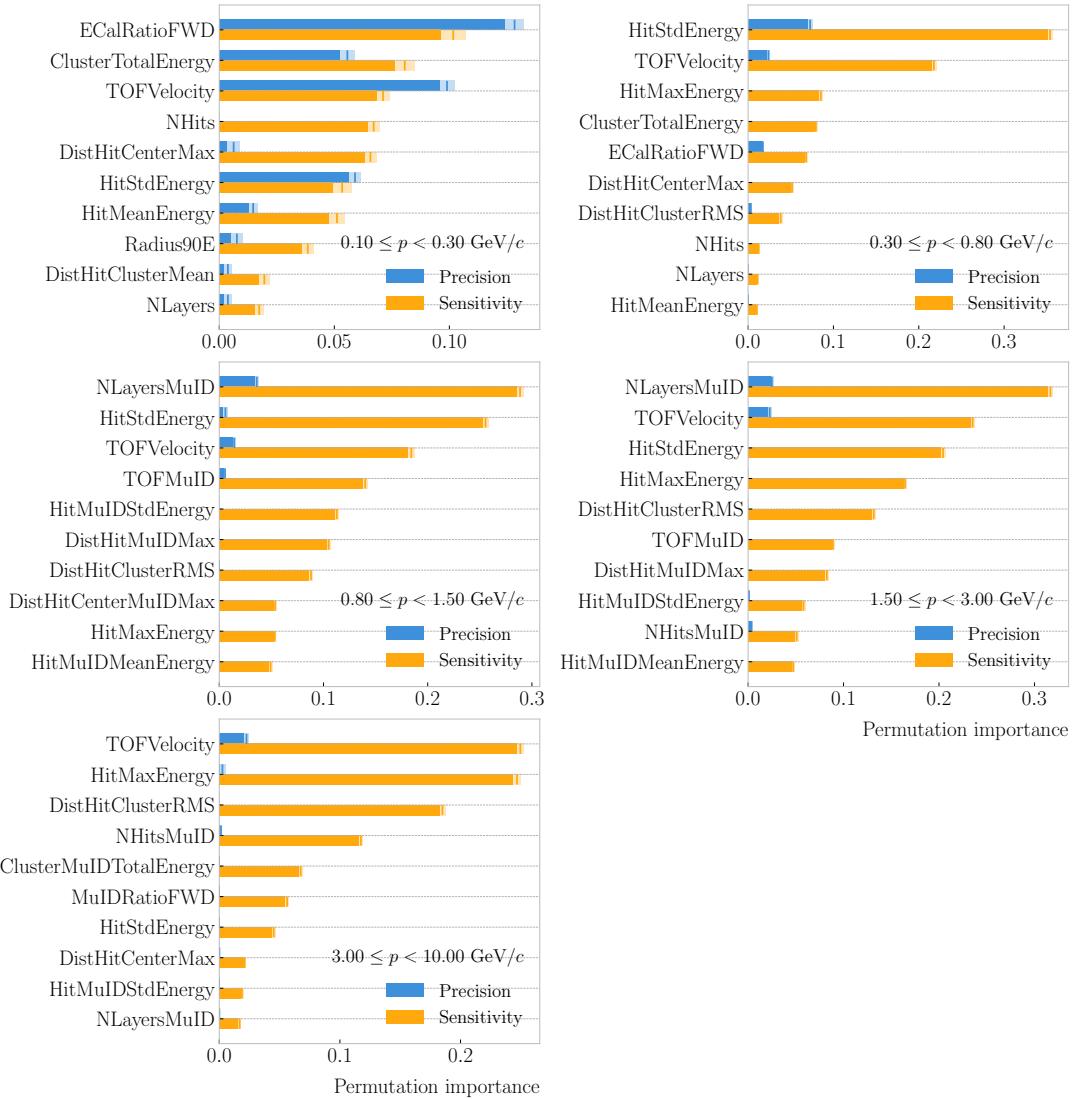


Figure 6.26: Permutation importances for the ten most important features in the different momentum ranges (from left to right, top to bottom, in increasing momentum order). The bars indicate the effect that permutations of each feature have on the purity (blue) and the sensitivity (yellow), the translucent regions representing one standard deviation around the central value.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

2953 ECal hit energy become more important in the last momentum ranges.

2954 The last step in the feature selection analysis is the feature permutation. This
2955 technique measures the contribution of each feature to the performance of a model by
2956 randomly shuffling its values and checking how some scores degrade. For the present
2957 case, I am interested in the precision or purity, and the sensitivity or efficiency, as these
2958 two are the most relevant metrics from a physics point of view. The `scikit-learn`
2959 module provides the user with a method to perform the permutation scans.

2960 The results of these are shown in Fig. 6.26. For the different momentum ranges
2961 I show the permutation importances for the ten most important features. For each
2962 of the variables I report the effect the permutations have on the precision (blue) and
2963 sensitivity (yellow) of the models. The bars indicate the importance value, with the
2964 lighter part representing one standard deviation around the mean (hinted as an additional
2965 vertical line). Something to notice is that, in the first momentum region, the feature
2966 permutations have an effect on both the precision and the sensitivity. However, for the
2967 rest the precision is almost unaffected, while the sensitivity changes are considerably
2968 larger.

2969 It is also interesting to see that most of the variables identified as important here
2970 are the same I found when looking at the Shapley values. The behaviour of these across
2971 the momentum ranges is also similar, with the same patterns of some features being
2972 important at low momenta and then dropping in importance for the high momentum
2973 ranges.

2974 Wit this, I conclude the study of the features. I have prepared the training and
2975 testing datasets and understood what features are likely to have the largest impact on
2976 the performance of the classifiers.

2977 6.3.4 Hyperparameter optimisation

2978 Any BDT requires the user to specify a number of parameters that will dictate its
2979 behaviour. They can be divided into two categories: (i) tree-specific parameters, which
2980 affect each individual tree in the model, and (ii) boosting parameters, which control the

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

2981 boosting operation in the model. The value of these so-called hyperparameters affect the
2982 performance and predictive power of the models. Therefore, one needs to carefully select
2983 their optimal values in order to extract as much information as possible from the data.

2984 From all the parameters used to define a tree in the `scikit-learn` implementation
2985 of the BDT classifier, I only consider a subset of them. This is due to the fact that some
2986 are mutually exclusive, but also because I noticed that others have little effect on the
2987 problem at hand. Therefore, the parameters I investigate are the following:

2988 • `min_samples_split`: defines the minimum number of samples required in a node
2989 to be considered for splitting. Higher values prevent a model from learning relations
2990 which might be highly specific to the particular sample, but may lead to under-fitting
2991 if the value is too low.

2992 • `min_samples_leaf`: defines the minimum samples required in a leaf node. For
2993 imbalanced problems it should take a low value, as there will not be many cases
2994 where the minority class dominates.

2995 • `max_depth`: maximum depth of a tree. Useful to prevent over-fitting, as higher
2996 depth will allow a model to learn relations specific to the training sample.

2997 In the case of the boosting parameters, the ones I look at are:

2998 • `learning_rate`: determines the impact of each tree on the final outcome. Low
2999 values make the model robust to the specific characteristics of a tree, and thus
3000 allow it to generalise well. However, that usually requires a large number of trees
3001 to model the data properly.

3002 • `n_estimators`: number of sequential trees to be trained. In general, BDTs are
3003 fairly robust at higher number of trees but it can still overfit at a point.

3004 • `subsample`: fraction of observations to be selected for each tree. Values slightly
3005 less than 1 make the model robust by reducing the variance.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

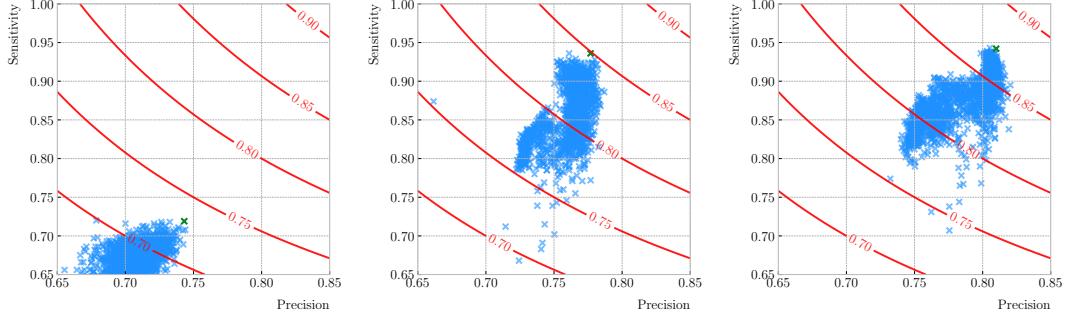


Figure 6.27: Values of the precision and sensitivity obtained for 10000 BDT hyperparameter configurations, for the momentum regions I, III and V. The red contours indicate the curves of equal F_1 -score, while the green crosses are the selected configurations.

3006 In general, hyperparameters depend on each other. Thus, it is not possible to
 3007 optimise them independently. In the literature, we find two main strategies to explore
 3008 the hyperparameter space. We could use a grid search, in which one discretises a
 3009 portion of the space of hyperparameters and evaluates the model at each point. Another
 3010 approach is the randomised search, where a certain number of random configurations of
 3011 hyperparameters are explored.

3012 In this case, I used the random search to scan the hyperparameter space. Also,
 3013 because it is not guaranteed that a set of hyperparameters can be efficiently applied
 3014 across different datasets, I perform the optimisation for each of the momentum ranges
 3015 considered. Table 6.3 shows the list of hyperparameters considered, and the range within
 3016 which I let them vary. I decided to fix the number of estimators to 400 in all cases, as
 3017 its value is correlated with that of the learning rate.

3018 I evaluate 10000 different hyperparameter configurations for each momentum range.
 3019 For the hyperparameter tuning, I used subsamples containing 10% of the full datasets,
 3020 keeping the original proportions between classes, in order to reduce the computational
 3021 load. The performance of the models was assessed using a stratified 3-fold cross-validation
 3022 with replacement. Cross-validation involves dividing the data in a number of subsets,
 3023 training the model using some of them, and testing it with the rest. In our case, I
 3024 divide the data in 3 equal-sized subsets, maintaining the class proportions of the original

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

Table 6.3: Optimal values of the hyperparameters used by the BDT, for each momentum range.

Hyperparameter	Range	Best value				
		I	II	III	IV	V
<code>min_samples_split</code>	[0.001, 1]	0.10	0.06	0.05	0.27	0.06
<code>min_samples_leaf</code>	[0.001, 1]	0.02	0.03	0.01	0.02	0.07
<code>max_depth</code>	{2, 3, ..., 8}	8	2	4	2	7
<code>learning_rate</code>	[0.05, 1]	0.10	0.23	0.07	0.13	0.09
<code>subsample</code>	[0.01, 1]	0.75	0.65	0.79	0.86	0.95

Table 6.4: Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges.

Metric	Value $\pm 1\sigma$				
	I	II	III	IV	V
Accuracy	0.779 ± 0.003	0.812 ± 0.003	0.846 ± 0.002	0.861 ± 0.003	0.874 ± 0.002
Precision	0.769 ± 0.003	0.752 ± 0.005	0.788 ± 0.002	0.805 ± 0.003	0.815 ± 0.003
Sensitivity	0.745 ± 0.009	0.921 ± 0.006	0.965 ± 0.002	0.967 ± 0.002	0.976 ± 0.001
F_1 -score	0.757 ± 0.004	0.828 ± 0.003	0.867 ± 0.002	0.879 ± 0.002	0.889 ± 0.002
ROC AUC	0.868 ± 0.003	0.865 ± 0.003	0.899 ± 0.002	0.902 ± 0.002	0.911 ± 0.001

dataset. Then, for 3 consecutive iterations, I train the models using 2 of the subsets while I compute the precision and sensitivity scores with the other. This approach provides a more robust estimate of the performance on unseen data.

Figure 6.27 shows the results in the precision versus sensitivity plane, for the momentum regions I, III and V (from left to right). The contours represent the curves of equal F_1 -score, i.e. the harmonic mean of the precision and the sensitivity. In order to select the optimal configurations (indicated in the plots with a green cross), I chose the point with the highest F_1 -score.

The results for the different momentum ranges are summarised in Tab. 6.3. One can see some consistency in hyperparameter choices, with models generally preferring small values for the tree-specific parameters, small learning rate, and relatively large subsample sizes.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

3037 Now that I have obtained the optimal values of the hyperparameters, I can train
3038 the different BDTs. In this case I use the complete datasets, keeping 20% of the data
3039 for testing. Table 6.4 shows the values of the different performance metrics obtained
3040 using the selected hyperparameters and 5-fold cross-validation. The last row indicates
3041 the value of the area under the receiver operating characteristic (ROC) curve. This
3042 represents the sensitivity of a model as a function of the false positive rate. I have
3043 included it here as it is a classic model metric used in the machine learning community.
3044 Overall, there is a clear trend of models performing better at higher momentum.

3045 6.3.5 Probability calibration

3046 So far, the trained BDTs are able to provide predictions of the class labels. Ideally,
3047 one would like the output of a classifier to give a confidence level about the prediction.
3048 However, it is not straightforward to interpret the outputs of our BDTs in terms of
3049 probabilities.

3050 A way to visualise how well the predictions of a classifier are calibrated is using
3051 reliability diagrams [161]. They represent the probability of the positive label versus the
3052 probability predicted by the classifier. These can be obtained by binning the predicted
3053 probabilities, and then compute the conditional probability $P(y_{true} = 1 | y_i \leq y_{pred} <$
3054 $y_{i+1})$ by checking the fraction of true positive instances in each bin. The reliability
3055 diagram of a perfectly calibrated classifier would be a diagonal line.

3056 In this case, I try to correct the raw response of the classifiers by applying a sigmoid
3057 function:

$$\sigma(x; A, B) = \frac{1}{1 + e^{Ax+B}}, \quad (6.18)$$

3058 where the parameters A and B are real numbers determined using the method of least
3059 squares.

3060 For each classifier, I perform a grid search to obtain the optimal values of A and B .
3061 For any pair, I compute the predicted probabilities as $y_{pred} = \sigma(y_{raw}; A, B)$, where y_{raw}

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

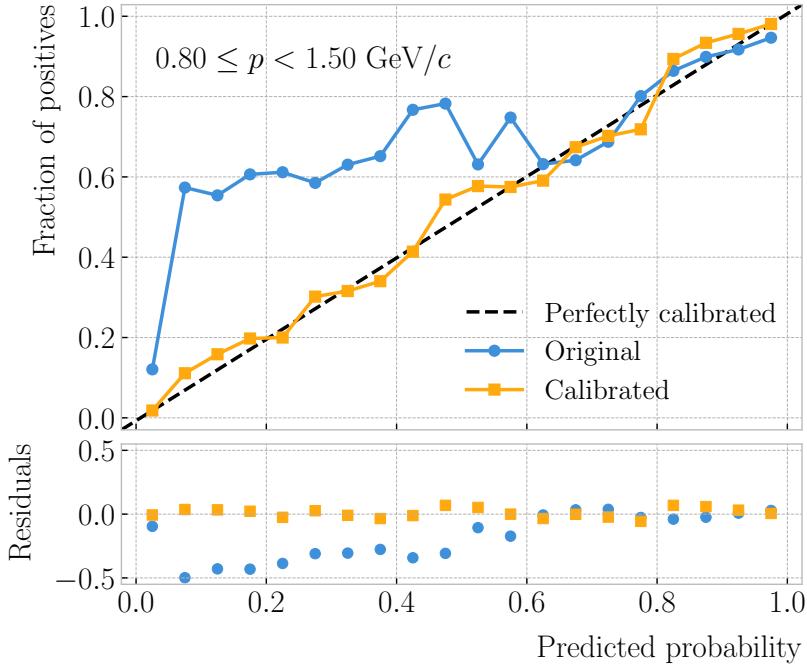


Figure 6.28: Reliability diagrams for the BDT classifier used in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$, both for the original (blue circles) and calibrated (yellow squares) responses. For reference, the response of a perfectly calibrated classifier is also shown (black dashed line).

3062 are the raw predictions of the classifier¹⁰. Then, I calculate the corresponding reliability
 3063 curve, and take the sum of the squared residuals between it and the response of the
 3064 perfectly calibrated classifier.

3065 Figure 6.28 shows the reliability diagrams for the original (blue) and calibrated
 3066 (yellow) probability predictions of the classifier for the III momentum range, $0.3 \leq p <$
 3067 $0.8 \text{ GeV}/c$. The original response of the classifier is given by $y_{pred} = \sigma(y_{raw}; -2, 0)$,
 3068 which is the transformation applied by `scikit-learn` to produce the probability estimate.
 3069 Notice how the calibrated prediction matches the ideal response much better than the
 3070 original, across all the probability range.

3071 One can also compare the responses of the uncalibrated and calibrated classifiers
 3072 broken down by true particle type, as shown in Fig. 6.29. It can be seen that the
 3073 distributions for both muons (blue) and charged pions (red) smoothen after calibration,

¹⁰In `scikit-learn` these correspond to the outputs of the `decision_function` method.

6.4. ECAL TIME-OF-FLIGHT

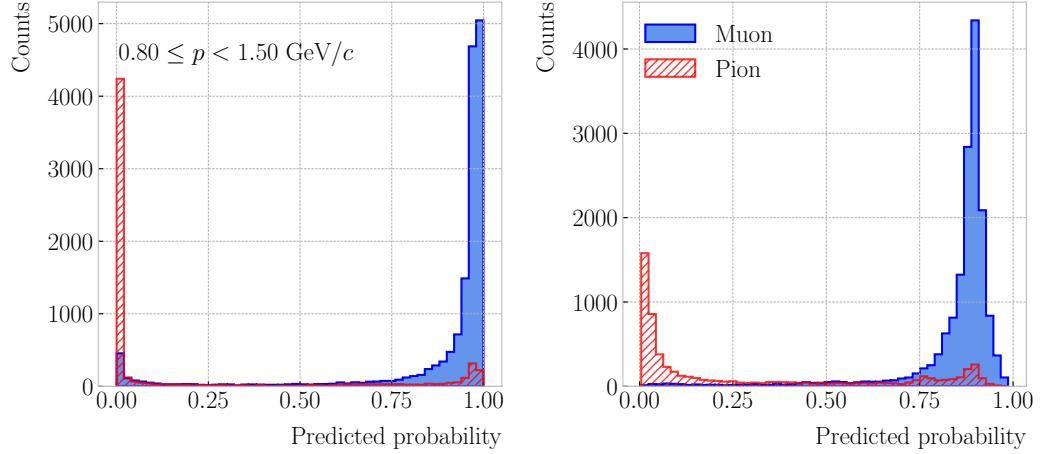


Figure 6.29: Uncalibrated (left panel) and calibrated (right panel) predicted probabilities assigned by the BDT classifiers for true muons (blue) and charged pions (red) in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$.

3074 but still the separating power of the classifier remains unchanged.

3075 At this point, having the trained classifiers and the probability calibration parameters,
3076 I am able to assess the performance of the classification strategy in a physics-relevant
3077 case.

3078 6.3.6 Performance

3079 6.4 ECal time-of-flight

3080 Looking at Fig. 6.20, it is clear that for momentum values in the range $1.0 - 3.0 \text{ GeV}/c$
3081 it is not possible to separate pions and protons using a $\langle dE/dx \rangle$ measurement in the
3082 HPgTPC. However, in the previous section I assumed that protons at those energies
3083 could be identified by other means, and therefore were not an issue for the muon and
3084 pion discrimination.

3085 Some detectors, like ALICE [162] or the ILD concept [163], complement the PID
3086 capabilities of their gaseous trackers with time-of-flight measurements. The use of
3087 fast timing silicon sensors, with hit time resolutions under 100 ps, would allow for the
3088 identification of charged hadrons via a ToF measurement up to $5.0 \text{ GeV}/c$. In the case
3089 of ND-GAr, one could think of using the inner layers of the ECal, the ones consisting of

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

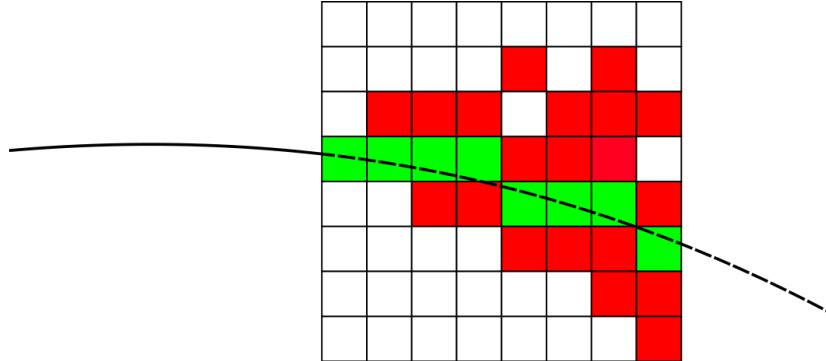


Figure 6.30: Schematic of the hit selection used for the ToF measurement. The grid represents the layers of the inner ECal, with coloured squares indicating the tiles with hits. Green squares indicate the selected hits.

3090 high-granularity tiles, to obtain a ToF-based PID, with some inputs from the TPC.

3091 Measuring the momentum and the velocity of a charged particle allows for a
3092 determination of the mass through the relativistic momentum formula:

$$m = \frac{p}{\beta} \sqrt{1 - \beta^2}. \quad (6.19)$$

3093 In our case, the momentum is measured in the TPC, using the curvature and the dip
3094 angle of the helix inside the magnetic field. The velocity of the particle can be written
3095 as:

$$\beta = \frac{\ell_{track}}{c\tau}, \quad (6.20)$$

3096 where ℓ_{track} is the length of the track, and τ the arrival time to the ECal.

3097 In GArSoft, the track length is computed at the Kalman filter stage. It is simply the
3098 sum of the line segments along the track, either in the forward or backward fit. In this
3099 case, because we are only interested in the particles that make it to the ECal, I choose
3100 the fit direction based on the results of the track-cluster associations.

3101 Additionally, because the last 30 cm of the TPC radius are uninstrumented¹¹, I need
3102 to correct for the length of the tracks. Using the track fit parameters to propagate the

¹¹Note to self: check this number.

6.4. ECAL TIME-OF-FLIGHT

3103 helix to its entry point in the ECal, one can write the total track length as:

$$\ell_{track} = \ell + \left| \frac{\phi_{EP} - \phi}{R^{-1}} \right| \sqrt{1 + \tan^2 \lambda}, \quad (6.21)$$

3104 where ϕ_{EP} is the angle of rotation at the entry point to the calorimeter, and ℓ , ϕ , R and
3105 λ are the track length, angle of rotation, radius of curvature and dip angle at the last
3106 point in the fit, respectively.

3107 To test the idea of performing a ToF measurement with the inner ECal, I generated
3108 two data samples. Each consists of 10000 single particle events, either charged pions or
3109 protons. Their momenta are uniformly distributed in the range $0.5 - 5.0$ GeV/ c , and
3110 their directions are isotropic. I process each sample using different values of the time
3111 resolution, from $\Delta\tau = 0$, the perfect time resolution case for comparison, to the current
3112 nominal value of $\Delta\tau = 0.7$ ns, and the worse scenario of $\Delta\tau = 1.0$ ns.

3113 6.4.1 Arrival time estimations

3114 In the simulation, the limited time resolution of the ECal is taken into account by
3115 applying a Gaussian smearing to the true hit times. Other effects, like the digitisation
3116 of the signals, are not taken into account and fall beyond the scope of this study. After
3117 the track-cluster, one ends up with a collection of ECal hits associated to each particle.
3118 From these, the arrival time of the particle to the ECal can be extracted.

3119 The simplest possibilities are to either take the time of the earliest hit or the hit
3120 closest to the entry point. Because these two coincide, in general, I focused only in
3121 the earliest hit time. However, this needs to be corrected, to account for the distance
3122 travelled from the entry point to the position of the hit:

$$\tau_{earliest} = \tau_{hit} - \frac{d_{EP-hit}}{c}, \quad (6.22)$$

3123 where τ_{hit} is the time of the earliest hit, and d_{EP-hit} is the distance between that hit
3124 and the entry point of the particle to the ECal. This is computed as the arc length

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

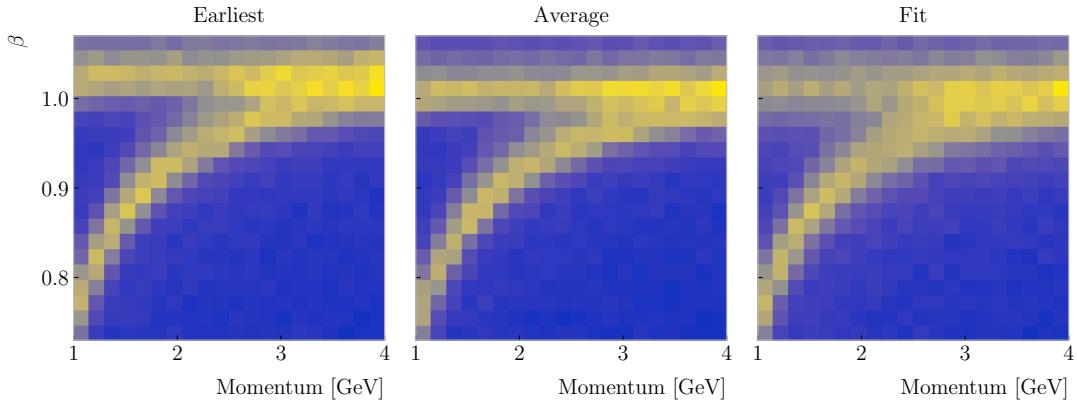


Figure 6.31: Particle velocity versus momentum measured with different ECal arrival time estimations. From left to right: earliest hit time, average hit time, and fitted hit time. In all cases the time resolution is $\Delta\tau = 0.1$ ns.

3125 between the entry point and the point of the extrapolated helix up to the layer of the
 3126 hit. This way of correcting the time assumes c for the propagation of the particle, which
 3127 may lead to biased estimates.

3128 I also tried to estimate the arrival times using information from the rest of the hits.
 3129 In order to do this, as a simplifying assumption, I approximate the hadronic shower
 3130 considering only its MIP component. For each layer, I keep only the hit in the tile closest
 3131 to the point of the extrapolated track up to that layer. Figure 6.30 shows an example of
 3132 how this hit selection works. The dashed line represents the extrapolated track, while
 3133 the coloured squares are the tiles containing hits. Green indicates the tiles closer to the
 3134 track in each layer (in the sketch they correspond to the grid columns).

3135 Now, I can use these collections of hits to estimate the arrival times. A possibility
 3136 is to take the average of the times of the selected hits, denoted $\tau_{average}$. For that to
 3137 work, one needs to correct these times, in a similar way as in Eq. (6.22), before taking
 3138 the average. However, as before, this correction assumes that the particle travels at the
 3139 speed of light inside the ECal. Another option is to perform a linear fit to the hit times
 3140 and the distances to the entry point. In that case, the arrival time would be the fitted
 3141 value of the intercept, τ_{fit} . This method would not assume a speed of light propagation.

3142 Figure 6.31 shows the velocity estimations as a function of the particle momentum,

6.4. ECAL TIME-OF-FLIGHT

for the earliest hit time (left panel), average hit time (middle panel), and fitted hit time (right panel). The two bands correspond to the π^\pm and the p particles. $\Delta\tau = 0.1$ ns. Notice how, for the earliest hit time method, the velocities are significantly biased towards larger values. For the multi-hit methods, the τ_{fit} estimate appears to produce a larger variance than when using the $\tau_{average}$ method.

6.4.2 Proton and pion separation

Once we have the velocities of the particles, one can estimate their masses through Eq. (6.19). The resulting mass spectra are shown in Fig. 6.32. I computed the masses for the three arrival time estimates discussed above, and three different values of the time resolution: $\Delta\tau = 0.00$ (perfect time resolution), $\Delta\tau = 0.10$ ns, and $\Delta\tau = 0.70$ ns. Although in all cases we have the same number of events, it appears as if the entries in the histograms decrease as the time resolution increases. Sometimes, the particles get unphysical values of $\beta > 1$, and in turn they do not contribute to the mass spectra. This is more likely to happen for higher values of $\Delta\tau$.

As noted before, the average hit time method produces the most robust estimates when increasing $\Delta\tau$. Intuitively this makes sense, as by taking the mean one averages out the effect of the Gaussian smearing. Going forward, I will use this arrival time estimator, as it appears to be the best performing one.

It is possible to use the velocity estimations to select a sample of protons. In this case, I do so by dividing the relevant momentum range in bins of 0.1 GeV/ c . For each momentum bin, I compute the expected velocity for the protons via the inverse of Eq. (6.19), and then take the fractional residuals of the measured velocities. Using that distribution, I choose the cut that maximises the F_1 -score of the proton selection.

The results can be seen in Fig. 6.33, for the case $\Delta\tau = 0.10$ ns. As expected from Fig. 6.31, the performance of the selection degrades rapidly with increasing momentum. However, the purity is still around 75% at 3.0 GeV/ c . This is likely to be sufficient, as we do not expect protons or charged pions with higher energies from the beam neutrino interactions.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

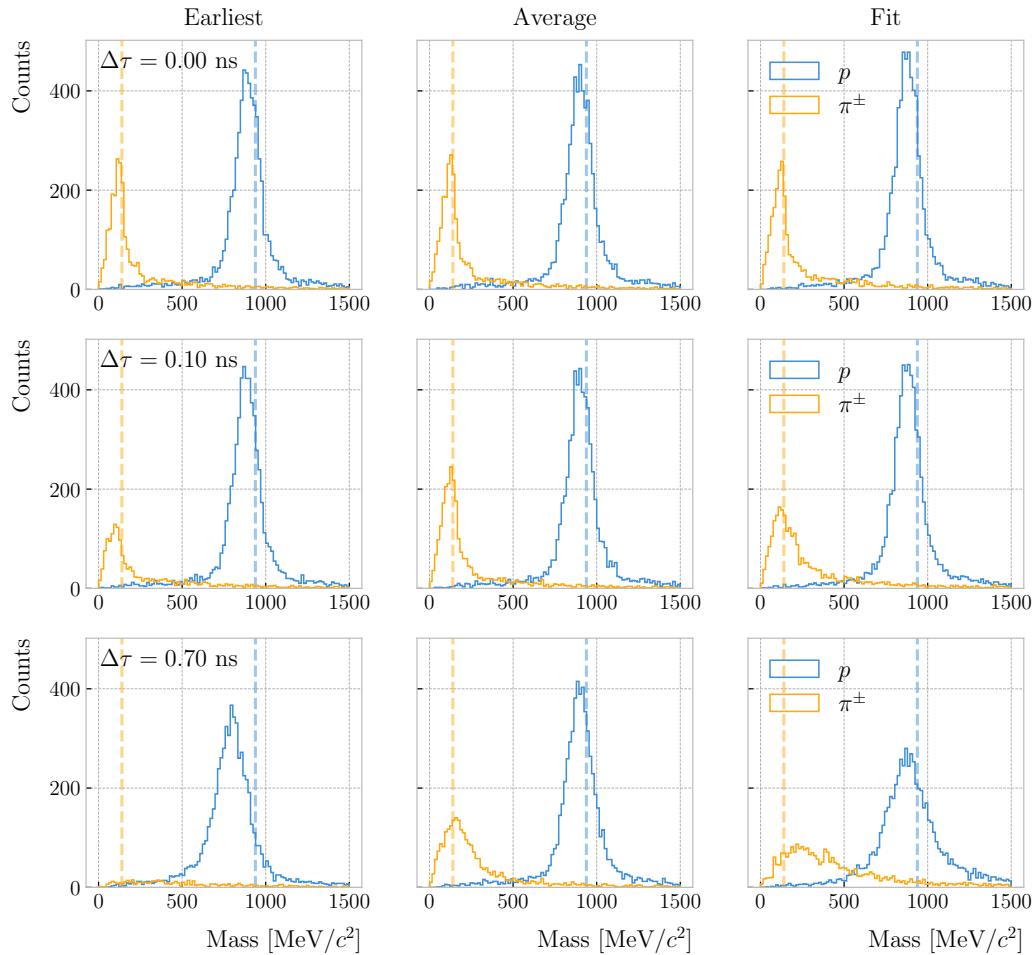


Figure 6.32: Mass spectra for p (blue) and π^\pm (yellow) particles, using different ECal time resolution values (from top to bottom, in ascending order), and arrival time estimates. From left to right: earliest hit time, average hit time, and fitted hit time. The dashed lines indicate the true masses of the particles.

3171

Figure 6.34

3172 6.5 Charged pion decay in flight

3173 As discussed previously, in GArSoft the TPC tracks are formed after a pattern recognition
 3174 algorithm and a Kalman filter are applied to the TPC clusters. These two steps can
 3175 find discontinuities in the track candidates (e.g. due to a particle decay) when these
 3176 so-called breakpoints are large enough. However, for some, more subtle, cases they may
 3177 miss them and form a single reconstructed track. It has been noted in the literature

6.5. CHARGED PION DECAY IN FLIGHT

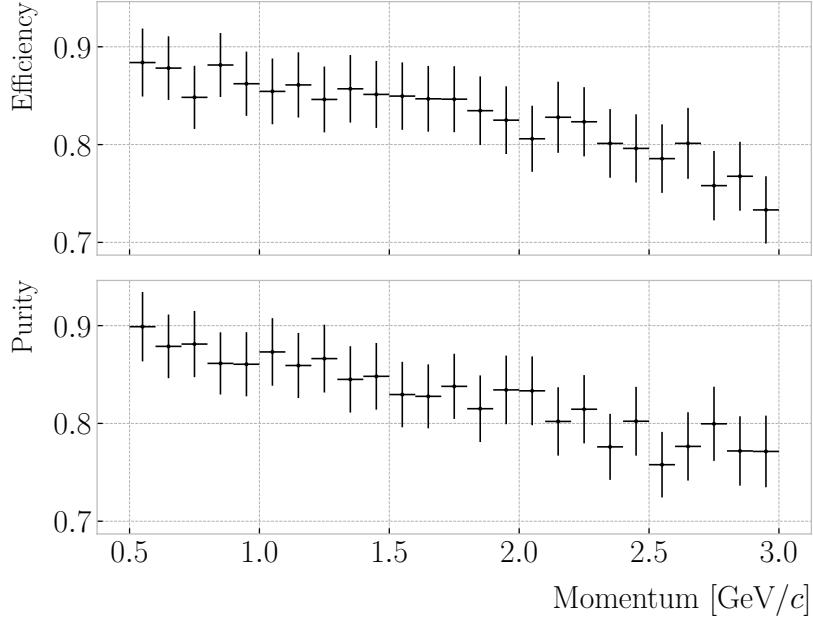


Figure 6.33: Efficiency (top panel) and purity (bottom panel) for the proton selection as a function of the momentum, for $\Delta\tau = 0.10$ ns.

that Kalman filters offer, as a by-product, additional information to form test statistics to identify these breakpoints [164, 165].

Considering the mean life of the charged pion, $\tau = (2.6033 \pm 0.0005) \times 10^{-8}$ s, one can estimate that about 12% of the pions with momentum $p \sim \mathcal{O}(500 \text{ MeV}/c)$ (roughly the peak of the pion momentum distribution in ν_μ CC interactions off argon) decay inside the TPC. Figure 6.35 (left panel) shows the amount of charged pions decaying in the full TPC and fiducial volumes from an isotropic, monoenergetic sample of 100000 negatively charged pions with $p = 500 \text{ MeV}/c$. We see that about 10% of those decayed, with more than half of them decaying inside the TPC fiducial volume.

Figure 6.35 (right panel) shows an example event display of a charged pion (magenta line) decays in flight inside the TPC, but because the angle of the muon (blue line) is small both were reconstructed as one single track (black line). In this case, the composite track reaches the ECal, where it undergoes a muon-like interaction, thus being classified as a muon.

A way to understand what decaying pion tracks were totally or partially reconstructed

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

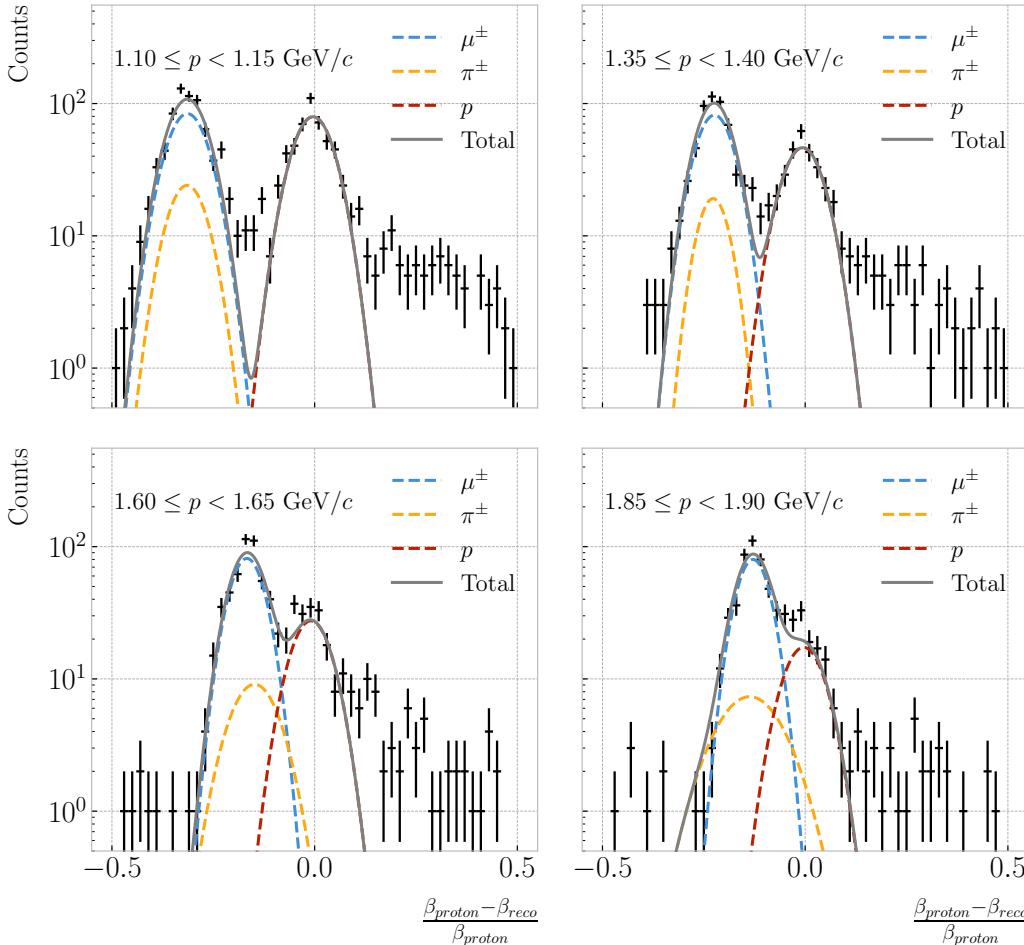


Figure 6.34: Distributions of the velocities measured by ToF with the inner ECal, for different momentum bins, in a FHC neutrino interaction sample. The Gaussian fits are performed around the maxima for each particle species.

3193 together with the daughter muon is looking at the relative energy contributions to the
 3194 reconstructed track. In order to select a sample of such events, I require that a minimum
 3195 50% of the total energy comes from the pion and at least 20% from the muon.

3196 6.5.1 Track breakpoints

3197 To identify potential decays we can use the information we obtain from the Kalman
 3198 filter at each step of the fitted track. The simplest test we can think about is computing

6.5. CHARGED PION DECAY IN FLIGHT

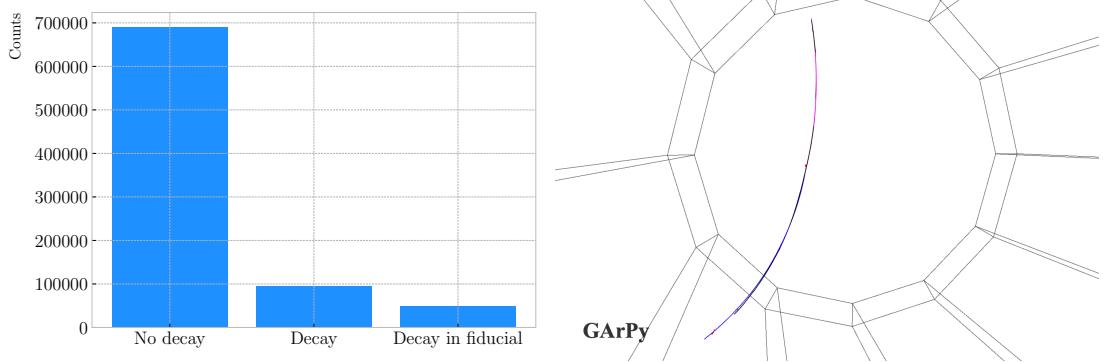


Figure 6.35: Left panel: number of non-decaying, decaying and decaying in the fiducial volume pions for a MC sample of 100000, $p = 500$ MeV/c isotropic positively charged pions inside the TPC. Right panel: event display for a positive pion decaying inside the fiducial volume, with a single reconstructed track for the pion and muon system.

3199 the χ^2 of the mismatch between all the parameters in the forward and the backward fits:

$$\chi_k^{2(FB)} = (\hat{x}_k^B - \hat{x}_k^F)^T [V^{(\hat{x}_k,B)} + V^{(\hat{x}_k,F)}]^{-1} (\hat{x}_k^B - \hat{x}_k^F), \quad (6.23)$$

3200 where \hat{x}_k^F , \hat{x}_k^B are the Kalman filter state vector estimates at step k in the forward and
3201 backward fits and $V^{(\hat{x}_k,F)}$, $V^{(\hat{x}_k,B)}$ the covariance matrices of \hat{x}_k^F and \hat{x}_k^B respectively.

3202 Using the values of the χ^2 at measurement k for the forward and backward fits we can
3203 compute another χ^2 value that characterises the overall track fit:

$$\chi_{track}^2 = \chi_k^{2(F)} + \chi_k^{2(B)} + \chi_k^{2(FB)}, \quad (6.24)$$

3204 which remains approximately constant for all k .

3205 An alternative approach proposed in the context of the NOMAD experiment was
3206 using a fit with a more elaborate breakpoint hypothesis, so we can perform a comparison
3207 of the χ^2 with and without breakpoints. This can be achieved by using some alternative
3208 parametrisation with extra parameters, which allows some of the track parameters to
3209 be discontinuous at certain points. A decay changes the momentum magnitude and
3210 direction, so we can use the new state vector:

$$\alpha = (y, z, 1/R_F, 1/R_B, \phi_F, \phi_B, \tan\lambda_F, \tan\lambda_B)^T. \quad (6.25)$$

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

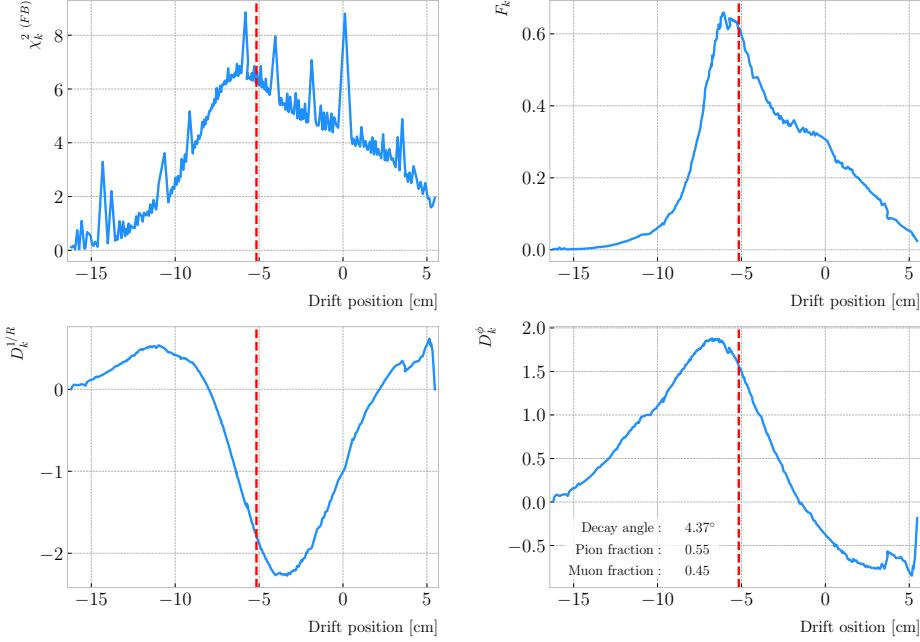


Figure 6.36: Values of $\chi_k^2(FB)$ (top left panel), F_k (top right panel), $D_k^{1/R}$ (bottom left panel) and D_k^ϕ (bottom right panel) versus position along the drift direction for a reconstructed track in a positive pion decay event. The vertical red dashed line indicates the true location of the decay point.

As we already have the estimates from the standard Kalman filter and their covariance matrices at each point, we do not need to repeat the Kalman fit for the new parametrisation. Instead, I can compute the values of α at each point k that minimise the χ^2 resulting from comparing them to $\{\hat{x}_k^B, \hat{x}_k^F\}$. Introducing the two 5×8 matrices:

$$H^F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad H^B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (6.26)$$

we can write this as:

$$\begin{aligned} \chi_k^2(FB)(\alpha) &= (\hat{x}_k^F - H^F \alpha)^T \left[V^{(\hat{x}_k, F)} \right]^{-1} (\hat{x}_k^F - H^F \alpha) \\ &\quad + (\hat{x}_k^B - H^B \alpha)^T \left[V^{(\hat{x}_k, B)} \right]^{-1} (\hat{x}_k^B - H^B \alpha). \end{aligned} \quad (6.27)$$

The minimum of $\chi_k^2(FB)(\alpha)$ is found when the measured new state vector takes the

6.5. CHARGED PION DECAY IN FLIGHT

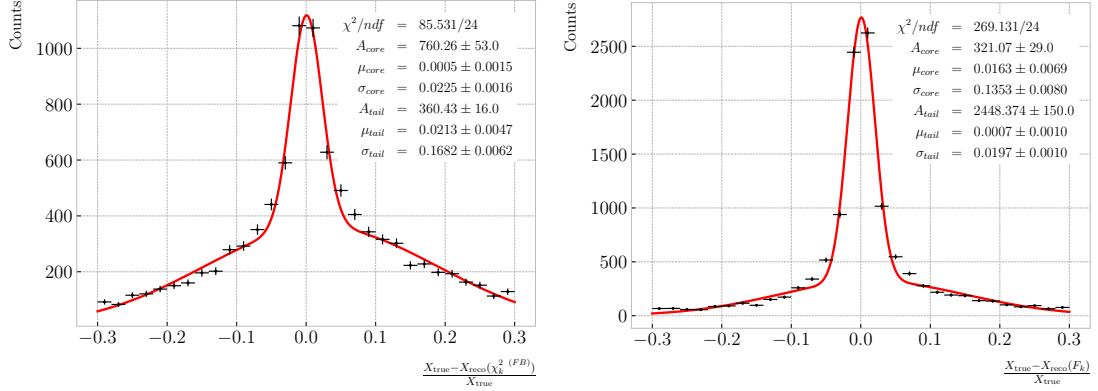


Figure 6.37: Fractional residual distributions of the true and reconstructed decay position along the drift coordinate, using the position of the maximum of $\chi_k^2(FB)$ (left panel) and F_k (right panel) as estimates of the decay position. Also shown are double Gaussian fits to these points (red lines).

3217 value:

$$\hat{\alpha}_k = V^{(\hat{\alpha}_k)} H^T (V^{(\hat{x}_k)})^{-1} \hat{X}, \quad (6.28)$$

3218 where $\hat{X} = \{\hat{x}_k^B, \hat{x}_k^F\}$, $V^{(\hat{x}_k)}$ is the block diagonal matrix formed by $V^{(\hat{x}_k, F)}$ and $V^{(\hat{x}_k, B)}$
3219 and $V^{(\hat{\alpha}_k)}$ is the covariance matrix of $\hat{\alpha}_k$, given by:

$$V^{(\hat{\alpha}_k)} = \left(H^T (V^{(\hat{x}_k)})^{-1} H \right)^{-1}. \quad (6.29)$$

3220 From these new fit estimates we can compute the F statistic, which tells us whether
3221 the model with breakpoint provides a statistically significant better fit:

$$F_k = \left(\frac{\chi_{\text{track},k}^2 - \chi_{\text{full},k}^2}{8 - 5} \right) / \left(\frac{\chi_{\text{full},k}^2}{N - 8} \right). \quad (6.30)$$

3222 One can also compute the signed difference of the duplicated variables divided by
3223 their standard deviation at each point. These represent how significant the discontinuity
3224 in each variable is. For any variable η we can write it as:

$$D_k^\eta = \frac{\hat{\eta}_k^B - \hat{\eta}_k^F}{\sqrt{\text{Var}[\hat{\eta}_k^F] + \text{Var}[\hat{\eta}_k^B] - 2\text{Cov}[\hat{\eta}_k^F, \hat{\eta}_k^B]}}. \quad (6.31)$$

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3225 In our case, the relevant ones to look at are $D_k^{1/R}$ and D_k^ϕ .

3226 Figure 6.36 shows the values of $\chi_k^2(FB)$, F_k , $D_k^{1/R}$ and D_k^ϕ as functions of the position
 3227 along the drift direction, for an example reconstructed track with 55.5% of the energy
 3228 coming from the charged pion and 45.5% from the daughter muon. The true position of
 3229 the decay is indicated (dashed red lines). Notice how $\chi_k^2(FB)$ and F_k , $D_k^{1/R}$ reach their
 3230 maxima near the decay point. In the former case this indicates a large forward-backward
 3231 difference in the track fit. In the later it represents that the extended state vector
 3232 improves the fit particularly around that point.

3233 I can estimate the decay position finding resolution by computing the difference
 3234 between the X position of the maxima of $\chi_k^2(FB)$ and F_k and the X position of the
 3235 true decay. Figure 6.37 represent the the fractional residual distributions for both
 3236 cases, from the sample of tracks containing pion decays. Fitting a double Gaussian to
 3237 the distributions (red lines) I find a resolution of $(3.31 \pm 0.15)\%$ and $(6.94 \pm 0.31)\%$
 3238 respectively.

3239 In principle, the F -statistic should follow a Fisher distribution with $(8 - 5)$ and
 3240 $(N - 8)$ degrees of freedom under the null hypothesis. In most of our cases $N \sim \mathcal{O}(100)$,
 3241 so the probability density functions will look very similar. In this case, it is safe to take
 3242 the limit $N \rightarrow \infty$ in the Fisher PDF:

$$\begin{aligned} \tilde{f}(x; a - b) &= \lim_{N \rightarrow \infty} f(x; a - b, N - a) \\ &= \frac{2^{-\frac{a-b}{2}}}{\Gamma(\frac{a-b}{2})} (a - b)^{\frac{a-b}{2}} x^{\frac{a-b}{2}-1} e^{-\frac{a-b}{2}x}. \end{aligned} \quad (6.32)$$

3243 In our case $a - b = 8 - 5 = 3$, so we would obtain a p-value of 0.05 at $x = 2.60$.

3244 Figure 6.38 contains the distributions of the maxima of $\chi_k^2(FB)$, F_k and D_k^ϕ and the
 3245 minima of $D_k^{1/R}$ for a sample of non-decaying pion tracks (blue) and another sample of
 3246 reconstructed tracks containing part of the pion and the daughter muon from a decay
 3247 inside the fiducial volume (red). Notice that, even though the values of $F_k^{(max)}$ for the
 3248 decay sample are typically larger than for the non-decaying one, just a small fraction of
 3249 the events go beyond the aforementioned value of $F = 2.60$. Therefore, from a practical

6.5. CHARGED PION DECAY IN FLIGHT

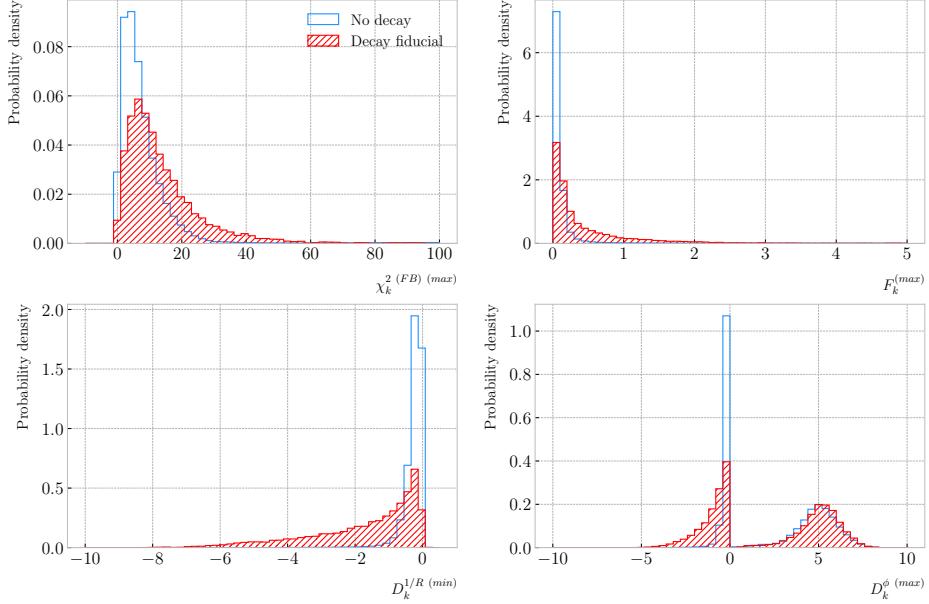


Figure 6.38: Distributions of the extreme values of $\chi_k^2(FB)$ (top left panel), F_k (top right panel), $D_k^{1/R}$ (bottom left panel) and D_k^ϕ (bottom right panel) for non-decaying reconstructed pion tracks (blue) and tracks which include the decay inside the fiducial volume (red).

3250 point of view, it is not the most efficient variable to use for selecting the decay events.

3251 However, looking at the $D_k^{1/R}(\min)$ distribution we can see there is a big difference
 3252 between non-decaying and decaying events in this variable. One can use a combination
 3253 of these four variables to distinguish between the pion decay events (signal) and the
 3254 non-decaying pions (background).

3255 An approach to this classification could be using a boosted decision tree (BDT). One
 3256 of the advantages of BDTs is that they are easy to interpret and identify the relative
 3257 importance of the different input variables. Training a BDT with 400 estimators and a
 3258 maximum depth of 4 I can obtain an efficient classification without overtraining. Figure
 3259 6.39 (left panel) shows the distribution of probabilities predicted by the BDT for a test
 3260 sample. The signal efficiency as a function of background acceptance, the so-called ROC
 3261 curve, is shown in Fig. 6.39 (right panel). With a relative importance of 0.83, the most
 3262 important variable turned out to be $D_k^{1/R}(\min)$.

3263 One thing we can check is how the resolution to the decay and the signal efficiency in

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

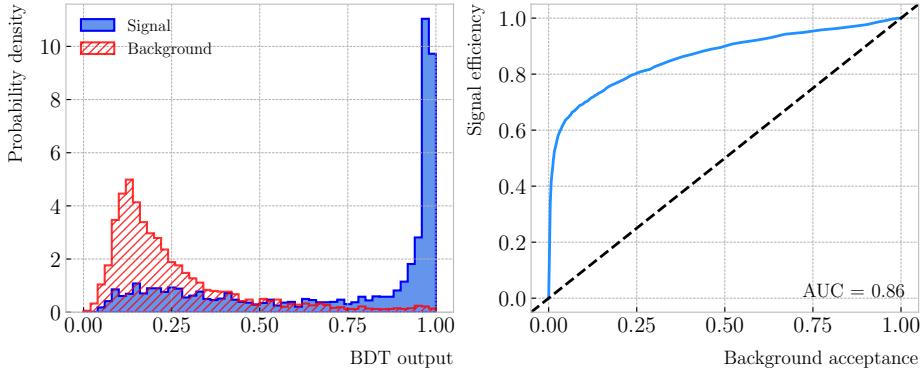


Figure 6.39: Left panel: distributions of the predicted probabilities assigned by the BDT classifier to a test sample of decaying pion+muon tracks (blue) and non-decaying pion tracks (red). Left: signal efficiency versus background acceptance (ROC curve) obtained from the BDT for the test sample.

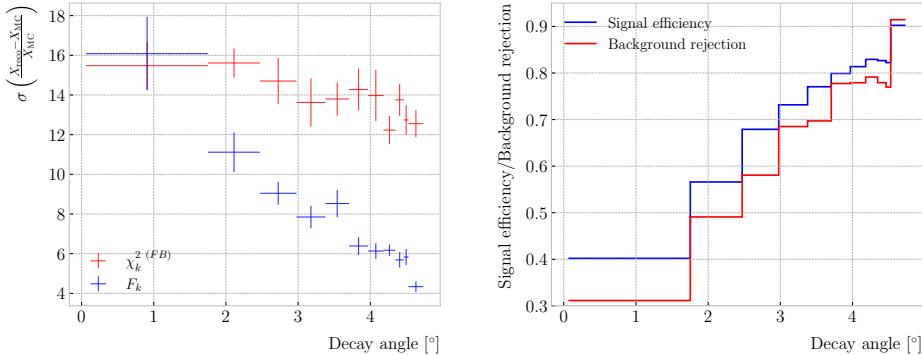


Figure 6.40: Left panel: dependence of the decay position finding resolution on the true value of the decay angle for the $\chi_k^2(FB)$ (red) and F_k (blue) methods. Right panel: signal efficiency (blue line) and background rejection (red line) from the BDT classifier versus true decay angle.

3264 the classification changes with the true decay angle. Using an equal-frequency binning
 3265 for the decay angles, we can repeat the previous steps for each bin.

3266 Figure 6.40 (left panel) shows the dependence on the decay angle of the decay finding
 3267 resolution. We can see that for the $\chi_k^2(FB)$ maximum location method the resolution
 3268 consistently lies between 12 to 16%. However, the $F_k^{(max)}$ approach gives a significantly
 3269 better resolution for high angle values, reaching the 4 – 6% range for decay angles $\geq 4^\circ$.

3270 For the classification dependence on the angle, I use the same classifier I trained
 3271 before but evaluating the test sample for each individual angular bin. I compute the
 3272 signal efficiency in each bin for a fixed value of the background rejection, in this case

6.6. NEUTRAL PARTICLE IDENTIFICATION

3273 90%. Similarly, for the background rejection estimation I use a fixed signal efficiency
3274 value of 90%. Figure 6.40 (right panel) represents the change in signal efficiency (blue)
3275 and background rejection (red) with the value of the true decay angles.

3276 6.6 Neutral particle identification

3277 6.6.1 ECal clustering

3278 Another important reconstruction item is the clustering algorithm of ECal hits in
3279 GArSoft. The default module features a NN algorithm that treats all hits in the same
3280 way, independently of the layer each hit comes from. However, the current ECal design
3281 of ND-GAr has two very different types of scintillator layers. The inner layers are made
3282 out of tiles, which provide excellent angular and timing resolutions. On the other hand,
3283 the outer layers are cross scintillator strips. That way, an algorithm that treats hits
3284 from both kinds of layers differently may be able to improve the current performance.

3285 Inspired by the reconstruction of T2K’s ND280 downstream ECal [166], the idea
3286 was to put together a clustering module that first builds clusters for the different ECal
3287 views (tiles, strips segmented in the X direction and strips segmented in Y direction),
3288 and then tries to match them together to form the final clusters.

3289 Working on a module-by-module basis, the algorithm first separates the hits depending
3290 on the layer type they come from. Then, it performs a NN clustering for the 3 sets of
3291 hits separately. For the tile hits it clusters together all the hits which are in nearest-
3292 neighbouring tiles and nearest-neighbouring layers, for strip hits it looks at nearest-
3293 neighbouring strips and next-to-nearest-neighbouring layers (as the layers with strips
3294 along the two directions are alternated). For strip clusters an additional cut in the
3295 direction along the strip length is needed.

3296 After this first clustering I then apply a recursive re-clustering for each collection
3297 of strip clusters based on a PCA method. In each case, we loop over the clusters with
3298 $N_{hits} \geq 2$, computing the centre of mass and three principal components. Propagating
3299 these axes up to the layers of the rest of the clusters, we check if the propagated point

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3300 and the centre of mass of the second cluster are within next-to-nearest-neighbouring
3301 strips. An additional cut in the direction along the strip length is also needed. Moreover,
3302 I require that the two closest hits across the two clusters are at most in next-to-nearest-
3303 neighbouring strips. I merge the clusters if these three conditions are satisfied. The
3304 re-clustering is repeated until no more cluster pairs pass the cuts.

3305 The clusters in each strip view are combined if their centres of mass are close enough
3306 and they point in the same direction. An alternative approach for the strip cluster
3307 merging could be to compute the overlap between the ellipsoids defined by the principal
3308 axes of the clusters, and then merge the pair if the overlap exceeds some threshold.
3309 Further study is needed to understand if this change would have an impact in the overall
3310 clustering performance.

3311 To merge the tile clusters to the combined strip clusters I propagate the principal
3312 axis of the strip cluster towards the inner layers, up to the centre of mass layer of the
3313 tile cluster. I merge the clusters if the distance between the propagated point and the
3314 centre of mass is below a certain cut.

3315 The last step is to check if clusters in neighbouring modules should be merged
3316 together, both across two barrel modules, across end cap modules and between barrel
3317 end cap modules. I check the distance between the two closest hits in the pair of clusters
3318 and merge them if it passes this and an additional direction cut.

3319 Figure ?? presents an example of the clustering steps relevant for strip layer hits, from
3320 the input hits (top left panel) to the NN clustering (top right panel) and re-clustering
3321 (bottom left panel) for each strip view and the final merging strip clusters (bottom
3322 right panel). It shows the hits from a single ECal barrel module in a ν_μ CC interaction
3323 event with a neutral pion and a proton in the final state. The two clusters on the left
3324 correspond to the photon pair from the π^0 decay and the one on the upper right corner
3325 is associated to the proton.

3326 This algorithm has a total number of eight free parameters that need to be optimised.
3327 I used a sample of 1000 ν_μ CC interactions in order to obtain the optimal configuration of
3328 clustering parameters. This sample was generated up to the default ECal hit clustering

6.6. NEUTRAL PARTICLE IDENTIFICATION

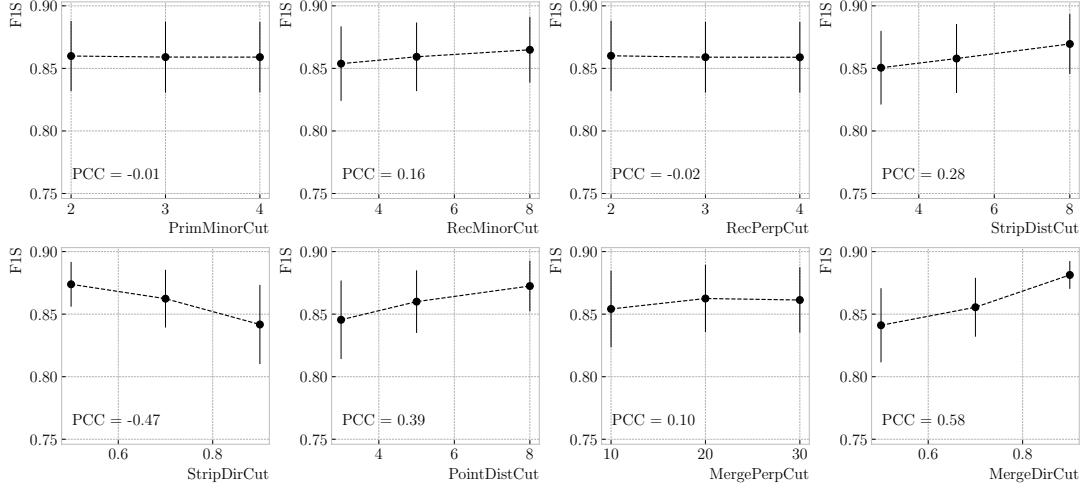


Figure 6.41: Mean values of the F_1 -score marginal distributions for the different free parameters of the new clustering algorithm, with the error bars representing one standard deviation around the mean. The F_1 -score values were computed for the 6561 possible parameter configurations using 1000 ν_μ CC interaction events.

Table 6.5: Summary of parameters and sampled values used in the optimisation of the clustering algorithm.

Name	Units	Sampled values	Description
PrimMinorCut	strips	2, 3, 4	Distance along strip length in NN clustering
RecMinorCut	strips	3, 5, 8	Distance between propagated point and CM along strip length in re-clustering
RecPerpCut	strips	2, 3, 4	Closest hit pair distance in re-clustering
StripDistCut	strips	3, 5, 8	Distance between CMs in strip cluster merging
StripDirCut	cos	0.5, 0.7, 0.9	Main axes direction cut in strip cluster merging
PointDistCut	tiles	3, 5, 8	Distance between propagated point and CM in strip-tile matching
MergePerpCut	cm	10, 20, 30	Closest hit pair distance in module merging
MergeDirCut	cos	0.5, 0.7, 0.9	Main axes direction cut in module merging

3329 level, so then I could run the new clustering algorithm each time with a different
 3330 configuration of parameters. As the number of parameters is relatively large, I only
 3331 performed a coarse-grained scan of the parameter space. Sampling each of the eight
 3332 parameters at three different points each I obtain 6561 different configurations. These
 3333 parameters, together with the used values, are summarised in Tab. 6.5.

3334 In order to measure the performance of the clustering, I use a binary classification
 3335 approach. For each formed cluster, I identify the Geant4 Track ID of the matching MC

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3336 particle and the energy fraction of each hit. Then, I assign to each cluster the Track ID
3337 with the highest total energy fraction. For each of the different Track IDs associated to
3338 the clusters, I select the cluster with the highest energy (only from the hits with the
3339 same Track ID). I identify such a cluster as the main cluster for that Track ID. I count
3340 as true positives (TPs) the hits with the correct Track ID in each main cluster. False
3341 positives (FPs) are the hits with the incorrect Track ID for the cluster they are in, not
3342 only main clusters. The false negatives (FNs) are the hits with the correct Track ID in
3343 clusters other than the main.

3344 Figure 6.41 shows the computed F_1 -score values for the different cuts. In each case,
3345 the central value represents the mean of the F_1 -score distribution for the specified value
3346 of the corresponding variable and the vertical error bar represents one standard deviation
3347 around the mean. Also shown are the Pearson correlation coefficients of these central
3348 values. We can see that five of the variables have a sizeable effect on the F_1 -score, with
3349 an absolute difference between the last and first values as big as 4%.

3350 The working configuration is obtained as follows. I first select all configurations
3351 with purity $\geq 90\%$. Among those, I choose the combinations that yield the maximum
3352 F_1 -score. If more than one configuration remains I select the one with the highest
3353 sensitivity. Doing so, I end up with a parameter configuration with an efficiency of 88%
3354 and a 90% purity. Compared with the default algorithm, which gives an efficiency of
3355 76% and a purity of 91% for the same sample, I have managed to improve the efficiency
3356 by a factor of 1.16.

3357 6.6.2 π^0 reconstruction

3358 One of the potential applications of the new ECal hit clustering is the reconstruction of
3359 neutral particles, in particular pions. Neutral pions decay promptly after being produced,
3360 through the $\pi^0 \rightarrow \gamma\gamma$ channel ($98.823 \pm 0.034\%$) of the time. The photon pair does
3361 not leave any traces in the HPgTPC (unless one or both of them converts into an
3362 electron-positron pair), but each of them will produce an electromagnetic shower in
3363 the ECal.

6.6. NEUTRAL PARTICLE IDENTIFICATION

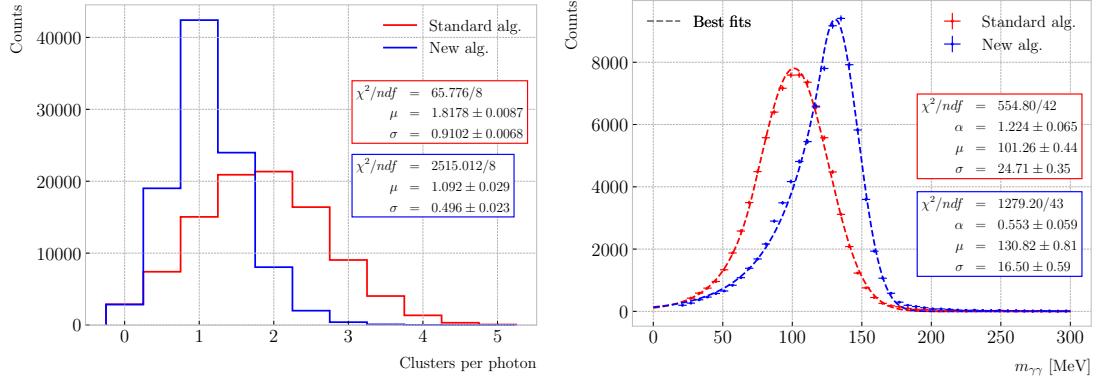


Figure 6.42: Left panel: distributions of the number of ECal clusters per photon from π^0 decays for the standard (red) and new (blue) clustering algorithms. Right panel: reconstructed invariant mass distributions for photon pairs from single π^0 events using the standard (red) and new (blue) ECal clustering algorithms.

3364 To test the potential impact of the new algorithm in π^0 reconstruction, I generated
 3365 a MC sample of single, isotropic neutral pions inside the HPgTPC. All pions were
 3366 generated with a momentum $p = 500$ MeV and their initial positions were uniformly
 3367 sampled inside a $2 \times 2 \times 2$ m box aligned with the centre of the TPC. I ran both the
 3368 default and the new clustering algorithms, using for the latter the optimised configuration
 3369 discussed above.

3370 The first thing to notice is that the number of clusters produced per photon has
 3371 decreased. Figure 6.42 (left panel) shows these distributions for the default (red) and
 3372 new (blue) algorithms. Using a simple Gaussian fit, we see that the mean number of
 3373 ECal clusters per photon went from 1.82 ± 0.01 to 1.09 ± 0.03 . This effectively means that
 3374 with the new algorithm the ECal activity of one true particle is typically reconstructed
 3375 as a single object. From the reconstruction point of view this can be an advantage. As
 3376 now most of the photon energy ends up in a single ECal cluster, I can simply use cluster
 3377 pairs to identify the π^0 decay.

3378 In general, one calculates the invariant mass of the photon pair as:

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos \theta)}, \quad (6.33)$$

3379 where E_i are the energies of the photons and θ the opening angle between them. In this

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

case I can use the energies deposited in the ECal and their incident directions. This quantity is computed for all possible pairs of clusters, using their position together with the true decay point. In a more realistic scenario, e.g. ν_μ CC interaction, one could use the position of the reconstructed primary vertex instead. I also tried to use the principal direction of the clusters, but that approach gave considerably worse results. For each event I only keep the pair with an invariant mass closer to the true π^0 mass value.

Figure 6.42 (right panel) shows the invariant mass distributions for the photon pairs we get using the default (red) and the new (blue) ECal clustering algorithms. For the fit I used a modified version of the Crystal Ball function [167], obtained by taking the limit where the parameter controlling the power-law tail goes to infinity:

$$f(x; N, \mu, \sigma, \alpha) = N \cdot \begin{cases} e^{\frac{\alpha(2x-2\mu+\alpha\sigma)}{2\sigma}}; & x \leq \mu - \alpha\sigma, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}; & x > \mu - \alpha\sigma. \end{cases} \quad (6.34)$$

Comparing the fitted mean and standard deviation values for the Gaussian cores, we see that the distribution for the new algorithm is a 67% narrower and also peaks much closer to the true m_{π^0} value, going from 101.3 ± 0.4 MeV to 130.8 ± 0.6 MeV.

6.7 Integration in GArSoft

All the additions and improvements to the reconstruction discussed in this Chapter had to be integrated in the GArSoft framework. This is necessary both to allow a more streamlined path for development, as this makes testing and adding features straightforward, as well as make the changes usable in future productions of simulated data. In this section, I outline the current status of the integration in GArSoft of the reconstruction work presented above.

The new track-cluster association code has been implemented in GArSoft, under the name of `TPCECALAssociation2`, and has now become the new default in the reconstruction. The structure of the module is similar to the previous implementation, and the data products they output are identical in form. Therefore, any existing code

6.7. INTEGRATION IN GArSOFT

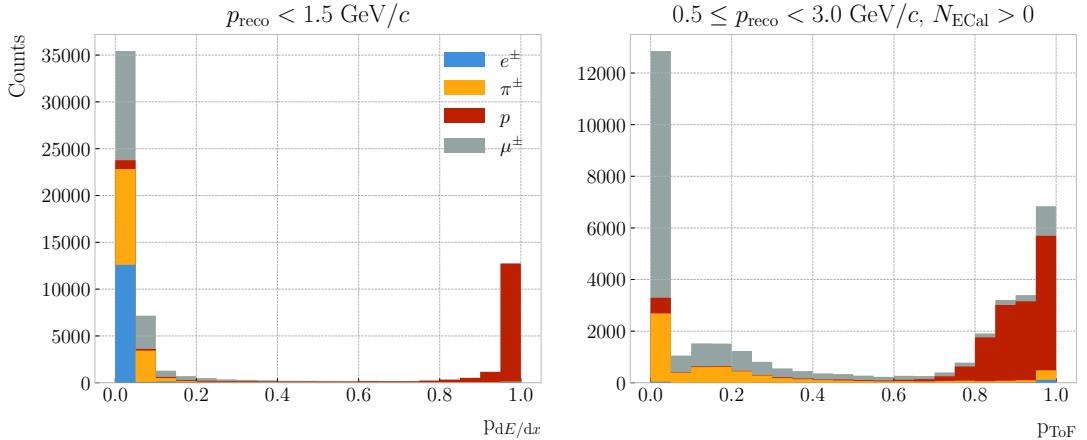


Figure 6.43: Distributions of proton dE/dx (left panel) and ToF (right panel) scores for a sample of 100000 FHC neutrino interactions in the HPgTPC. The distributions are broken down by the true particle type associated to the reconstructed particle.

3404 using the association objects does not need to be modified.

3405 The computation of the truncated mean dE/dx of the tracks, the evaluation of
 3406 the muon score for muon and pion separation, and the estimation of the velocity from
 3407 time-of-flight are all orchestrated by the `CreateRecoParticles` module. Each one of
 3408 these is implemented as a separate algorithm, which is then called by the parent module.
 3409 This generates the `gar::rec::RecoParticle` products, a new high-level data object in
 3410 GArSoft. These combine the information from the HPgTPC, ECal, and μ ID to create
 3411 an object useful for analysers. At the moment, these data products are only generated
 3412 for charged particles. However, in the future the module can be extended to incorporate
 3413 other algorithms used for the identification of neutral particles, like neutral pions and
 3414 neutrons.

3415 Additionally, analogous to the muon score, the `gar::rec::RecoParticle` objects
 3416 contain two other scores based on the $\langle dE/dx \rangle$ and ToF estimates which measure the
 3417 “protoness” of a reconstructed particle. These are obtained in a number of momentum
 3418 bins, and are a measure of the distance to the point in the corresponding distribution
 3419 that maximises the F_1 -score for the proton separation. This distance is then transformed
 3420 applying a sigmoid function, which produces a score in the 0 – 1 range, with coefficients
 3421 obtained following a procedure similar to the one used to calibrate the response of

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

3422 the muon score. The dE/dx proton score is defined for all particles with momenta
3423 $p_{\text{reco}} < 1.5 \text{ GeV}/c$, whereas the ToF proton score is available for the particles with at least
3424 one associated hit in the inner ECal and momentum in the range $0.5 \leq p_{\text{reco}} < 3.0 \text{ GeV}/c$.
3425 As an example, Fig. 6.43 shows the distributions of the dE/dx (left panel) and ToF
3426 (right panel) proton scores for the reconstructed particles in a 100000 FHC neutrinos
3427 sample.

3428 The calculation of the track breakpoint variables for pion decay identification is
3429 currently implemented as an analysis module in GArSoft. It would be interesting to add
3430 this information to the `gar::rec::RecoParticle` products, possibly calling the code as
3431 an additional algorithm in the `CreateRecoParticles` module. However, the best way
3432 to propagate the information to the high-level objects is still unclear.

3433 About the new ECal clustering algorithm, it is still in a development phase, and
3434 as such it has not replaced the current clustering module. At the moment, its latest
3435 version is integrated in GArSoft as the `CaloClustering2` module. The algorithm used
3436 is implemented separately, and then invoked in the main code. The module can be
3437 run standalone on the outputs of the reconstruction, creating a second instance of the
3438 `gar::rec::Cluster` collection. In the future it may replace the current algorithm as
3439 the default in the reconstruction chain. However, more work is needed in order to
3440 understand its performance in all the different use cases.

3441

3442

Event selection in ND-GAr

3443 *You have power over your mind, not outside events. Realise this, and you will
 3444 find strength.*

3445

– Marcus Aurelius, *Meditations*

3446 As discussed previously, it is necessary to evaluate the capabilities of ND-GAr at
 3447 identifying different particles. In the context of the LBL analysis, we want ND-GAr to
 3448 provide data samples containing events of specific topologies, like ν_μ CC $1\pi^\pm$, ν_μ CC
 3449 $1p1\pi^\pm$, etcetera. Thus, developing a strategy for the event selection using the current
 3450 reconstruction is required.

3451 In this Chapter, I present the results of a number of preliminary studies focused on
 3452 the event selection in ND-GAr, particularly the ν_μ CC selection and the pion tagging
 3453 strategies. I also investigate the neutrino energy reconstruction, as well as the systematic
 3454 uncertainties relevant for our detector.

3455 7.1 Data sample

3456 For the event selection studies I used a MC sample consisting of 10^5 FHC neutrino
 3457 interaction events inside the HPgTPC volume. The version of **GENIE** used was v3_04_00,
 3458 with the G18 tune. This is a preliminary version of the re-tune produced from CCQE,
 3459 CC1 π , CC2 π , and CC inclusive bubble chamber cross section data [168]. It uses the local
 3460 Fermi gas as a description of the nuclear model. The quasielastic-like events are described
 3461 by the Nieves quasielastic [169] and Valencia 2p2h [170] models. The Berger-Seghal

CHAPTER 7. EVENT SELECTION IN ND-GAR

3462 model [171, 172] is used for the resonant and coherent pion production. As in all the
3463 GENIE tunes, the Bodek-Yang model [173] describes the DIS interactions. Finally, the
3464 FSI are described using the effective intranuclear transport model in INTRANUKE.

3465 For this sample, I used `GArG4` instead of `edep-sim` for the particle propagation.
3466 Because both `Geant4` wrappers use different configurations for the simulation, the results
3467 obtained are different. The default `edep-sim` configuration used by the DUNE ND
3468 is appropriate for ND-LAr, where thresholds for particle production are higher. In
3469 the case of ND-GAr, these parameters need to be adjusted accordingly. For the time
3470 being, in these first productions of analysis files, we will use our standalone `Geant4`
3471 implementation.

3472 The detector simulation and reconstruction used was GArSoft version v02_21_00. I
3473 made use of the standard routines for the readout simulation and the reconstruction,
3474 which include the additions described in section 6.7. A summary of the GArSoft outputs
3475 is extracted in the form of a plain `ROOT TTree`. These are then used, together with the
3476 GENIE output files, to produce the files known within DUNE as common analysis files
3477 (CAFs). The version of the CAF format used in this analysis is `duneanaobj v3_07_00`.

3478 This sample only includes single interaction events. In the future, we will move
3479 to simulate full neutrino spills. Also, we will need to include neutrino interactions in
3480 the other detector volumes (ECal, magnet, . . .), as well as rock muons making it to
3481 ND-GAr. However, this will require a significant amount of work to go into the so-called
3482 interaction slicer, the part of the reconstruction in charge of splitting the reconstructed
3483 events.

3484 Looking forward, these sort of small samples are useful to prepare for launching a
3485 full production of ND-GAr events. In the original DUNE TDR LBL analysis, the event
3486 rates are calculated with a 1.1×10^{21} POT/year assumption, which assumes a combined
3487 uptime and efficiency of the accelerator complex and the LBNF beamline of 57% [49].
3488 If we have one spill every 1.2 s, that translates into 7.5×10^{13} POT/spill. Therefore,
3489 assuming that the POT/spill scales linearly with beam power, in Phase II we will have
3490 1.3×10^{14} POT/spill for the for the 2.1 MW beam. Or equivalently, 1.9×10^{21} POT/year

7.2. ν_μ CC SELECTION

Table 7.1: Estimated event rates in ND-GAr, divided by interaction type and pion multiplicity, for two different values of the POT/year.

Process	Events/ton/year	
	1.1×10^{21} POT/year	1.9×10^{21} POT/year
All ν_μ -CC	1.60×10^6	2.83×10^6
CC 0π	5.28×10^5	9.35×10^5
CC $1\pi^\pm$	3.02×10^5	5.34×10^5
CC $1\pi^0$	1.65×10^5	2.92×10^5
CC 2π	3.18×10^5	5.63×10^5
CC 3π	1.36×10^5	2.41×10^5
CC other	1.52×10^5	2.69×10^5
All $\bar{\nu}_\mu$ -CC	7.54×10^4	1.33×10^5
All NC	5.50×10^5	9.73×10^5
All ν_e -CC	2.70×10^4	4.78×10^4

3491 using the same efficiency. The event rates per year in ND-GAr computed for these two
 3492 possible values of the POT/year are shown in Tab. 7.1.

3493 The latest PRISM plan requires 1.50 POT · years of data on-axis, followed by
 3494 0.25 POT · years at each off-axis position ($2, 4, 8, 12, 16, 20, 24$, and 28 m), both for
 3495 FHC and RHC mode. This implies that a full on-axis ND-GAr production will require
 3496 a total of 2.85×10^{21} POT for both horn currents. The production of these samples
 3497 is necessary to understand the impact of ND-GAr on the LBL sensitivities, and the
 3498 studies presented here should be considered as a first step towards the realisation of
 3499 such analysis.

3500 7.2 ν_μ CC selection

3501 In a ν_μ CC inclusive selection, the signal topology we look for is a neutrino-induced
 3502 muon with or without other final state particles. Here, I also require the neutrino vertex
 3503 to be located inside the fiducial volume (FV) of ND-GAr.

3504 The FV is defined as a smaller cylinder within the cylindrical volume of the HPgTPC.

CHAPTER 7. EVENT SELECTION IN ND-GAR

3505 The FV has a radius R_{FV} and a half-length L_{FV} . For a particle position to lie within
 3506 the FV it must satisfy:

$$\vec{x}_i \in \left\{ \vec{x} \in \mathbb{R}^3 \mid |x_0| \leq L_{\text{FV}} \text{ \& } \sqrt{x_1^2 + x_2^2} \leq R_{\text{FV}} \right\}, \quad (7.1)$$

3507 in the reference frame of the HPgTPC. For convenience, I define:

$$\begin{aligned} \Delta R_{\text{FV}} &= R_{\text{HPgTPC}} - R_{\text{FV}}, \\ \Delta L_{\text{FV}} &= L_{\text{HPgTPC}} - L_{\text{FV}}, \end{aligned} \quad (7.2)$$

3508 where R_{HPgTPC} and L_{HPgTPC} refer to the radius and the half-length of the HPgTPC,
 3509 respectively. Figure 7.1 shows the HPgTPC volume with the FV inside of it. In that
 3510 representation, the FV is defined as $\Delta L_{\text{FV}} = 30.0$ cm and $\Delta R_{\text{FV}} = 30.0$ cm. Also shown
 3511 is the HPgTPC reference frame, with x being the drift direction and z aligned along the
 3512 beam direction.

3513 In some cases, it is interesting to divide the signal events in different categories
 3514 based on their true interaction mode. In this work, I will distinguish between charged-
 3515 current quasi-elastic (CCQE), coherent (CCCOH), resonant (CCRES), and deep-inelastic
 3516 (CCDIS) interactions. I also use a separate category for the interactions not included in
 3517 any of the other categories (CCOther).

3518 Any other events are considered backgrounds. For this selection, I use the following
 3519 categorisation of background events:

- 3520 • Out of FV: if the true neutrino vertex lies outside the defined FV.
- 3521 • NC: if the event is a true neutral-current event.
- 3522 • $\bar{\nu}_\mu$ CC: if the true neutrino candidate is of muon antineutrino flavour.
- 3523 • Other: if the event is not signal nor falls in any of the other background categories.

3524 The key to the CC selection is the identification of a primary muon candidate.
 3525 Typically, this is the longest track in the event. However, sometimes protons and pions

7.2. ν_μ CC SELECTION

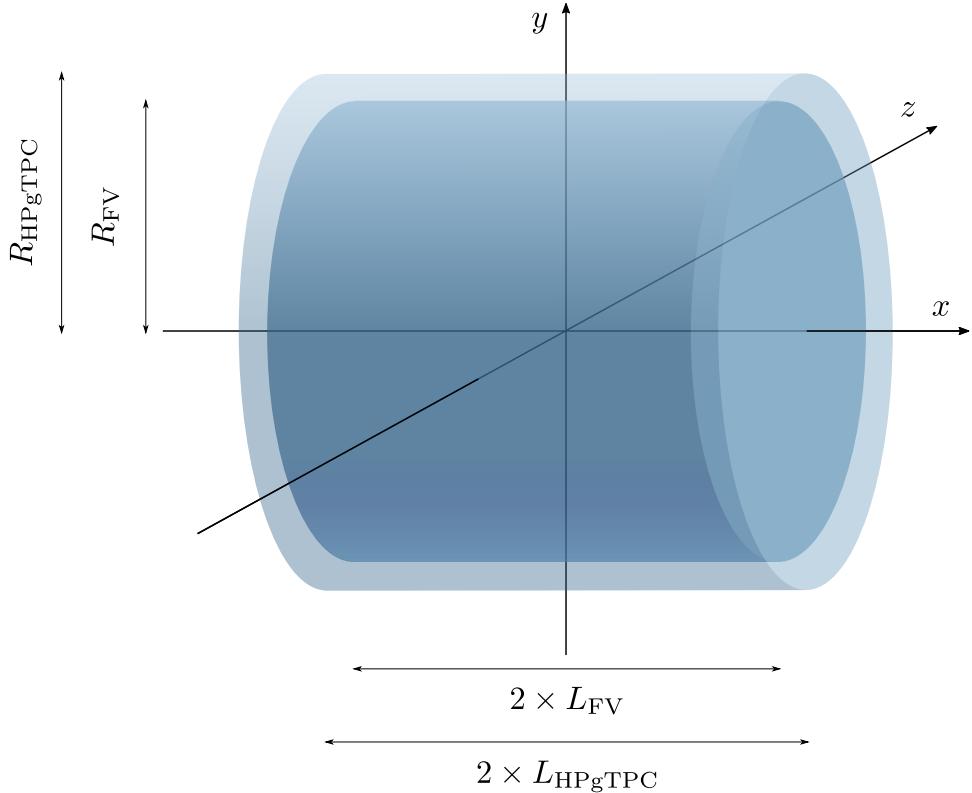


Figure 7.1: Schematic diagram of the HPgTPC including the fiducial volume (FV). In this case the FV is given by $\Delta L_{\text{FV}} = 30.0$ cm and $\Delta R_{\text{FV}} = 30.0$ cm.

3526 leave tracks longer than that of the muon. This is particularly important in the GAr
 3527 medium, considerably less dense than the LAr. For this reason, the muon identification
 3528 in ND-GAr relies heavily on the capabilities of the ECal.

3529 The selection strategy proposed combines the information coming from the three
 3530 main detection systems of ND-GAr: the HPgTPC charge readout, and the ECal and
 3531 μ ID detectors. It consists of five steps:

- 3532 1. Event contains reconstructed particles.
 3533 2. Select particles with reconstructed negative charge, $q_{\text{reco}} = -1$.
 3534 3. Select particles passing the muon score cut, $\mu_{\text{score}} \geq \mu_{\text{score}}^{\text{cut}}$.
 3535 4. Keep reconstructed particle with the highest momentum, $\max [p_{\text{reco}}]$.
 3536 5. Check that the remaining particle starts within the FV.

CHAPTER 7. EVENT SELECTION IN ND-GAR

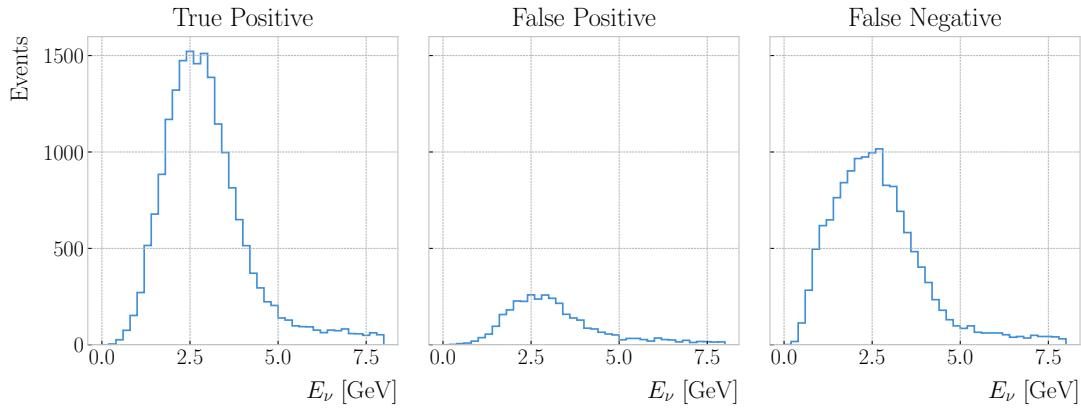


Figure 7.2: True positive (left panel), false positive (middle panel), and false negative (right panel) true neutrino energy distributions for the ν_μ CC selection given by a muon score cut of $\mu_{\text{score}}^{\text{cut}} = 0.75$, and a FV defined as $\Delta L_{\text{FV}} = 30.0$ cm and $\Delta R_{\text{FV}} = 30.0$ cm.

3537 All the events passing these cuts are classified as signal, and the selected particle is
 3538 regarded as the primary muon candidate.

3539 **7.2.1 Selection optimisation**

3540 I performed an optimisation of this selection, comparing the performance of a number of
 3541 configurations. For the muon selection, I varied the value of $\mu_{\text{score}}^{\text{cut}}$ from 0.05 to 0.95,
 3542 using a step size of 0.05. Additionally, to optimise the FV, I systematically explored a
 3543 number of different parameter configurations, moving within the 10.0 – 70.0 cm range for
 3544 ΔL_{FV} and 25.0 – 75.0 cm for ΔR_{FV} , in increments of 10.0 cm and 5.0 cm respectively.

3545 For each parameter configuration, I extract three different true neutrino energy
 3546 distributions. These are built combining the results of the selection described previously,
 3547 which we can refer to as the “reco” selection, and a “true” selection. The later identifies
 3548 the true ν_μ CC events using the GENIE event records, and checks that the true neutrino
 3549 vertices are contained in the FV.

3550 The first distribution consists of the events passing both selections, i.e., these are
 3551 the true ν_μ CC events which pass the “reco” selection. The second distribution contains
 3552 the events passing the “reco” selection but failing the “true” selection. These are
 3553 the background events that the selection misidentifies. Finally, the third distribution

7.2. ν_μ CC SELECTION

3554 corresponds to the events picked by the “true” selection but not by the “reco” one. In
 3555 other words, these are the true ν_μ CC events that our selection misses. In analogy to
 3556 the machine learning jargon, I refer to these distributions as the true positive (TP),
 3557 false positive (FP), and false negative (FN) spectra, respectively. Figure 7.2 shows an
 3558 example of these three distributions for the case $\mu_{\text{score}}^{\text{cut}} = 0.75$, $\Delta L_{\text{FV}} = 30.0$ cm, and
 3559 $\Delta R_{\text{FV}} = 30.0$ cm.

3560 By making different combinations of these distributions one can compute a series of
 3561 performance metrics. Using the full information from the spectra allows to obtain the
 3562 scores as a function of the true neutrino energy, whereas the totals can be obtained by
 3563 integrating the histograms. This way, the efficiency of the selection is given by:

$$\begin{aligned} \text{Efficiency} &= \frac{\text{Selected true } \nu_\mu \text{ CC events}}{\text{Total true } \nu_\mu \text{ CC events}} \\ &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \end{aligned} \quad (7.3)$$

3564 while the purity can be written as:

$$\begin{aligned} \text{Purity} &= \frac{\text{Selected true } \nu_\mu \text{ CC events}}{\text{Total selected events}} \\ &= \frac{\text{TP}}{\text{TP} + \text{FP}}. \end{aligned} \quad (7.4)$$

3565 Another scoring metric typically used when quantifying the performance of a selection
 3566 is the significance. It is defined as:

$$\text{Significance} = \frac{S}{\sqrt{S+B}} = \frac{\text{TP}}{\sqrt{\text{TP}+\text{FP}}}. \quad (7.5)$$

3567 The significance measures the relative size of the true signal within the selection, $S = \text{TP}$
 3568 with respect to one standard deviation of the counting experiment. Assuming Poisson
 3569 statistics, the variance is equal to the number of observations, and therefore the standard
 3570 deviation equals to $\sqrt{N} = \sqrt{S+B} = \sqrt{\text{TP}+\text{FP}}$. I use this metric to

3571 Figure 7.3 shows the change in efficiency (blue) and purity (red) of the ν_μ CC
 3572 selection as a function of the different cuts. From left to right, I vary $\mu_{\text{score}}^{\text{cut}}$, ΔL_{FV} ,

CHAPTER 7. EVENT SELECTION IN ND-GAR

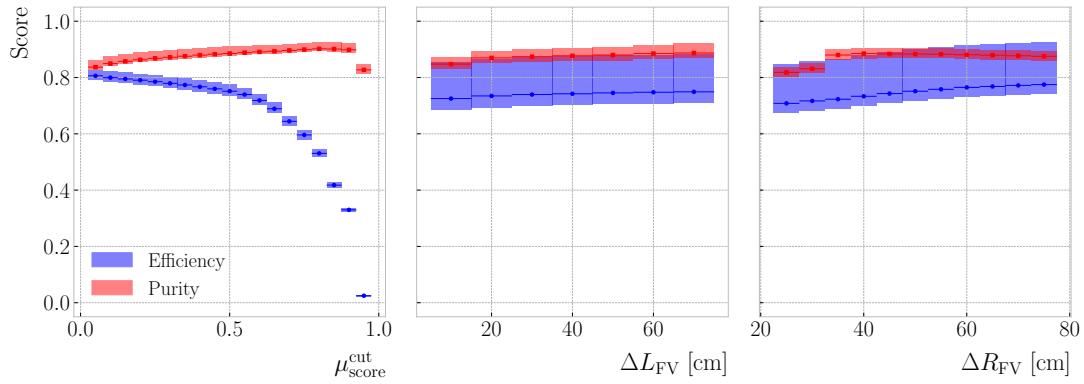


Figure 7.3: Efficiency (blue) and purity (red) for the ν_μ CC selection as a function of the muon score cut (left panel), FV half-length cut (middle panel), and radial cut (right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the horizontal line corresponds to the median.

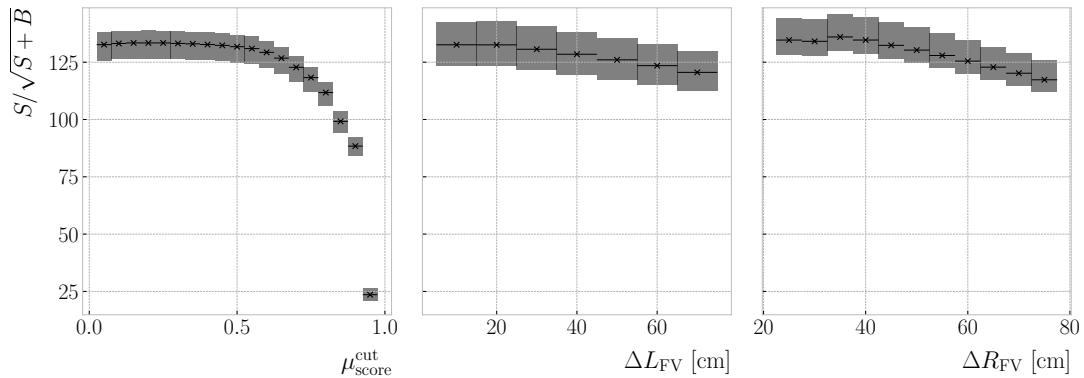


Figure 7.4: Significance for the ν_μ CC selection as a function of the muon score cut (left panel), FV half-length cut (middle panel), and radial cut (right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the horizontal line corresponds to the median.

3573 and ΔR_{FV} . For each value of the cuts, I compute the median and IQR (represented
 3574 by the horizontal lines and the heights of the boxes, respectively) of the corresponding
 3575 conditional distributions of efficiency and purity. This representation is useful to get
 3576 an idea of the general trend the scores follow with the cuts, as well as the spread. It
 3577 is clear that the muon score cut has the biggest impact on the efficiency, which ranges
 3578 between 0.05 to 0.80, whereas the purity remains stable with values around 0.85.

3579 A similar depiction of the significance can be found in Fig. 7.4. In this case, one can
 3580 see that the $S/\sqrt{S+B}$ decreases as the cuts grow tighter. However, there are hints of

7.2. ν_μ CC SELECTION

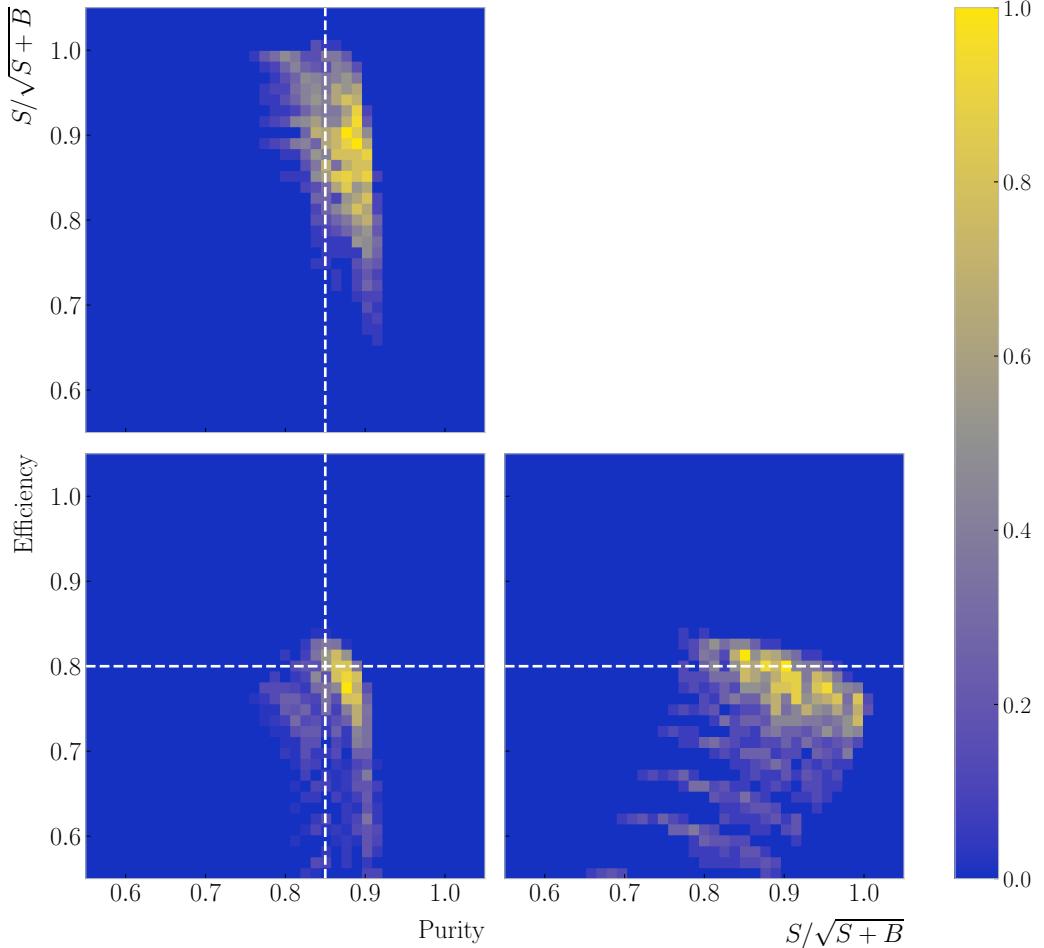


Figure 7.5: Normalised 2D distributions of efficiency, purity and significance for the ν_μ CC selection. The $S/\sqrt{S+B}$ is normalised to the highest value achieved. The vertical dashed line indicates a purity value of 0.85, whereas the horizontal one corresponds to an efficiency of 0.80.

3581 local maxima at intermediate values.

3582 Selecting the cut configuration with the highest significance, 147 ± 11 for the parameter
 3583 values explored here, results in an efficiency and purity of 0.754 ± 0.006 and 0.833 ± 0.007 ,
 3584 respectively. Figure 7.5 shows the 2D distributions resulting when combining pairs of
 3585 efficiency, purity and significance, obtained for the cut configurations explored. The
 3586 significance is normalised to the highest value obtained in the parameter scan. Looking
 3587 at this, it is clear that a selection with highest efficiency and purity can be achieved,
 3588 maintaining a similar significance level.

CHAPTER 7. EVENT SELECTION IN ND-GAR

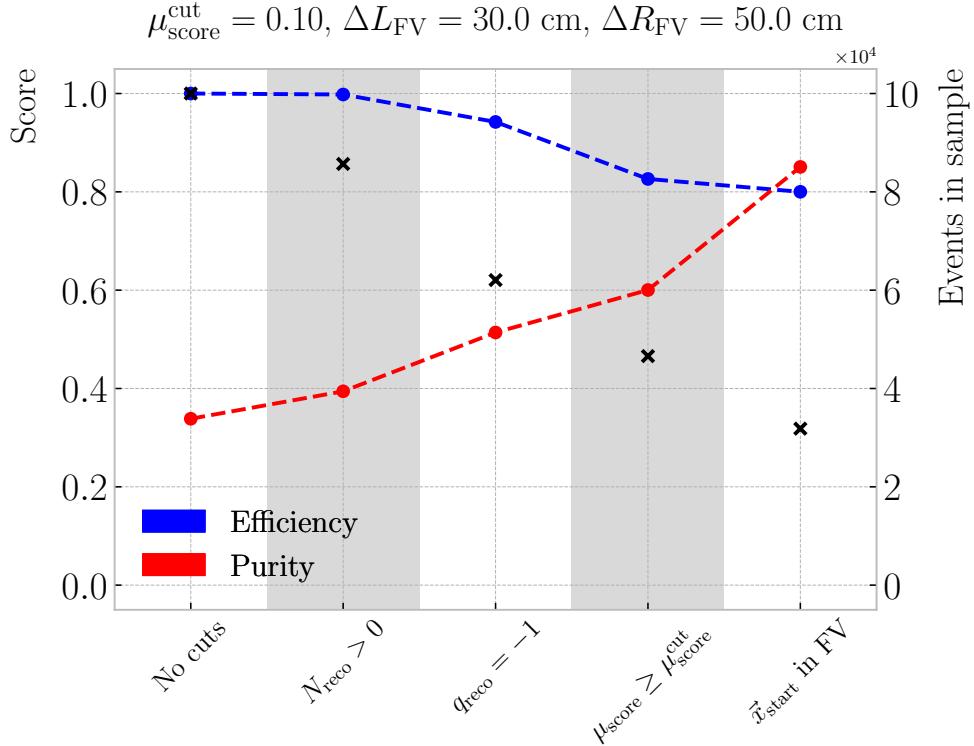


Figure 7.6: Cumulative efficiency (blue) and purity (red) of the ν_μ CC selection. The secondary axis indicates the number of events in the sample after each cut (black crosses).

Table 7.2: Step-by-step ν_μ CC selection cuts and cumulative passing rates. Relative passing rates are indicated in parentheses.

Cut #	Selection cut	Events	Passing rates
0	Total number of events (No cuts)	100000	100.00% (100.00%)
1	At least one reconstructed particle	85680	85.68% (85.68%)
2	Negatively charged particles only	62054	62.05% (72.43%)
3	$\mu_{\text{score}} \geq \mu_{\text{score}}^{\text{cut}}$	46585	46.59% (75.07%)
4	Candidate \vec{x}_{start} in FV	31834	31.83% (68.34%)

3589 Therefore, to get a more refined selection, I first select the configurations with a
 3590 purity and an efficiency higher than 0.85 and 0.80, respectively. After that, I select the
 3591 tuple of cuts yielding the highest significance. The resulting value for the muon score
 3592 cut is $\mu_{\text{score}}^{\text{cut}} = 0.10$, and the FV is given by $\Delta L_{\text{FV}} = 30.0 \text{ cm}$ and $\Delta R_{\text{FV}} = 50.0 \text{ cm}$.
 3593 With these, one obtains a total efficiency of 0.800 ± 0.007 and purity of 0.851 ± 0.008 ,

7.2. ν_μ CC SELECTION

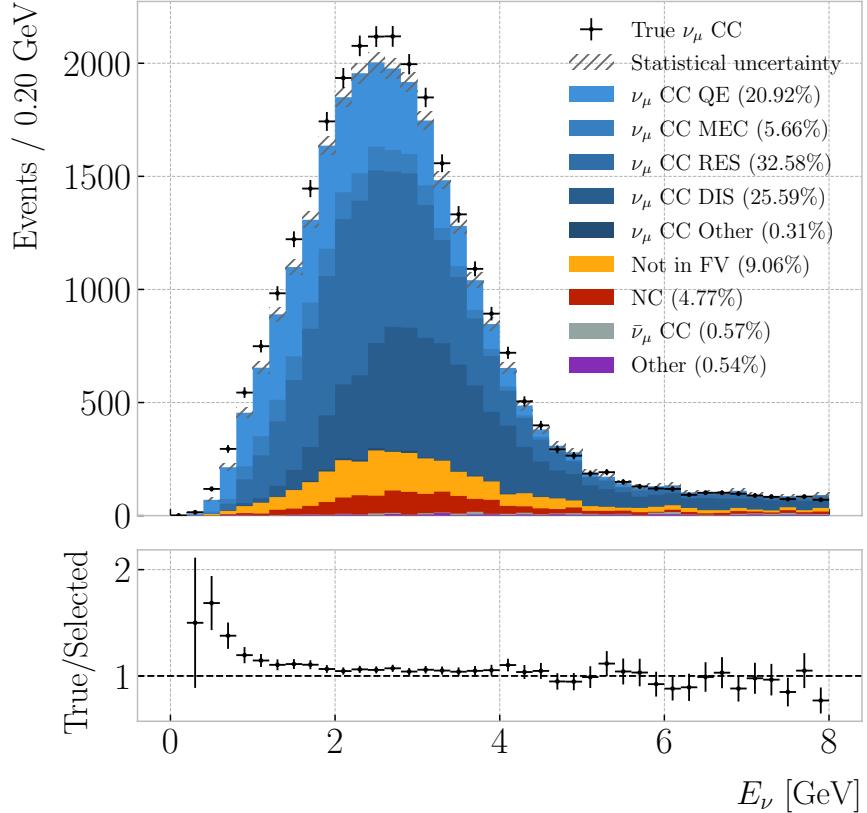


Figure 7.7: True neutrino energy spectra for the ν_μ CC selection. The selected events correspond to the coloured stacked histogram, broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray. The true distribution is also shown with the black data points. The bottom panel shows the ratio between the number of true and selected ν_μ CC events per bin.

with a significance of 138 ± 11 . Hereafter, I use this optimised selection cuts, unless specified otherwise.

A summary of the selection can be found in Tab. 7.2. It shows the number of events in the selected sample after each selection cut, as well as the absolute and relative passing rates. Figure 7.6 shows the overall efficiencies (blue) and purities (red) after each cut in the event selection is applied. As expected, the efficiency drops while the purity increases with the successive cuts.

Notice how, out of the cuts prior to the FV constraint, the sign selection produces the highest increase in purity. This is one of the advantages of having a magnetised TPC, and can also be used for a $\bar{\nu}_\mu$ CC selection when running in RHC mode.

CHAPTER 7. EVENT SELECTION IN ND-GAR

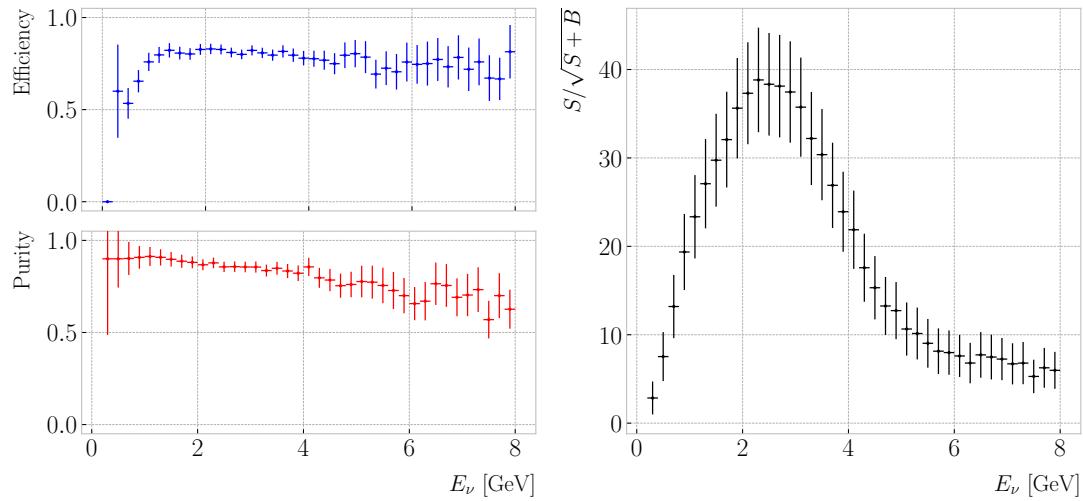


Figure 7.8: Left panel: efficiency (top panel) and purity (bottom panel) for the ν_μ CC selection as a function of the true neutrino energy. Right panel: significance for the ν_μ CC selection as a function of the true neutrino energy

3604 7.2.2 Selection performance

3605 Using the stored spectra discussed above, the true neutrino energy distribution for the
 3606 selected events can be recovered doing TP + FP. Similarly, the combination TP + FN
 3607 gives the true spectrum. Figure 7.7 shows the true (black data points) and selected
 3608 (coloured stacked histogram) E_ν distributions for the optimised ν_μ CC selection. The
 3609 colours in the selected spectrum indicate the different signal categories and backgrounds,
 3610 with the overall statistical uncertainty represented by the gray hatched mess. The ratio
 3611 between the true and selected events is also shown. One can see that it sits around 1 for
 3612 most of the energy range. However, for energies ≤ 1 GeV there is a significant deficit of
 3613 selected events.

3614 These spectra also allow to compute the efficiency and purity of the selection as
 3615 a function of the true neutrino energy, as shown in Fig. 7.8 (left panel). As it could
 3616 be expected from the previous ratio plot, the efficiency is low at low neutrino energies.
 3617 Nonetheless, it raises quickly with the energy, until it stabilises around a value of 0.80.
 3618 Looking at the purity, one may notice that, although it starts at around 0.90, there is a
 3619 significant decrease towards the high end of the spectrum. Figure 7.8 (right panel) also

7.2. ν_μ CC SELECTION

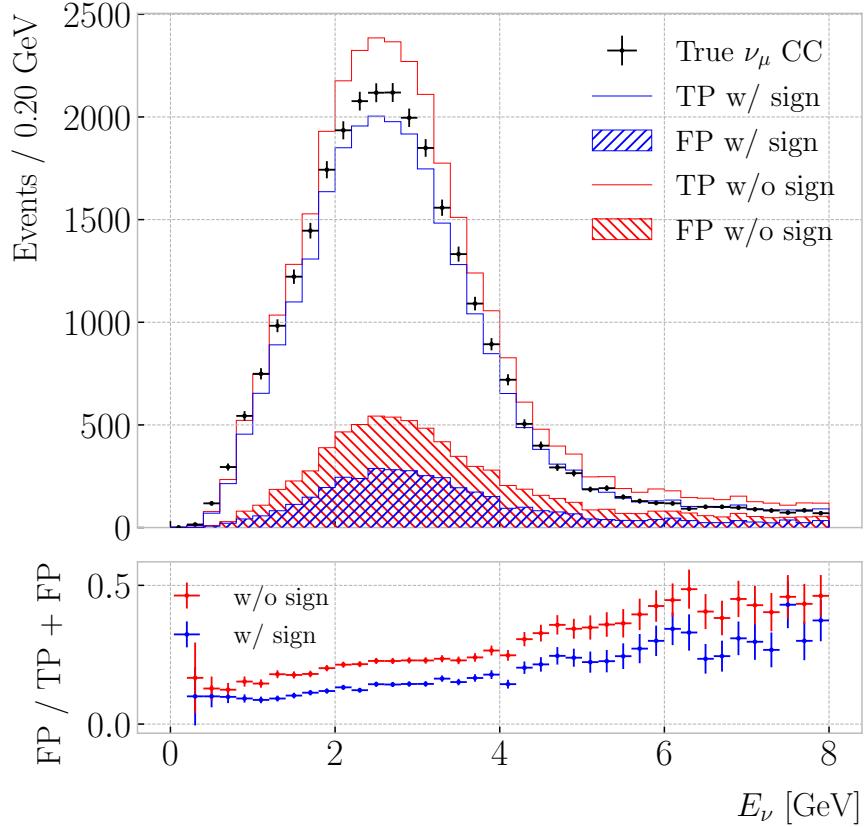


Figure 7.9: True neutrino energy spectra for the ν_μ CC selection with (blue) and without (red) sign selection. The selected events are broken down by true positives (signal) and false positives (background). The true distribution is also shown (black data points). The bottom panel shows the ratios between the number of false positives and total selected events per bin.

3620 shows the significance as a function of the energy. In this case, the highest $S/\sqrt{S+B}$ is
 3621 achieved around the energies where the spectrum peaks.

3622 A variation of the ν_μ CC selection one can try is to apply it without the reconstructed
 3623 charge cut. Figure 7.9 (top panel) shows the E_ν distributions corresponding to the
 3624 selection with (blue stacked histogram) and without (red stacked histogram) the sign
 3625 selection. In the former case, the out of FV contamination amounts to 9.06% of the
 3626 total, while the NC contamination results 4.77% and the wrong-sign contamination
 3627 0.57%. For the later, these backgrounds account for the 10.01%, 10.82%, and 2.18%
 3628 of the selected events, respectively. As expected, removing the positive particles does
 3629 not change the FV-related effects noticeably. However, the sign selection proves its

CHAPTER 7. EVENT SELECTION IN ND-GAR

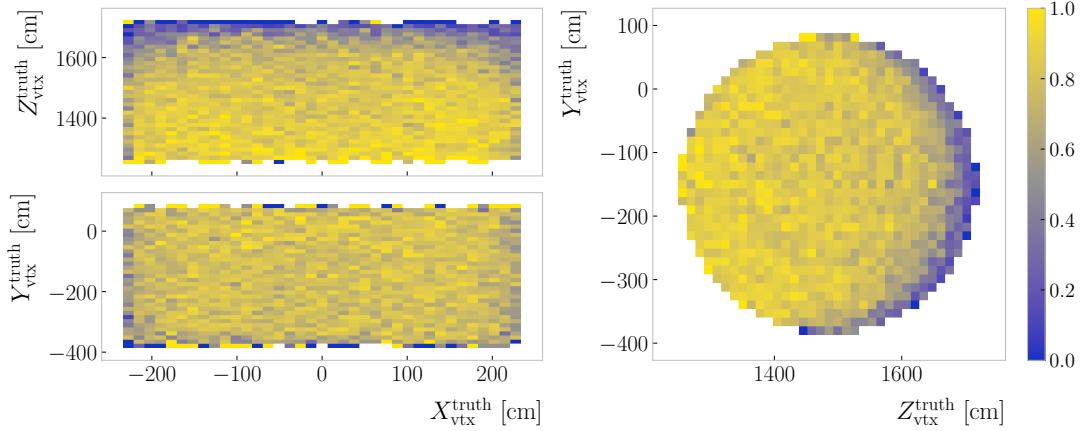


Figure 7.10: Efficiency 2D distributions for the ν_μ CC selection given the true position of the interaction vertex.

3630 worth in the rejection of $\bar{\nu}_\mu$ CC events, which drop almost by one order of magnitude.

3631 Additionally, the charge selection cuts the NC events in half, as it reduces the chances

3632 of misidentifying a positively charged hadron for a muon.

3633 As an additional check, I explored how the performance of the ν_μ CC selection
 3634 depends on the position of the neutrino interaction within the HPgTPC. Maps of the
 3635 selection efficiency for the X, Z (top left panel), X, Y (bottom left panel), and Z, Y
 3636 (right panel) true vertex position pairs are given in Fig. 7.10. It can be seen that the
 3637 efficiency remains stable along the drift direction, only slightly degrading close to the
 3638 edges of the FV. Regarding the radial direction, it is clear that an important number of
 3639 events with high $Z_{\text{vtx}}^{\text{truth}}$ are not being selected. Intuitively, the muons arising from these
 3640 interactions will leave short tracks. As their directions are typically aligned with the
 3641 beam direction, they enter the ECal shortly after production. This is likely to affect
 3642 the tracking, and therefore their identification. As a result, the regions with the lowest
 3643 efficiency are the downstream corners of the HPgTPC, i.e. the areas with high $|X_{\text{vtx}}^{\text{truth}}|$
 3644 and $Z_{\text{vtx}}^{\text{truth}}$.

3645 7.2.3 Primary muon kinematics

3646 This ν_μ CC selection relies on the identification of the a primary muon, meaning that
 3647 for each selected event a particle is picked out as the muon candidate. It is because of

7.2. ν_μ CC SELECTION

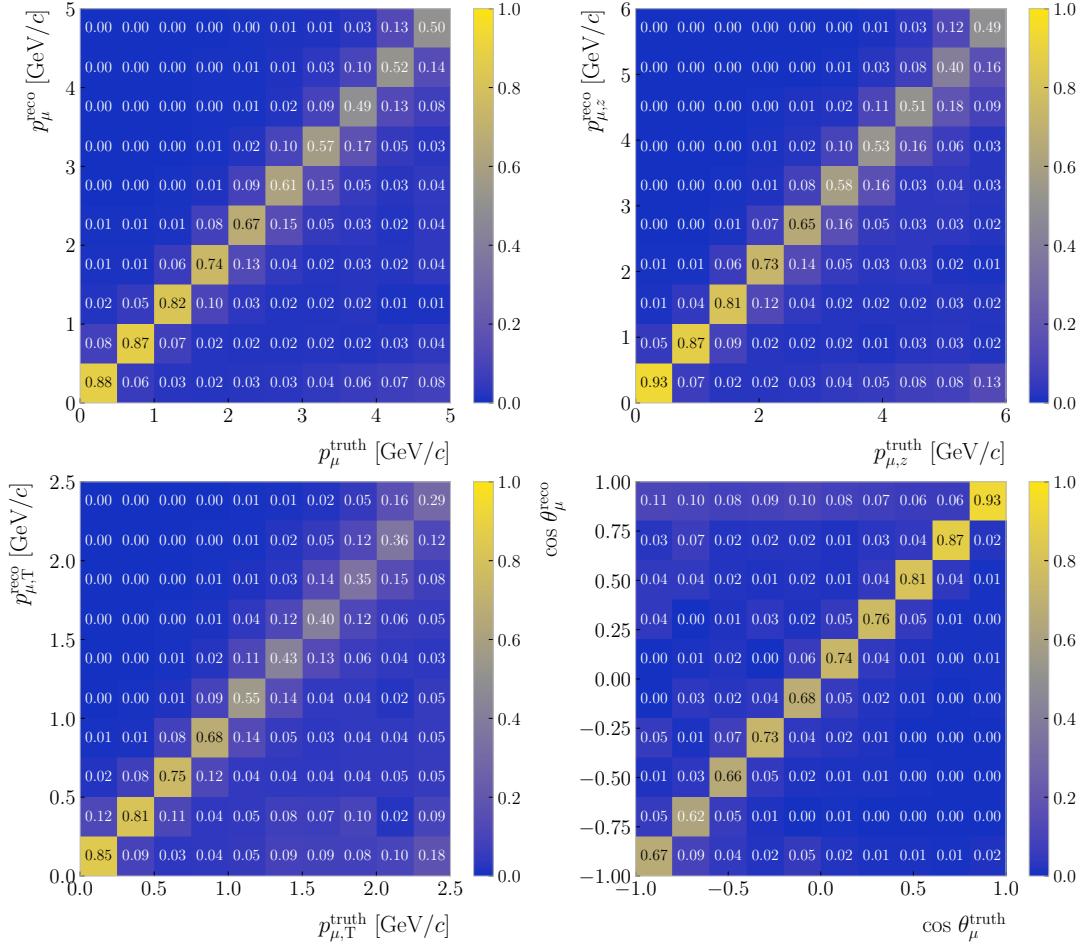


Figure 7.11: Distributions for the reconstructed versus truth generated primary muon momentum (top left panel), longitudinal momentum (top right panel), transverse momentum (bottom left panel), and beam angle (bottom right panel). The reconstructed values correspond to the selected primary muon candidate, whereas the truth values come from the true primary muon in the event.

3648 this that one can study the kinematics of these selected primary muons.

3649 Figure 7.11 shows a comparison between some of the reconstructed and truth primary
 3650 muon kinematic variables. From top to bottom, left to right, we have muon momentum,
 3651 longitudinal momentum, transverse momentum and beam angle. The histograms are
 3652 column-normalised, and so the diagonal entries give an idea of the resolution for the
 3653 different variables. The match between truth and reconstructed values can only be done
 3654 for the selected true ν_μ CC events, as the others do not have a primary muon. However,
 3655 for this comparison I do not require the events to start inside the FV.

CHAPTER 7. EVENT SELECTION IN ND-GAR

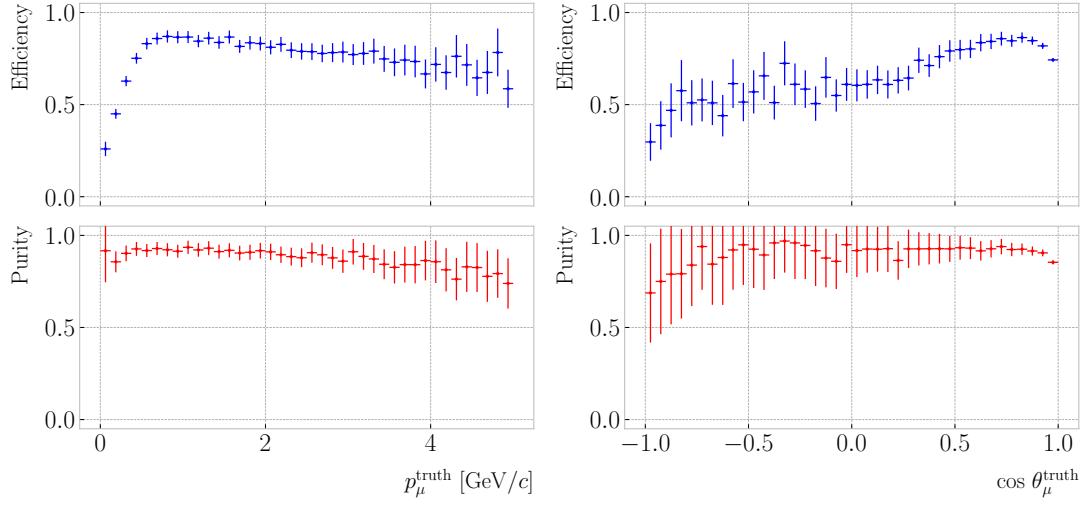


Figure 7.12: Efficiency (blue) and purity (red) of the ν_μ CC selection as a function of the primary muon true momentum (left panel) and beam angle (right panel).

3656 Notice that, for the reconstructed values, the variables do not necessarily come
 3657 from a reconstructed particle that matches the true primary muon. In other words,
 3658 sometimes, even though the event was correctly identified, the primary muon may have
 3659 been confused with another particle. That means that in these distributions include
 3660 both reconstruction and selection deficiencies.

3661 I also studied the performance of the ν_μ CC selection as a function of the kinematic
 3662 variables of the primary muon. As before, these metrics are only possible to compute for
 3663 true ν_μ CC events. The efficiency (top panels) and purity (bottom panels) as a function
 3664 of the truth muon momentum (left) and beam angle (right) are shown in Fig. 7.12. One
 3665 can see that there are some similarities in the behaviour of both metrics between the
 3666 true neutrino energy and the muon momentum cases. This is to be expected, as these
 3667 two variables are highly correlated. For the efficiency, there is a rapid increase at low
 3668 momentum values until it peaks at around 1 GeV/*c*, after which it starts decreasing
 3669 slowly. The purity remains relatively constant, with a slight drop towards high p_μ^{truth}
 3670 values. In the case of the muon angle, the decrease in efficiency at high $\theta_\mu^{\text{truth}}$ is more
 3671 noticeable. However, note that the number of events with backward-going muons is
 3672 much smaller than those aimed towards the forward direction, as can be seen from the

7.2. ν_μ CC SELECTION

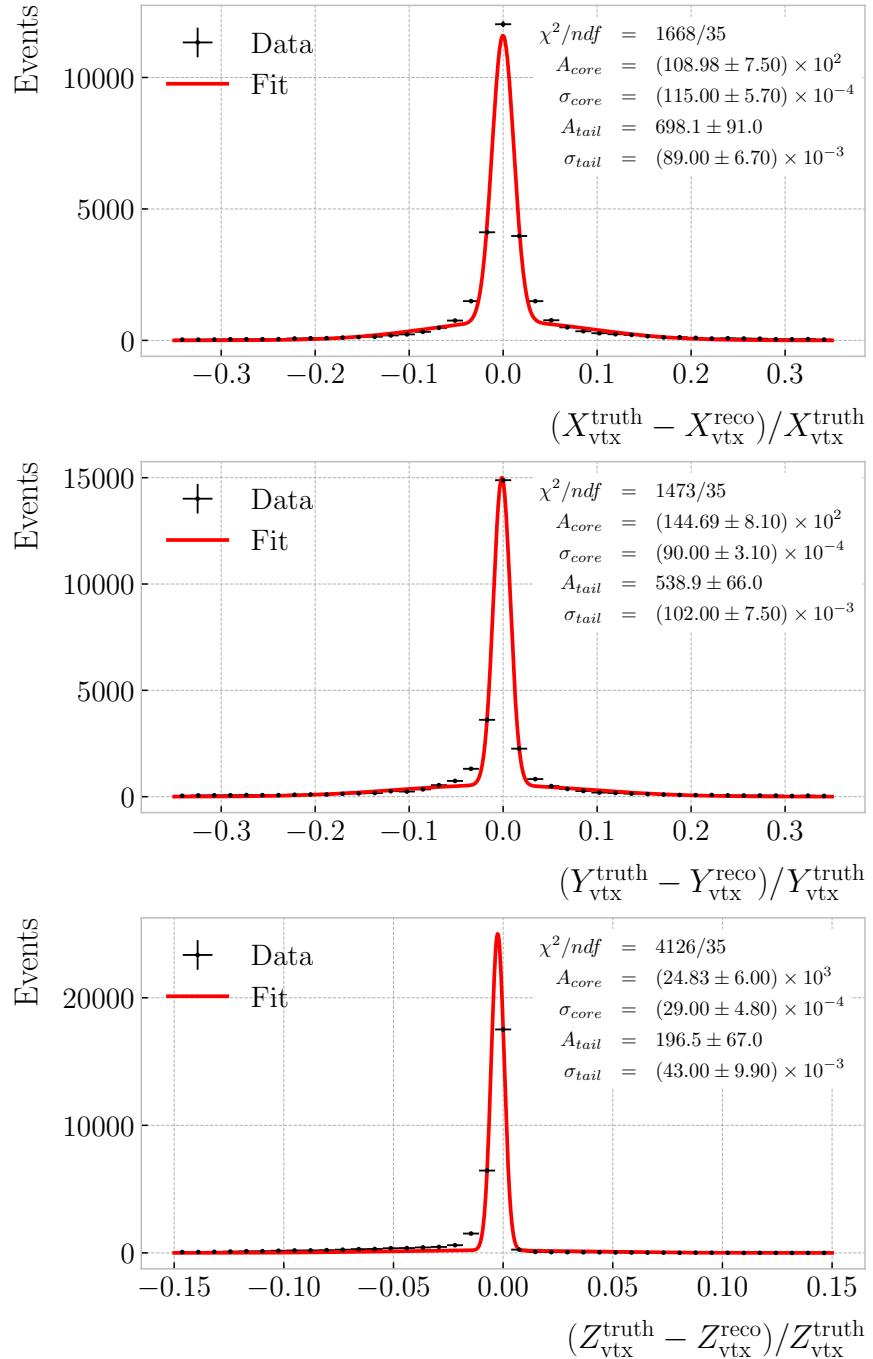


Figure 7.13: Fractional residual distributions for the position of the primary vertex in the ν_μ CC selection. The best fits to a double Gaussian function are also shown (red lines).

CHAPTER 7. EVENT SELECTION IN ND-GAR

3673 size of the vertical error bars. There is also a decline in the purity with the beam angle,
3674 but this effect is much smaller.

3675 A byproduct of selecting the primary lepton in the interaction is the position
3676 of the reconstructed neutrino vertex candidate. Checking how the position of the
3677 selected reconstructed primary vertex and the true vertex position compare is needed to
3678 understand the validity of our method. Figure 7.13 shows the distributions of fractional
3679 residuals between the truth and reconstructed vertex positions in the X (top panel),
3680 Y (middle panel), and Z (bottom panel) directions. Performing a double Gaussian fit
3681 to the distributions (red lines), I estimate the reconstructed vertex resolution achieved
3682 with this method to be $1.62 \pm 0.08\%$, $1.23 \pm 0.05\%$, and $0.32 \pm 0.05\%$ for the X , Y ,
3683 and Z directions, respectively. As expected, the resolution along the drift direction.
3684 However, the significant difference in resolution between the two transverse directions is
3685 worth noting. Not only the resolution is better for the Z direction, but the layout of the
3686 residual distribution is highly asymmetrical. This may be related to the variability in
3687 the selection efficiency along that direction.

3688 7.3 Charged pion identification

3689 Now that I have checked the robustness of the proposed ν_μ CC selection, it can be
3690 used as a starting point for other, more convoluted, selections. One of the priorities
3691 of ND-GAr, as mentioned previously, is the identification of pions. With its lower
3692 tracking thresholds, ND-GAr is expected to do better regarding π^\pm identification than
3693 the traditional LArTPCs, like ND-LAr. Moreover, it can make use of the different
3694 detector subcomponents to tag the charged pions.

3695 The ν_μ CC selection provides a starting point for the pion identification. The first
3696 thing one can do is rule out the selected primary muon candidate. Then, by looking at
3697 the properties of the rest of the reconstructed particles, one can start the counting of
3698 the charged pions.

3699 The two proton scores, the one based on the dE/dx in the HPgTPC and the one

7.3. CHARGED PION IDENTIFICATION

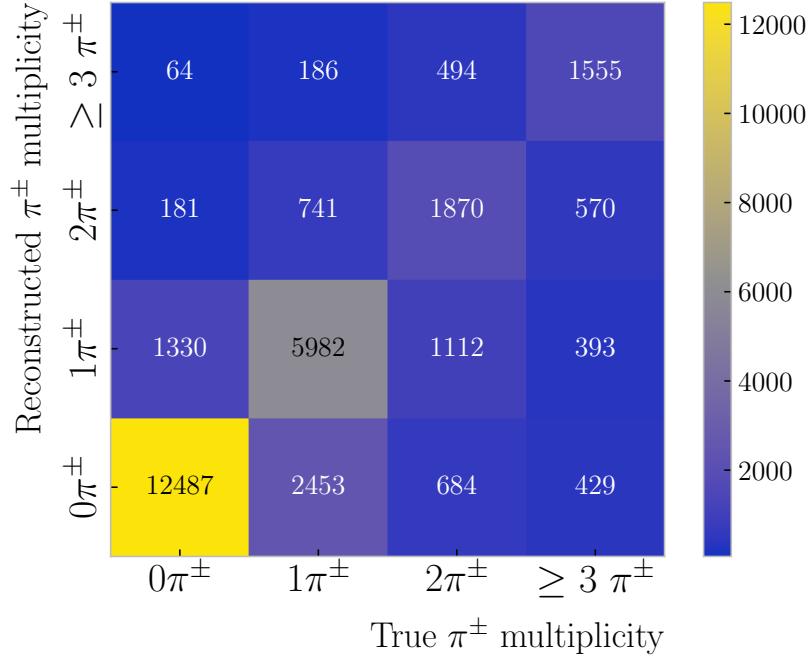


Figure 7.14: Distribution of events given their true and reconstructed π^\pm multiplicity, for the selection given by $p_{dE/dx}^{\text{cut}} = p_{\text{ToF}}^{\text{cut}} = 0.50$, $\Delta_{dE/dx}^{\pi^\pm} = 0.20$, and $d_\mu^{\text{cut}} = 50.0$ cm.

3700 obtained from the ToF measurement in the ECal, can be used to separate the protons
 3701 from the sample of charged pions. By providing appropriate cuts for these, a good
 3702 separation can be achieved.

3703 Another source of information available is the dE/dx of the track associated to the
 3704 reconstructed particle. To select the charged pions, we can require that the measured
 3705 mean dE/dx is compatible with the expectation for a true π^\pm , in other words:

$$\left\langle \frac{dE}{dx} \right\rangle_{\pi^\pm} \left(1 - \Delta_{dE/dx}^{\pi^\pm} \right) \leq \left\langle \frac{dE}{dx} \right\rangle_{\text{meas.}} < \left\langle \frac{dE}{dx} \right\rangle_{\pi^\pm} \left(1 + \Delta_{dE/dx}^{\pi^\pm} \right), \quad (7.6)$$

3706 where the parameter $\Delta_{dE/dx}^{\pi^\pm}$ measures the fractional variation one allows around the
 3707 theoretical expectation. To obtain the expected mean dE/dx of a charged pion with a
 3708 given momentum, I use the ALEPH parametrisation with the parameter values obtained
 3709 previously.

3710 Also, as we are only interested in the primary pions, and because these are by
 3711 definition close to the interaction vertex, one can apply an additional distance cut. Using

CHAPTER 7. EVENT SELECTION IN ND-GAR

3712 the start position of the muon candidate, we can restrict the starting point of pions to a
3713 certain volume around the vertex.

3714 Combining all these ideas, I propose the following procedure to identify the charged
3715 pions in an event:

- 3716 1. Apply ν_μ CC selection.
- 3717 2. Disregard particle selected as primary muon.
- 3718 3. Remove particles with momentum below threshold.
- 3719 4. Select particles with proton dE/dx score below threshold.
- 3720 5. Select particles with proton ToF score below threshold.
- 3721 6. Select particles with mean dE/dx around the expected value for a pion.
- 3722 7. Remove particles with a distance between the start of the track and the primary
3723 vertex greater than the cut.

3724 The remaining particles after all these cuts are taken to be charged pion candidates.

3725 This counting method depends on four cuts, denoted by $p_{dE/dx}^{\text{cut}}$, $p_{\text{ToF}}^{\text{cut}}$, $\Delta_{dE/dx}^{\pi^\pm}$, and
3726 d_μ^{cut} in order of appearance. The momentum threshold is necessary to compare with
3727 the true multiplicity. For values of the kinetic energy lower than 10 – 20 MeV, we
3728 do not expect to be able to tag individual pions. Such low energy particles just leave
3729 small traces in the TPC which, together with the busy environment of the neutrino
3730 interaction vertex, leaves one with no other option but to only account for their energy
3731 calorimetrically. As such, the true pion counting also features this momentum threshold.

3732 I performed an optimisation of the charged pion counting by scanning the space of
3733 possible cut configurations. For the two proton scores, I let them vary between 0.10 to
3734 0.90, in increments of 0.10. Similarly, the parameter $\Delta_{dE/dx}^{\pi^\pm}$ takes values in the range
3735 0.05 – 0.25, with a step size of 0.05. Finally, the distance cut changes in 10 cm steps,
3736 from 10 to 120 cm.

7.3. CHARGED PION IDENTIFICATION

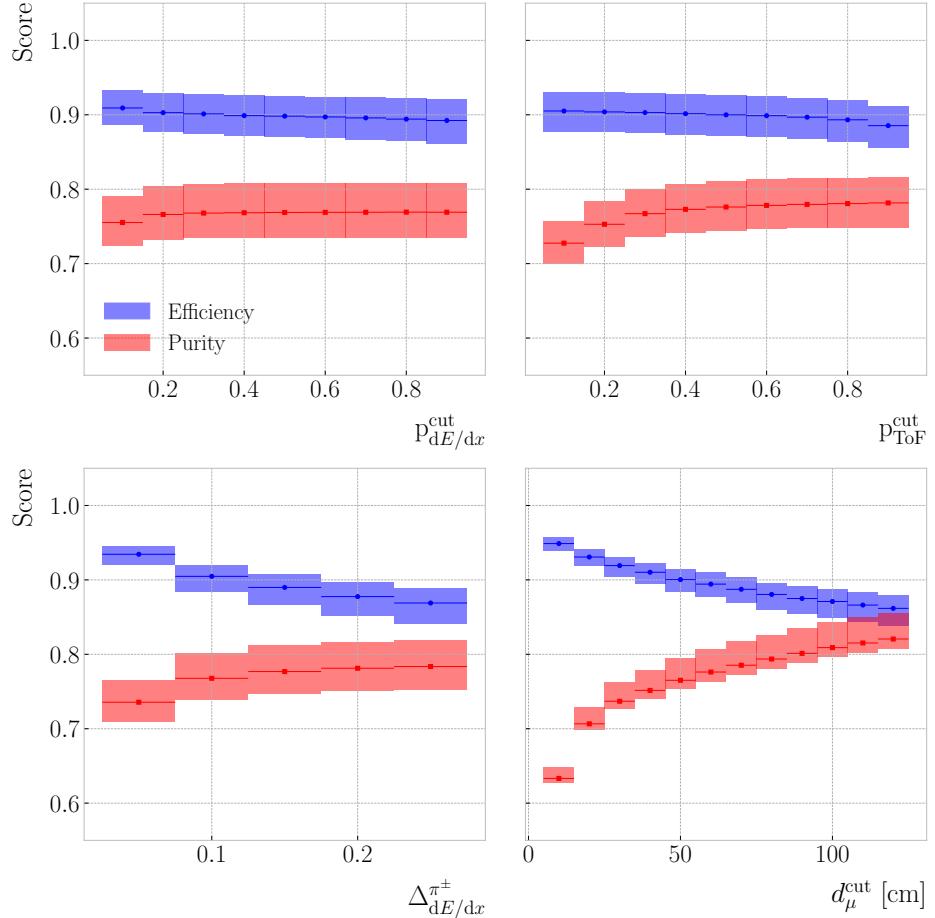


Figure 7.15: Efficiency (blue) and purity (red) for the ν_μ CC $0\pi^\pm$ selection as a function of the proton dE/dx score cut (top left panel), proton ToF score cut (top right panel), pion dE/dx cut (bottom left panel), and distance to muon cut (bottom right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the line corresponds to the median.

3737 For each combination of selection cuts, I compare the true charged pion multiplicity
 3738 given by GENIE with the number of charged pion candidates I count with this method,
 3739 hereafter referred to as the reconstructed π^\pm multiplicity. The result of this comparison
 3740 is a matrix, with columns and rows indicating true and reconstructed charged pion
 3741 multiplicity, respectively. An example of one of these matrices, obtained for a certain
 3742 configuration of cuts, can be seen in Fig. 7.14. From these multiplicity matrices one can
 3743 extract performance metrics, like efficiency, purity, and significance.

3744 Given a multiplicity matrix \mathbf{M} , the efficiency for the i -th multiplicity value can be

CHAPTER 7. EVENT SELECTION IN ND-GAR

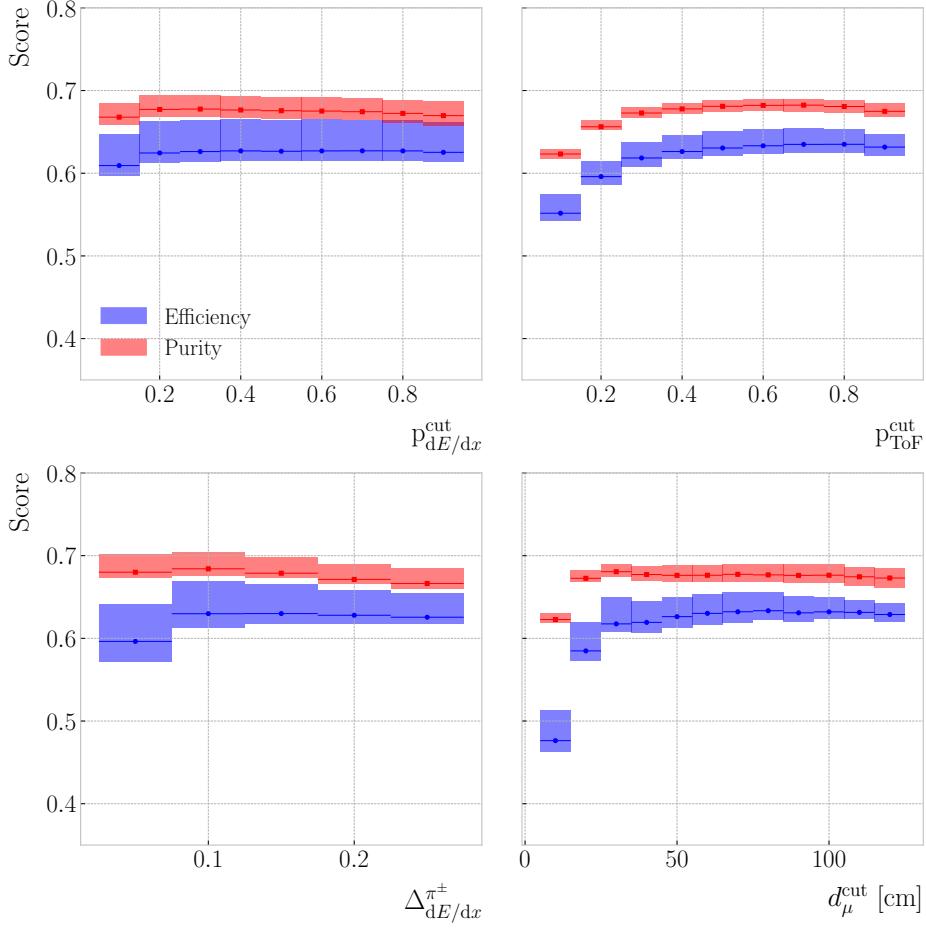


Figure 7.16: Efficiency (blue) and purity (red) for the ν_μ CC $1\pi^\pm$ selection as a function of the proton dE/dx score cut (top left panel), proton ToF score cut (top right panel), pion dE/dx cut (bottom left panel), and distance to muon cut (bottom right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the line corresponds to the median.

3745 computed as:

$$\text{Efficiency}|_i = \frac{M_{ii}}{\sum_j M_{ij}}, \quad (7.7)$$

3746 or in other words, dividing the corresponding diagonal entry by the sum of all the entries

3747 in the same column. On the other hand, the purity is given by:

$$\text{Purity}|_i = \frac{M_{ii}}{\sum_j M_{ji}}, \quad (7.8)$$

3748 which is just the ratio between the diagonal entry and the sum of the entries in the

7.3. CHARGED PION IDENTIFICATION

3749 corresponding row. Similarly, the significance is obtained by taking the square root of
 3750 the denominator in the previous expression:

$$\text{Significance}|_i = \frac{S}{\sqrt{S+B}}|_i = \frac{M_{ii}}{\sqrt{\sum_j M_{ji}}}. \quad (7.9)$$

3751 Figures 7.15 and 7.16 show the efficiency (blue) and the purity (red) for the ν_μ
 3752 CC $0\pi^\pm$ and $1\pi^\pm$ selections, respectively, as a function of the different cut values. In
 3753 the figures, each box represents the IQR of the conditional distribution for the fixed
 3754 value of the corresponding cut, and the horizontal lines correspond to the medians. The
 3755 first thing one notices is that the efficiency is always higher than the purity in the $0\pi^\pm$
 3756 selection, while the opposite is true for the $1\pi^\pm$ selection. Also, it is clear that the range
 3757 within these metrics fluctuate in the $0\pi^\pm$ selection is significantly higher than it is for
 3758 the $1\pi^\pm$ case. This shows that it is easier to assess that no charged pions are present in
 3759 the event than actually tagging them.

3760 For the ν_μ CC $0\pi^\pm$ selection, the performance metrics follow the expected tendency.
 3761 As the purity grows with a cut value, the efficiency decreases. Interestingly, this is not
 3762 the case for the $1\pi^\pm$ selection, where both efficiency and purity follow roughly the same
 3763 trends along the different cuts. This makes sense when one comprehends that this is not
 3764 a traditional cut-based selection, but more of a counting exercise. Some restrictive cut
 3765 configurations will not tag any particles as pions. On the contrary, loose cuts will render
 3766 every particle as a π^\pm . Therefore, when looking at a specific multiplicity, the relation
 3767 between the cut value and the performance metrics is not obvious. Thus, sometimes
 3768 efficiency and purity can both increase, as the cuts refine the definition of a reconstructed
 3769 pion.

3770 To have a working point for our studies, I chose the cut configuration that yields
 3771 the maximum significance for the ν_μ CC $1\pi^\pm$ selection. Of course, other cuts would be
 3772 more appropriate in certain scenarios. However, this provides us with a starting point
 3773 to understand the performance of the selection. A significance of 66 ± 7 for the $1\pi^\pm$
 3774 selection is achieved for the cut values $p_{dE/dx}^{\text{cut}} = 0.30$, $p_{\text{ToF}}^{\text{cut}} = 0.70$, $\Delta_{dE/dx}^{\pi^\pm} = 0.10$, and

CHAPTER 7. EVENT SELECTION IN ND-GAR

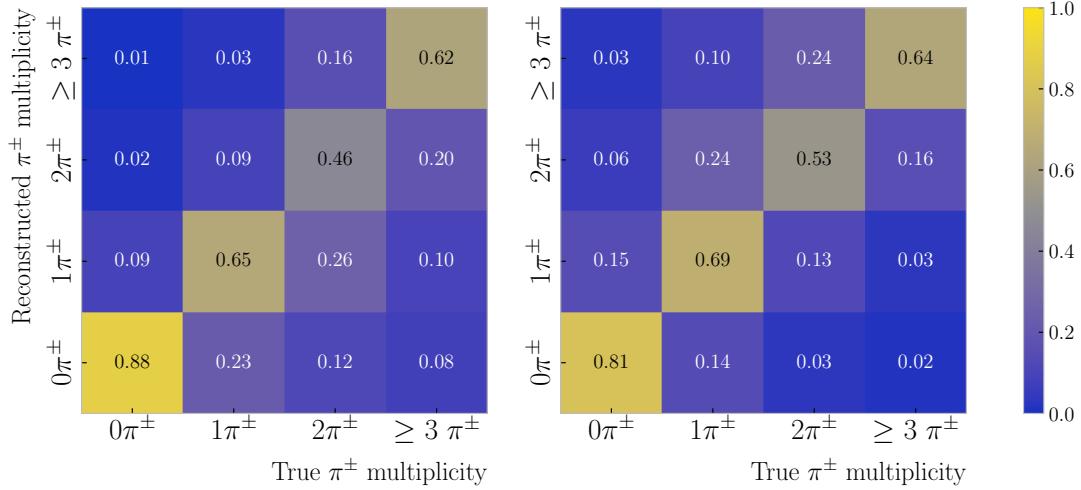


Figure 7.17: Distribution of events given their true and reconstructed π^\pm multiplicity, both column-wise (left panel) and row-wise (right panel) normalised, for the selection that maximises the significance of the ν_μ CC $1\pi^\pm$ selection. The column normalisation yields the efficiency in the diagonal entries, whereas the row normalisation reveals the purity.

3775 $d_\mu^{\text{cut}} = 110.0$ cm.

3776 Figure 7.17 shows the multiplicity matrices resulting from this optimised ν_μ CC $1\pi^\pm$
 3777 selection. Although both matrices are produced with the same selection cuts, one is
 3778 column normalised (left panel), whereas the other is row normalised (right panel). It
 3779 follows from the definitions in Eqs. (7.7) and ((7.8)) that the diagonal entries of these
 3780 matrices correspond to the efficiencies and the purities, respectively, for each of the
 3781 possible charged pion multiplicity selections.

3782 An additional check to make is understand how this configuration performs when
 3783 applied to the other selections, like ν_μ CC $0\pi^\pm$, and how it compares to the other
 3784 possible configurations. A comparison between the different pion multiplicity selections,
 3785 indicated with colours, in the purity versus efficiency space is shown in Fig. 7.18. For
 3786 each of the possible multiplicity choices, the performance obtained for the $1\pi^\pm$ optimised
 3787 selection is indicated by an outlined point. From this, one can see that the selected
 3788 configuration performs reasonably well, within the limits of what can be achieved in
 3789 each case, across the different multiplicities.

3790 At this point, one can study the charged pion selection performance as a function of

7.3. CHARGED PION IDENTIFICATION

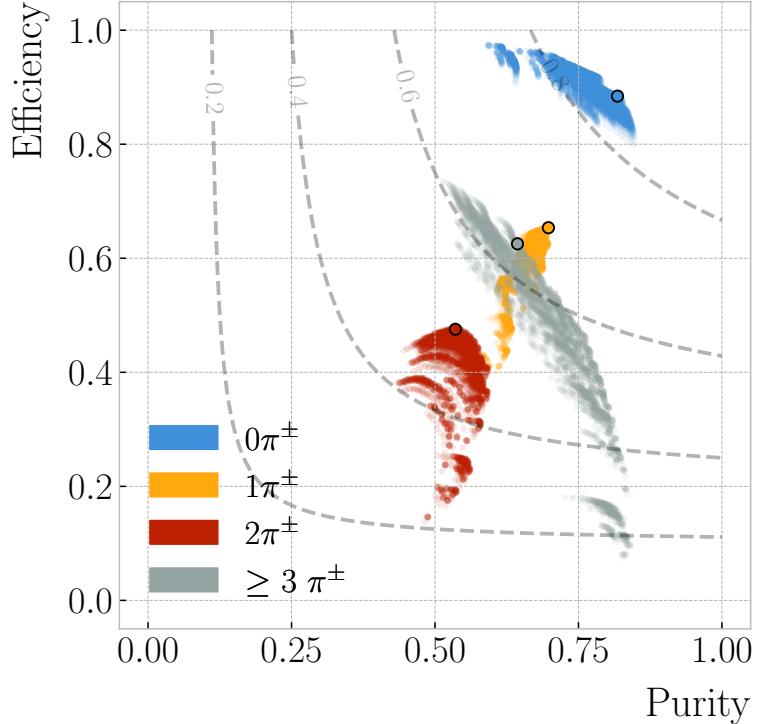


Figure 7.18: Purity versus efficiency achieved for the different cut configurations explored separated by the various ν_μ CC $N\pi^\pm$ selections. The outlined points indicate the state for each possible multiplicity when using the configuration that maximises the significance of the ν_μ CC $1\pi^\pm$ selection. The contours indicate the surfaces of equal F_1 -score.

3791 other quantities of interest. A natural variable to check is the hadronic invariant mass.

3792 It is defined as:

$$W = \sqrt{-Q^2 + 2m_n q_0 + m_n^2} \quad (7.10)$$

3793 where Q^2 is the momentum transfer from the neutrino to the primary muon, q_0 the
 3794 energy transfer, and m_n the mass of the nucleon. This quantity is related to the elasticity
 3795 of the neutrino interaction, and defines the transitions between the QEL, RES and DIS
 3796 regions. An interesting invariant mass range for DUNE is the one that extends between
 3797 the mass of the Δ resonance, even though it is typically extended down to $m_p + m_{\pi^\pm}$,
 3798 and 2.0 GeV. It is estimated that roughly 7 in every 10 events in our ND will take
 3799 place in this region. Although the RES production dominates at these W values, this
 3800 range also includes the transition to the DIS regime. Thus, it is often called the shallow

CHAPTER 7. EVENT SELECTION IN ND-GAR

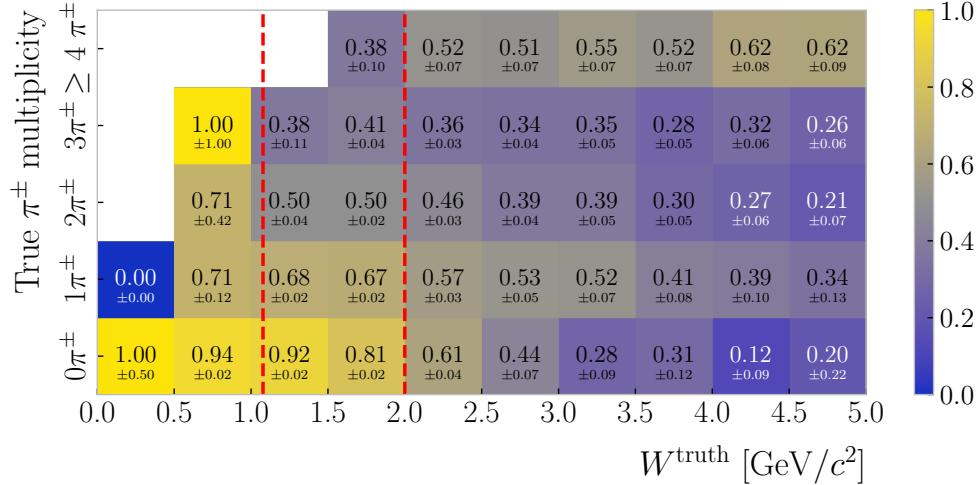


Figure 7.19: Efficiency of the various ν_μ CC $N\pi^\pm$ selections as a function of the true hadronic invariant mass. The dashed vertical lines correspond (from left to right) to the values $m_p + m_{\pi^\pm}$ and $2.0 \text{ GeV}/c^2$, which define the shallow inelastic scattering region.

3801 inelastic scattering (SIS) region.

3802 Within these boundaries, the resonant events produce either 1 or 2 charged pions,
 3803 whereas the multipion events are typically associated to non-resonant production.
 3804 Therefore, our ability of correctly select events with $\geq 2\pi^\pm$ in the SIS region will
 3805 impact our ability of constraining this poorly understood regime. Figure 7.19 shows the
 3806 efficiency of the various charged pion multiplicity selections in a number of hadronic
 3807 invariant mass bins. The two red dashed lines indicate the boundaries of the SIS region.
 3808 One can see that, although not as good as the single pion selection, the efficiency for the
 3809 multipion events is reasonable in the relevant invariant mass range. The total efficiency
 3810 for the ν_μ CC $\geq 2\pi^\pm$ selection in the SIS regime is estimated to be 0.65 ± 0.02 .

3811 7.3.1 ν_μ CC $1\pi^\pm$ selection

3812 By focusing on the $1\pi^\pm$ selection, one can study the kinematics of the selected pion.
 3813 This allows one to understand how well the charged pions are tagged. This is difficult
 3814 to do only using the multiplicity matrices, as with them one can only check that the
 3815 number of charged pions is the same as in the truth. Sometime, even if the estimated
 3816 pion multiplicity is correct, the identified particles may not be true pions.

7.3. CHARGED PION IDENTIFICATION

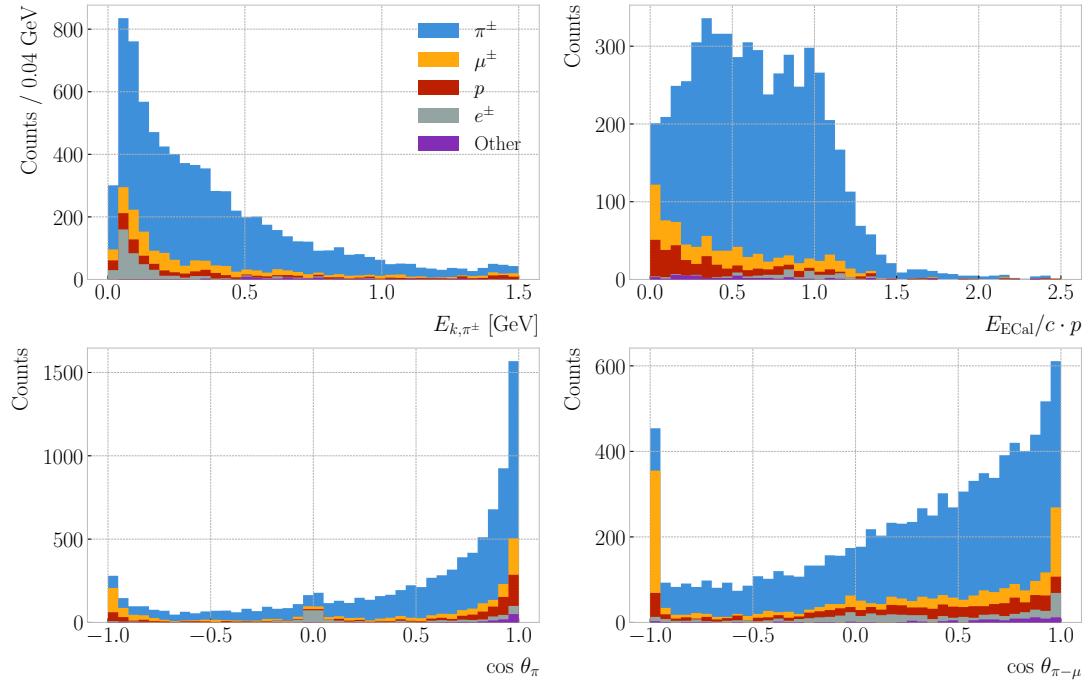


Figure 7.20: Reconstructed kinematic distributions for the pion candidate in the ν_μ CC $1\pi^\pm$ selection, broken down by true ID of the particle. From left to right, top to bottom, we have the kinetic energy given the pion hypothesis, the ratio between the energy deposited in the ECal and the momentum, the pion angle, and the angle between the muon and pion candidates.

3817 Figure 7.20 displays the distributions of various reconstructed kinematic variables
 3818 for the selected pion candidate. The different colours indicate the ID of the true particle
 3819 associated to the reconstructed pion.

3820 First, we have the kinetic energy distribution. For this set of reconstructed particles,
 3821 because they have been tagged as charged pions, the kinetic energy is computed using their
 3822 momentum assuming the pion hypothesis. One can see that most of the contaminants
 3823 sit in low energy range, up to around 0.2 GeV.

3824 The next distribution presents the ratio between the energy deposited in the ECal
 3825 associated to the particle over the momentum measured in the HPgTPC. This variable is
 3826 restricted to particles with at least one associated hit in the ECal. It is interesting to see
 3827 two peak structure in the true pion distribution. The first one presumably corresponds
 3828 to the pions punching-through the ECal, while the latter is probably due to the ones

CHAPTER 7. EVENT SELECTION IN ND-GAR

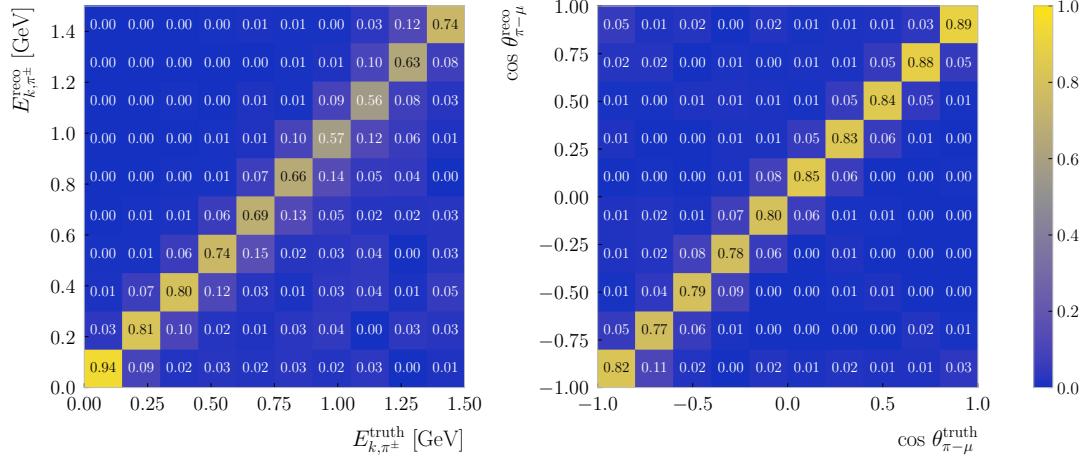


Figure 7.21: Distributions for the reconstructed versus truth generated pion kinetic energy (left panel) and cosine of the angle between the pion and muon (right panel). The reconstructed values correspond to the selected primary muon and pion candidates in the ν_μ CC $1\pi^\pm$ selection, whereas the truth values come from the true primary muon and pion in the events.

3829 stopping in it. On the other hand, the misidentified particles, other than the electrons,
 3830 tend to lower ratios. This is expected for protons, as this could not be higher than 0.5
 3831 for momenta ≤ 1 GeV/ c even if they stopped, but for the muons it may point to a
 3832 misreconstruction.

3833 The following distribution shows the angle of the pion candidates with respect to the
 3834 beam direction. Although most of them are aimed in the forward direction, it can be
 3835 noted that an important number of the misidentified muons seem to be backward-going.
 3836 This is likely a reconstruction artifact, produced by broken tracks that got assigned the
 3837 wrong propagation direction. Also, there is a sizeable number of true electrons with
 3838 directions perpendicular to the beam, probably delta electrons from the primary muon.

3839 Finally, I included the reconstructed pion-muon angular distribution. Even though
 3840 it shares some similarities with the previous distribution, as the primary muon typically
 3841 goes forward, the pion distribution is not as prominently forward-going in this case.
 3842 Also, it may be noted that approximately 25% of the muons misidentified as pions have
 3843 $\cos \theta_{\pi-\mu} \leq -0.95$. Therefore, putting an additional angular cut improves the purity of
 3844 the charged pion selection from 0.74 ± 0.01 to 0.77 ± 0.01 , while not loosing a substantial

7.4. NEUTRAL PION IDENTIFICATION

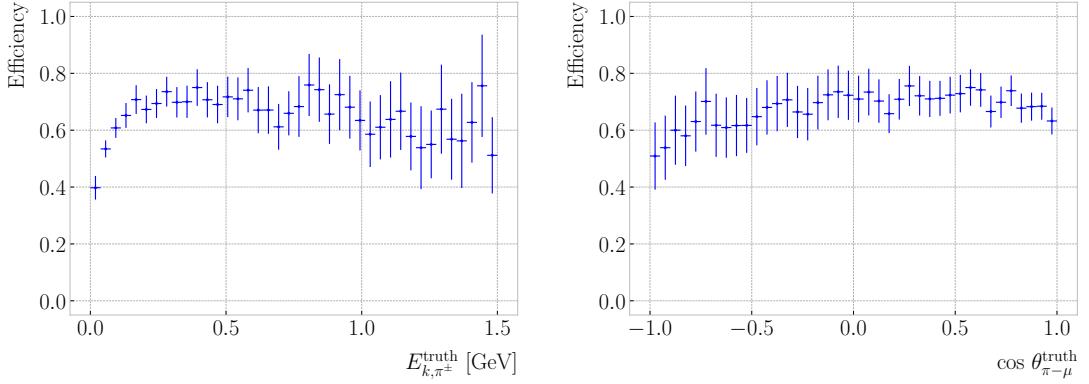


Figure 7.22: Efficiency of the ν_μ CC $1\pi^\pm$ selection as a function of the true pion kinetic energy (left panel) and pion-muon angle (right panel).

3845 amount of true pions.

3846 A comparison between the true and the reconstructed values of the pion kinetic
 3847 energy (left panel) and pion-muon angle (right panel) is shown in Fig. 7.21. The
 3848 distributions are column normalised, which allows to see the fraction of events in the
 3849 correct bins. For this, I selected the events where only one reconstructed pion and
 3850 one true pion were identified, as that is the only case where a pairing of the variables is
 3851 possible. It showcases the excellent agreement between the reconstruction and the truth
 3852 information.

3853 One can also study the performance of the pion selection as a function of the
 3854 truth pion kinematics. Fig. 7.22 shows the selection efficiency versus the true kinetic
 3855 energy (left panel) and the angle between the true primary pion and muon (right panel).
 3856 The efficiency is computed from the events with a single true and reconstructed pion,
 3857 comparing their number to the total of events with one true pion. Notice how the
 3858 efficiency, although it starts with relatively low values, plateaus around 0.70 quickly
 3859 after 0.20 GeV. In terms of the pion-muon angle, the efficiency looks relatively flat, only
 3860 dropping slightly towards the back-to-back case.

CHAPTER 7. EVENT SELECTION IN ND-GAR

3861 7.4 Neutral pion identification

3862 The ν_μ CC selection can also be used as a stepping stone for the identification of
3863 neutral pions. Although these particles do not leave any traces in the HPgTPC, using a
3864 combination of the different detectors within ND-GAr. Being able to tag the neutral
3865 pions is a valuable asset for the estimation of the reconstructed neutrino energy, as both
3866 their kinetic and mass components can then be added in the calculation.

3867 In the case that both photons from the π^0 decay do not undergo pair production
3868 of a e^+e^- pair, they will reach the ECal where they will produce an electromagnetic
3869 shower. This activity inside the ECal will not be associated to any charged particle track
3870 inside the HPgTPC, unless there is a reconstruction failure. Thus, having a neutrino
3871 interaction vertex candidate from the ν_μ CC selection, one can reconstruct the mass of
3872 the π^0 using the energy and position of the photons. I already used this same technique
3873 in section 6.6 for a single π^0 sample. However, here I apply it to neutrino interaction
3874 events, and the vertex position is not cheated but selected from the reconstruction
3875 products.

3876 The idea is to look for all the ECal clusters that were not associated to tracks in
3877 each event. Then, if two or more were identified, compute the invariant mass for all
3878 possible combinations. At this point, I select the pair whose invariant mass is closest to
3879 m_{π^0} , remove the pairs containing any of the two selected clusters from the collection,
3880 and iterate until no more pairs can be formed.

3881 I repeat this procedure for the events with 0, 1, 2 and 3 or more true neutral pions.
3882 For each of them, I extract the invariant mass of the first three cluster pair candidates
3883 (in order of proximity to m_{π^0}), in case they can be formed. If the number of the cluster
3884 pair is lower than the true neutral pion multiplicity of the event, that entry will be
3885 counted as signal. The additional candidates for an event of a given multiplicity are
3886 considered background. The resulting distribution is shown in Fig. 7.23 (black data
3887 points).

7.4. NEUTRAL PION IDENTIFICATION

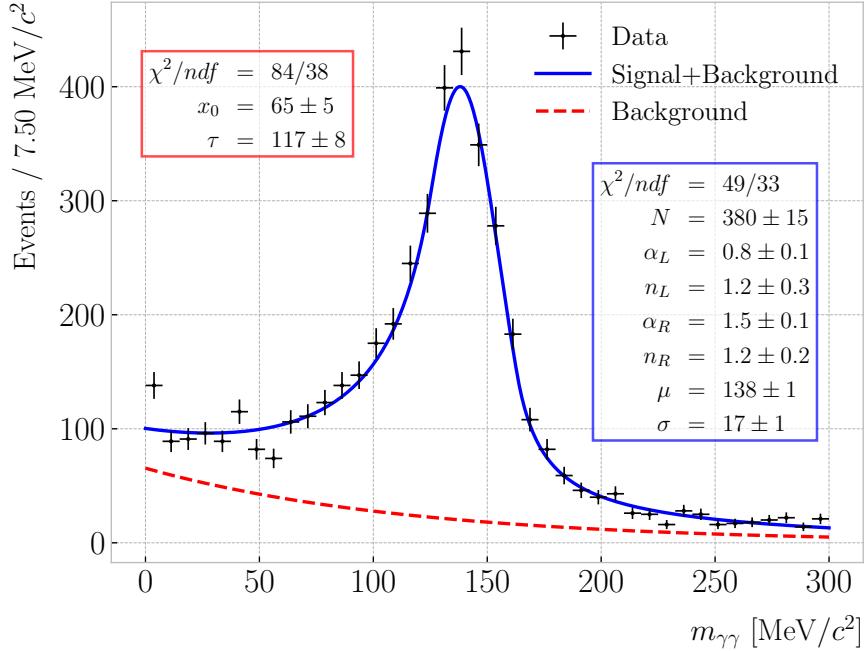


Figure 7.23: Invariant mass distribution obtained for the unassociated ECal cluster pair in the event with the value closest to the true π^0 mass. The best fits for signal plus background (solid blue line) and background only (dashed red line) are also shown.

3888

I fit the signal distribution to a double-sided Crystal Ball function:

$$f_s(x; \alpha_{L,R}, n_{L,R}, \mu, \sigma, N) = N \begin{cases} A_L \left(B_L - \frac{x-\mu}{\sigma} \right)^{-n_L}, & \text{for } \frac{x-\mu}{\sigma} < -\alpha_L, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & \text{for } -\alpha_L \leq \frac{x-\mu}{\sigma} \leq \alpha_R, \\ A_R \left(B_R + \frac{x-\mu}{\sigma} \right)^{-n_R}, & \text{for } \frac{x-\mu}{\sigma} > \alpha_R, \end{cases} \quad (7.11)$$

3889

where $A_{L,R}$ and $B_{L,R}$ are given by:

$$A_{L,R} = \left(\frac{n_{L,R}}{|\alpha_{L,R}|} \right)^{n_{L,R}} e^{-\frac{|\alpha_{L,R}|^2}{2}}, \quad (7.12)$$

$$B_{L,R} = \frac{n_{L,R}}{|\alpha_{L,R}|} - |\alpha_{L,R}|.$$

3890 The tails of this distribution accommodate the asymmetric shape of the misreconstruction
 3891 effects. The values obtained for the best fit parameters are indicated in Fig. 7.23 (blue
 3892 box).

CHAPTER 7. EVENT SELECTION IN ND-GAR

3893 The background is characterised by an exponential function of the form:

$$f_b(x; x_0, \tau) = x_0 e^{-x/\tau}. \quad (7.13)$$

3894 Similarly, the best fit values can be seen in Fig. 7.23 (red box).

3895 Figure 7.23 also shows the results of the fits for the signal plus background (blue line)
3896 and the background only (dashed red line) cases. Using these, I estimate the tagging
3897 efficiency of this method to be 0.90 ± 0.01 with a purity of 0.85 ± 0.01 , when selecting
3898 the candidates with an invariant mass in the range $54.0 - 288.0 \text{ GeV}/c^2$.

3899 This is a robust method to identify the photon pair from the π^0 decay. However,
3900 this approach is not enough to efficiently identify all the events containing neutral pions
3901 in the sample. A quick calculation reveals that only 20% of the ν_μ CC $1\pi^0$ events can
3902 be correctly identified with it.

3903 This approach can be complemented with the identification of the secondary vertices
3904 from the e^+e^- conversions. This will make it possible to cover the cases when either
3905 one or both photons convert in the HPgTPC. In those cases, one can try pairing the
3906 e^+e^- with unassociated activities in the ECal, or matching pairs of secondary vertices.
3907 However, this will require further work on the reconstruction, and thus falls out of the
3908 scope of this analysis.

3909 7.5 Neutrino energy reconstruction

3910 In a neutrino-nucleus CC interaction, where alongside the charged lepton N nucleons
3911 where knocked out and M mesons produced, the reconstructed neutrino energy can be
3912 computed as:

$$E_{\text{rec}} = S_n + E_\ell + \sum_{i=0}^N E_{k,n_i} + \sum_{j=0}^M E_{m_j}, \quad (7.14)$$

3913 where S_n is the average single-nucleon separation energy, E_ℓ the energy of the primary
3914 lepton, E_{k,n_i} is the kinetic energy of the i -th knocked-out nucleon and E_{m_j} the total
3915 energy of the j -th produced meson.

7.5. NEUTRINO ENERGY RECONSTRUCTION

3916 This represents the ideal scenario, where all the kinetic energy of the nucleons is
 3917 visible in the detector and one can identify all mesons produced in the interaction. In a
 3918 real experiment, some of these energy components will not be , and this needs to be
 3919 accounted for in any estimation of the reconstructed energy.

3920 For instance, in ND-GAr neutrons are complicated to account for, as they do not
 3921 produce tracks in the TPC. They may be identified either from scatterings off Ar nuclei
 3922 in the HPgTPC, or performing a ToF measurement in the ECal. However, these methods
 3923 are not fully mature in the current reconstruction, and their development is beyond the
 3924 scope of this study. So, in the following, I will completely ignore the contribution of
 3925 neutrons.

3926 Also, with a real detector we can not expect to tag all the charged pions irrespective
 3927 of their energy. This is why one has to introduce detection thresholds in the energy
 3928 estimation. Thus, in the reconstructed energy calculation I will add only the kinetic
 3929 energy for the charged pions below the threshold, and the total energy for the pions
 3930 above the threshold.

3931 Likewise, the identification of all neutral pions in the sample is challenging. As
 3932 discussed in the previous section, with our ECal we are able to identify the photons
 3933 from the π^0 decays, but that selection still needs to be completed with other methods.
 3934 Therefore, for this first study I do not take into account the energy contribution of the
 3935 neutral pions.

3936 With all this in mind, using the truth information from the events I compute the
 3937 reconstructed neutrino energy as:

$$E_{\text{rec}}^{\text{truth}} = S_n + E_\ell + \sum_{i=0}^{N_p} E_{k,p_i} + \sum_{j=0}^{M_{\pi^\pm}^<} E_{k,\pi_j^\pm} + \sum_{k=0}^{M_{\pi^\pm}^>} E_{\pi_k^\pm}, \quad (7.15)$$

3938 where N_p is the number of protons, and $M_{\pi^\pm}^<$ and $M_{\pi^\pm}^>$ the number of charged pions
 3939 below and above the threshold, respectively. As before, I assume a kinetic energy
 3940 threshold of 20 MeV for the charged pions.

3941 At the reconstruction level, I use the energy of the primary muon candidate, computed

CHAPTER 7. EVENT SELECTION IN ND-GAR

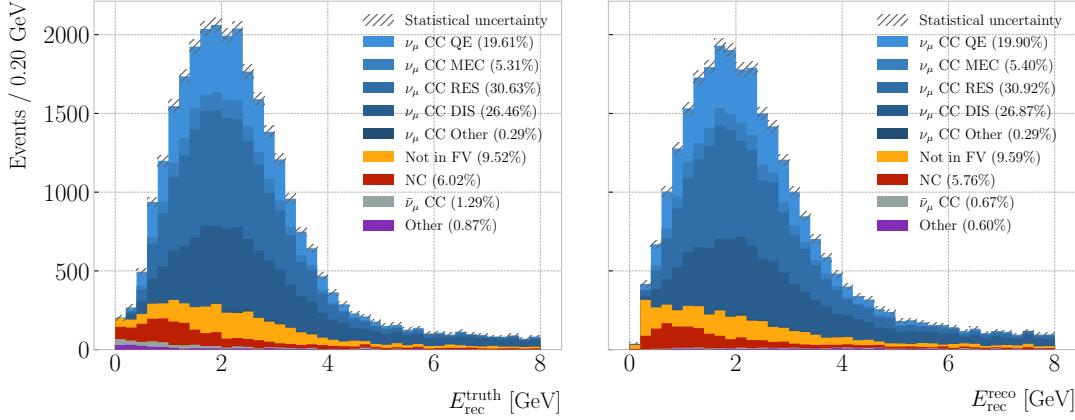


Figure 7.24: Reconstructed neutrino energy spectra for the ν_μ CC selection obtained using the truth (left panel) and reconstructed (right panel) information. The selected events are broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray.

from its momentum, as the starting point for the neutrino energy calculation. Then, I add the total energy contributions from the identified charged pions, again using their momenta. After that, I try to identify the protons by looking at the two proton scores. If any of them are above threshold (here the thresholds used are the same as for the pion identification), the kinetic energy of the particle is added to the total. Finally, I check if any of the remaining particles are fully contained within the FV. I add their kinetic contributions using the total energy they deposited in the HPgTPC.

Figure 7.24 shows the resulting distributions of reconstructed neutrino energy obtained from the truth (left panel) and reconstructed (right panel) particle collections. The overall shape of the distributions is similar, with the reconstructed one having a slightly larger high energy tail. Note also that the background events from outside the FV tend to have a smaller energy in the reconstructed case. This is likely due to a misreconstruction of the primary muon, which clearly does not affect the other computation.

I also compared the reconstructed energies to the true energy of the neutrino. Figure 7.25 displays the ratio of the energy residuals to the true energy for the truth (left panel) and reconstructed (right panel) cases. As expected, using the true particles one never overestimates the neutrino energy. Also, using the reconstructed objects one is more

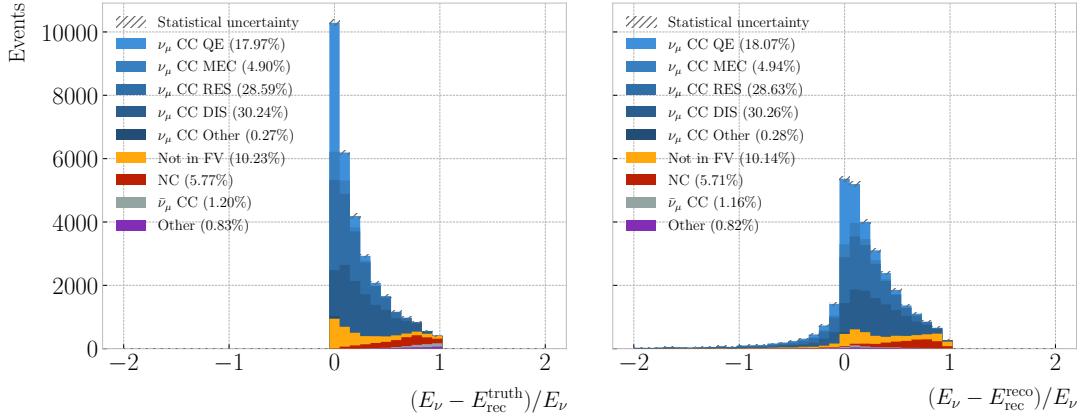


Figure 7.25: Neutrino energy residuals distributions for the ν_μ CC selection obtained using the truth (left panel) and reconstructed (right panel) information. The selected events are broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray.

3960 prone to underestimate the neutrino energy, due to deficiencies in the reconstruction.

3961 7.6 Systematic uncertainties

3962 Although the implementation and study of the systematic uncertainties relevant for
 3963 ND-GAr is out of the scope of this preliminary analysis, in this section I give an extended
 3964 overview of the topic. can be classified in three categories: neutrino flux uncertainties,
 3965 neutrino-nucleus interaction model uncertainties, and detector response uncertainties.

3966 7.6.1 Flux uncertainties

3967 The neutrino flux prediction is affected by systematic uncertainties arising from two
 3968 sources: the uncertainties in the production of hadrons in the target and the uncertainties
 3969 in the design parameters of the beamline itself. These fluxes and their uncertainties are
 3970 generated with the G4LBNF simulation [73], a Geant4 implementation of the LBNF
 3971 beamline, and the Package to Predict the FluX (PPFX) framework, originally developed
 3972 for MINERvA [174].

3973 The hadron production uncertainties are associated to the kinematic distributions
 3974 of the hadrons produced when the protons interact with the carbon target, as well

CHAPTER 7. EVENT SELECTION IN ND-GAR

3975 as the possible interactions of the hadrons with the beamline materials. The PPFX
3976 package estimates these uncertainties by performing a number of random throws of the
3977 production model parameters [175]. This way, different predictions of the LBNF flux
3978 are generated, which can be compared to the nominal prediction to build a matrix of
3979 the covariances between neutrino energies, flavours and running modes (either FHC or
3980 RHC). The resulting hadron production uncertainties are described by the eigenvectors
3981 associated to the largest eigenvalues in this matrix, obtained performing a PCA analysis.

3982 The other set of uncertainties affecting the neutrino flux prediction come from the
3983 limited precision with which we know the parameters of the different components in the
3984 beamline. These include the specifications of the target, the dimensions of the decay
3985 pipe, and the current and alignment of the magnetic horns. The effects on the flux
3986 predictions of these uncertainties are estimated using the G4LBNF simulation. For each
3987 of the parameters, the simulation runs with said parameter shifted by $\pm 1\sigma$ from the
3988 nominal value, and the resulting flux prediction is compared to the nominal one.

3989 7.6.2 Cross section uncertainties

3990 As discussed previously in section 2.6, the neutrino-nucleus interaction model is of
3991 great importance for neutrino experiments, as it maps the true neutrino energy to the
3992 kinematics of the final state particles. The uncertainties on the cross section model are
3993 implemented in three ways: varying the parameters used in the GENIE simulation, using
3994 weights that parametrise cross section effects not accounted for in GENIE, and comparing
3995 the GENIE predictions to other interaction models.

3996 Within the DUNE TDR LBL analysis, the default interaction model was that
3997 implemented in GENIE v2_12_10 [108]. A summary of the cross section systematic
3998 parameters present in GENIE used in that analysis is presented in Tab. 7.3. The
3999 additional systematic parameters used in the analysis are described in Tab. 7.4.

4000 In this default GENIE configuration, the initial state of the nucleons is described
4001 by the Bodek-Ritchie global Fermi gas model [176]. The model is known give a poor
4002 agreement whe compared to neutrino-nucleon data [177]. Because of the limitations of

7.6. SYSTEMATIC UNCERTAINTIES

Table 7.3: Neutrino interaction systematic parameters implemented in GENIE used in the DUNE TDR LBL analysis. Events with low W that are not QE are mainly RES, whereas DIS events dominate at high W . The initials BY refer to the Bodek-Yang model. Table adapted from Ref. [108].

Systematic	1σ value
Quasielastic	
Axial mass for CCQE	$+0.25_{-0.15}$ GeV
CCQE vector form factor shape	N/A
Fermi surface momentum for Pauli blocking	$\pm 30\%$
Low W	
Axial mass for CC resonance	± 0.05 GeV
Vector mass for CC resonance	$\pm 10\%$
θ_π distribution for Δ decay	N/A
High W (BY model)	
A_{HT}	$\pm 25\%$
B_{HT}	$\pm 25\%$
C_{v1u}	$\pm 30\%$
C_{v2u}	$\pm 40\%$
Other neutral current	
Axial mass for NC resonance	$\pm 10\%$
Vector mass for NC resonance	$\pm 5\%$
Intra-nuclear	
Nucleon charge exchange	$\pm 50\%$
Nucleon elastic reaction	$\pm 30\%$
Nucleon inelastic reaction	$\pm 40\%$
Nucleon absorption	$\pm 20\%$
Nucleon π -production	$\pm 20\%$
π charge exchange	$\pm 50\%$
π elastic reaction	$\pm 10\%$
π inelastic reaction	$\pm 40\%$
π absorption	$\pm 20\%$
π π -production	$\pm 20\%$

CHAPTER 7. EVENT SELECTION IN ND-GAR

Table 7.4: Neutrino interaction systematic parameters used in the DUNE TDR LBL analysis not present in **GENIE**. I have omitted the parameters only relevant for the FD. Table adapted from Ref. [108].

Systematic	Mode	Description
BeRPA	$1p1h/\text{QE}$	Nuclear model suppression
ArC2p2h	$2p2h \text{ Ar/C}$	Electron scattering SRC pairs
E_{2p2h}	$2p2h$	$2p2h$ energy dependence
CC non-resonant	CC DIS 1π	$\nu + n/p \rightarrow \ell + 1\pi$
Other non-resonant	DIS $N\pi$	$1 < W < 5 \text{ GeV}/c^2$
NC normalisation	NC	$\pm 20\%$ to all NC events at the ND

4003 the model, the current versions of **GENIE** use the local Fermi gas approach, which takes
 4004 into account the correlation between the momentum of the nucleons and their location
 4005 within the nucleus.

4006 For the CCQE events, the dominant model uncertainties arise from the axial form
 4007 factor of the nucleon, for which a dipole parametrisation is used, and the nuclear
 4008 correlation effects computed using the random phase approximation (RPA). In the
 4009 analysis, a parametrisation of the Valencia RPA effect [170] is used. This consists of
 4010 a third-order Bernstein polynomial up to $Q^2 = 1.2 \text{ GeV}^2$ followed by an exponential
 4011 decay (BeRPA), originally proposed by the T2K collaboration [178].

4012 The $2p2h$ interactions are included using the Valencia model [170], with an additional
 4013 correction following the observation of an underprediction of these events in MINERvA
 4014 [179]. Additional uncertainties for the energy dependence of the missing strength were
 4015 added. Also, the uncertainties in the scaling from carbon to argon are included, based
 4016 on measurements of electron scattering off short-range correlated (SRC) nucleon pairs
 4017 on multiple targets [180].

4018 In this version of **GENIE**, the Rein-Sehgal model describes the single pion resonant
 4019 production events [181]. It includes 16 different resonances, with no interference between
 4020 them. Two parameters account for the uncertainties on the axial and vector masses of
 4021 the resonances. In subsequent **GENIE** tunes, like the one used in the studies presented in
 4022 this Chapter, the Berger-Sehgal model is used [171]. This is an improved version of the

7.6. SYSTEMATIC UNCERTAINTIES

4023 Rein-Sehgal model, which includes the lepton mass effects in the calculations.

4024 The Bodek-Yang parametrisation is used to describe the DIS events [173]. The
4025 parameters A_{HT} and B_{HT} account for higher twist effects in the scaling variable, while
4026 C_{v1u} and C_{v2u} control the form of the valence quark K factors. For the analysis, the
4027 uncertainties on the values of these parameters are taken into consideration. Also, due to
4028 the difficulties of GENIE at describing the transition region between RES and DIS events,
4029 a set of systematic parameters affecting the different non-resonant pion production
4030 channels were developed, following the example of NOvA [182]. There are independent
4031 parameters for the interactions on protons and neutrons, except for the CC DIS 1π case
4032 where they are merged. All start with an uncertainty of 50% for $W \leq 3 \text{ GeV}/c^2$, which
4033 linearly decreases until reaching a 5% at $W = 5 \text{ GeV}/c^2$.

4034 For the TDR analysis, an additional 20% normalisation uncertainty was added to all
4035 NC events in the ND. It was implemented to understand if the NC events passing the
4036 selection cuts affected the results of the analysis [108].

4037 Finally, the effective intranuclear transport model (often denoted as hA) is a part
4038 of GENIE, implemented in the INTRANUKE module. GENIE features a large number of
4039 parameters for the uncertainties on the intranuclear cascade model, which are summarised
4040 in the last portion of Tab. 7.3. In following GENIE releases, updated versions of the
4041 INTRANUKE model are used.

4042 Although part of this cross section systematic treatment is outdated, as the tunes
4043 currently used feature different models, it gives a good idea of what systematic effects
4044 are relevant for the different measurements we may want to perform in the future. At
4045 the moment, a significant effort is channeled to the creation of new tunes specifically
4046 tailored for DUNE, including the development of parametrisations particularly relevant
4047 for ND-GAr.

4048 7.6.3 Detector uncertainties

Conclusion and outlook

4051 *Our plans miscarry because they have no aim. When a man does not know
4052 what harbour he is making for, no wind is the right wind.*

4053 – Lucio Anneo Seneca, *Epistulae morales ad Lucilium*

A

4054

An appendix

4055

Bibliography

- 4057 [1] P.F. de Salas, D.V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena,
 4058 C.A. Ternes et al., *2020 global reassessment of the neutrino oscillation picture*,
 4059 *JHEP* **02** (2021) 071 [[2006.11237](#)]. 33, 57, 65
- 4060 [2] W. Pauli, *Dear radioactive ladies and gentlemen*, *Phys. Today* **31N9** (1978) 27. 39
- 4061 [3] F. Reines and C.L. Cowan, *Detection of the free neutrino*, *Phys. Rev.* **92** (1953)
 4062 830. 39
- 4063 [4] S.L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** 579. 39
- 4064 [5] S. Weinberg, *A model of leptons*, *Physical Review Letters* **19** 1264. 39
- 4065 [6] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968)
 4066 367. 39
- 4067 [7] A. Pich, *The Standard Model of Electroweak Interactions*, in *2010 European
 4068 School of High Energy Physics*, pp. 1–50, 1, 2012 [[1201.0537](#)]. 40
- 4069 [8] ALEPH, DELPHI, L3, OPAL, SLD, LEP ELECTROWEAK WORKING GROUP,
 4070 SLD ELECTROWEAK GROUP, SLD HEAVY FLAVOUR GROUP collaboration,
 4071 *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006)
 4072 257 [[hep-ex/0509008](#)]. 43
- 4073 [9] R. Davis, D.S. Harmer and K.C. Hoffman, *Search for neutrinos from the sun*,
 4074 *Physical Review Letters* **20** 1205. 44

BIBLIOGRAPHY

- 4075 [10] J.N. Bahcall, A.M. Serenelli and S. Basu, *New solar opacities, abundances,*
4076 *helioseismology, and neutrino fluxes, Astrophys. J. Lett.* **621** (2005) L85
4077 [[astro-ph/0412440](#)]. 44, 128, 129, 130
- 4078 [11] J.N. Bahcall, N.A. Bahcall and G. Shaviv, *Present Status of the Theoretical*
4079 *Predictions for the ^{37}Cl Solar-Neutrino Experiment, . 44*
- 4080 [12] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,
4081 *Measurement of the solar electron neutrino flux with the Homestake chlorine*
4082 *detector, Astrophys. J.* **496** (1998) 505. 44
- 4083 [13] SAGE collaboration, *Measurement of the solar neutrino capture rate with gallium*
4084 *metal. III: Results for the 2002–2007 data-taking period, Phys. Rev. C* **80** (2009)
4085 015807 [[0901.2200](#)]. 44, 56
- 4086 [14] F. Kaether, W. Hampel, G. Heusser, J. Kiko and T. Kirsten, *Reanalysis of the*
4087 *GALLEX solar neutrino flux and source experiments, Phys. Lett. B* **685** (2010) 47
4088 [[1001.2731](#)]. 44, 56
- 4089 [15] Q.R. Ahmad, R.C. Allen, T.C. Andersen, J.D. Anglin, G. Bühler, J.C. Barton
4090 et al., *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by*
4091 *^8B Solar Neutrinos at the Sudbury Neutrino Observatory, . 45*
- 4092 [16] Q.R. Ahmad, R.C. Allen, T.C. Andersen, J. D. Anglin, J.C. Barton, E.W. Beier
4093 et al., *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current*
4094 *Interactions in the Sudbury Neutrino Observatory, . 45*
- 4095 [17] T.K. Gaisser and M. Honda, *Flux of atmospheric neutrinos, . 45*
- 4096 [18] K. Hirata, T. Kajita, M. Koshiba, M. Nakahata, S. Ohara, Y. Oyama et al.,
4097 *Experimental study of the atmospheric neutrino flux, . 46*
- 4098 [19] D. Casper, R. Becker-Szendy, C.B. Bratton, D.R. Cady, R. Claus, S.T. Dye et al.,
4099 *Measurement of atmospheric neutrino composition with the imb-3 detector, . 46*

BIBLIOGRAPHY

- 4100 [20] M. Ambrosio, R. Antolini, C. Aramo, G. Auriemma, A. Baldini, G. C. Barbarino
 4101 et al., *Measurement of the atmospheric neutrino-induced upgoing muon flux using*
 4102 *macro*, . 46
- 4103 [21] W. Allison, G. Alner, D. Ayres, W. Barrett, C. Bode, P. Border et al.,
 4104 *Measurement of the atmospheric neutrino flavour composition in soudan 2*, . 46
- 4105 [22] SUPER-KAMIOKANDE collaboration, *Evidence for oscillation of atmospheric*
 4106 *neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [[hep-ex/9807003](#)]. 46
- 4107 [23] P. Minkowski, $\mu \rightarrow e\gamma$ at a Rate of One Out of 10^9 Muon Decays?, *Phys. Lett. B*
 4108 **67** (1977) 421. 49
- 4109 [24] M. Gell-Mann, P. Ramond and R. Slansky, *Complex Spinors and Unified Theories*,
 4110 *Conf. Proc. C* **790927** (1979) 315 [[1306.4669](#)]. 49
- 4111 [25] T. Yanagida, *Horizontal gauge symmetry and masses of neutrinos*, *Conf. Proc. C*
 4112 **7902131** (1979) 95. 49
- 4113 [26] R.N. Mohapatra and G. Senjanovic, *Neutrino Mass and Spontaneous Parity*
 4114 *Nonconservation*, *Phys. Rev. Lett.* **44** (1980) 912. 49
- 4115 [27] J. Schechter and J.W.F. Valle, *Neutrino Masses in $SU(2) \times U(1)$ Theories*, *Phys.*
 4116 *Rev. D* **22** (1980) 2227. 49
- 4117 [28] B. Pontecorvo, *Mesonium and anti-mesonium*, *Sov. Phys. JETP* **6** (1957) 429. 49
- 4118 [29] M. Gell-Mann and A. Pais, *Behavior of neutral particles under charge conjugation*,
 4119 . 49
- 4120 [30] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of*
 4121 *Leptonic Charge*, *Zh. Eksp. Teor. Fiz.* **53** (1967) 1717. 50
- 4122 [31] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, *Zh.*
 4123 *Eksp. Teor. Fiz.* **34** (1957) 247. 51

BIBLIOGRAPHY

- 4124 [32] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary*
4125 *particles, Prog. Theor. Phys.* **28** (1962) 870. 51
- 4126 [33] PARTICLE DATA GROUP collaboration, *Review of particle physics, Phys. Rev. D*
4127 **110** (2024) 030001. 53
- 4128 [34] L. Wolfenstein, *Neutrino Oscillations in Matter, Phys. Rev. D* **17** (1978) 2369. 54
- 4129 [35] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,
4130 *Measurement of the solar electron neutrino flux with the Homestake chlorine*
4131 *detector, Astrophys. J.* **496** (1998) 505. 56
- 4132 [36] G. Bellini et al., *Precision measurement of the ^{7}Be solar neutrino interaction rate*
4133 *in Borexino, Phys. Rev. Lett.* **107** (2011) 141302 [[1104.1816](#)]. 56
- 4134 [37] SUPER-KAMIOKANDE collaboration, *Solar neutrino measurements in*
4135 *super-Kamiokande-I, Phys. Rev. D* **73** (2006) 112001 [[hep-ex/0508053](#)]. 56
- 4136 [38] SNO collaboration, *Combined Analysis of all Three Phases of Solar Neutrino*
4137 *Data from the Sudbury Neutrino Observatory, Phys. Rev. C* **88** (2013) 025501
4138 [[1109.0763](#)]. 56
- 4139 [39] SUPER-KAMIOKANDE collaboration, *Atmospheric neutrino oscillation analysis*
4140 *with external constraints in Super-Kamiokande I-IV, Phys. Rev. D* **97** (2018)
4141 072001 [[1710.09126](#)]. 56, 159
- 4142 [40] ICECUBE collaboration, *Measurement of Atmospheric Neutrino Oscillations at*
4143 *6–56 GeV with IceCube DeepCore, Phys. Rev. Lett.* **120** (2018) 071801
4144 [[1707.07081](#)]. 56
- 4145 [41] KAMLAND collaboration, *Reactor On-Off Antineutrino Measurement with*
4146 *KamLAND, Phys. Rev. D* **88** (2013) 033001 [[1303.4667](#)]. 57
- 4147 [42] RENO collaboration, *Measurement of Reactor Antineutrino Oscillation Amplitude*
4148 *and Frequency at RENO, Phys. Rev. Lett.* **121** (2018) 201801 [[1806.00248](#)]. 57

BIBLIOGRAPHY

- 4149 [43] DAYA BAY collaboration, *Measurement of the Electron Antineutrino Oscillation*
 4150 *with 1958 Days of Operation at Daya Bay*, *Phys. Rev. Lett.* **121** (2018) 241805
 4151 [1809.02261]. 57
- 4152 [44] K.J. Kelly, P.A.N. Machado, S.J. Parke, Y.F. Perez-Gonzalez and R.Z. Funchal,
 4153 *Neutrino mass ordering in light of recent data*, *Phys. Rev. D* **103** (2021) 013004.
 4154 57
- 4155 [45] P. Dunne, “Latest Neutrino Oscillation Results from T2K.” Jul, 2020. 57
- 4156 [46] MINOS collaboration, *Combined analysis of ν_μ disappearance and $\nu_\mu \rightarrow \nu_e$*
 4157 *appearance in MINOS using accelerator and atmospheric neutrinos*, *Phys. Rev. Lett.* **112** (2014) 191801 [1403.0867]. 57
- 4159 [47] OPERA collaboration, *Final Results of the OPERA Experiment on ν_τ*
 4160 *Appearance in the CNGS Neutrino Beam*, *Phys. Rev. Lett.* **120** (2018) 211801
 4161 [1804.04912]. 57
- 4162 [48] K2K collaboration, *Measurement of Neutrino Oscillation by the K2K Experiment*,
 4163 *Phys. Rev. D* **74** (2006) 072003 [hep-ex/0606032]. 57
- 4164 [49] DUNE collaboration, *Long-baseline neutrino oscillation physics potential of the*
 4165 *DUNE experiment*, *Eur. Phys. J. C* **80** (2020) 978 [2006.16043]. 57, 236
- 4166 [50] HYPER-KAMIOKANDE collaboration, *Physics potential of Hyper-Kamiokande for*
 4167 *neutrino oscillation measurements*, *PoS NuFact2019* (2019) 040. 57
- 4168 [51] SUPERNEMO collaboration, *Probing New Physics Models of Neutrinoless Double*
 4169 *Beta Decay with SuperNEMO*, *Eur. Phys. J. C* **70** (2010) 927 [1005.1241]. 58
- 4170 [52] SNO+ collaboration, *Current Status and Future Prospects of the SNO+*
 4171 *Experiment*, *Adv. High Energy Phys.* **2016** (2016) 6194250 [1508.05759]. 58
- 4172 [53] NEXT collaboration, *Sensitivity of a tonne-scale NEXT detector for neutrinoless*
 4173 *double beta decay searches*, *JHEP* **2021** (2021) 164 [2005.06467]. 58

BIBLIOGRAPHY

- 4174 [54] P. Coloma and P. Huber, *Impact of nuclear effects on the extraction of neutrino*
4175 *oscillation parameters*, *Phys. Rev. Lett.* **111** (2013) 221802 [[1307.1243](#)]. 59
- 4176 [55] P. Coloma, P. Huber, C.-M. Jen and C. Mariani, *Neutrino-nucleus interaction*
4177 *models and their impact on oscillation analyses*, *Phys. Rev. D* **89** (2014) 073015
4178 [[1311.4506](#)]. 59
- 4179 [56] U. Mosel, *Neutrino Interactions with Nucleons and Nuclei: Importance for*
4180 *Long-Baseline Experiments*, *Ann. Rev. Nucl. Part. Sci.* **66** (2016) 171
4181 [[1602.00696](#)]. 59
- 4182 [57] J.A. Formaggio and G.P. Zeller, *From ev to eev: Neutrino cross sections across*
4183 *energy scales*, *Rev. Mod. Phys.* **84** (2012) 1307 [[1305.7513](#)]. 60
- 4184 [58] L. Bathe-Peters, S. Gardiner and R. Guenette, *Comparing generator predictions*
4185 *of transverse kinematic imbalance in neutrino-argon scattering*, [2201.04664](#). 61
- 4186 [59] R.A. Smith and E.J. Moniz, *Neutrino reactions on nuclear targets*, *Nucl. Phys. B*
4187 **43** (1972) 605. 61
- 4188 [60] H. Nakamura and R. Seki, *Quasi-elastic neutrino-nucleus scattering and spectral*
4189 *function*, *Nuclear Physics B - Proceedings Supplements* **112** 197. 61
- 4190 [61] V. Pandey, N. Jachowicz, T. Van Cuyck, J. Ryckebusch and M. Martini,
4191 *Low-energy excitations and quasielastic contribution to electron-nucleus and*
4192 *neutrino-nucleus scattering in the continuum random-phase approximation*, *Phys.*
4193 *Rev. C* **92** (2015) 024606 [[1412.4624](#)]. 61
- 4194 [62] A. Nikolakopoulos, R. González-Jiménez, N. Jachowicz, K. Niewczas, F. Sánchez
4195 and J.M. Udías, *Benchmarking intranuclear cascade models for neutrino*
4196 *scattering with relativistic optical potentials*, *Phys. Rev. C* **105** (2022) 054603
4197 [[2202.01689](#)]. 61

BIBLIOGRAPHY

- 4198 [63] MINIBooNE collaboration, *First Measurement of the Muon Neutrino Charged*
 4199 *Current Quasielastic Double Differential Cross Section*, *Phys. Rev. D* **81** (2010)
 4200 092005 [[1002.2680](#)]. 62
- 4201 [64] MINERvA collaboration, *Measurements of the Inclusive Neutrino and*
 4202 *Antineutrino Charged Current Cross Sections in MINERvA Using the Low- ν Flux*
 4203 *Method*, *Phys. Rev. D* **94** (2016) 112007 [[1610.04746](#)]. 62
- 4204 [65] MICROBooNE collaboration, *First Measurement of Energy-Dependent Inclusive*
 4205 *Muon Neutrino Charged-Current Cross Sections on Argon with the MicroBooNE*
 4206 *Detector*, *Phys. Rev. Lett.* **128** (2022) 151801 [[2110.14023](#)]. 62
- 4207 [66] SBND collaboration, *Neutrino cross-section measurement prospects with SBND*,
 4208 *PoS NuFact2017* (2018) 067. 62
- 4209 [67] NOvA collaboration, *Measurement of the double-differential muon-neutrino*
 4210 *charged-current inclusive cross section in the NOvA near detector*, *Phys. Rev. D*
 4211 **107** (2023) 052011 [[2109.12220](#)]. 62
- 4212 [68] T2K collaboration, *Measurement of the ν_μ charged-current cross sections on*
 4213 *water, hydrocarbon, iron, and their ratios with the T2K on-axis detectors*, *PTEP*
 4214 **2019** (2019) 093C02 [[1904.09611](#)]. 62
- 4215 [69] \$\\\$PRISM collaboration, *ν PRISM: A new way of probing neutrino*
 4216 *interactions*, *PoS NUFAC2014* (2015) 046. 62
- 4217 [70] C. Hasnip, *DUNE-PRISM – A New Method to Measure Neutrino Oscillations*,
 4218 Ph.D. thesis, Oxford U., 2023. 62
- 4219 [71] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE)*, *Far*
 4220 *Detector Technical Design Report, Volume I Introduction to DUNE*, *JINST* **15**
 4221 (2020) T08008 [[2002.02967](#)]. 63, 64, 65, 70, 71, 81, 82, 83, 84
- 4222 [72] DUNE collaboration, *Snowmass Neutrino Frontier: DUNE Physics Summary*,
 4223 [2203.06100](#). 65, 66

BIBLIOGRAPHY

- 4224 [73] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*
4225 *Detector Technical Design Report, Volume II: DUNE Physics*, 2002.03005. 65, 67,
4226 68, 75, 135, 144, 148, 269
- 4227 [74] SUPER-KAMIOKANDE collaboration, *Search for Proton Decay via $p \rightarrow e^+ \pi_0$ and*
4228 *$p \rightarrow \mu^+ \pi_0$ in a Large Water Cherenkov Detector*, *Phys. Rev. Lett.* **102** (2009)
4229 141801 [0903.0676]. 66
- 4230 [75] S. Raby, *Grand Unified Theories*, in *2nd World Summit: Physics Beyond the*
4231 *Standard Model*, 8, 2006 [[hep-ph/0608183](#)]. 66
- 4232 [76] KAMIOKANDE-II collaboration, *Observation of a Neutrino Burst from the*
4233 *Supernova SN 1987a*, *Phys. Rev. Lett.* **58** (1987) 1490. 67
- 4234 [77] R.M. Bionta et al., *Observation of a Neutrino Burst in Coincidence with*
4235 *Supernova SN 1987a in the Large Magellanic Cloud*, *Phys. Rev. Lett.* **58** (1987)
4236 1494. 67
- 4237 [78] J. Strait, E. McCluskey, T. Lundin, J. Willhite, T. Hamernik, V. Papadimitriou
4238 et al., *Long-baseline neutrino facility (lbnf) and deep underground neutrino*
4239 *experiment (dune) conceptual design report volume 3: Long-baseline neutrino*
4240 *facility for dune june 24, 2015*, 1601.05823. 67, 68
- 4241 [79] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE) Near*
4242 *Detector Conceptual Design Report, Instruments* **5** (2021) 31 [2103.13910]. 69,
4243 71, 73, 74, 76, 77
- 4244 [80] DUNE collaboration, *DUNE Phase II: Scientific Opportunities, Detector*
4245 *Concepts, Technological Solutions*, 2408.12725. 72, 78
- 4246 [81] DUNE collaboration, *A Gaseous Argon-Based Near Detector to Enhance the*
4247 *Physics Capabilities of DUNE*, 2203.06281. 74

BIBLIOGRAPHY

- 4248 [82] F. Sauli, *Gem: A new concept for electron amplification in gas detectors*, *Nuclear*
4249 *Instruments and Methods in Physics Research Section A: Accelerators,*
4250 *Spectrometers, Detectors and Associated Equipment* **386** 531. 76, 80
- 4251 [83] A. Ritchie-Yates et al., *First operation of an ALICE OROC operated in high*
4252 *pressure Ar-CO₂ and Ar-CH₄*, *Eur. Phys. J. C* **83** (2023) 1139 [2305.08822]. 79
- 4253 [84] ALICE TPC collaboration, *The upgrade of the ALICE TPC with GEMs and*
4254 *continuous readout*, *JINST* **16** (2021) P03022 [2012.09518]. 79
- 4255 [85] F. Sauli, *The gas electron multiplier (gem): Operating principles and applications*,
4256 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,*
4257 *Spectrometers, Detectors and Associated Equipment* **805** 2. 80
- 4258 [86] C. Lippmann, *A continuous read-out tpc for the alice upgrade*, *Nuclear*
4259 *Instruments and Methods in Physics Research Section A: Accelerators,*
4260 *Spectrometers, Detectors and Associated Equipment* **824** 543. 80
- 4261 [87] C. Calabria, *Large-size triple gem detectors for the cms forward muon upgrade*,
4262 *Nuclear and Particle Physics Proceedings* **273–275** 1042. 80
- 4263 [88] DUNE collaboration, *The DUNE Far Detector Vertical Drift Technology*,
4264 *Technical Design Report*, 2312.03130. 85, 86
- 4265 [89] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE)*, *Far*
4266 *Detector Technical Design Report, Volume IV: Far Detector Single-phase*
4267 *Technology*, *JINST* **15** (2020) T08010 [2002.03010]. 87
- 4268 [90] DUNE DAQ, “dtp-firmware.”
4269 <https://gitlab.cern.ch/dune-daq/readout/dtp-firmware>, 2020. 90
- 4270 [91] DUNE DAQ, “dtp-simulation.”
4271 <https://gitlab.cern.ch/dune-daq/readout/dtp-simulation>, 2020. 92

BIBLIOGRAPHY

- 4272 [92] DUNE DAQ, “dtpemulator.”
4273 https://github.com/DUNE-DAQ/dtpemulator/tree/fmlopez/filter_ana,
4274 2022. 92
- 4275 [93] J. McClellan and T. Parks, *A personal history of the Parks-McClellan algorithm*,
4276 *IEEE Signal Processing Magazine* **22** (2005) 82. 93
- 4277 [94] G. Turin, *An introduction to matched filters*, *IEEE Transactions on Information*
4278 *Theory* **6** (1960) 311. 97
- 4279 [95] J.W. Goodman, *Statistical Optics*, Wiley (1985). 99
- 4280 [96] B. Dwork, *Detection of a pulse superimposed on fluctuation noise*, *Proceedings of*
4281 *the IRE* **38** (1950) 771. 100
- 4282 [97] L. Wainstein and V. Zubakov, *Extraction of Signals from Noise*, Prentice-Hall
4283 (1962). 100
- 4284 [98] E.D. Church, *Larsoft: A software package for liquid argon time projection drift*
4285 *chambers*, **1311.6774**. 103, 104
- 4286 [99] S.V. Stehman, *Selecting and interpreting measures of thematic classification*
4287 *accuracy*, *Remote Sensing of Environment* **62** (1997) 77. 116
- 4288 [100] A.A. Taha and A. Hanbury, *Metrics for evaluating 3d medical image*
4289 *segmentation: analysis, selection, and tool*, *BMC Medical Imaging* **15** (2015) . 117
- 4290 [101] J. Silk, K.A. Olive and M. Srednicki, *The Photino, the Sun and High-Energy*
4291 *Neutrinos*, *Phys. Rev. Lett.* **55** (1985) 257. 125
- 4292 [102] M. Srednicki, K.A. Olive and J. Silk, *High-Energy Neutrinos from the Sun and*
4293 *Cold Dark Matter*, *Nucl. Phys. B* **279** (1987) 804. 125
- 4294 [103] J.S. Hagelin, K.W. Ng and K.A. Olive, *A High-energy Neutrino Signature From*
4295 *Supersymmetric Relics*, *Phys. Lett. B* **180** (1986) 375. 125

BIBLIOGRAPHY

- 4296 [104] T.K. Gaisser, G. Steigman and S. Tilav, *Limits on Cold Dark Matter Candidates*
 4297 *from Deep Underground Detectors*, *Phys. Rev. D* **34** (1986) 2206. 125
- 4298 [105] N. Bernal, J. Martín-Albo and S. Palomares-Ruiz, *A novel way of constraining*
 4299 *WIMPs annihilations in the Sun: MeV neutrinos*, *JCAP* **08** (2013) 011
 4300 [1208.0834]. 125, 126, 132
- 4301 [106] C. Rott, J. Siegal-Gaskins and J.F. Beacom, *New Sensitivity to Solar WIMP*
 4302 *Annihilation using Low-Energy Neutrinos*, *Phys. Rev. D* **88** (2013) 055005
 4303 [1208.0827]. 125, 132
- 4304 [107] C. Rott, S. In, J. Kumar and D. Yavlali, *Dark Matter Searches for Monoenergetic*
 4305 *Neutrinos Arising from Stopped Meson Decay in the Sun*, *JCAP* **11** (2015) 039
 4306 [1510.00170]. 125
- 4307 [108] DUNE collaboration, *Searching for solar KDAR with DUNE*, *JCAP* **10** (2021)
 4308 065 [2107.09109]. 125, 270, 271, 272, 273
- 4309 [109] G. Busoni, A. De Simone and W.-C. Huang, *On the Minimum Dark Matter Mass*
 4310 *Testable by Neutrinos from the Sun*, *JCAP* **07** (2013) 010 [1305.1817]. 126
- 4311 [110] R. Garani and S. Palomares-Ruiz, *Dark matter in the Sun: scattering off electrons*
 4312 *vs nucleons*, *JCAP* **05** (2017) 007 [1702.02768]. 126, 127, 130, 131, 132, 155
- 4313 [111] V.A. Bednyakov and F. Simkovic, *Nuclear spin structure in dark matter search:*
 4314 *The Zero momentum transfer limit*, *Phys. Part. Nucl.* **36** (2005) 131
 4315 [hep-ph/0406218]. 127
- 4316 [112] A. Gould, *WIMP Distribution in and Evaporation From the Sun*, *Astrophys. J.*
 4317 **321** (1987) 560. 127
- 4318 [113] J.N. Bahcall, M.H. Pinsonneault and S. Basu, *Solar models: Current epoch and*
 4319 *time dependences, neutrinos, and helioseismological properties*, *Astrophys. J.* **555**
 4320 (2001) 990 [astro-ph/0010346]. 130

BIBLIOGRAPHY

- 4321 [114] T. Golan, J.T. Sobczyk and J. Zmuda, *NuWro: the Wroclaw Monte Carlo*
 4322 *Generator of Neutrino Interactions*, *Nucl. Phys. B Proc. Suppl.* **229-232** (2012)
 4323 499. 134
- 4324 [115] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara and S. Midorikawa,
 4325 *Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model*,
 4326 *Phys. Rev. D* **92** (2015) 023004 [[1502.03916](#)]. 135, 136
- 4327 [116] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for*
 4328 *likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [[1007.1727](#)].
 4329 135
- 4330 [117] T. Kaluza, *Zum Unitätsproblem der Physik, Sitzungsber. Preuss. Akad. Wiss.*
 4331 *Berlin (Math. Phys.)* **1921** (1921) 966 [[1803.08616](#)]. 137
- 4332 [118] O. Klein, *Quantum Theory and Five-Dimensional Theory of Relativity. (In*
 4333 *German and English)*, *Z. Phys.* **37** (1926) 895. 137
- 4334 [119] T. Appelquist, H.-C. Cheng and B.A. Dobrescu, *Bounds on universal extra*
 4335 *dimensions*, *Phys. Rev. D* **64** (2001) 035002 [[hep-ph/0012100](#)]. 137
- 4336 [120] N. Arkani-Hamed, S. Dimopoulos and G.R. Dvali, *The Hierarchy problem and new*
 4337 *dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263 [[hep-ph/9803315](#)]. 137
- 4338 [121] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra*
 4339 *dimension*, *Phys. Rev. Lett.* **83** (1999) 3370 [[hep-ph/9905221](#)]. 137
- 4340 [122] G. Servant and T.M.P. Tait, *Is the lightest Kaluza-Klein particle a viable dark*
 4341 *matter candidate?*, *Nucl. Phys. B* **650** (2003) 391 [[hep-ph/0206071](#)]. 137
- 4342 [123] H.-C. Cheng, K.T. Matchev and M. Schmaltz, *Radiative corrections to*
 4343 *Kaluza-Klein masses*, *Phys. Rev. D* **66** (2002) 036005 [[hep-ph/0204342](#)]. 139
- 4344 [124] M. Blennow, J. Edsjö and T. Ohlsson, *Neutrinos from WIMP annihilations using*
 4345 *a full three-flavor Monte Carlo*, *JCAP* **01** (2008) 021 [[0709.3898](#)]. 139, 142

BIBLIOGRAPHY

- 4346 [125] J. Edsjö, J. Elevant and C. Niblaeus, *WimpSim Neutrino Monte Carlo.* 139, 142
- 4347 [126] M. Colom i Bernadich and C. Pérez de los Heros, *Limits on Kaluza-Klein dark*
4348 *matter annihilation in the Sun from recent IceCube results, Eur. Phys. J. C,* **80** 2
4349 (*2020*) **129** **80** (2019) [[1912.04585](#)]. 140
- 4350 [127] ANTARES collaboration, *Search for Dark Matter in the Sun with the ANTARES*
4351 *Neutrino Telescope in the CMSSM and mUED frameworks, Nucl. Instrum. Meth.*
4352 *A* **725** (2013) 76 [[1204.5290](#)]. 140
- 4353 [128] N. Deutschmann, T. Flacke and J.S. Kim, *Current LHC Constraints on Minimal*
4354 *Universal Extra Dimensions, Phys. Lett. B* **771** (2017) 515 [[1702.00410](#)]. 140, 141
- 4355 [129] U. Haisch and A. Weiler, *Bound on minimal universal extra dimensions from*
4356 *anti-B —> X(s)gamma, Phys. Rev. D* **76** (2007) 034014 [[hep-ph/0703064](#)]. 141
- 4357 [130] A. Freitas and U. Haisch, *Anti-B —> X(s) gamma in two universal extra*
4358 *dimensions, Phys. Rev. D* **77** (2008) 093008 [[0801.4346](#)]. 141
- 4359 [131] C. Rott, D. Jeong, J. Kumar and D. Yaylali, *Neutrino Topology Reconstruction at*
4360 *DUNE and Applications to Searches for Dark Matter Annihilation in the Sun,*
4361 *JCAP* **07** (2019) 006 [[1903.04175](#)]. 144
- 4362 [132] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al.,
4363 *Scikit-learn: Machine learning in Python, Journal of Machine Learning Research*
4364 **12** (2011) 2825. 150
- 4365 [133] ICECUBE collaboration, *Search for GeV-scale dark matter annihilation in the Sun*
4366 *with IceCube DeepCore, Phys. Rev. D* **105** (2022) 062004 [[2111.09970](#)]. 152, 153
- 4367 [134] C.-S. Chen, F.-F. Lee, G.-L. Lin and Y.-H. Lin, *Probing Dark Matter*
4368 *Self-Interaction in the Sun with IceCube-PINGU, JCAP* **10** (2014) 049
4369 [[1408.5471](#)]. 152, 153

BIBLIOGRAPHY

- 4370 [135] N.F. Bell, M.J. Dolan and S. Robles, *Searching for dark matter in the Sun using*
4371 *Hyper-Kamiokande*, *JCAP* **11** (2021) 004 [[2107.04216](#)]. 152, 153
- 4372 [136] E. Behnke et al., *Final Results of the PICASSO Dark Matter Search Experiment*,
4373 *Astropart. Phys.* **90** (2017) 85 [[1611.01499](#)]. 152, 153
- 4374 [137] PICO collaboration, *Dark Matter Search Results from the Complete Exposure of*
4375 *the PICO-60 C₃F₈ Bubble Chamber*, *Phys. Rev. D* **100** (2019) 022001
4376 [[1902.04031](#)]. 152, 153
- 4377 [138] J. Kopp, V. Niro, T. Schwetz and J. Zupan, *DAMA/LIBRA and leptonically*
4378 *interacting Dark Matter*, *Phys. Rev. D* **80** (2009) 083502 [[0907.3159](#)]. 154
- 4379 [139] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2020**
4380 (2020) 083C01. 155
- 4381 [140] M. Beltran, D. Hooper, E.W. Kolb and Z.C. Krusberg, *Deducing the nature of*
4382 *dark matter from direct and indirect detection experiments in the absence of*
4383 *collider signatures of new physics*, *Phys. Rev. D* **80** (2009) 043509 [[0808.3384](#)].
4384 155
- 4385 [141] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron.*
4386 *Astrophys.* **641** (2020) A6 [[1807.06209](#)]. 156
- 4387 [142] XENON collaboration, *Light Dark Matter Search with Ionization Signals in*
4388 *XENON1T*, *Phys. Rev. Lett.* **123** (2019) 251801 [[1907.11485](#)]. 156, 157
- 4389 [143] DARKSIDE collaboration, *Search for Dark Matter Particle Interactions with*
4390 *Electron Final States with DarkSide-50*, *Phys. Rev. Lett.* **130** (2023) 101002
4391 [[2207.11968](#)]. 156, 157
- 4392 [144] PANDAX-II collaboration, *Search for Light Dark Matter-Electron Scatterings in*
4393 *the PandaX-II Experiment*, *Phys. Rev. Lett.* **126** (2021) 211803 [[2101.07479](#)].
4394 156, 157

BIBLIOGRAPHY

- 4395 [145] C. Principato, *An Indirect Search for Weakly Interacting Massive Particles in the Sun Using Upward-going Muons in NOvA*, Ph.D. thesis, Virginia U., 2021.
 4396
 4397 10.18130/x5z2-1466. 159
- 4398 [146] G. Wikström and J. Edsjö, *Limits on the wimp-nucleon scattering cross-section from neutrino telescopes*, 0903.2986. 159
- 4400 [147] SUPER-KAMIOKANDE collaboration, *Search for neutrinos from annihilation of captured low-mass dark matter particles in the Sun by Super-Kamiokande*, *Phys. Rev. Lett.* **114** (2015) 141301 [1503.04858]. 159
 4401
 4402
- 4403 [148] C. Rott, T. Tanaka and Y. Itow, *Enhanced sensitivity to dark matter self-annihilations in the sun using neutrino spectral information*, 1107.3182. 159
 4404
- 4405 [149] M.M. Boliev, S.V. Demidov, S.P. Mikheyev and O.V. Suvorova, *Search for muon signal from dark matter annihilations inthe Sun with the Baksan Underground Scintillator Telescope for 24.12 years*, *JCAP* **09** (2013) 019 [1301.1138]. 159, 160
 4406
 4407
- 4408 [150] M. Honda, T. Kajita, K. Kasahara, S. Midorikawa and T. Sanuki, *Calculation of atmospheric neutrino flux using the interaction model calibrated with atmospheric muon data*, *Phys. Rev. D* **75** (2007) 043006 [[astro-ph/0611418](#)]. 160
 4409
 4410
- 4411 [151] C. Green, J. Kowalkowski, M. Paterno, M. Fischler, L. Garren and Q. Lu, *The Art Framework*, *J. Phys. Conf. Ser.* **396** (2012) 022020. 161
 4412
- 4413 [152] H. Bethe, *Zur theorie des durchgangs schneller korpuskularstrahlen durch materie*, *Annalen der Physik* **397** (1930) 325. 167
 4414
- 4415 [153] E. Fermi, *The ionization loss of energy in gases and in condensed materials*, *Physical Review* **57** (1940) 485. 167
 4416
- 4417 [154] R. Sternheimer, M. Berger and S. Seltzer, *Density effect for the ionization loss of charged particles in various substances*, *Atomic Data and Nuclear Data Tables* **30** (1984) 261. 167
 4418
 4419

BIBLIOGRAPHY

- 4420 [155] W.W.M. Allison and J.H. Cobb, *Relativistic Charged Particle Identification by*
4421 *Energy Loss, Ann. Rev. Nucl. Part. Sci.* **30** (1980) 253. 167
- 4422 [156] W. Blum, L. Rolandi and W. Riegler, *Particle detection with drift chambers,*
4423 *Particle Acceleration and Detection* (2008), 10.1007/978-3-540-76684-1. 168
- 4424 [157] ALICE TPC collaboration, *Particle identification of the ALICE TPC via dE/dx,*
4425 *Nucl. Instrum. Meth. A* **706** (2013) 55. 168
- 4426 [158] L. Landau, *On the energy loss of fast particles by ionization, J. Phys. (USSR)* **8**
4427 (1944) 201. 171
- 4428 [159] W. Ulmer and E. Matsinos, *Theoretical methods for the calculation of Bragg*
4429 *curves and 3D distributions of proton beams, The European Physical Journal*
4430 *Special Topics* **190** (2010) 1. 172
- 4431 [160] E. Aprile, A.E. Bolotnikov, A.L. Bolozdynya and T. Doke, *Noble Gas Detectors,*
4432 Wiley (Oct., 2008), 10.1002/9783527610020. 174
- 4433 [161] D.S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press. 211
- 4434 [162] ALICE collaboration, *Production of pions, kaons and protons in pp collisions at*
4435 $\sqrt{s} = 900 \text{ gev with alice at the lhc}$, **1101.4110**. 213
- 4436 [163] U. Einhaus, *Charged hadron identification with de/dx and time-of-flight at future*
4437 *higgs factories*, **2110.15115**. 213
- 4438 [164] R. Frühwirth, *Application of filter methods to the reconstruction of tracks and*
4439 *vertices in events of experimental high energy physics*, Ph.D. thesis, Technischen
4440 Universität Wien, 1988. 219
- 4441 [165] P. Astier, A. Cardini, R.D. Cousins, A. Letessier-Selvon, B.A. Popov and
4442 T. Vinogradova, *Kalman filter track fits and track breakpoint analysis, Nuclear*
4443 *Instruments and Methods in Physics Research Section A: Accelerators,*
4444 *Spectrometers, Detectors and Associated Equipment* **450** (2000) 138. 219

BIBLIOGRAPHY

- 4445 [166] T2K UK collaboration, *The Electromagnetic Calorimeter for the T2K Near*
4446 *Detector ND280, JINST* **8** (2013) P10019 [[1308.3445](#)]. 227
- 4447 [167] J.E. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ' ,*
4448 Ph.D. thesis, Stanford University, 1982. 232
- 4449 [168] GENIE collaboration, *Neutrino-nucleon cross-section model tuning in GENIE v3,*
4450 *Phys. Rev. D* **104** (2021) 072009 [[2104.09179](#)]. 235
- 4451 [169] J. Nieves, J.E. Amaro and M. Valverde, *Inclusive quasi-elastic neutrino reactions,*
4452 *Phys. Rev. C* **70** (2004) 055503 [[nucl-th/0408005](#)]. 235
- 4453 [170] J. Nieves, I. Ruiz Simo and M.J. Vicente Vacas, *Inclusive Charged-Current*
4454 *Neutrino-Nucleus Reactions, Phys. Rev. C* **83** (2011) 045501 [[1102.2777](#)]. 235,
4455 272
- 4456 [171] C. Berger and L.M. Sehgal, *Lepton mass effects in single pion production by*
4457 *neutrinos, Phys. Rev. D* **76** (2007) 113004 [[0709.4378](#)]. 236, 272
- 4458 [172] C. Berger and L.M. Sehgal, *PCAC and coherent pion production by low energy*
4459 *neutrinos, Phys. Rev. D* **79** (2009) 053003 [[0812.2653](#)]. 236
- 4460 [173] A. Bodek and U.K. Yang, *Modeling deep inelastic cross-sections in the few GeV*
4461 *region, Nucl. Phys. B Proc. Suppl.* **112** (2002) 70 [[hep-ex/0203009](#)]. 236, 273
- 4462 [174] T. Golan, L. Aliaga and M. Kordosky, *Minerva's flux prediction, in Proceedings of*
4463 *the 10th International Workshop on Neutrino-Nucleus Interactions in Few-GeV*
4464 *Region (NuInt15), Journal of the Physical Society of Japan, DOI.* 269
- 4465 [175] A. Bashyal, H. Schellman and L. Fields, *Ppfx implementation in deep underground*
4466 *neutrino experiment, .* 270
- 4467 [176] A. Bodek and J.L. Ritchie, *Fermi Motion Effects in Deep Inelastic Lepton*
4468 *Scattering from Nuclear Targets, Phys. Rev. D* **23** (1981) 1070. 270

BIBLIOGRAPHY

- 4469 [177] C. Wilkinson et al., *Testing charged current quasi-elastic and multinucleon*
4470 *interaction models in the NEUT neutrino interaction generator with published*
4471 *datasets from the MiniBooNE and MINERνA experiments*, *Phys. Rev. D* **93**
4472 (2016) 072010 [1601.05592]. 270
- 4473 [178] T2K collaboration, *Search for CP Violation in Neutrino and Antineutrino*
4474 *Oscillations by the T2K Experiment with 2.2×10^{21} Protons on Target*, *Phys. Rev.*
4475 *Lett.* **121** (2018) 171802 [1807.07891]. 272
- 4476 [179] MINERvA collaboration, *Identification of nuclear effects in neutrino-carbon*
4477 *interactions at low three-momentum transfer*, *Phys. Rev. Lett.* **116** (2016) 071802
4478 [1511.05944]. 272
- 4479 [180] C. Colle, O. Hen, W. Cosyn, I. Korover, E. Piasetzky, J. Ryckebusch et al.,
4480 *Extracting the mass dependence and quantum numbers of short-range correlated*
4481 *pairs from $a(e, e' p)$ and $a(e, e' pp)$ scattering*, *Phys. Rev. C* **92** (2015) 024604. 272
- 4482 [181] D. Rein and L.M. Sehgal, *Neutrino Excitation of Baryon Resonances and Single*
4483 *Pion Production*, *Annals Phys.* **133** (1981) 79. 272
- 4484 [182] M. Sanchez, “NOvA Results and Prospects.” June, 2018. 273