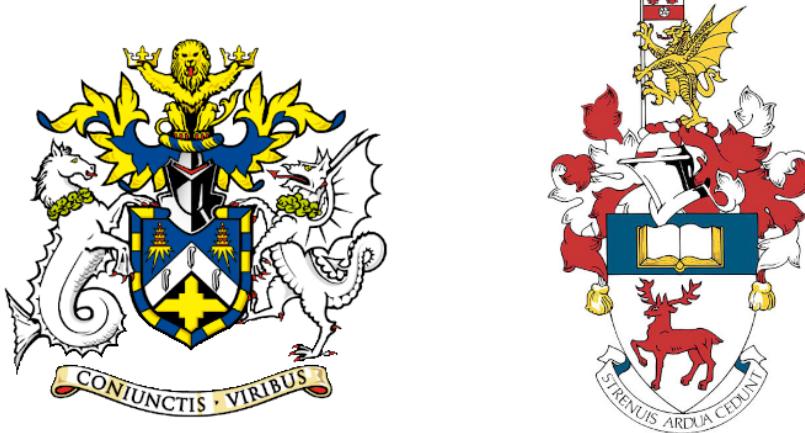


¹ EXPANDING THE PHYSICS REACH
² OF DUNE IN THE NEAR AND
³ FAR DETECTORS



⁴

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	14
List of Figures	15
List of Tables	29
List of Abbreviations	31
1 Introduction	33
2 Neutrino physics	37
2.1 Neutrinos in the SM	37
2.2 Trouble in the neutrino sector	41
2.2.1 The solar neutrino problem	41
2.2.2 The atmospheric neutrino problem	43
2.3 Massive neutrinos	45
2.4 Neutrino oscillation formalism	47
2.4.1 Oscillations in vacuum	49
2.4.2 Oscillations in matter	52
2.4.3 Current status of neutrino oscillations	54
2.5 Open questions in the neutrino sector	55
2.6 Neutrino interactions	57

CONTENTS

3 The Deep Underground Neutrino Experiment	61
3.1 Overview	61
3.2 Physics goals of DUNE	63
3.3 LBNF beamline	66
3.4 Near Detector	66
3.4.1 ND-LAr	68
3.4.2 TMS/ND-GAr	69
3.4.3 PRISM	70
3.4.4 SAND	72
3.5 A More Capable Near Detector	72
3.5.1 Requirements	73
3.5.2 Reference design	73
3.5.3 R&D efforts	76
3.6 Far Detector	79
3.6.1 Horizontal Drift	80
3.6.2 Vertical Drift	82
3.6.3 FD Data Acquisition System	84
4 Matched Filter approach to Trigger Primitives	87
4.1 Motivation	87
4.2 Signal-to-noise ratio definition	89
4.3 Low-pass FIR filter design	91
4.4 Matched filters	94
4.5 Monte Carlo studies	101
4.5.1 Angular dependence	106
4.5.2 Distortion and peak asymmetry	108
4.5.3 Hit sensitivity	112
4.6 VD ColdBox data taking	120

CONTENTS

5 Dark Matter searches with neutrinos from the Sun	123
5.1 Gravitational capture of DM by the Sun	123
5.2 Neutrino flux from DM annihilations	130
5.3 Computing limits from solar neutrino fluxes	131
5.4 Example: Kaluza-Klein Dark Matter	135
5.5 High energy DM neutrino signals	139
5.5.1 DIS-like events	141
5.5.2 Single proton QE-like events	146
5.5.3 Results	150
5.6 Example: Leptophilic Dark Matter	151
5.7 Systematic uncertainties	156
5.7.1 Systematic uncertainties in the solar WIMP signal	156
5.7.2 Systematic uncertainties in the atmospheric background	157
5.7.3 Common systematic uncertainties	158
6 Particle identification in ND-GAr	159
6.1 GArSoft	159
6.1.1 Event generation	160
6.1.2 Detector simulation	161
6.1.3 Reconstruction	162
6.2 dE/dx measurement in the TPC	164
6.2.1 Energy calibration	166
6.2.2 Truncated dE/dx mean	176
6.2.3 Mean dE/dx parametrisation	179
6.3 Muon and pion separation in the ECal and MuID	184
6.3.1 Track-ECal matching	186
6.3.2 Classification strategy	191
6.3.3 Feature selection and importance	195
6.3.4 Hyperparameter optimisation	206

CONTENTS

6.3.5	Probability calibration	210
6.3.6	Performance	212
6.4	ECal time-of-flight	214
6.4.1	Arrival time estimations	216
6.4.2	Proton and pion separation	217
6.5	Charged pion decay in flight	220
6.5.1	Track breakpoints	222
6.6	Neutral particle identification	228
6.6.1	ECal clustering	228
6.6.2	π^0 reconstruction	232
6.7	Integration in GArSoft	233
7	Event selection in ND-GAr	237
7.1	Data sample	237
7.2	ν_μ CC selection	239
7.2.1	Selection optimisation	242
7.2.2	Selection performance	248
7.2.3	Primary muon kinematics	250
7.3	Charged pion identification	254
7.3.1	ν_μ CC $1\pi^\pm$ selection	262
7.4	Neutral pion identification	266
7.5	Neutrino energy reconstruction	268
7.6	Systematic uncertainties	271
7.6.1	Flux uncertainties	271
7.6.2	Cross section uncertainties	272
7.6.3	Detector uncertainties	275
8	Conclusion and outlook	279
Bibliography		283

List of Figures

2.1	Solar neutrino fluxes for the solar model BS05(OP).	42
2.2	Zenith angle distributions for the selected ν_e and ν_μ events in the SK detector.	44
2.3	$K^0 \rightleftharpoons \bar{K}^0$ mixing through W^\pm exchange.	48
2.4	Feynman diagrams of the four most relevant CC interactions to long baseline neutrino oscillation experiments.	57
2.5	Total muon neutrino CC cross section per nucleon as a function of the neutrino energy.	58
2.6	Schematic representation of a ν_μ CCQE interaction with a neutron inside a nucleus.	59
3.1	Schematic diagram of the DUNE experiment and the LBNF beamline. .	62
3.2	Schematic longitudinal section of the LBNF beamline at Fermilab. . .	65
3.3	Predicted neutrino fluxes at the FD in FHC mode and RHC mode. . .	67
3.4	Representation of the ND hall in Phase II, showing the different subcomponents. .	68
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.	69
3.6	Schematic view of the TMS detector, highlighting its main parts. . . .	70
3.7	Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet.	71
3.8	Predicted beam muon neutrino flux at the ND location for different off-axis positions.	72

LIST OF FIGURES

3.9	Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers.	74
3.10	Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers.	75
3.11	Photographs of the TOAD pressure vessel at RHUL.	77
3.12	Electron microscope image and schematic diagram of a GEM electrode.	78
3.13	Schematic diagram showing the operating principle of a LArTPC with wire readout.	79
3.14	Proposed design for the FD-2 module following the HD principle.	80
3.15	Schematic representation of an APA frames showing the U, V, X and G wires.	81
3.16	A PDS module containing 24 X-ARAPUCAs and the location of the modules on the APAs.	82
3.17	Proposed design for the FD-1 and FD-3 modules following the VD principle.	83
3.18	Schematic representation of the electrode strip configuration for a top and bottom CRU.	84
3.19	Detailed diagram of the DUNE FD DAQ system.	85
4.1	Schematic representation of the Trigger Primitive Generation chain in the DUNE FD.	88
4.2	Example unfiltered waveform from a ProtoDUNE-SP raw data capture.	90
4.3	Example filtered waveform from a ProtoDUNE-SP raw data capture.	92
4.4	Power spectra for the current low-pass FIR filter and the optimal filter obtained using the Parks-McClellan algorithm.	93
4.5	Relative change in the S/N of the ProtoDUNE-SP raw data capture for different values of the cutoff frequency f_c and the transition width δf	94
4.6	Distribution of the relative change of the S/N on the different wire planes after the optimal FIR filter was applied.	95

LIST OF FIGURES

4.7 Example matched filtered waveform from a ProtoDUNE-SP raw data capture.	96
4.8 Relative change in the S/N for the ProtoDUNE-SP raw data capture for different values of δ and σ from the matched filter parametrisation.	99
4.9 Distribution of the relative change of the S/N on the induction planes from the ProtoDUNE-SP raw data capture after their respective optimal matched filters were applied.	101
4.10 Distributions of the particles track length in the liquid argon for the generated single particle monoenergetic samples.	102
4.11 Schematic representation of the two rotated reference frames used in the analysis of the MC filter performance.	104
4.12 Distributions of the mean S/N change per event for the different MC samples after applying the matched filters.	105
4.13 Angular dependence of the mean S/N and the S/N improvement for the monoenergetic muon sample.	106
4.14 Angular dependence of the mean S/N and the S/N improvement for the monoenergetic electron sample.	107
4.15 Waveforms for two muon events, one parallel to the APA and to the wires in the U plane and other normal to the APA plane and perpendicular to the U plane wires.	109
4.16 Distribution of the peak asymmetry for the monoenergetic muon sample and dependence of the S/N change on the mean peak asymmetry of the event.	110
4.17 Raw event display for an electron event showing the true, standard, and matched filter hits produced.	112
4.18 Dependence of the precision, sensitivity, and F_1 scores on the threshold values used in the hit finder for the FIR and matched filters.	114

LIST OF FIGURES

4.19 Dependence of the averaged hit sensitivity on the kinetic energy of the events for the matched filter and standard hits, for the case of the muon and electron samples.	116
4.20 Distributions of the hit sensitivity in the U and V planes versus the sensitivity in the X plane for the standard hits and the matched filter hits in the electron sample.	117
4.21 Standard residuals and quantile-quantile plots for the hit sensitivity of the X and U planes.	118
4.22 Schematic diagram of the vertical drift ColdBox setup at CERN.	120
4.23 Event display of the data taken with the matched filter and HMA trigger at the VD ColdBox.	121
4.24 Comparison between firmware-produced and simulated TP quantities for a matched filter run at the VD ColdBox.	122
5.1 Input solar parameters used in the capture rate computation as a function of the solar radius.	127
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.	128
5.3 NuWro computed $\nu_\mu - {}^{40}\text{Ar}$ charged-current scattering cross section as a function of the neutrino energy.	132
5.4 Expected atmospheric neutrino flux as a function of the neutrino energy E_ν at Homestake at solar minimum.	134
5.5 Feynman diagrams for $B^1 B^1$ annihilation into SM fermions.	135
5.6 Feynman diagrams for $B^1 B^1$ annihilation into a Higgs boson pair.	136
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}	137
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} . 138	

LIST OF FIGURES

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.	139
5.10 Distribution of the muon neutrino energies from the $\tau^+\tau^-$ and $b\bar{b}$ annihilation channels.	141
5.11 Angular distributions for the $b\bar{b}$ DIS sample with $m_{\text{DM}} = 10$ GeV and the atmospheric background.	141
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.	143
5.13 Signal efficiencies for the $\tau^+\tau^-$ and $b\bar{b}$ DIS samples as functions of the DM mass.	144
5.14 Angular distributions for the $\tau^+\tau^-$ QE sample with $m_{\text{DM}} = 5$ GeV and the atmospheric background.	145
5.15 Performance of the MLP classifier for the $\tau^+\tau^-$ QE signal with $m_{\text{DM}} = 5$ GeV.	147
5.16 Signal efficiencies for the $\tau^+\tau^-$ and $b\bar{b}$ single proton QE samples as functions of the DM mass.	148
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	149
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.	154
6.1 Schematic diagram showing the different modules involved in the ND-GAr production.	163
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.	168

LIST OF FIGURES

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 $N_{group} = 4$ reclustering.	169
6.4	Distribution of the Geant4-simulated energy losses per unit length versus residual range for the stopping proton sample.	170
6.5	Fitted most probable dQ/dx values for each dE/dx bin, together with best fit to the logarithmic calibration function.	171
6.6	Fitted most probable dQ/dx values for each dE/dx bin for three different ADC bit limits.	173
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.	174
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.	176
6.9	Estimated values of the mean dE/dx bias and resolution for the stopping proton sample at different values of the truncation factor.	177
6.10	Examples of the truncated mean dE/dx LanGauss fits for various $\beta\gamma$ bins, from a simulated FHC neutrino sample.	179
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.	181
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.	182
6.13	Distribution of the 60% truncated mean dE/dx versus reconstructed momentum for the FHC neutrino sample.	183

LIST OF FIGURES

6.14 Estimated values of the mean dE/dx bias and resolution obtained for the true protons in a FHC neutrino sample.	184
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.	185
6.16 Distributions of energy deposits in the ECal for a muon and a charged pion with similar momenta.	187
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).	188
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, $t_{cluster}$	190
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.	192
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.	193
6.21 Example ECal feature distributions for muons and charged pions in the five different momentum ranges considered.	197

LIST OF FIGURES

6.22 Example MuID feature distributions for muons and charged pions in the three different momentum ranges considered.	198
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 $0.3 \leq p < 0.8 \text{ GeV}/c$	200
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 $0.8 \leq p < 1.5 \text{ GeV}/c$	201
6.25 Evolution of the SHAP importance for the top six most important features across all five momentum ranges.	203
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.	205
6.27 Values of the precision and sensitivity obtained for 10000 BDT hyperparameter configurations, for the momentum regions I, III and V.	208
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).	210

LIST OF FIGURES

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 $0.3 \leq p < 0.8$ GeV/c	211
6.30 Calibrated predicted probabilities assigned by the BDT classifiers for the true muons and charged pions in a FHC neutrino sample.	212
6.31 Efficiency and purity of the muon selection as a function of the reconstructed momentum for the FHC neutrino sample.	213
6.32 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.	215
6.33 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.	216
6.34 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
6.35 Efficiency (top panel) and purity (bottom panel) for the proton selection as a function of the momentum, for $\Delta\tau = 0.10$ ns.	219
6.36 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
6.37 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

LIST OF FIGURES

LIST OF FIGURES

6.44 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
6.45 Distributions of proton dE/dx and ToF scores for a sample of 100000 FHC neutrino interactions in the HPgTPC.	234
7.1 Schematic diagram of the HPgTPC including the fiducial volume.	241
7.2 True positive, false positive, and false negative true neutrino energy distributions for a ν_μ CC selection.	242
7.3 Efficiency and purity for the ν_μ CC selection as a function of the different cuts.	244
7.4 Significance for the ν_μ CC selection as a function of the different cuts.	244
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.	245
7.6 Cumulative efficiency and purity for the ν_μ CC selection.	246
7.7 True neutrino energy spectra for the ν_μ CC selection.	247
7.8 True neutrino energy spectra for the ν_μ CC selection.	248
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.	249
7.10 Efficiency 2D distributions for the ν_μ CC selection given the true position of the interaction vertex.	250

LIST OF FIGURES

7.11 Distributions for the reconstructed versus truth primary muon momentum, longitudinal momentum, transverse momentum, and angle.	251
7.12 Efficiency and purity of the ν_μ CC selection as a function of the primary muon true momentum and beam angle.	252
7.13 Fractional residual distributions for the position of the primary vertex in the ν_μ CC selection.	253
7.14 Distribution of events given their true and reconstructed π^\pm multiplicity, for a given selection.	255
7.15 Efficiency and purity for the ν_μ CC $0\pi^\pm$ selection as a function of the different cuts.	257
7.16 Efficiency and purity for the ν_μ CC $1\pi^\pm$ selection as a function of the different cuts.	258
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.	260
7.18 Purity versus efficiency achieved for the different cut configurations explored separated by the various ν_μ CC $N\pi^\pm$ selections.	261
7.19 Efficiency of the ν_μ CC $N\pi^\pm$ selections as a function of the hadronic invariant mass.	262
7.20 Reconstructed kinematic distributions for the pion candidate in the ν_μ CC $1\pi^\pm$ selection.	263
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.	264
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.	265
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.	267
7.24 Reconstructed neutrino energy spectra for the ν_μ CC selection obtained using the truth and reconstructed information	270

LIST OF FIGURES

7.25 Neutrino energy residuals distributions for the ν_μ CC selection obtained using the truth and reconstructed information	271
7.26 Estimated dE/dx resolution as a function of the track length for true muons and protons in a ν_μ CC sample.	276

List of Tables

2.1	Values of T_3 and $Y/2$ assigned to the first generation of fermions.	39
2.2	Neutral current couplings.	40
2.3	Summary of neutrino oscillation parameters determined in the Neutrino Global Fit of 2020.	55
3.1	Summary of the two-phased plan for DUNE.	63
3.2	Exposure and time required to achieve the different physics milestones of the two phases.	64
4.1	Characteristic parameters of the two monoenergetic muon events selected for the peak asymmetry study.	109
5.1	Systematic uncertainties for the solar WIMP signal events.	156
5.2	Systematic uncertainties for the solar WIMP atmospheric background events.	157
6.1	Calibration parameters obtained from the fit of the ND-GAr simulated stopping proton sample to the calibration function, for different ADC limits.	173
6.2	Momentum ranges and description of the PID approach assumed for the muon and pion classification task.	194
6.3	Optimal values of the hyperparameters used by the BDT, for each momentum range.	209
6.4	Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges.	209

LIST OF TABLES

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.	239
7.2	Step-by-step ν_μ CC selection cuts and cumulative passing rates.	246
7.3	Neutrino interaction systematic parameters implemented in GENIE used in the DUNE TDR LBL analysis.	273
7.4	Neutrino interaction systematic parameters used in the DUNE TDR LBL analysis not present in GENIE	274

List of Abbreviations

ADC	Analog to Digital Converter	LBNF	Long Baseline Neutrino Facility
ALEPH	Apparatus for LEP Physics	MEC	Meson-Exchange Current
ALICE	A Large Ion Collider Experiment	MuID	Muon IDentification system
BDT	Boosted Decision Tree	NC	Neutral Current
CAF	Common Analysis File	ND	Near Detector
CC	Charged Current	ND-GAr	Near Detector Gaseous Argon
DIS	Deep Inelastic Scattering	ND-LAr	Near Detector Liquid Argon
DM	Dark Matter	PDG	Particle Data Group
DUNE	Deep Underground Neutrino Experiment	POT	Protons On Target
ECal	Electromagnetic Calorimeter	QE	QuasiElastic
FD	Far Detector	RES	RESonant
FHC	Forward Horn Current	RHC	Reverse Horn Current
GAr	Gaseous Argon	SIS	Shallow Inelastic Scattering
HPgTPC	High Pressure gaseous Time Projection Chamber	SM	Standard Model
LAr	Liquid Argon	SNB	Supernova Neutrino Burst
LBL	Long BaseLine	TP	Trigger Primitive
		WIMP	Weakly Interacting Massive Particle

Introduction

The beginning is the most important part of any work.

³⁸ – Plato, *The Republic*

The Standard Model (SM) of particle physics [1–3] has provided a deep understanding of the electromagnetic, weak and strong interactions, and over the past decades it has passed all kind of precision tests [4]. However, the SM by itself can not explain certain observed phenomena, such as the baryon asymmetry of the universe [5], the existence of Dark Matter (DM) [6], or the origin of neutrino masses [7].

One of the biggest puzzles in physics nowadays is how the universe came to be matter-dominated. Following the Big Bang, matter and antimatter were created in equal amounts. A. D. Sakharov described what are the necessary conditions to generate a matter-antimatter asymmetry in the early universe [8]. One of them is the existence of interactions that violate the CP symmetry. It has already been established that the amount of CP violation in the quark sector is not enough to generate the baryon asymmetry [9]. Leptons could contribute to the CP violation through the neutrino oscillation mechanism [10]. However, there is no experimental evidence for this so far.

Another yet to be solved mystery of modern physics concerns the nature of DM. From astrophysical observations (see Ref. [6] and references therein), we are aware of the existence of some unknown matter which only interacts gravitationally with other particles. Usually, extensions of the SM include feasible DM candidates. These are usually very stable, heavy particles with small interactions (if any) with SM particles. These

CHAPTER 1. INTRODUCTION

57 states are known as weakly interacting massive particles (WIMPs) [11, 12]. Experiments
58 looking for DM have set the interaction cross section between DM and SM particles to
59 be very small for DM masses below 1 TeV [13].

60 Among other next generation particle experiments, the Deep Underground Neutrino
61 Experiment (DUNE) stands out. Conceived as a neutrino oscillation experiment, it will
62 provide definitive answers to different open questions in the neutrino sector. Its main
63 goals are the discovery of CP violation in the leptonic sector and the determination
64 of the neutrino mass ordering [14]. It will also provide precision measurements of the
65 oscillation parameters within the three-flavour picture.

66 The DUNE far detector (FD) will also search for baryon-number violation and
67 neutrinos originated from supernova explosions. Moreover, its near detector (ND)
68 complex will sit next to the most powerful neutrino beam to date, allowing for a
69 rich neutrino cross section programme. This broad physics range requires a superb
70 performance from the detectors, which can also be used to look for other BSM phenomena.

71 In this thesis, I explore three different aspects of DUNE. Focusing on the data
72 acquisition system of the far detector, I start by proposing a method to enhance the
73 sensitivity of the online processing to low energy events. The idea is to modify the
74 processing chain in order to have more information available to form trigger decisions.
75 I motivate this new approach using both ProtoDUNE data and Monte Carlo (MC)
76 samples, as well as with the results from a test in a real detector setup.

77 Then, I investigate the potential of detecting neutrino fluxes from DM annihilations
78 inside the Sun with DUNE. Although this is the territory of the large volume neutrino
79 telescopes, a detector with the high resolution and pointing capabilities of the DUNE
80 FD can provide complementary information in certain regimes. I present here the results
81 of a preliminary analysis, showing the projected sensitivities for the general case and
82 two particular DM scenarios.

83 Finally, I discuss my work on the reconstruction of ND-GAr, the gaseous argon
84 component of the DUNE ND. These efforts were focused towards the development of
85 the particle identification strategy in the detector. Following a series of additions and

86 upgrades in the reconstruction, I make use of that to perform the first event selection
87 studies with fully reconstructed events in this detector.

88 This thesis opens with an overview of the status of neutrino physics in Chapter
89 2. I start summarising the role that neutrinos play in the SM, to then focus on the
90 developments that lead to the discovery of neutrino oscillations and how to accommodate
91 massive neutrinos in the model. I then discuss the phenomenology of the neutrino
92 oscillations, as well as the current experimental landscape and open questions. In the
93 final section, I review the basics of the neutrino-nucleus interaction modelling, which is
94 of great importance for DUNE.

95 Chapter 3 introduces DUNE, its physics programme and various components. I give
96 detail descriptions of the LBNF beamline, the near detector and the far detector designs.
97 I also discuss the current staging plans for DUNE. This leads to the of ND-GAr, the
98 more capable near detector planned for DUNE Phase II.

99 In Chapter 4 I start by reviewing how the trigger primitives (TPs), the basic building
100 blocks of the DUNE far detector trigger chain, are formed. I then motivate how to
101 use the filtering to enhance the TP generation in the induction channels. I describe
102 the concept of matched filter, and how to optimise it using ProtoDUNE-SP data. I
103 use different MC samples to study its performance, and assess how it improves the hit
104 finding. Finally, I present the results of the tests we performed at the VD ColdBox setup
105 at CERN, were for the first time we collected TP data with a matched filter.

106 The solar DM analysis is presented in Chapter 5. After reviewing the theoretical
107 basis for the solar DM capture and how capture and annihilation rates are related,
108 I introduce the analysis framework used. I then focus on the event selection studies
109 based on two topologies: high-energy DIS events and low-energy $1\mu 1p$ QE events. I
110 use these to extract the projected sensitivities for the DM-nucleon scattering cross
111 section, and compare them to the current status of other direct and indirect DM searches.
112 Additionally, I discuss the potential of DUNE in two specific DM models. I end with a
113 discussion of the systematic uncertainties relevant for this analysis.

114 Chapter 6 starts with a description of GArSoft, the simulation and reconstruction

CHAPTER 1. INTRODUCTION

software of ND-GAr. Then, I describe the charge calibration procedure I implemented using a MC sample of stopping protons. I use this to compute the mean ionisation loss per unit length of the tracks, and show how this procedure can be used for particle separation. Next, I summarise my investigations on the muon and pion separation using the information from the calorimeter. I outline the strategy I followed for the training and testing of the classifiers, commenting on the achieved performance using a neutrino interaction sample. Following this, I introduce the possibility of using the fast timing of the calorimeter to perform a time-of-flight measurement. It will allow to separate pions and protons in a momentum range not accessible to other methods. Additionally, I present a method to identify the decays of charged pions in the TPC, when the decay angle is too small and the pion and muon get merged into a single track. I construct a collection of variables from the track fit that allow to locate the position of the decay. Lastly, I propose a new clustering algorithm optimised for our calorimeter. I then demonstrate its impact in the context of the neutral pion reconstruction. The Chapter finishes with an overview of the integration of these reconstruction items in GArSoft.

The event selection studies are covered in Chapter 7. I start by describing the MC neutrino interaction sample I use for the studies. Then, I focus on the ν_μ CC selection, which includes an optimisation of the fiducial volume. I also explore the kinematics of the selected primary muon and the location of the neutrino vertex. Next, I study the performance of the selections based on the reconstructed charged pion multiplicity, paying special attention to the $1\pi^\pm$ selection. I briefly discuss the possibility of adding the neutral pions in the analysis. Following that, I present the results on the energy reconstruction for the selected charged-current events. I finish with a detailed discussion of the different sources of systematic error relevant for ND-GAr. These include flux and neutrino interaction modelling uncertainties, as well as detector effects.

Eventually, the thesis concludes with Chapter 8. There, I summarise the main results presented in this work, and discuss future plans for the different projects.

2

143

144

Neutrino physics

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

148

– Terry Pratchett, *Sourcery*

149 Ever since they were postulated in 1930 by Wolfgang Pauli to explain the continuous
150 β decay spectrum [15], and later found by F. Reines and C. Cowan at the Savannah
151 River reactor in 1953 [16], neutrinos have had a special place among all other elementary
152 particles. They provide a unique way to probe a wide range of physics, from nuclear
153 physics to cosmology, from astrophysics to colliders. Moreover, there is compelling
154 evidence to believe that the study of neutrinos may be key to unveil different aspects of
155 BSM physics, difficult to test elsewhere.

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

159 2.1 Neutrinos in the SM

160 The SM of fundamental interactions was initially proposed in 1967 by S. Glashow,
161 S. Weinberg and A. Salam[1–3]. This theoretical framework describes the dynamics
162 of leptons and quarks, by introducing a collection of mediating gauge vector bosons
163 and one scalar particle, known as the Higgs boson. It assumes that the local $SU(3) \times$

CHAPTER 2. NEUTRINO PHYSICS

164 $SU(2)_L \times U(1)_Y$ gauge symmetry is an internal symmetry of the system, with $SU(3)$
165 describing quantum chromodynamics, and $SU(2)_L \times U(1)_Y$ being the gauge groups of
166 the electroweak sector. For a detailed overview of the SM of electroweak interactions,
167 see Ref. [17].

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

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

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

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

173 The fact that there are no right-handed neutrino fields implies that neutrinos are
174 strictly massless within the SM. This restriction follows from the experimental observation
175 that all neutrinos produced via weak interactions are pure left-handed helicity states
176 (and similarly antineutrinos are pure right-handed states). The hypothetical existence
177 of right-handed neutrinos could be indirectly inferred from the observation of non-zero
178 neutrino masses. Nevertheless, the existence of neutrino masses is not a sufficient
179 condition for the existence of such fields.

180 Left and right-handed fermions transform differently under $SU(2)_L \times U(1)_Y$ rotations,
181 as the right-handed particles are singlets under $SU(2)_L$. Applying a local transformation,
182 they change as:

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

183 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

184 $\alpha_a(x)$ are the parameters of the rotation.

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

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

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

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

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

200 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,
201 respectively, and g and g' are the corresponding gauge couplings. It can be shown that
202 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

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

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

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

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

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

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

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

2.2. TROUBLE IN THE NEUTRINO SECTOR

218 with the currents defined as:

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

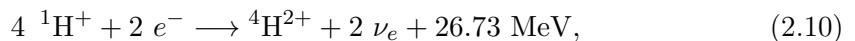
219 where f denotes any SM fermion, ℓ and ν_ℓ a charged lepton and a neutrino of any flavour,
220 and u_f and d_f an up-like and a down-like quarks of any flavour. For the NC case, the
221 values of the v_f and a_f couplings are given in Tab. 2.2.

222 As seen in Eq. (2.9), in the electroweak theory neutrinos are coupled to the Z boson
223 in a universal way. Therefore, by measuring the so-called invisible decay width of the Z
224 boson we have an estimate of the number of light (i.e. lighter than the Z boson) neutrino
225 flavours. This number was measured by LEP in a combined analysis of $e^+ e^- \rightarrow \mu^+ \mu^-$
226 and $e^+ e^- \rightarrow$ hadrons to be $N_\nu = 2.9840 \pm 0.0082$ [18].

227 2.2 Trouble in the neutrino sector

228 2.2.1 The solar neutrino problem

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



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

CHAPTER 2. NEUTRINO PHYSICS

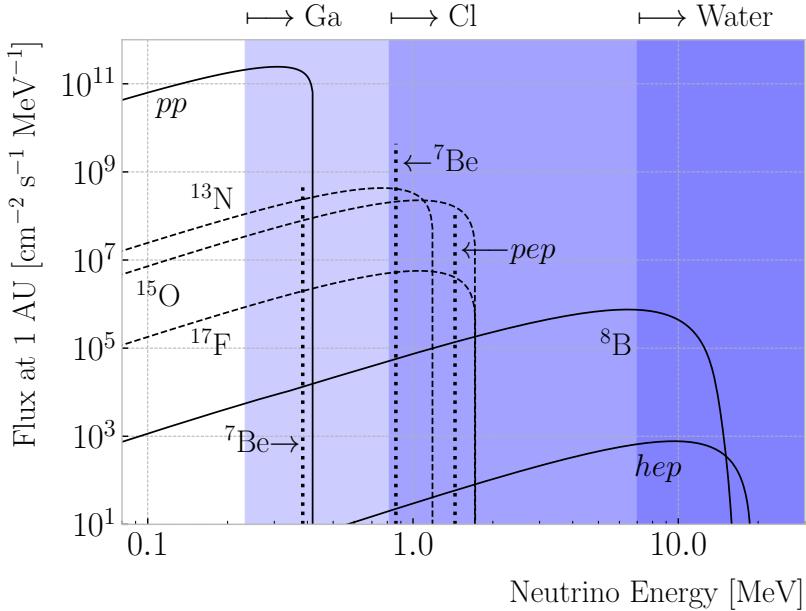


Figure 2.1: Solar neutrino fluxes for the solar model BS05(OP). The detection thresholds for Gallium, Chlorine and water-based experiments are also shown. Figure adapted from Ref. [20].

In the late 1960s, the Brookhaven Solar Neutrino Experiment, led by R. Davis, started

data taking with the goal of measuring the solar neutrino flux [19]. The experiment

used a tank containing 380 m^3 of tetrachloroethene (C_2Cl_4), a liquid commonly used

in dry-cleaning, located 1.5 km underground in the Homestake mine in Lead, South

Dakota. The incoming neutrinos would get captured following the reaction:

$$\nu_e + {}^{37}\text{Cl} \longrightarrow {}^{37}\text{Ar} + e^-, \quad (2.11)$$

therefore allowing to measure the neutrino flux by counting the ${}^{37}\text{Ar}$ isotopes. The

threshold for this reaction is 0.814 MeV, just below the 0.862 MeV line from the ${}^7\text{Be}$

ground state transition.

The results of the experiment were compared to the theoretical predictions made by

J. Bahcall [21]. During its operation from 1968 to 2002, the experiment observed a solar

ν_e flux that was approximately a third of the total prediction [22].

In the early 1990s, the SAGE [23] and GALLEX [24] experiments started operations.

2.2. TROUBLE IN THE NEUTRINO SECTOR

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

253 In the early 2000s, the SNO experiment put an end to the solar neutrino puzzle
 254 [25, 26]. Thanks to its directionality capabilities, being a Cherenkov light detector, as
 255 well as to its heavy water target, SNO measured the total solar neutrino flux through
 256 the NC process:

$$\nu_\alpha + d \longrightarrow n + p + \nu_\alpha, \quad (2.12)$$

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

$$\nu_e + d \longrightarrow p + p + e^-, \quad (2.13)$$

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

261 2.2.2 The atmospheric neutrino problem

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

$$\begin{aligned} \pi^+ &\longrightarrow \mu^+ + \nu_\mu, \\ &\mu^+ \longrightarrow e^+ + \bar{\nu}_\mu + \nu_e, \end{aligned} \quad (2.14)$$

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

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

266 of produced neutrinos and antineutrinos is, in good approximation, equal to two [27].

CHAPTER 2. NEUTRINO PHYSICS

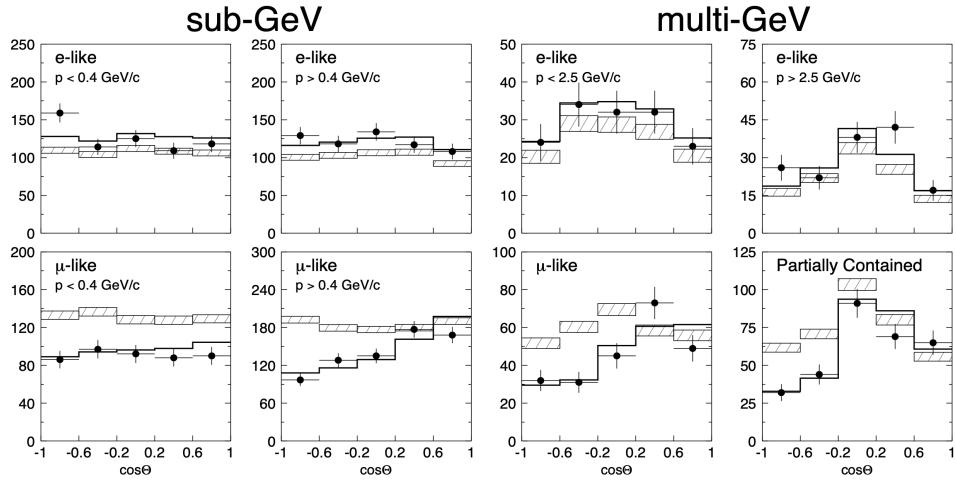


Figure 2.2: Zenith angle distributions for the selected ν_e (top row) and ν_μ (bottom row) events in the SK detector. The hatched region corresponds to the expectation in the case of no oscillations, whereas the solid line indicates the best-fit in the case of $\nu_\mu \rightarrow \nu_\tau$ oscillations. Figure taken from Ref. [32].

267 During the 1980s, several proton decay experiments, like Kamiokande [28], IMB [29],

268 MACRO [30], and Soudan-2 [31], measured the flux of atmospheric neutrinos. This was

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

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

271 lower than the predictions.

272 A few years before the SNO discovery, in 1998, the Super-Kamiokande collaboration

273 (SK) measured the atmospheric ν_e and ν_μ spectra as a function of the zenith angle

274 [32]. Upward-going particles have negative zenith angle, $\cos \Theta < 0$, indicating that they

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

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

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

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

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

280 energy and the path length of the neutrino.

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

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

283 groundbreaking discoveries of modern physics and has acted as driving force for BSM

2.3. MASSIVE NEUTRINOS

284 physics. The minimal extension of the SM we can make to address this phenomenon is
285 introducing different masses for at least two of the neutrinos. This way, we are left with
286 three neutrino mass eigenstates ν_1 , ν_2 , and ν_3 , with masses m_1 , m_2 , and m_3 respectively,
287 which in general will not coincide with the flavour eigenstates, ν_e , ν_μ , and ν_τ .

288 2.3 Massive neutrinos

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

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

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

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

302 If one imposes lepton number symmetry conservation, the Majorana term must
303 banish, $M_N = 0$. In this case, if $m = 3$ we can identify the sterile neutrinos as the
304 right-handed component of the neutrino field. The Dirac mass matrix can be diagonalised
305 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

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

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

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

308 with:

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

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

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

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

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

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

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

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

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

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

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

2.4. NEUTRINO OSCILLATION FORMALISM

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

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

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

327 where the two mass matrices are given by:

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

328 with V_l and V_h two unitary matrices.

329 This scenario represents the so-called see-saw mechanism [33–37]. The name comes
 330 from the fact that the masses of the heavy states are proportional to M_N , whereas for
 331 the light states they are proportional to M_N^{-1} . While both the heavy and the light
 332 neutrinos are Majorana particles, it can be shown that the heavy states are mainly
 333 right-handed, whereas the light ones are mostly left-handed.

334 2.4 Neutrino oscillation formalism

335 Neutrino oscillations were first proposed in 1958 by B. Pontecorvo [38], inspired by the
 336 neutral kaon oscillation phenomenon [39]. Neutral kaons, K^0 and \bar{K}^0 , have opposite
 337 strangeness (± 1) and are produced in strong processes. It was observed that, when
 338 having a beam initially pure of neutral kaons of one type, these would transition into
 339 their antiparticles while propagating. Because the weak interaction does not conserve
 340 strangeness, neutral kaons can change their identity via the processes shown in Fig. 2.3.

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

CHAPTER 2. NEUTRINO PHYSICS

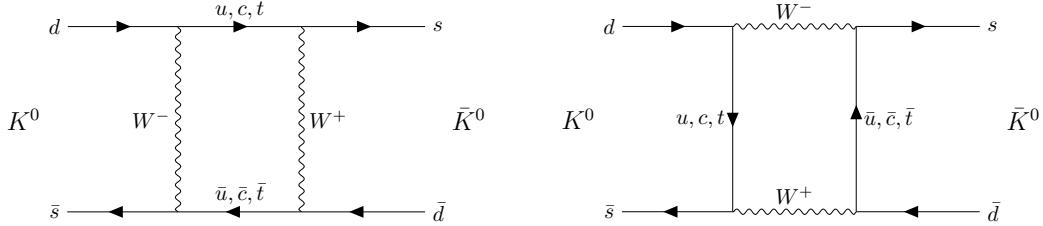


Figure 2.3: $K^0 \rightleftharpoons \bar{K}^0$ mixing through W^\pm exchange.

342 antineutrino states, as only one neutrino flavour was known at the time. After the
 343 discovery of the muon neutrino, the mixing between flavours was also explored [40].

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

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

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

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

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

2.4. NEUTRINO OSCILLATION FORMALISM

361 Nakagawa-Sakata (PMNS) matrix [41, 42], relates the set of active neutrinos and the
362 three mass eigenstates as:

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

363 where the Greek index α denotes the flavour $\{e, \mu, \tau\}$ and the Latin index i the mass state
364 $\{1, 2, 3\}$. This leptonic mixing matrix may be parametrized in terms of 6 parameters,
365 of which are mixing angles θ_{12} , θ_{13} and θ_{23} , one CP-violating phase δ_{CP} , and 2 Majorana
366 phases α and β :

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

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

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

376 2.4.1 Oscillations in vacuum

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

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

379 in the plane wave approximation, as the mass eigenstates are also eigenstates of the free
380 Hamiltonian.

CHAPTER 2. NEUTRINO PHYSICS

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

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

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

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

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

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

388 where Δm_{ij}^2 is the difference of the squared masses of the j th and i th neutrino mass
389 eigenvalues. A usual way to write the phase responsible for the oscillations is:

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

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

2.4. NEUTRINO OSCILLATION FORMALISM

³⁹³ CP symmetry violation:

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

³⁹⁴ Assuming that CPT invariance holds, then the following relation must be true:

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

³⁹⁵ as these two process are related by the CPT symmetry. From the definition of probability,

³⁹⁶ we also must have:

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

³⁹⁷ where the sum includes all flavours (also α). From these two constraints, one can probe

³⁹⁸ that:

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

³⁹⁹ and in particular:

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

⁴⁰⁰ A direct consequence of this last relation is that there are no observable CP-violating

⁴⁰¹ effects in the so-called disappearance experiments. One needs to perform appearance

⁴⁰² experiments, where the flavour detected is different from the original flavour, in order

⁴⁰³ to measure the CP asymmetry. Neutrino experiments often report the amount of CP-

⁴⁰⁴ violation through the Jarlskog invariant. In terms of the parametrisation typically used

⁴⁰⁵ to write the PMNS matrix, it is given by:

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

⁴⁰⁶ The Jarlskog invariant can be used to compare the amount of CP-violation in the lepton

⁴⁰⁷ and the quark sectors, where $J = 3.12^{+0.13}_{-0.12} \times 10^{-5}$ in the latter [43].

CHAPTER 2. NEUTRINO PHYSICS

408 2.4.2 Oscillations in matter

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

415 The first proposed model to account for neutrino oscillations in matter was proposed
 416 by Mikhaev, Smirnov and Wolfenstein (MSW) [44]. It relies on the fact that, as the
 417 only charged lepton present in ordinary matter is the electron, electron neutrinos can
 418 undergo both charged and neutral-current interactions with matter whereas for muon
 419 and tau neutrinos just neutral currents are possible.

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

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

423 with a vacuum Hamiltonian given by:

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

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

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

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

2.4. NEUTRINO OSCILLATION FORMALISM

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

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

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

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

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

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

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

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

441 with:

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

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

CHAPTER 2. NEUTRINO PHYSICS

443 $\nu_e \rightarrow \nu_\mu$ (in the two flavour approximation) reads:

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

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

449 2.4.3 Current status of neutrino oscillations

450 A wide range of neutrino experiments provide experimental input to the neutrino
451 oscillation framework, both using natural or man-made neutrino sources. The results
452 from one of the neutrino global fit analyses, shown in Tab. 2.3 ¹, summarise well our
453 current understanding of the different oscillation parameters.

454 **Solar neutrino experiments** detect neutrinos produced in thermonuclear reactions
455 inside the Sun, mainly from the so-called *pp* chain and the CNO cycle. These neutrinos
456 have a typical energy in the range from 0.1 to 20 MeV. These experiments (Homestake
457 [45], GALLEX [24], SAGE [23], Borexino [46], Super-Kamiokande [47] and SNO [48])
458 provide the best sensitivities to θ_{12} and Δm_{21}^2 .

459 **Atmospheric neutrino experiments** detect the neutrino flux produced when
460 cosmic rays scatter with particles in Earth's atmosphere. These collisions generate particle
461 showers that eventually produce electron and muon neutrinos (and antineutrinos). Their
462 energies range from few MeV to about 10^9 GeV. Experiments like Super-Kamiokande
463 [49] and IceCube [50] use atmospheric neutrinos to measure oscillations, and are specially
464 sensitive to θ_{23} and Δm_{32}^2 .

465 **Reactor neutrino experiments** look for the $\bar{\nu}_e$ spectrum produced by nuclear
466 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 [61].

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

467 long-baseline experiments like KamLAND [51] are sensitive to the solar mass splitting
468 Δm_{21}^2 , whereas much shorter baseline experiments such as RENO [52] or DayaBay [53]
469 measure θ_{13} and Δm_{31}^2 .

470 **Accelerator experiments** measure neutrino interactions from beams generated
471 by particle accelerators. Mesons are produced when the protons from the accelerator
472 collide into a target. These are focused into a beam, then some of them decay to muon
473 neutrinos while the rest are removed from the beam by an absorber. Depending on the
474 configuration one can obtain a beam made of mostly neutrinos or antineutrinos. The
475 typical energies of these neutrinos are in the GeV range. Experiments such as NOvA
476 [54], T2K [55], MINOS [56], OPERA [57] and K2K [58] (and in the future DUNE [59])
477 are primarily sensitive to θ_{13} , θ_{23} and Δm_{32}^2 . Also, in the coming years DUNE [59] and
478 Hyper-Kamiokande [60] will be sensitive to δ_{CP} .

479 2.5 Open questions in the neutrino sector

480 A crucial question that remains open these days, and is of vital importance for the
481 oscillation phenomenology, is whether the mass eigenstate ν_3 is the heaviest (what we

CHAPTER 2. NEUTRINO PHYSICS

482 call normal ordering) or the lightest (referred to as inverted ordering) of the mass states.
483 In other words, this means that we do not know the sign of Δm_{32}^2 , so we can either have
484 $m_1 < m_2 < m_3$ (NO) or $m_3 < m_1 < m_2$ (IO).

485 Also relevant to the neutrino oscillations, there is the problem regarding the θ_{23}
486 octant. Previous experimental results were compatible with its value being close to
487 maximal, $\theta_{23} \sim 45^\circ$ [62, 63]. However, global data fits indicate a deviation from the
488 maximal mixing, giving rise to two degenerate solutions, one in the lower octant $\theta_{23} < 45^\circ$
489 and another in the higher octant $\theta_{23} > 45^\circ$ (see e.g. Ref. [61]).

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

497 These three questions, because of their nature, could be understood thanks to the
498 next generation of oscillation experiments, like DUNE.

499 Notwithstanding, there are other mysteries that can not be unveiled just by conducting
500 oscillation experiments, as certain quantities do not influence these phenomenon. Among
501 these there is the question of the absolute values of the neutrino masses. Depending
502 on the value of the lightest of the neutrino masses we can have different mass spectra,
503 from hierarchical $m_1 \ll m_2 < m_3$ (NO) or $m_3 \ll m_1 < m_2$ (IO) to quasi-degenerate
504 $m_1 \simeq m_2 \simeq m_3$.

505 Other open question concerns the nature itself of the neutrinos. If neutrinos are Dirac
506 particles then their mass term can be generated through the usual Higgs mechanism
507 by adding right-handed neutrino fields. However, if they are Majorana particles and
508 therefore their own antiparticles, there is no need to add extra fields to have the mass
509 term in the Lagrangian. Experiments like SuperNEMO [64], SNO+ [65], and NEXT
510 [66], which search for neutrino-less double beta decay, will be able to determine whether

2.6. NEUTRINO INTERACTIONS

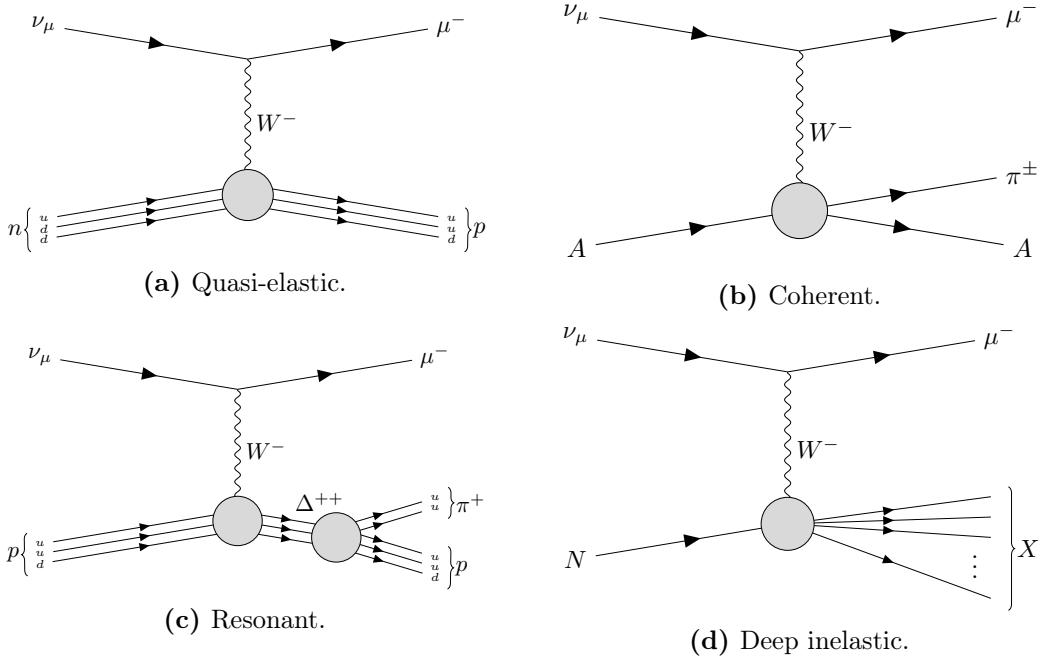


Figure 2.4: Feynman diagrams of the four most relevant CC interactions to long baseline neutrino oscillation experiments.

511 neutrinos are Dirac or Majorana.

512 2.6 Neutrino interactions

513 The study of neutrino-nucleus interactions is of great importance for long baseline
514 neutrino oscillation experiments. The interaction model provides a mapping between
515 the energy of the incoming neutrino and the final state particles after the interaction.
516 Because in this kind of experiments neutrinos are obtained as secondary decay products
517 of mesons, typically charged pions and kaons, their energies are not known a priori. Not
518 only that, but the kinematics of the interacting nucleon are also unknown. Therefore, we
519 rely on the neutrino interaction models to provide this relation between the observables
520 in the detector and the true kinematics of the neutrino. Interaction modelling is expected
521 to be the one of the leading sources of systematic uncertainties in the next generation of
522 long baseline experiments [67–69].

523 In the case of neutrino interactions with nuclei, at the energies relevant for long

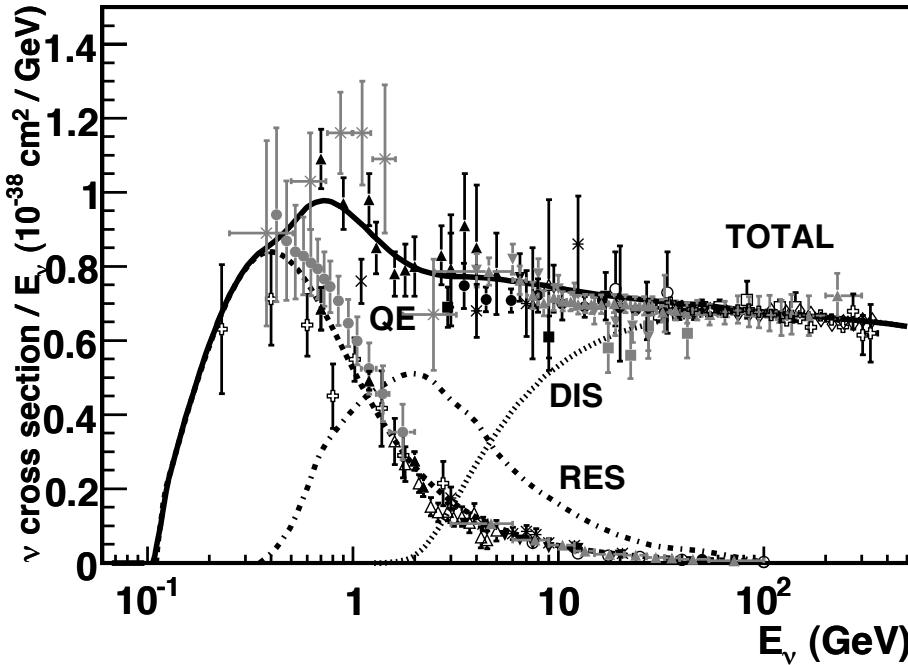


Figure 2.5: Total ν_μ CC cross section per nucleon as a function of the neutrino energy. The contributions from the different channels are shown. Figure taken from Ref. [70].

524 baseline oscillation experiments, around the GeV-scale, the process is dominated by
 525 the interaction between the neutrino and a single nucleon within the nuclear medium.
 526 Figure 2.4 shows examples of the four most common neutrino CC interactions. In this
 527 diagrams *A* indicated that the interaction happened with the nucleus as a whole, whereas
 528 *N* denotes a single nucleon.

529 At low energies, below 1 GeV, quasi-elastic (QE) interactions dominate. In a CCQE
 530 interaction a neutrino (antineutrino) interacts with a neutron (proton), converting it into
 531 a proton (neutron) which is then ejected from the nucleus together with the resulting
 532 charged lepton. Neutrinos can also scatter off bound states of nucleons inside the nucleus.
 533 These interactions are typically denoted as $n p n h$, as the neutrino interacts with a bound
 534 state of n nucleons and leave n holes in the nucleus. Particularly relevant are the $2p2h$
 535 events, as distinguishing them from pure CCQE events is challenging. The $2p2h$ neutrino
 536 interactions are dominated by meson exchange current events (MEC), where the nucleons
 537 are bound by the exchange of a virtual meson.

2.6. NEUTRINO INTERACTIONS

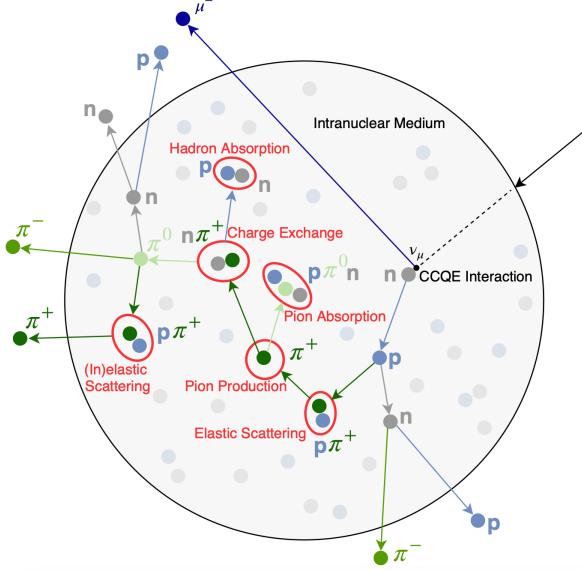


Figure 2.6: Schematic representation of a ν_μ CCQE interaction with a neutron inside a nucleus. The reaction produces a muon and a proton, which travel through the nuclear medium. The outgoing proton undergoes various kinds of hadronic FSIs on its way out. Figure taken from Ref. [71].

538 At energies above 1 GeV the neutrino is able to excite the nucleons into a baryonic
 539 resonance, which promptly decays into a nucleon and a pion. These are the so-called
 540 resonant (RES) interactions. Neutrinos also interact with entire nucleus coherently, in
 541 the process known as coherent (COH) interaction. This kind of reactions also produce a
 542 single pion in the final state. At high neutrino energies, above 5 GeV, deep inelastic
 543 scattering (DIS) takes place. In these processes, the neutrino interacts with a single
 544 quark within the nucleon, breaking the nucleon and producing a hadronic shower.

545 Figure 2.5 shows a compilation of measurements of the total ν_μ CC cross section
 546 (see Ref. [70] for the details of the different experimental results). Also shown are the
 547 contributions from the different interaction modes. The contribution of the CCCOH
 548 interaction is omitted, as it is negligible compared to the others. This shows how the
 549 interaction model needs to accurately predict the neutrino-nucleon cross section for the
 550 different interaction modes across a broad energy range, to obtain the correct relative
 551 contributions.

552 Nuclear effects alter the neutrino cross section, as well as the multiplicities of the

CHAPTER 2. NEUTRINO PHYSICS

final state particles. Therefore, the interaction models need to account for the effects introduced by the nuclei. There are several models available to describe the initial state of the nucleus, like the relativistic Fermi gas model [72], spectral functions [73] or the random phase approximation [74]. The other main effect that interaction models have to deal with are the so-called final state interactions (FSI). These are the interactions of the particles produced in the neutrino-nucleon scattering as they travel through the nuclear medium. Typically, the lepton exits the nucleus without interacting. However, hadrons tend to get scattered, absorbed or re-emitted. These effects are usually described by means of intra-nuclear cascade models [75]. Figure 2.6 illustrates the effects of FSI on the observable particle content in the detector after a ν_μ CCQE interaction.

There exists a rich experimental programme dedicated to the measurement of neutrino cross sections. The list of such experiments in the recent years include MiniBooNE [76], MINERvA [77], MicroBooNE [78] and SBND [79]. Additionally, thanks to their near detectors, long baseline experiments can perform cross section measurements. Some recent examples are NOvA [80] or T2K [81]. Future oscillation experiments will greatly benefit from these measurements, as the measurement of the oscillation parameters depends on the cross section modelling. However, there are alternative data-driven approaches to extract the oscillation probabilities without relying on a neutrino interaction model, which are planned to be explored in the next generation of experiments [82, 83].

574

The Deep Underground Neutrino Experiment

576

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

578

– Frank Herbert, *Dune*

579

The Deep Underground Neutrino Experiment (DUNE) is a next generation long-baseline neutrino oscillation experiment [14]. It will address several questions in neutrino physics, study neutrinos from astrophysical sources and search for beyond the standard model physics.

583

This Chapter reviews the main goals of DUNE, the operating principle of the LBNF beamline, the role that the near detector plays in the oscillation measurement, and the design of the far detector modules and their data acquisition (DAQ) system.

586

3.1 Overview

587

The main physics goals of DUNE are:

588

- measure the neutrino mass hierarchy, the amount of CP violation in the leptonic sector and the θ_{23} octant,

590

- detect rare low energy neutrino events, like neutrinos from supernova bursts, and

591

- search for proton decay and other beyond the standard model phenomena.

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

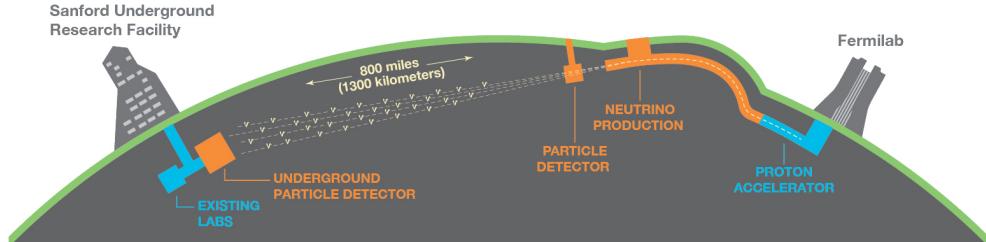


Figure 3.1: Schematic diagram of the DUNE experiment and the LBNF beamline [14].

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 diagram with 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 programme, 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. [84].

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

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

614 DUNE is planned to be built using a staged approach consisting on two phases,
 615 which are summarised in Tab. 3.1. Phase I consists of a FD with 50% of the total
 616 fiducial mass, a reduced version of the ND complex and a 1.2 MW proton beam. It will
 617 be sufficient to achieve some early physics goals, like the determination of the neutrino
 618 mass ordering. For its Phase II, DUNE will feature the full four FD modules, a more
 619 capable ND and a 2.4 MW proton beam. The physics milestones for the two phases are
 620 given in Tab. 3.2, in a staging scenario which assumes that Phase II is completed after
 621 6 years of operation.

622 A summary of the DUNE science programme can be found in the DUNE FD
 623 Technical Design Report (TDR) Volume I [14]. For a detailed discussion on the two-
 624 phased approach the reader is referred to the DUNE Snowmass 2021 report [84].

625 3.2 Physics goals of DUNE

626 As noted in the literature (see for instance Ref. [61] for a review), the parameter space of
 627 the neutrino oscillation phenomena within the three-flavour picture is quite constrained
 628 by current experimental data. However, there are still crucial open questions, like the
 629 mass ordering, the value of δ_{CP} or the θ_{23} octant. One of the main goals of DUNE is to
 630 determine precisely the values of these parameters [85].

631 To address these questions DUNE can look to the subdominant oscillation channel
 632 $\nu_\mu \rightarrow \nu_e$ ($\bar{\nu}_\mu \rightarrow \bar{\nu}_e$) and study the energy dependence of the ν_e ($\bar{\nu}_e$) appearance probability.
 633 When we focus on the antineutrino channel $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ there is a change in the sign of δ_{CP} ,
 634 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. [84].

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

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

Another of the main physics goals of DUNE is the search for baryon-number violating processes. Specifically, it will try to answer the question of whether protons are stable or not. There is no symmetry argument that forbids protons from decaying, but its apparent stability seems to suggest that baryon number is conserved [86]. However, proton decay is a usual feature of grand-unified theories, where electromagnetic, weak and strong interactions are unified above a certain energy scale [87].

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

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

3.3. LBNF BEAMLINE

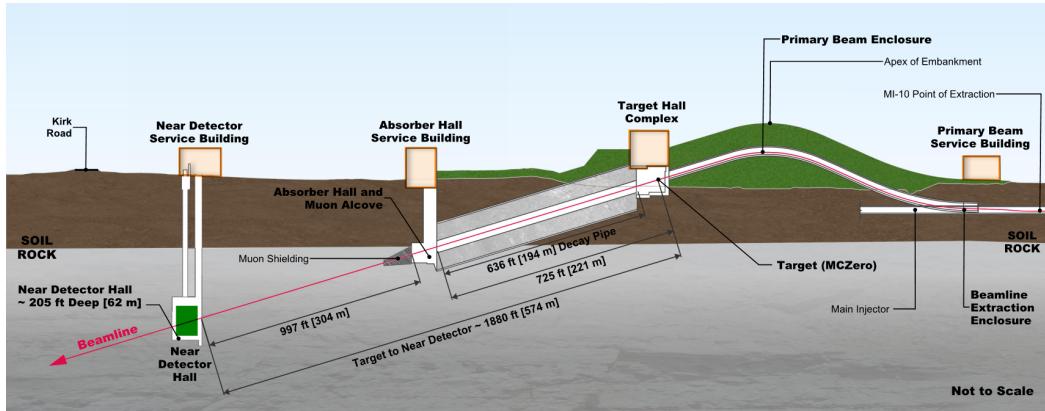


Figure 3.2: Schematic longitudinal section of the LBNF beamline at Fermilab (not to scale). Figure taken from Ref. [90].

653 events from the 1987A supernova located in the Magellanic Cloud, 50 kpc away from
 654 Earth [88,89].

655 DUNE aims to collect SNB events. Although these are quite rare, as the expected
 656 supernovae explosion events are about one every few decades for our galaxy and
 657 Andromeda, the long lifetime of the experiment (around a couple of decades as well)
 658 makes it reasonable to expect some. Nowadays the main sensitivity to SNB of most
 659 experiments is to the $\bar{\nu}_e$ flux through inverse beta decay. One of the advantages of
 660 DUNE is its expected sensitivity to MeV-scale ν_e events, since the dominant channel
 661 will be ν_e CC scattering.

662 Moreover, due to the stringent requirements that the main physics goals set for
 663 DUNE, it will allow also to perform searches for all kind of BSM physics. Among
 664 others, DUNE will be able to look for: active-sterile neutrino mixing, non-unitarity of
 665 the PMNS matrix, non-standard interactions, Lorentz and CPT violations, neutrino
 666 trident production, light-mass DM, boosted DM and heavy neutral leptons. The reader
 667 is referred to the DUNE FD TDR Volume II [85] for a full discussion of the physics
 668 scope of DUNE.

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

669 3.3 LBNF beamline

670 The LBNF project is responsible for producing the neutrino beam for the DUNE detectors.
671 A detailed discussion of the LBNF programme can be found in the DUNE/LBNF CDR
672 Volume III [90].

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

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

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

689 3.4 Near Detector

690 To estimate the oscillation parameters we measure the neutrino energy spectra at the FD.
691 This reconstructed energy arises from a convolution of the neutrino flux, cross section,
692 detector response and the oscillation probability. Using theoretical and empirical models
693 to account for the other effects, one can extract the oscillation probability using the
694 measurement. However, these models have associated a number of uncertainties that

3.4. NEAR DETECTOR

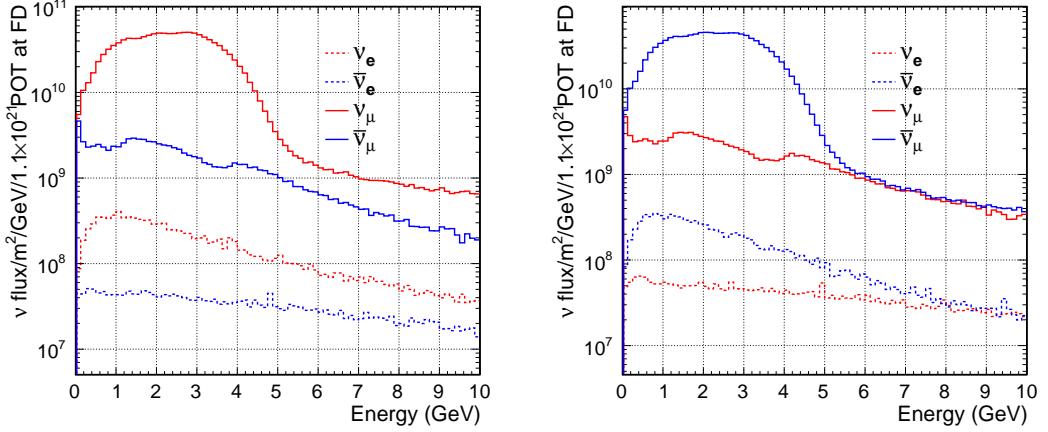


Figure 3.3: Predicted neutrino fluxes at the FD in FHC mode (left panel) and RHC mode (right panel). Figures taken from Ref. [85].

695 are then propagated to the oscillation parameters.

696 One of the main roles of the ND is to measure the neutrino interaction rates before
 697 the oscillation effects become relevant, i.e. close to the production point. By measuring
 698 the ν_μ and ν_e energy spectra, and that of their corresponding antineutrinos, at the ND
 699 we can constrain the model uncertainties. A complete cancellation of the uncertainties
 700 when taking the ratio between the FD and ND measurements is not possible, as that
 701 would require both detectors to have identical designs and the neutrino fluxes seen by
 702 them to be the same. Because of the distance, the flux probed by the FD will have a
 703 different energy and flavour composition than that at the ND, as neutrinos oscillate and
 704 the beam spreads. The differences in the flux also determine the design of the detectors,
 705 therefore the ND is limited in its capability to match the FD design.

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

711 Additionally, the ND will have a physics programme of its own. In particular, it will
 712 measure neutrino cross sections that will then be used to constrain the model used in

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

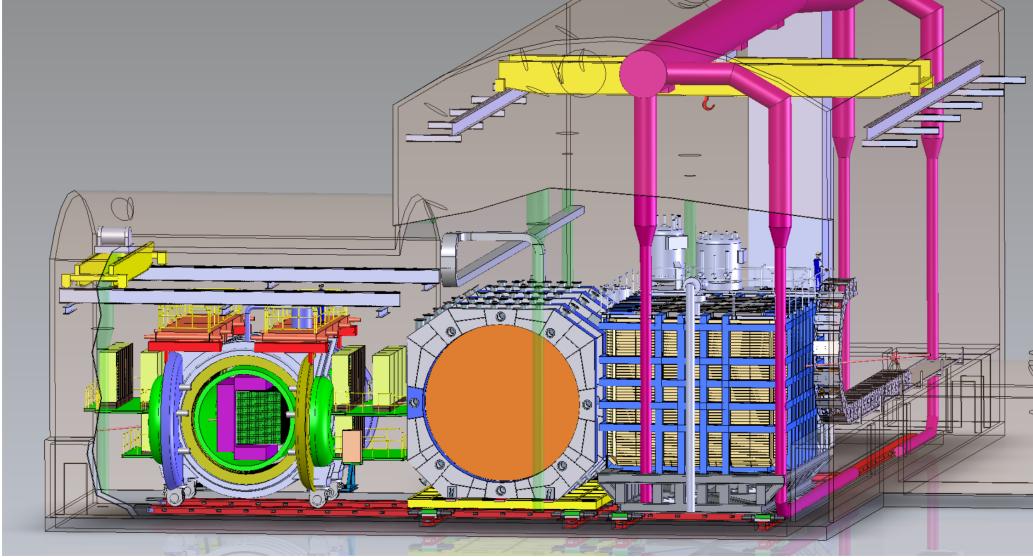


Figure 3.4: Representation of the ND hall in Phase II, showing the different subcomponents. From right to left, in the direction of the beam, we have ND-LAr, ND-GAr and SAND. Figure taken from Ref. [91].

713 the long-baseline oscillation analysis. It will also be used to search for BSM phenomena
 714 such as heavy neutral leptons, dark photons, millicharged particles, etc.

715 The DUNE ND can be divided in three main components, a LArTPC known as ND-
 716 LAr, a magnetised muon spectrometer, which will be the Temporary Muon Spectrometer
 717 (TMS) in Phase I and ND-GAr in Phase II, and the System for on-Axis Neutrino
 718 Detection (SAND). The layout of the Phase II DUNE ND can be seen in Fig. 3.4. The
 719 first two components of the ND will be able to move off-axis, in what is called the
 720 Precision Reaction-Independent Spectrum Measurement (PRISM) concept. More details
 721 on the purpose and design of the ND can be found in the DUNE ND Conceptual Design
 722 Report (CDR) [91].

723 3.4.1 ND-LAr

724 ND-LAr is a LArTPC, as the ND needs a LAr component to reduce cross section and
 725 detector systematic uncertainties in the oscillation analysis. However, its design differs
 726 significantly from those proposed for the FD modules. Because of the high event rates
 727 at the ND, approximately 55 neutrino interaction events per 10 μ s spill, ND-LAr will be

3.4. NEAR DETECTOR

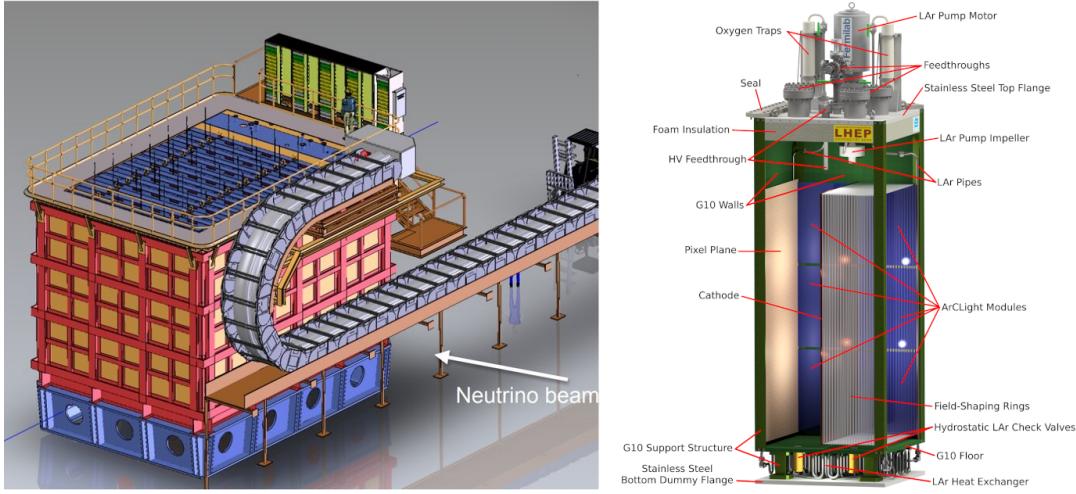


Figure 3.5: Schematic representation of the external components of ND-LAr, including the cryostat and the PRISM movable system (left) and detailed drawing of one ArgonCube module (right). Figure adapted from Ref. [14].

728 built in a modular way. Each of the modules, based on the ArgonCube technology, is a
 729 fully instrumented, optically isolated TPC with a pixelated readout [92]. The pixelisation
 730 allows for a fully 3D reconstruction and the optical isolation reduces the problems due
 731 to overlapping interactions. Figure 3.5 shows a representation of the external parts of
 732 ND-LAr (left) and a detailed diagram of an ArgonCube module (right).

733 With a fiducial mass of 67 t and dimensions 7 m (w) \times 3 m (h) \times 5 m (l), ND-LAr
 734 will be able to provide high statistics and contain the hadronic systems from the beam
 735 neutrino interactions, but muons with a momentum higher than 0.7 GeV will exit the
 736 detector.

737 3.4.2 TMS/ND-GAr

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

741 In Phase I that role will be fulfilled by TMS. It is a magnetised sampling calorimeter,
 742 with alternating steel and plastic scintillator layers. Figure 3.6 shows a schematic view
 743 of the TMS detector. The magnetic field allows a precise measurement of the sign of the

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

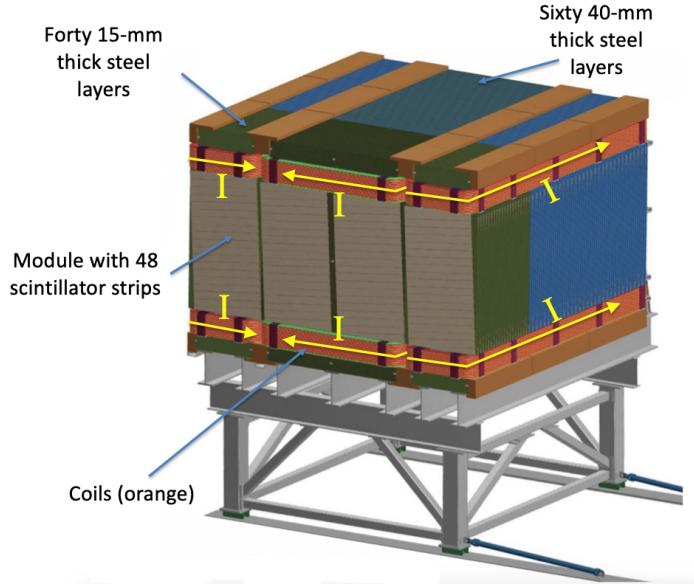


Figure 3.6: Schematic view of the TMS detector, highlighting its main parts. Figure adapted from Ref. [14].

744 muon, so one can distinguish between neutrino and antineutrino interactions.

745 After the Phase II upgrade, TMS will be replaced with a more capable near detector.

746 The current technology considered is ND-GAr. This detector is a magnetised, high-
 747 pressure gaseous argon (GAr) TPC (often denoted as HPgTPC) surrounded by an
 748 electromagnetic calorimeter (ECal) and a muon tagger. A cross section of its geometry
 749 can be seen in Fig. 3.7. ND-GAr will be able to measure the momenta of the outgoing
 750 muons while also detect neutrino interactions inside the GAr volume. This allows
 751 ND-GAr to constrain the systematic uncertainties even further, as it will be able to
 752 accurately measure neutrino interactions at low energies thanks to the lower tracking
 753 thresholds of GAr.

754 3.4.3 PRISM

755 In general, the observed peak neutrino energy of a neutrino beam decreases as the
 756 observation angle with respect to the beam direction increases. This feature has been
 757 used in other long-baseline neutrino experiments, like T2K (2.5° off-axis) and NOvA
 758 (0.8° off-axis), to achieve narrower energy distributions. The DUNE PRISM concept

3.4. NEAR DETECTOR

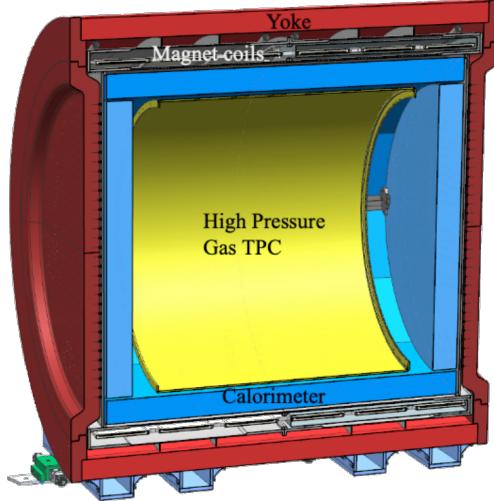


Figure 3.7: Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet. Figure adapted from Ref. [93].

759 exploits this effect using a movable ND. Within PRISM both ND-LAr and the muon
 760 spectrometer (TMS in Phase I and ND-GAr in Phase II) can be moved up to 3.2°
 761 off-axis, equivalent to moving the detectors 30.5 m laterally through the ND hall.

762 This allows us to record additional data samples with different energy compositions.
 763 Figure 3.8 compares the on-axis muon neutrino flux at the ND with the fluxes at different
 764 off-axis positions. As the off-axis position increases the neutrino flux becomes closer to
 765 a monoenergetic beam with a lower peak energy. These samples can be used to perform
 766 a data-driven determination of the relation between true and reconstructed neutrino
 767 energy, to reduce the dependence on the interaction model. The off-axis samples are
 768 linearly combined to produce a narrow Gaussian energy distribution centered on a target
 769 true energy. From the combination coefficients one can build a sample of reconstructed
 770 neutrino events that will determine the energy mapping.

771 The PRISM samples will also be used to form a flux at the ND location similar in
 772 shape to the oscillated flux measured by the FD. This method can be used to extract
 773 the oscillation parameters with minimal input from the neutrino interaction model [83].

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

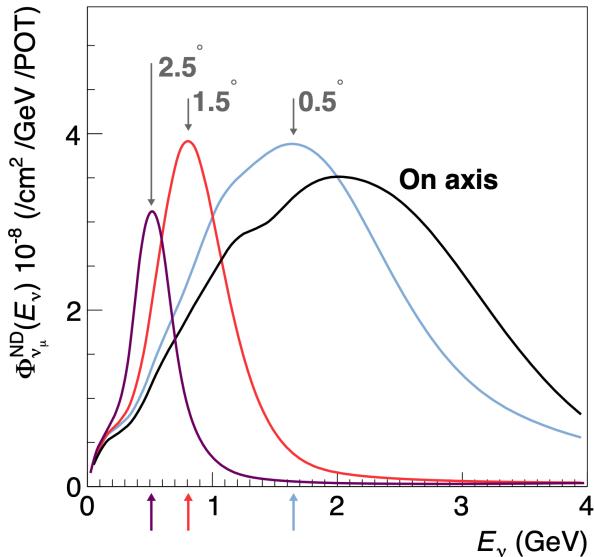


Figure 3.8: Predicted beam muon neutrino flux at the ND location for different off-axis positions. Figure taken from Ref. [91].

3.4.4 SAND

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

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

3.5 A More Capable Near Detector

In DUNE Phase II, a more capable near detector is needed to achieve the ultimate physics goals of the experiment. The current leading proposal for this detector is ND-GAr. As mentioned previously, it will fulfill the role of TMS, measuring the momentum and sign of the charged particles exiting ND-LAr. Additionally, it will be able to measure

3.5. A MORE CAPABLE NEAR DETECTOR

788 neutrino interactions inside the HPgTPC, achieving lower energy thresholds than those
789 of the ND and FD LArTPCs. It will also provide a uniform event acceptance, similar
790 to the FD, which could not be achieved by ND-LAr + TMS. By doing so, ND-GAr
791 will allow to constrain the relevant systematic uncertainties for the LBL analysis even
792 further. A detailed discussion on the requirements, design, performance and physics of
793 ND-GAr can be found in the DUNE ND CDR [91] and the ND-GAr white paper [94].

794 **3.5.1 Requirements**

795 The primary requirement for ND-GAr is to measure the momentum and charge of
796 muons from ν_μ and $\bar{\nu}_\mu$ CC interactions in ND-LAr, in order to measure their energy
797 spectrum. To achieve the sensitivity to the neutrino oscillation parameters described
798 in the DUNE FD TDR Volume II [85], ND-GAr should be able to constrain the muon
799 energy within a 1% uncertainty or better.

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

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

812 **3.5.2 Reference design**

813 The final design of ND-GAr is still under preparation. However, a preliminary baseline
814 design was in place at the time of the ND CDR. This section summarises the main

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

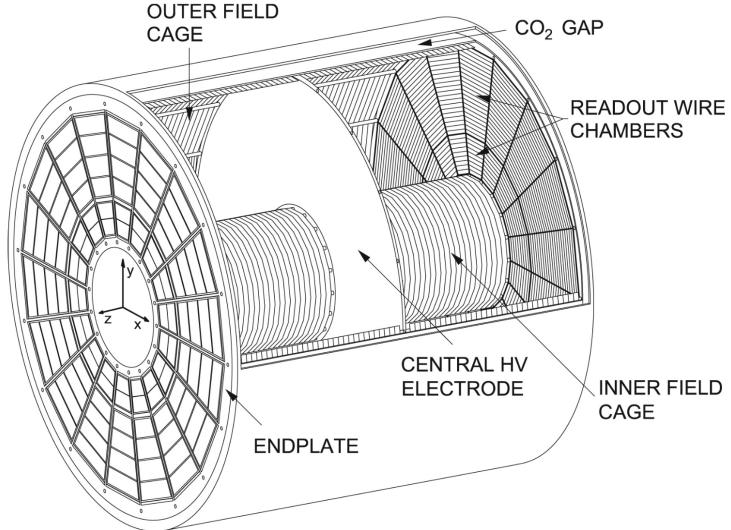


Figure 3.9: Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers. Figure taken from Ref. [91].

815 features of that design, as it is also the one used by default in our simulation. The
 816 different options under consideration for the ND-GAr design are further discussed in the
 817 DUNE Phase II white paper [93].

818 **HPgTPC**

819 The reference design for the ND-GAr HPgTPC follows closely that of the ALICE TPC
 820 [95]. It is a cylinder with a central high-voltage cathode, generating the electric field
 821 for the two drift volumes, with a maximum drift distance of 2.5 m each. The anodes
 822 will be instrumented with charge readout chambers. The original design repurposed
 823 the multi-wire proportional readout chambers (MWPCs) of ALICE. However, some of
 824 the current R&D efforts focus on a gas electron multiplier (GEM) [96] option instead.
 825 Figure 3.9 shows a schematic diagram of the ALICE TPC design. The basic ND-GAr
 826 geometry will resemble this, except for the inner field cage.

827 It will use a 90:10 molar fraction Ar:CH₄ mixture at 10 bar. With this baseline gas
 828 mixture light collection is not possible, as the quenching gas absorbs most of the VUV
 829 photons. Additional R&D efforts are underway, to understand if different mixtures allow
 830 for the light signal to be used to provide a t_0 while maintaining stable charge gain.

3.5. A MORE CAPABLE NEAR DETECTOR

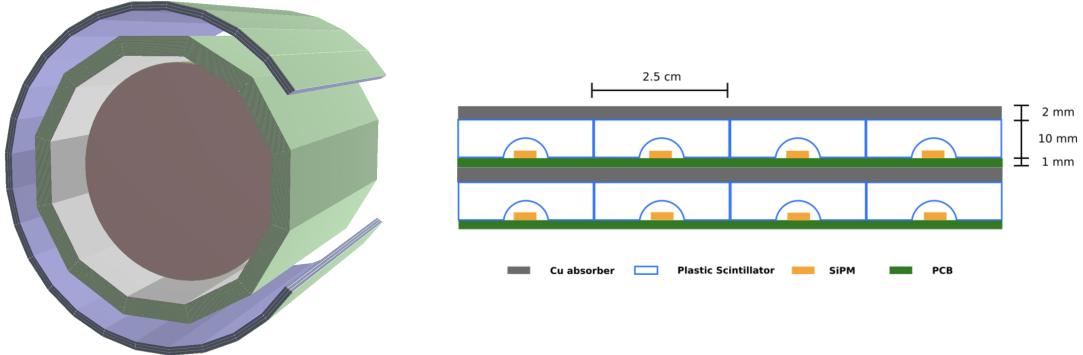


Figure 3.10: View of the 12-sided ECal barrel and outer muon tagger geometries (left) and layout of the ECal tile layers for the 2 mm Cu, 10 mm scintillator option (right). Figure adapted from Ref. [91].

831 ECal

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

838 The ECal design features three independent subdetectors, two end caps at each side
 839 and a barrel surrounding the HPgTPC. Each of the detectors is divided in modules,
 840 which combine alternating layers of plastic scintillator and absorber material readout
 841 by SiPMs. The inner scintillator layers consist of $2.5 \times 2.5 \text{ cm}^2$ high-granularity tiles,
 842 whereas the outer layers are made out of 4 cm wide cross-strips spanning the whole
 843 module length. The current barrel geometry consists of 8 tile layers and 34 strip layers,
 844 while the end caps feature 6 and 36 respectively. The thickness of the scintillator layers
 845 is 7 mm and 5 mm for the Pb absorber layers. The 12-sided geometry of the ECal barrel
 846 (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

847 Magnet

848 The ND-GAr magnet design, know as the Solenoid with Partial Yoke (SPY), consists
849 of two coupled solenoids with an iron return yoke [97]. The idea behind the design is
850 to have a solenoid as thin as possible, as well as a return yoke mass distribution that
851 minimises the material budget between ND-LAr and ND-GAr. The magnet needs to
852 provide a 0.5 T field in the direction perpendicular to the beam, parallel to the drift
853 electric field. It needs to host the pressure vessel and the surrounding ECal, which points
854 to a inner diameter of ~ 6.4 m.

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

861 Muon system

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

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

868 3.5.3 R&D efforts

869 There are several ND-GAr-related prototypes, mostly focused on the TPC charge
870 readout and electronics. The priority is to test the full readout chain, in a high-pressure
871 environment, using a gas mixture with high argon fraction. A detailed summary of these
872 can be found in the DUNE Phase II white paper [93].

3.5. A MORE CAPABLE NEAR DETECTOR

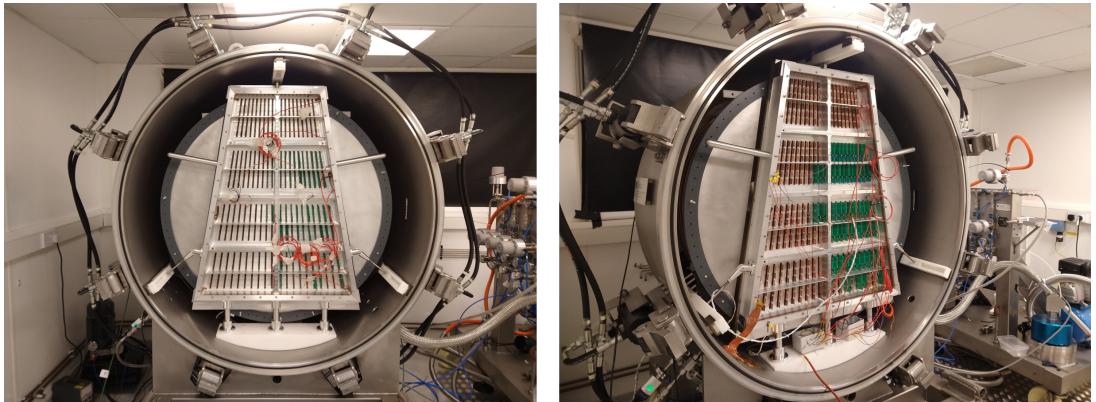


Figure 3.11: Photographs of the TOAD pressure vessel at RHUL. The TPC is mounted inside the vessel, and the OROC is supported by an aluminium frame. Figure taken from Ref. [98].

873 Multi-Wire Proportional Chambers

874 As mentioned before, the original ND-GAr design repurposes the MWPCs of the ALICE
875 TPC, which became available after the recent upgrade [99]. These were operated using
876 a 90:10:5 Ne:CO₂:N₂ gas mixture at 1 atm. Therefore, their performance needed to be
877 studied in an argon gas environment at high pressure.

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

882 The Test stand of an Overpressure Argon Detector (TOAD) tested an ALICE outer
883 MWPC, also known as OROC, up to 5 atm. During its time at RHUL, it was used to
884 study the achievable gas gain of the OROC [98]. At the moment, it is being commissioned
885 at Fermilab for a full detector test of the readout electronics and the DAQ.

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

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

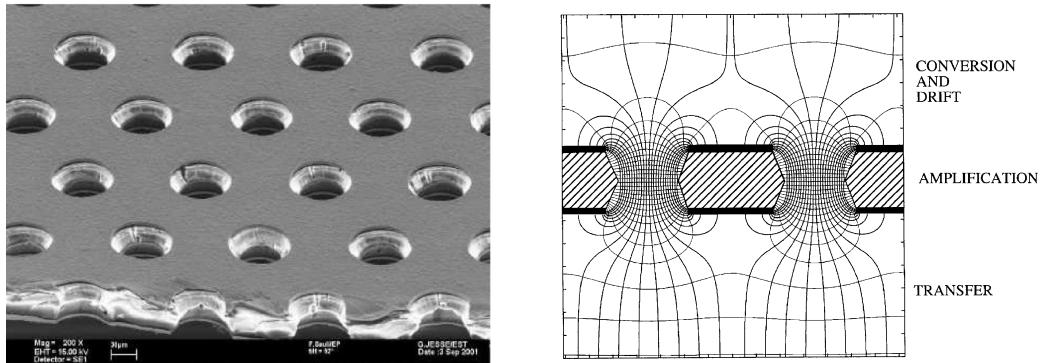


Figure 3.12: Left panel: electron microscope image of a $50\ \mu\text{m}$ thick GEM electrode, with hole pitch and diameter of 140 and $70\ \mu\text{m}$, respectively. Right panel: Schematics of a GEM electrode cross section, showing the electric field lines around the holes. Figures taken from Ref. [100].

889 Gas Electron Multiplier

890 An alternative to the MWPC option is the use of GEMs. These are a type of micro-patter
 891 detector, where the ionisation electrons passing through the holes in the GEM layers
 892 are accelerated by a high intensity electric field. The acceleration causes the electrons
 893 to ionise the medium, resulting in an avalanche which increase the signal exponentially
 894 [96]. GEMs are used in numerous experiments that need a high spatial resolution, like
 895 ALICE [101] and CMS [102] after their upgrades.

896 Figure 3.12 (left panel) shows an electron microscope picture of a $50\ \mu\text{m}$ thick GEM
 897 electrode, with a pitch between neighbouring holes of $140\ \mu\text{m}$ and a hole diameter of
 898 $70\ \mu\text{m}$. A schematic representation of the cross section of a GEM layer is shown in Fig.
 899 3.12 (left panel).

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

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

²Spanish for cat.

³Persian for wolf.

3.6. FAR DETECTOR

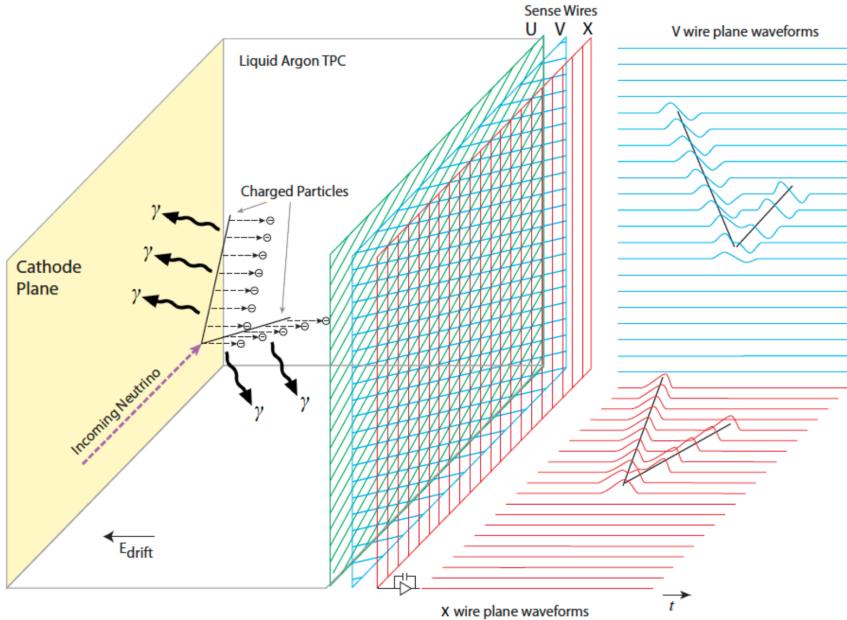


Figure 3.13: Schematic diagram showing the operating principle of a LArTPC with wire readout. Figure taken from Ref. [14].

906 3.6 Far Detector

907 The DUNE FD complex will sit 1300 km away from the beam target and 1.5 km
 908 underground at SURF, South Dakota. Two caverns will host the four FD modules, two
 909 of them per cavern, each embedded in cryostats of dimensions 18.9 m (w) \times 17.8 m (h) \times
 910 65.8 m (l). A central, smaller cavern will host the cryogenic system.

911 Three out of the four modules are confirmed to be LArTPC detectors, with a LAr
 912 fiducial mass of at least 10 kt each. The first and third FD modules, FD-1 and FD-3,
 913 will use a Vertical Drift (VD) technology, whereas the second module, FD-2, will have
 914 a Horizontal Drift (HD) direction. The technology for the fourth module is still to be
 915 decided.

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

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

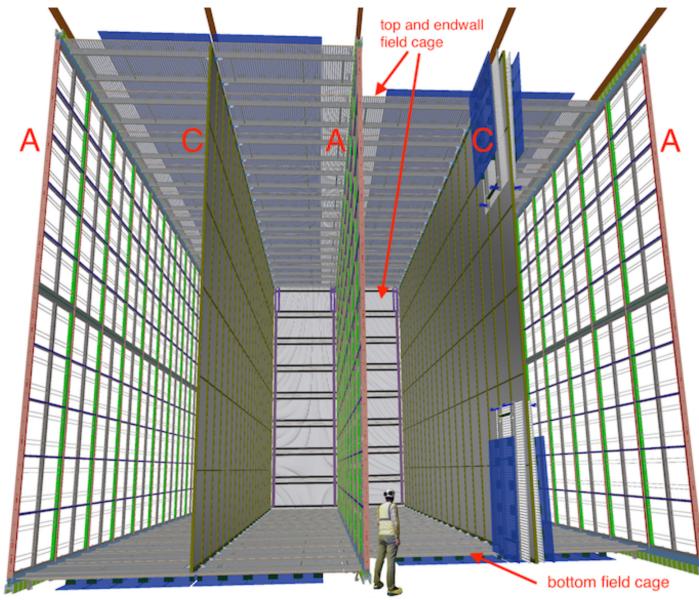


Figure 3.14: Proposed design for the FD-2 module following the HD principle. Figure taken from Ref. [14].

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

926 3.6.1 Horizontal Drift

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

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

3.6. FAR DETECTOR

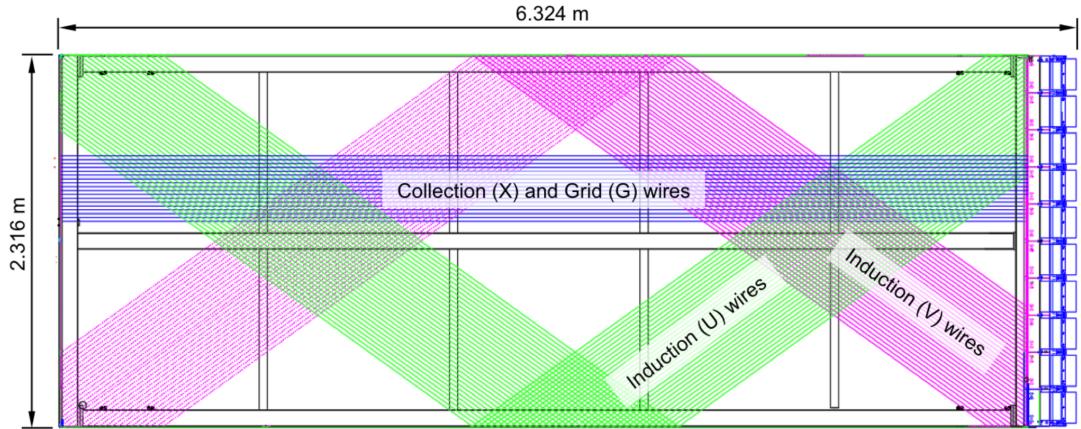


Figure 3.15: Schematic representation of an APA. The black lines represent the APA steel frame. The green and magenta lines correspond to the direction of the U and V induction wires respectively. The blue lines indicate the direction of the X collection wires and the wire shielding G. Figure taken from Ref. [14].

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

939 Each APA is made of 2560 active wires arranged in three layers, plus an extra grid
 940 layer, wrapped around a metal frame. The two induction wire planes, U and V, sit at
 941 $\pm 35.7^\circ$ to the vertical on each side of the APA. The collection and shielding plane wires,
 942 X and G, run parallel to the vertical direction. The ionisation electrons drift past the
 943 induction planes, generating bipolar signals on those wires, and are collected by the
 944 collection plane, producing a monopolar positive signal. The spacing between the wires
 945 is ~ 5 mm, and it defines the spatial resolution of the APA.

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

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

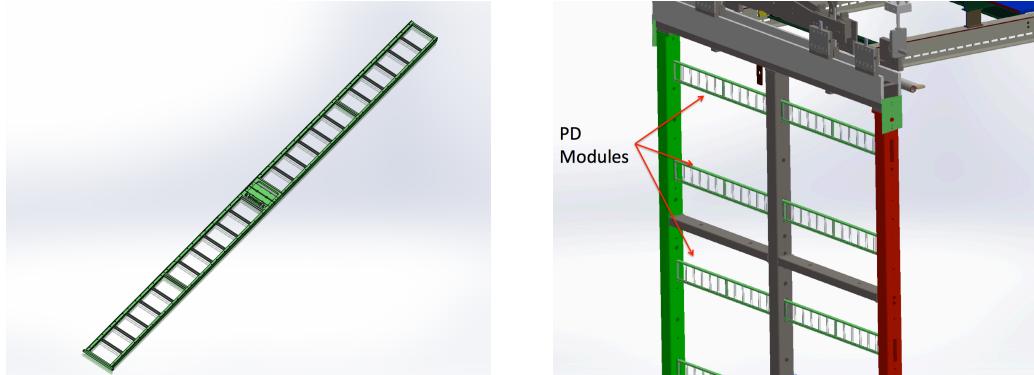


Figure 3.16: A PDS module containing 24 X-ARAPUCAs (left) and the location of the modules on the APAs (right). Figure taken from Ref. [14].

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

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

962 3.6.2 Vertical Drift

963 In the VD case the ionisation electrons will drift vertically until they meet a printed
 964 circuit board-based (PCB) readout plane. It is based on the original dual-phase (DP)
 965 design deployed at CERN, in the detector known as ProtoDUNE-DP, which used a
 966 vertical drift design with an additional amplification of the ionisation electrons using a
 967 GAr layer above the liquid phase. The VD module incorporates the positive features of
 968 the DP design without the complications of having the LAr-GAr interface.

969 The current design of the FD VD module consists of two drift chambers with
 970 a maximum drift distance of 6.5 m. A cathode plane splits the detector volume
 971 perpendicular to the drift direction, while the two anode planes are connected to the

3.6. FAR DETECTOR

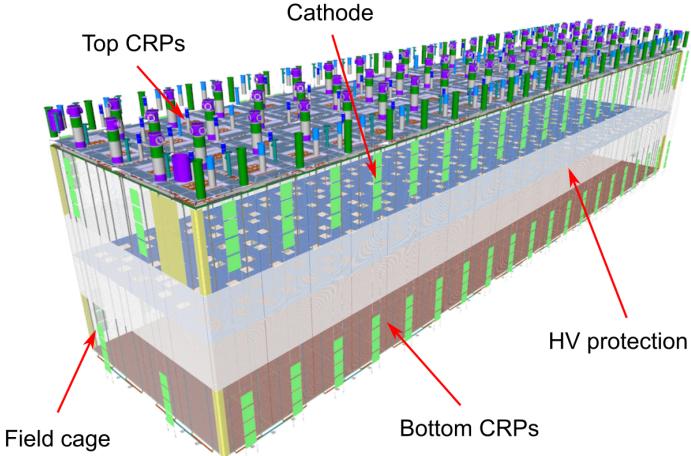


Figure 3.17: Proposed design for the FD-1 and FD-3 modules following the VD principle. Figure adapted from Ref. [103].

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

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

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

The PDS in the VD module will use the same X-ARAPUCA technology developed 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

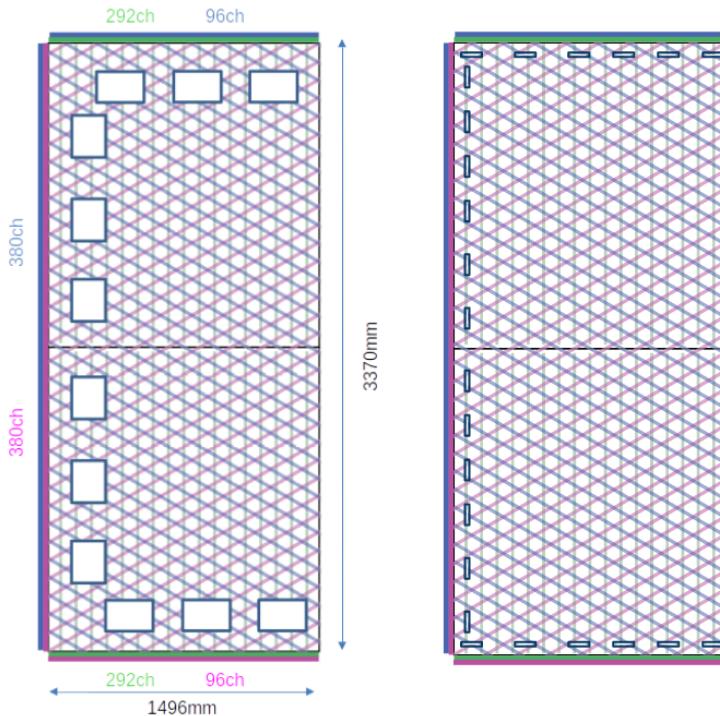


Figure 3.18: Schematic representation of the electrode strip configuration for a top (left) and bottom (right) CRU. Figure taken from Ref. [103].

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

990 3.6.3 FD Data Acquisition System

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

995 The enormous sample rate and the number of channels in TPC and PD readouts
 996 will produce a very large volume of data. These pose really strong requirements and
 997 challenges to the DUNE FD DAQ architecture. It will be required to read out data of
 998 the order of ten thousand or more channels at rates of a few MHz. To cope with the
 999 huge data volume, segmented readouts and compression algorithms are used to reduce
 1000 the data rate to manageable levels.

1001 The DAQ system of the DUNE FD is composed of five different subsystems. The

3.6. FAR DETECTOR

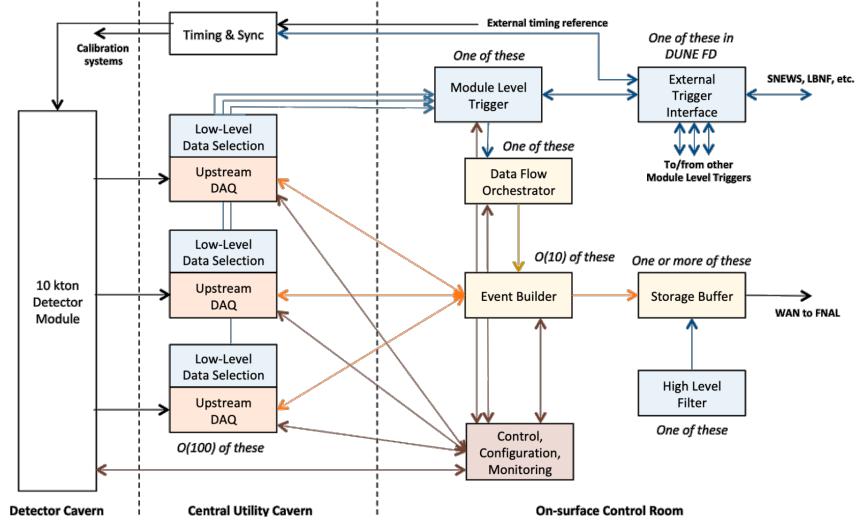


Figure 3.19: Detailed diagram of the DUNE FD DAQ system. Figure taken from Ref. [104].

1002 first one is the upstream DAQ, which receives the raw data from the detector, buffers it
 1003 and performs some low-level pre-processing. The minimally processed data is then fed
 1004 into a hierarchical data selection system, which then performs a module level trigger
 1005 decision. In case of a positive decision, a trigger command is produced and executed by
 1006 the data flow orchestrator, located in the back-end (BE) DAQ subsystem. Subsequently
 1007 the DAQ BE retrieves the relevant data from the buffers located in the upstream DAQ,
 1008 adds all the data into a cohesive record and saves it to permanent storage. Watching
 1009 over all the other subsystems we also have the control, configuration and monitoring
 1010 subsystem and the time and synchronization subsystem. Figure 3.19 shows a schematic
 1011 diagram of the DAQ system, showing the different subsystems and their relations.

1012 A notorious challenge for the DUNE DAQ system comes from its broad physics goals.
 1013 We must be prepared to process events spanning a wide range of time windows from
 1014 5 ms in the case of beam and cosmic neutrinos and nucleon decay events, to 100 s in the
 1015 case of SNBs. This requires a continuous readout of the detector modules. Moreover,
 1016 because of the off-beam measurements, we need to ensure the capabilities of online data
 1017 processing and self-triggering. Having this into account, together with the technical
 1018 constraints, the DUNE FD DAQ faces a series of challenges: it needs to be fault tolerant

CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

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

1023

1024

Matched Filter approach to Trigger

1025

Primitives

1026

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

1028

– Arthur Conan Doyle, *A scandal in Bohemia*

1029

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

1033

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

1037

4.1 Motivation

1038

The lowest-level objects that are formed within the DUNE FD DAQ system are the so-called trigger primitives (TPs) [105]. These represent the hits on a channel, and are used as input to the rest of the DAQ trigger chain. The TPs are formed in the hit finder chain. A schematic representation of it is shown in Fig. 4.1. This chain takes the raw ADC data from the detector, removes the constant pedestal of the signal using a

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES



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

1043 dynamical median estimation method, applies a filter to the waveform, and tries to find
 1044 peaks over a certain threshold. These peaks form the TPs, which contain information
 1045 such as the start and end times over the threshold, the maximum ADC value and the
 1046 corresponding ADC integral. Currently, there are two implementations of the hit finder
 1047 chain, one firmware-based and other software-based.

1048 The filter implemented in the firmware of the upstream DUNE FD DAQ is a 32nd-
 1049 order low-pass finite impulse-response (FIR) filter. The output of such filter for a discrete
 1050 system can be written as:

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

1051 where N is the order of the filter, y is the output sequence, x is the input sequence and
 1052 h is the set of filter coefficients. The current implementation within `dtp-firmware` [106]
 1053 uses a set of 16 non-zero integer coefficients. For the software case, only a 5th-order
 1054 filter is used, as the filtering is the most CPU-expensive part of the software hit finder.

1055 Filtering is a vital step in the hit finder chain. It helps suppressing the noise and
 1056 enhances the signal peaks with respect to the noiseless baseline. A good filtering strategy
 1057 allows us to use lower thresholds when forming the TPs, thus increasing the sensitivity
 1058 of our detector to low energy physics events. In such events, the hits produced by the
 1059 ionisation electrons tend to have lower amplitudes than those of interest to the LBL
 1060 physics programme of the DUNE experiment.

1061 This is particularly important for the induction planes. In general, signal peaks in
 1062 the induction channels have smaller amplitude than the ones in the collection plane.
 1063 This, together with the fact that the pulse shapes are bipolar, reduces our capacity to

4.2. SIGNAL-TO-NOISE RATIO DEFINITION

1064 detect the hits on these channels. The inefficiency of detecting TPs in the induction
1065 planes (denoted as U and V planes) leads trigger algorithms to focus mainly on the
1066 TPs from the collection plane (so-called X plane). As a result, the possibility of making
1067 trigger decisions based on the coincidence of TPs across the three wire planes remains
1068 nowadays unexploited in DUNE. This will be beneficial for low energy events, as it
1069 adds redundancy to the algorithms, as well as for other physics that requires online
1070 directionality information, like the supernova pointing.

1071 A possible improvement of the current hit finder chain may require optimising the
1072 existing or choosing a new filter implementation. A filter strategy which benefits the
1073 induction signals may be able to enhance the detection efficiency of TPs from the
1074 induction planes and ideally make it comparable to that of the collection plane.

1075 The goal is to implement a better finite-impulse response filter and to evaluate its
1076 performance relative to the current filter. To do so, I need to take into account the
1077 limitations of the firmware: the FIR filter shall have maximum 32 coefficients (so-called
1078 taps) whose values are 12-bit unsigned integers. Although it is technically possible to
1079 include non-integer coefficients, it would be a technical challenge. For instance, in the
1080 HD design there are 40 FIR instances per APA, as there are 4 FIR blocks per optical
1081 link and 10 optical links per APA. Therefore, the impact of increasing the complexity of
1082 the filter will be amplified forty times in the FPGA load. With these restrictions, the
1083 task is to provide a set of 32 coefficients which yield an optimal filter performance for the
1084 induction channels. A solution compatible with the software hit finder implementation
1085 is not considered, due to its current limitations concerning the filtering stage.

1086 4.2 Signal-to-noise ratio definition

1087 In the following, I use the signal to noise ratio (S/N) as a measure of the FIR filter
1088 performance. The S/N metrics allow us to compare different filter implementations
1089 and serve as a basis for more detailed studies presented later in this Chapter. Here,
1090 I demonstrate how to extract its value for a set of ProtoDUNE-SP data. Specifically,

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

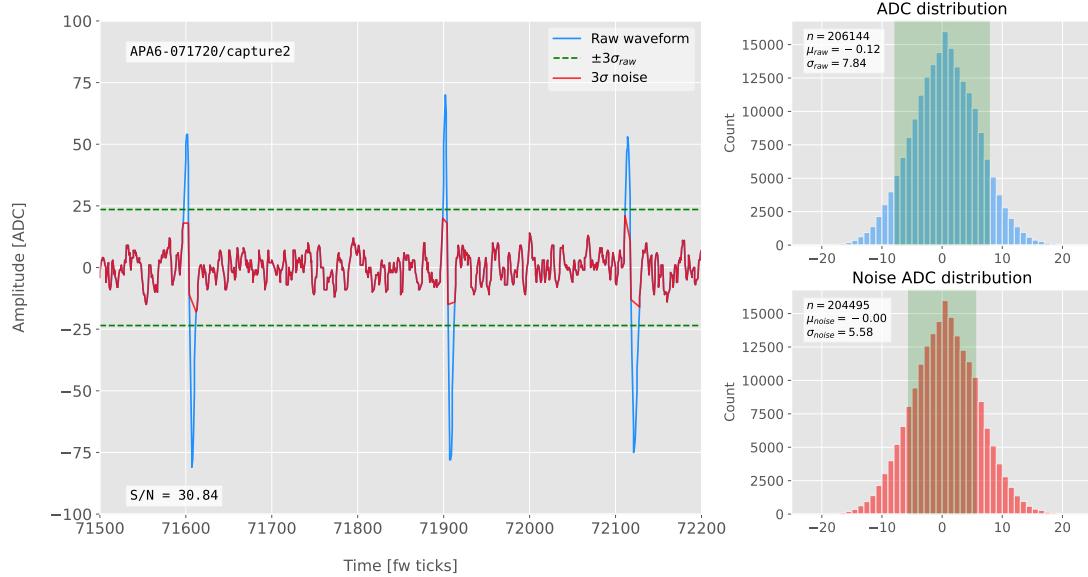


Figure 4.2: Left panel: Zoomed unfiltered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The green dashed lines mark the region $\pm 3\sigma_{raw}$. The resulting noise waveform is also shown (red line). Top right panel: ADC distribution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

1091 I use the ADC capture `felix-2020-07-17-21:31:44`, a raw data capture taken for
 1092 firmware validation purposes. I define the S/N of a channel as the height of the signal
 1093 peaks relative to the size of the noise. To quantify this, I first estimate the standard
 1094 deviation of the ADC data for each channel, σ_{ADC} . Then, I define the corresponding
 1095 noise waveform to be the ADC values in the range $\pm 3 \sigma_{ADC}$. From this new noise data
 1096 I compute the mean and standard deviation, μ_{noise} and σ_{noise} , so I can write the S/N
 1097 for any given channel as:

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

1098 where $\max [ADC]$ is simply the maximum ADC value found in the corresponding channel.
 1099 As an example, I apply this definition of the S/N to a waveform from one of the
 1100 channels of the data capture. Figure 4.2 shows a zoomed region of the waveform
 1101 corresponding to channel 7840 (blue line), where one can clearly see three signal peaks

4.3. LOW-PASS FIR FILTER DESIGN

and continuous additive noise¹. I estimated the standard deviation of this raw waveform to be $\sigma_{raw} = 7.84$ ADC, and from this I am able to define the noise waveform (red line) as the ADC values in the range ± 23.52 ADC. This way, I obtain $\mu_{noise} = 0.01$ ADC and $\sigma_{noise} = 5.58$ ADC, which gives S/N = 30.84.

I repeat this calculation now for the corresponding filtered waveform, using the current firmware FIR filter. Figure 4.3 shows 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 noise waveform (red line) is formed by selecting the ADC values in the ± 32.91 ADC range, which gives $\mu_{noise} = -0.47$ 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

To optimise the frequency response of a digital filter, we can use the Parks-McClellan algorithm, where one finds a set of N real coefficients that give the best response for the specified pass-band and order of the filter [107].

Taking the detector ticks as the time unit, the Nyquist frequency will simply be $1/2$ ticks⁻¹. The current implementation of the filter seems to have as pass-band the range $[0, 0.1]$ ticks⁻¹. This can be seen in Fig. 4.4, where I show the power spectrum, in decibels, of that filter implementation (blue solid line). The Park-McClellan algorithm finds the optimal Chebyshev FIR filter [108] taking as input the boundaries of the target pass-band and stop-band, which can be written in the form:

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

¹There are actually 6 peaks, 3 positive and 3 negative, but, because by design for induction channels the expected signal pulse shapes are bipolar, we treat them as a collection of 3 individual signals.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

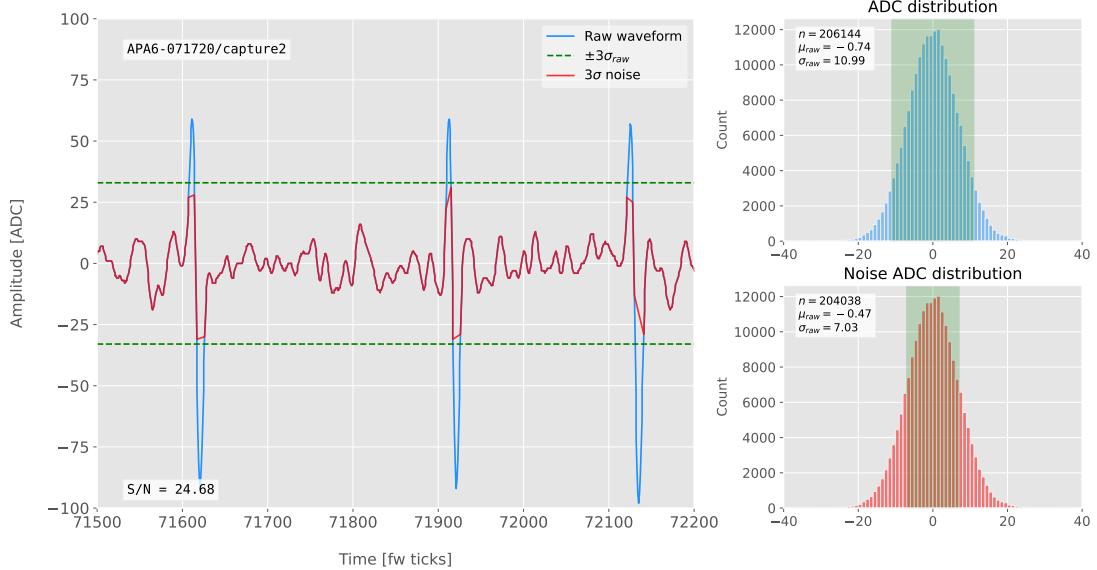


Figure 4.3: Left panel: Zoomed filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The filter used was the current implementation of the low-pass FIR filter in `dtp-firmware`. The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$. 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 $\pm \sigma_{\text{noise}}$.

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

At this point, I tried to improve the performance of the FIR filter using the Parks-McClellan method, i.e. maximise the overall S/N, using the available data captures. I did so by varying the values of the two quantities that parametrise the pass-band and stop-band, the cut-off frequency f_c and the transition width δf .

4.3. LOW-PASS FIR FILTER DESIGN

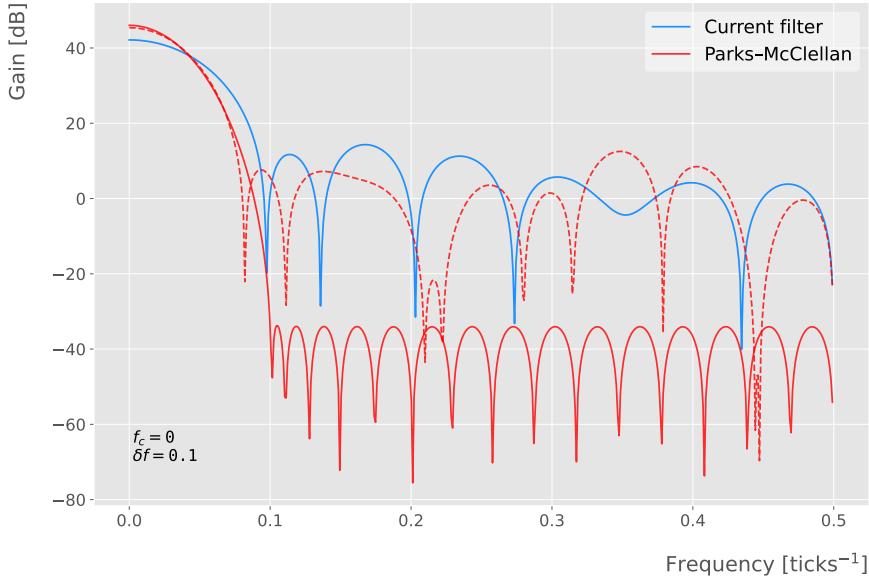


Figure 4.4: Power spectrum in decibels for the current implementation of the low-pass FIR filter in `dtp-firmware` (blue line), compared to the response of an optimal filter obtained using the Parks-McClellan algorithm for the same pass-band (red line). Also for comparison I include the spectrum of the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

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

As it can be seen in Fig. 4.5 (bottom right panel) the configuration which gives the best mean performance for the three planes is $f_c = 0.068$ ticks⁻¹ and $\delta f = 0.010$ ticks⁻¹. We can use these to see how the filter affects the different channels. Figure 4.6 shows the distribution of the S/N improvement values for all the channels in the raw ADC capture `felix-2020-07-17-21:31:44`, separated by wire plane, after the optimal Chebyshev filter was applied. One can see that there is a clear improvement for both U and V

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

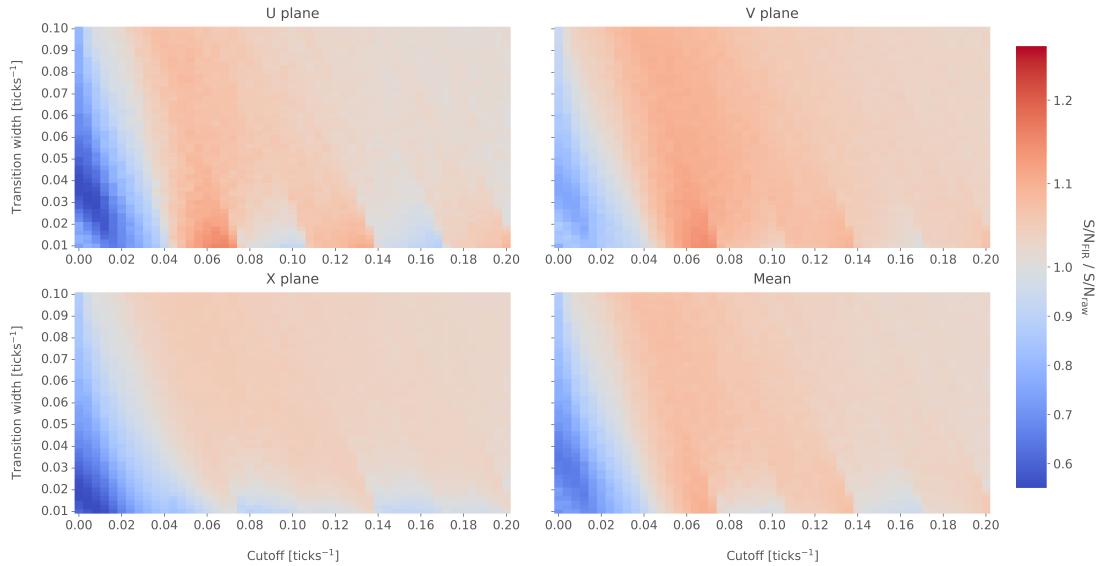


Figure 4.5: Relative change in the S/N for the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44`, using different values of the cutoff frequency f_c and the transition width δf . The optimal Chebyshev filters were applied using just the integer part of the coefficients given by the Parks-McClellan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

induction planes, obtaining a mean change of 1.25 and 1.30 for them, respectively. However, in the case of the X collection plane the distribution peaks around 1, meaning that an important fraction of channels in that plane get a slightly worse S/N after the filter is applied. This is not a big issue, as the S/N for collection channels is usually much higher than the one for induction channels.

The results I obtained optimising the low pass filter with the Parks-McClellan method are promising. Nonetheless, the improvement found is rather marginal. Thus, I explored alternative approaches to the filtering problem, which may yield better outputs. This way, I found a possible solution in matched filters. By construction, this kind of filters offer the best improvement on the S/N.

4.4 Matched filters

In the context of signal processing, a matched filter is the optimal linear filter for maximising the S/N in the presence of additive noise. It is obtained by convolving a conjugated time-reversed known template with an unknown signal to detect the presence

4.4. MATCHED FILTERS

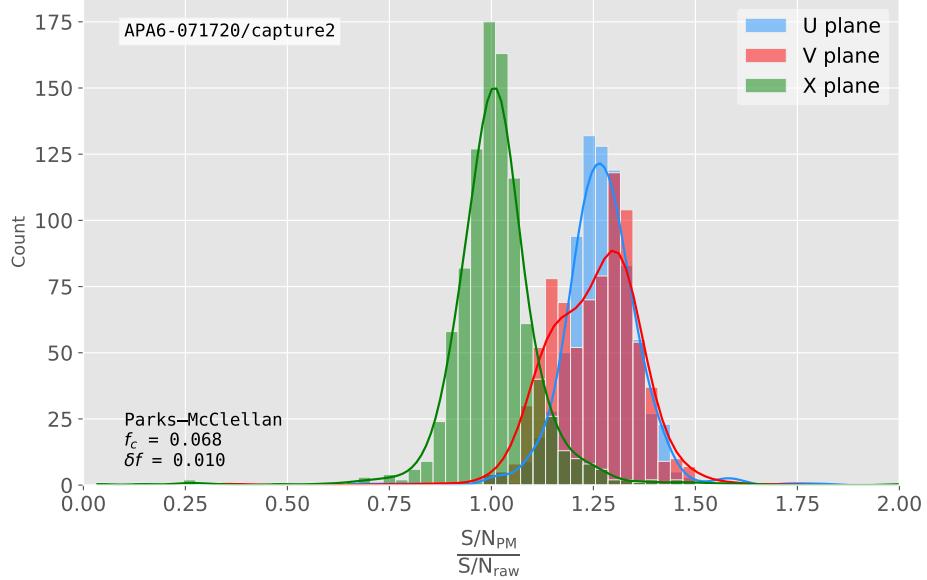


Figure 4.6: Distribution of the relative change of the S/N on the different wire planes from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` after the optimal Chebyshev filter was applied. The filter was computed with the Parks-McClellan algorithm using a cutoff of $f_c = 0.068$ ticks $^{-1}$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

1166 of the template in the signal [109].

1167 Given a known signal sequence $s(t)$ and another (a priori unknown) noise sequence
1168 $n(t)$, the input signal can be written as:

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

1169 Now, considering a linear time-invariant filter, whose impulse-response function I
1170 will refer to as $h(t)$, one can write the output signal as:

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

1171 where $y_s(t)$ and $y_n(t)$ are simply the outputs of the filter due to the signal and the noise
1172 components respectively.

1173 The goal of the matched filter is to detect the presence of the signal $s(t)$ in the input

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

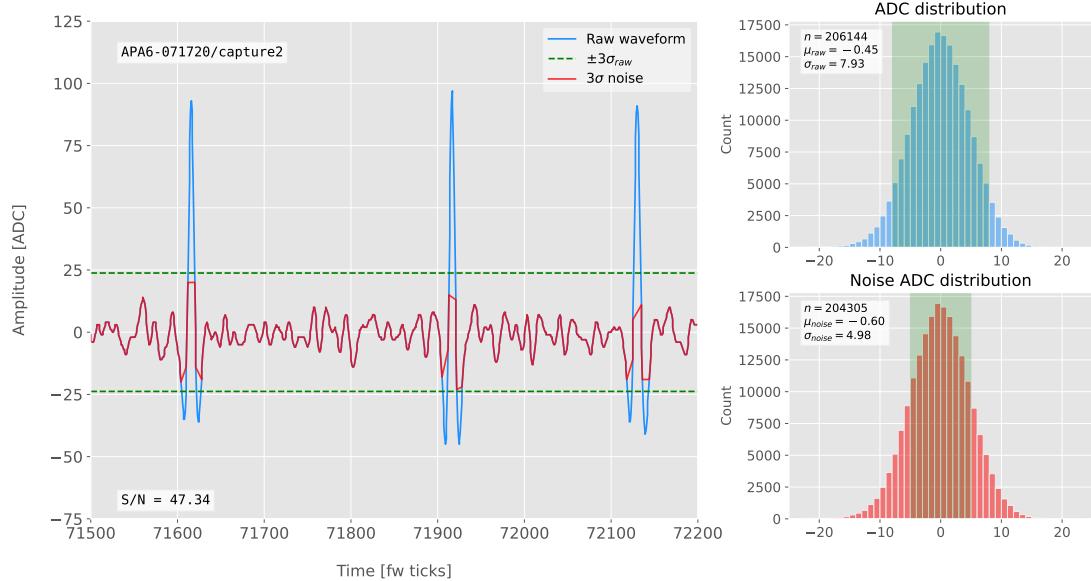


Figure 4.7: Left panel: Zoomed match filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` (blue line). The filter used was directly extracted from the data, being the 32 values around the first peak in the original waveform. The green dashed 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 $\pm \sigma_{\text{noise}}$

sample $x(t)$ at a certain time t_0 , which effectively means that we need to maximise the S/N at that given time. This way, what one wants is to have a filter which gives a much bigger output when the known signal is present than when it is not. Putting it in other words, the instantaneous power of the signal output $y_s(t)$ should be much larger than the average power of the noise output $y_n(t)$ at some time t_0 .

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

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

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

Now, focusing on the noise part, we can use the Wiener-Khinchin theorem [110] to

4.4. MATCHED FILTERS

1184 write the mean power of the noise after filtering as:

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

1185 where $S_n(\omega)$ is the power spectral density of the noise.

1186 Having these, one can write the instantaneous S/N at time t_0 as:

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

1187 Once we have this expression, we need to find its upper limit to determine what would
1188 be the optimal choice for the transfer function. For this, we use the Cauchy-Schwarz
1189 inequality, which in the present case takes the form:

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

1190 for any two analytical functions $f(x)$ and $g(x)$. One can prove that making the choice:

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

1191 leads to the following upper bound for the S/N:

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

1192 From Eqs. (4.8), (4.9) and (4.10) one can also derive the form of the transfer function
1193 such that the upper bound is exactly reached [111]:

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

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1194 From this last expression we can clearly see the way the matched filter acts. As the
 1195 transfer function is proportional to the Fourier transform of the signal it will try to only
 1196 pick the frequencies present in the signal [112].

1197 The matched filter transfer function can be greatly simplified if the input noise is
 1198 Gaussian. In that case, the power spectral density of the noise is a constant, so it can be
 1199 re-absorbed in the overall normalisation of the transfer function. Moreover, considering
 1200 that the input signal is a real function, one can simply set $S^*(\omega) = S(-\omega)$, which gives:

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

1201 For a discrete signal, one can think of the input and impulse-response sequences
 1202 as vectors. Then, the matched filter tries to maximise the inner product of the signal
 1203 and the filter while minimising the output due to the noise by choosing a filter vector
 1204 orthogonal to the latter. In the case of additive noise, that leads to the impulse-response
 1205 vector:

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

1206 where s is a reversed signal template sequence of length N equal to the order of the filter
 1207 and R_n is the covariance matrix associated with the noise sequence n . For the Gaussian
 1208 noise case, the covariance matrix is simply the unit matrix, so the above expression
 1209 simplifies again to:

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

1210 To test whether this choice of filter is appropriate one needs to choose a signal
 1211 template. As an example of how a matched filter would affect our signal, I simply took
 1212 the matched filter coefficients to be the 32 ADC values around a signal peak present in
 1213 the data. In Fig. 4.7 (left panel) I plotted a zoomed region for channel 7840 in the raw
 1214 data capture `felix-2020-07-17-21:31:44`, after applying the matched filter described
 1215 before (blue line). When compared to the raw and FIR filtered case (see Figs. 4.2 and
 1216 4.3), after applying the matched filter the standard deviation of the noise waveform (red

4.4. MATCHED FILTERS

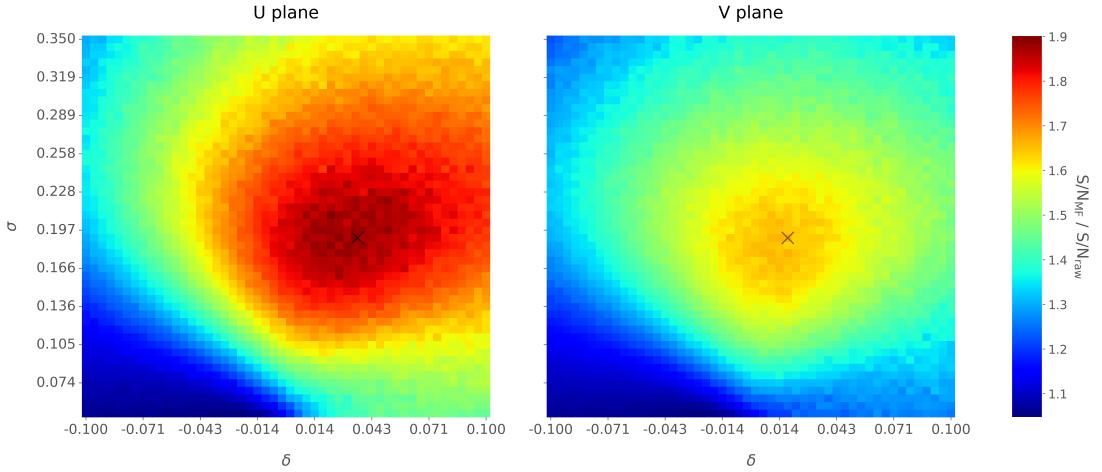


Figure 4.8: Relative change in the S/N for the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44` for different values of δ and σ from the matched filter parametrisation in Eq. (4.16). The black crosses in both panels denote the location of the maximum ratio value.

line) decreases and at the same time the signal peaks are enhanced. This leads to an improvement of the S/N by a factor of 1.92 when compared to the raw waveform.

To obtain the matched filter that is more suitable for our data, I explored different configurations of signal templates. I parametrised the signal using the bipolar function:

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

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

As this parametrisation is only adequate for bipolar signals I will focus exclusively on the induction channels. Also, to achieve the best possible performance, I optimise the coefficients for the U and V planes separately. However, as I will discuss, the differences are not very pronounced. In case it is not technically possible to separate channels in the firmware according to the plane they are coming from and use different sets of filter coefficients for them, we can just find a common set of coefficients. In such case, I do not expect the results to change drastically.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1231 Figure 4.8 presents the results of the parameter scan, for channels in the induction
1232 planes U (left panel) and V (right panel). For each configuration of σ and δ the resulting
1233 matched filter was applied to all channels in the corresponding plane within the data
1234 capture `felix-2020-07-17-21:31:44`. The change in S/N is computed with respect to
1235 the raw waveforms, and then the mean value for all channels is kept as a score for each
1236 filter. One can see that the improvements obtained for the U plane are in general higher
1237 than the ones for the V plane. However, these ratios are substantially higher than the
1238 ones obtained for the low-pass FIR filters. For the optimal configurations, I attained
1239 improvements up to a factor of 1.85 for the U plane and 1.65 for the V plane.

1240 The sets of optimal matched filter coefficients were obtained for the parameters
1241 $\delta = 0.035$, $\sigma = 0.191$ for the U plane and $\delta = 0.018$, $\sigma = 0.191$ for the V plane. I
1242 show these two sets of coefficients in Fig. 4.9 (left panel). Figure 4.9 (right panel)
1243 shows the distribution of the S/N improvement after the optimal match filters for
1244 the U and V were applied to the corresponding channels in the raw data capture
1245 `felix-2020-07-17-21:31:44`. As mentioned before, the mean improvement achieved
1246 for the U plane channels is slightly higher than the one for the V channels. Note, however,
1247 that the spread of the distribution for the V plane is smaller than the one for the U
1248 plane.

1249 Overall, one can see that the improvements on the S/N are much more significant in
1250 the case of the matched filter than they were for the low-pass FIR filters. The analysis
1251 of the raw data captures from ProtoDUNE-SP suggests that matched filters increase the
1252 S/N of induction channels by a factor of 1.5 more than the optimal low-pass FIR filters.

1253 Although these results are by themselves great points in favour of the matched
1254 filter, more studies are needed to completely assess the robustness of this approach. I
1255 proceeded then to test the matched filter with simulated data samples.

4.5. MONTE CARLO STUDIES

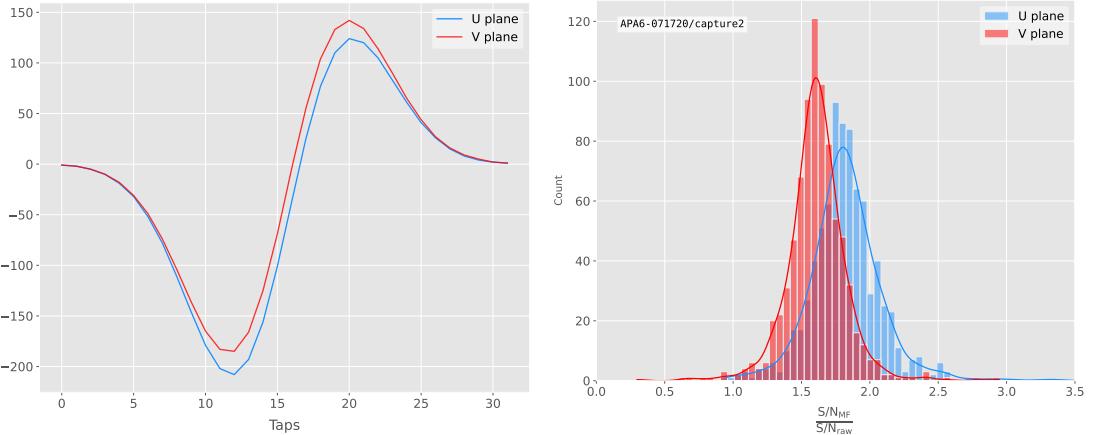


Figure 4.9: Left panel: Optimal matched filter coefficients for the U (blue line) and V (red line) planes. The filters were computed with our parametrisation in Eq. (4.16) for the parameter values $\delta = 0.035$, $\sigma = 0.191$ and $\delta = 0.018$, $\sigma = 0.191$ respectively. Right panel: Distribution 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.

1256 4.5 Monte Carlo studies

1257 To further test the matched filter, the next step is to generate and process data samples
 1258 using LArSoft [113], the simulation and reconstruction software of the DUNE FD. In this
 1259 way, one can control the particle content of the samples, the orientation of the tracks
 1260 and their energy, and therefore see how the matched filter behaves in various situations.

1261 To begin with, I prepared different monoenergetic and isotropic samples containing
 1262 a single particle per event. Each sample contains a different particle species, namely
 1263 electrons, muons, protons and neutral pions, all with a kinetic energy of $E_k = 100$ MeV.
 1264 I chose these because of the fairly different topologies they generate in the liquid argon,
 1265 ranging from shower-like to track-like.

1266 The event were generated with the single particle gun, and the Geant4 stage of the
 1267 LArSoft simulation [113] was performed with the standard configuration for the DUNE
 1268 FD HD design.

1269 For simplicity, I restricted the particles to start drifting in a single TPC volume²,

²A TPC volume is defined as the drift region between a single APA and the cathode. Therefore, for one drift volume of a HD module, there are twice as many TPC volumes as there are APAs in the

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

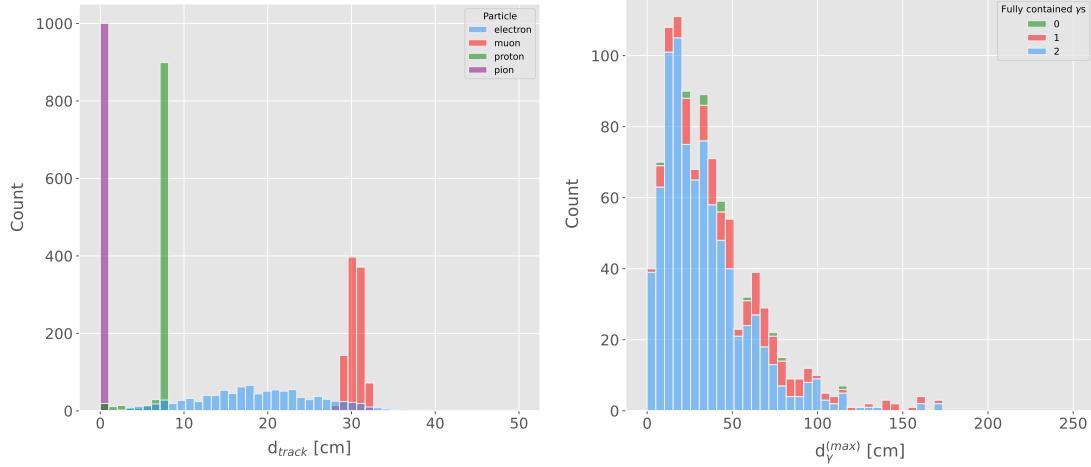


Figure 4.10: Left panel: distributions of the particles track length in the liquid argon for the generated $E_k = 100$ MeV monoenergetic samples, electrons (blue), muons (red), protons (green) and neutral pions (purple). Right panel: distribution of the length of the longest photon in the neutral pion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

so I can focus exclusively on the signals coming from one APA. The chosen kinetic energy for all the particles is $E_k = 100$ MeV, as this produce tracks which are typically contained in one TPC volume. Figure 4.10 (left panel) shows the distributions of the track lengths in the liquid argon of all generated particles with $E_k = 100$ MeV. One can see that, in the case of the track-like particles (i.e. muons and protons), their length distributions are quite sharp and centered at relatively low distances (30 and 8 cm, respectively). For electrons, the distribution is quite broad but it does not extend past ~ 30 cm. The case of the neutral pions can be misleading. As they decay promptly, the track length associated to the true MC particle is always < 1 cm. In Fig. 4.10 (right panel) I show the effective length distribution of the longest photon after the pion decays as $\pi^0 \rightarrow \gamma\gamma$, highlighting the number of fully contained photons in the TPC volume per event (either zero, one, or two). One can see that the vast majority of events have both photons contained in the TPC volume, whereas just a negligible fraction of them have none. However, for the sake of caution, I keep only the pion events with both photons contained.

The next step is to process the sample through the detector simulation. To make corresponding anode.

4.5. MONTE CARLO STUDIES

1286 adequate estimations of the noise levels, one needs to turn off the default zero-suppression
1287 of the waveforms produced by the simulation. At this stage I am only interested int
1288 the waveforms with noise added, so I keep the noise addition option as true in the
1289 configuration. However, for studies related to the hit finder performance one also needs
1290 to store the noiseless waveforms, to retrieve the truth information of the hits. I will
1291 discuss this approach next.

1292 To reduce the amount of data that will go for processing, I used the information from
1293 the Geant4 step of the simulation to select only the active channels, i.e. the channels
1294 where some ionisation electrons arrived. Moreover, I only extract the waveforms from
1295 one APA and exclusively the ones coming from induction channels. The resulting **ROOT**
1296 file contains a **TTree** with two branches, one containing the waveforms for each event
1297 and channel and the other with the corresponding offline channel numbers.

1298 Finally, I extract the truth values for the orientation of the tracks and the energies
1299 of the particles to use them in the analysis. These are stored in a **ROOT** file with a single
1300 **TTree**, containing several branches with information such as the components of the
1301 initial momentum of the particles, initial and final positions, track length, etc.

1302 For the analysis of the resulting waveforms and truth values I used a custom analysis
1303 code independent of LArSoft. Among other functionality, it allows the user to read the
1304 **ROOT** files, export the raw data as **pandas** objects, apply the filters and compute the
1305 S/N of both the raw and filtered signals. The default configuration for the filtering uses
1306 the set of optimal matched filter coefficients that I found using the ProtoDUNE-SP data
1307 samples.

1308 Additionally, for the analysis of the samples it was necessary to use two different
1309 reference frames, to study separately the signals coming from the U and V induction
1310 planes. Focussing on a single APA, the U and V channels have a different orientation in
1311 the yz plane. In the case of U channels, these are tilted 35.7° clockwise from the vertical
1312 (y direction), whereas the V channels are at the same angle but in the counter-clockwise
1313 direction. Because of this, the best option is to deal with two new coordinate systems
1314 rotated by $\pm 35.7^\circ$ along the x axis, so the new y' and y'' directions are aligned with the

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

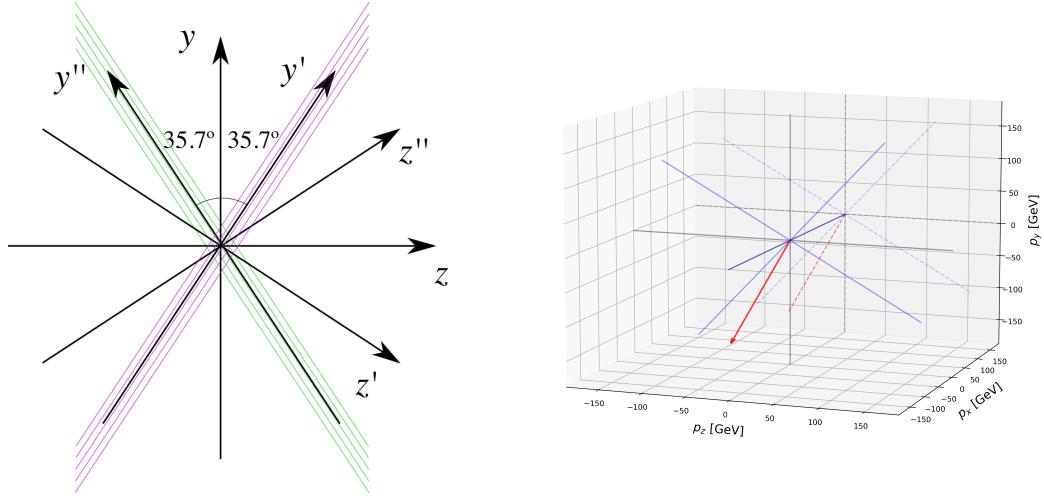


Figure 4.11: Left panel: schematic representation of the two new rotated reference frames used in this analysis (denoted as prime and double prime), viewed from the yz plane. The magenta stack of lines represent the wires in the U plane, whereas the green lines correspond to the wires in the V plane. 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.

1315 U and V induction channels, respectively. Figure 4.11 (left panel) shows a schematic
 1316 representation of the original reference frame together with the two rotated ones (denoted
 1317 by primed and double primed). This way, one can easily understand how parallel was a
 1318 track to the channels in the two induction planes. Figure 4.11 (right panel) shows a 3D
 1319 representation of the momentum of a track (red arrow) in the original reference frame
 1320 (black lines), along with the new reference frame for the U plane (blue lines). I added
 1321 the projections onto the yz plane of these, to show the usefulness of the new reference
 1322 frame to tell whether a track is parallel or perpendicular to the channels in a induction
 1323 plane.

1324 Figure 4.12 shows the distribution of the average S/N change per event when I apply
 1325 the optimised matched filters. I produce separate distributions for the channels in the U
 1326 (red) and V (blue) induction planes. Notice that the S/N distributions for the track-like
 1327 particles, i.e. muons (top left panel) and protons (bottom left panel), have significantly
 1328 larger mean values than the distributions of the shower like particles, i.e. electrons (top

4.5. MONTE CARLO STUDIES

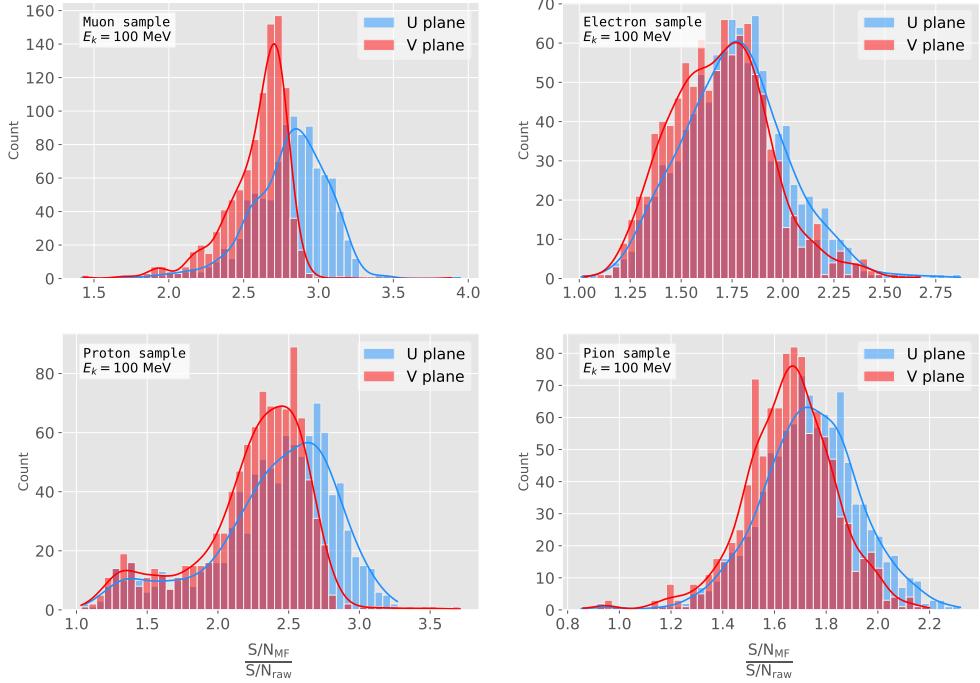


Figure 4.12: Distributions of the mean S/N change per event for the different MC samples after applying the matched filters. Here I separated the change in the U plane (blue) and the V plane (red) channels. From top left to the right: muon, electron, proton and neutral pion. All the events have a fixed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

right panel) and neutral pions (bottom right panel). An important difference between these results and the ones obtained before for the ProtoDUNE-SP data is that the overall improvements that I get with simulated data are more significant. This could be due to an underestimation of the noise levels in the LArSoft simulation. 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 follow to present the results results, in the case of the raw and filtered S/N of each event I take the average of these quantities over all the active channels in the event. That is, if a certain event has N_{chan} active channels the two S/N values are computed as:

$$(S/N_{fir})_{event} = \frac{\sum_{i=0}^{N_{chan}} (S/N_{fir})_i}{N_{chan}}, \quad (4.17)$$

$$(S/N_{raw})_{event} = \frac{\sum_{i=0}^{N_{chan}} (S/N_{raw})_i}{N_{chan}}.$$

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

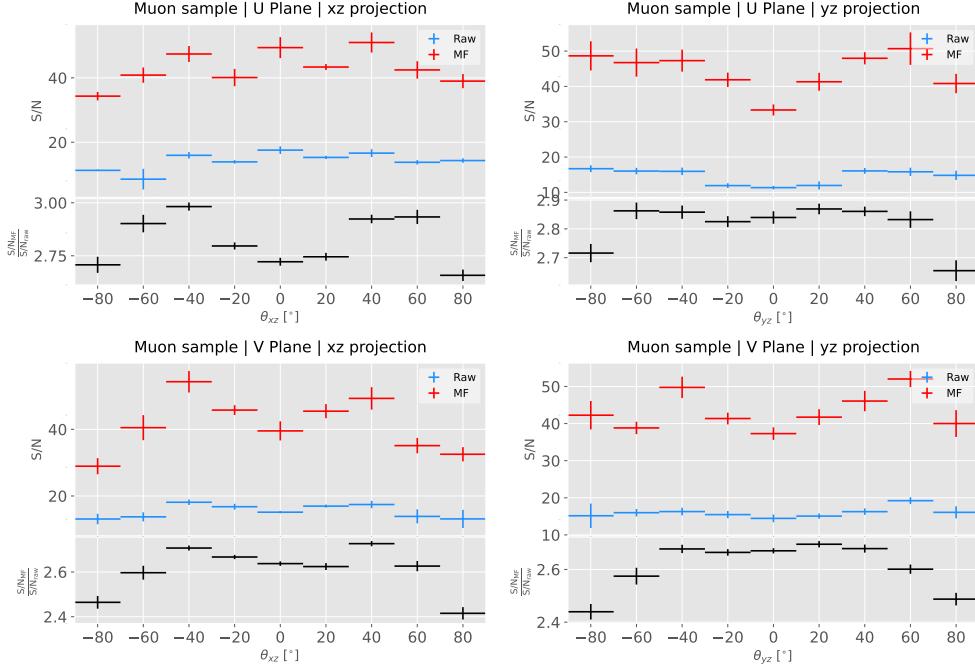


Figure 4.13: Angular dependence of the mean S/N and the S/N improvement for the monoenergetic muon sample. The top and bottom rows correspond to the U and V planes, respectively. The top subplots show the mean S/N for raw (blue) and filtered (red) waveforms whereas the bottom subplots depict the averaged S/N improvement (black).

1339 However, for the ratio of the raw and filtered S/N (what I call the S/N change) per
 1340 event I do not take the ratio of the previous two quantities but compute the average of
 1341 the individual ratios per channel in the event:

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

1342 therefore:

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

1343 4.5.1 Angular dependence

1344 Having these monoenergetic samples, one can study the angular dependence of the
 1345 matched filter performance. This is an important point, as it is a well established
 1346 fact that for certain track configurations the S/N is much lower than average as the

4.5. MONTE CARLO STUDIES

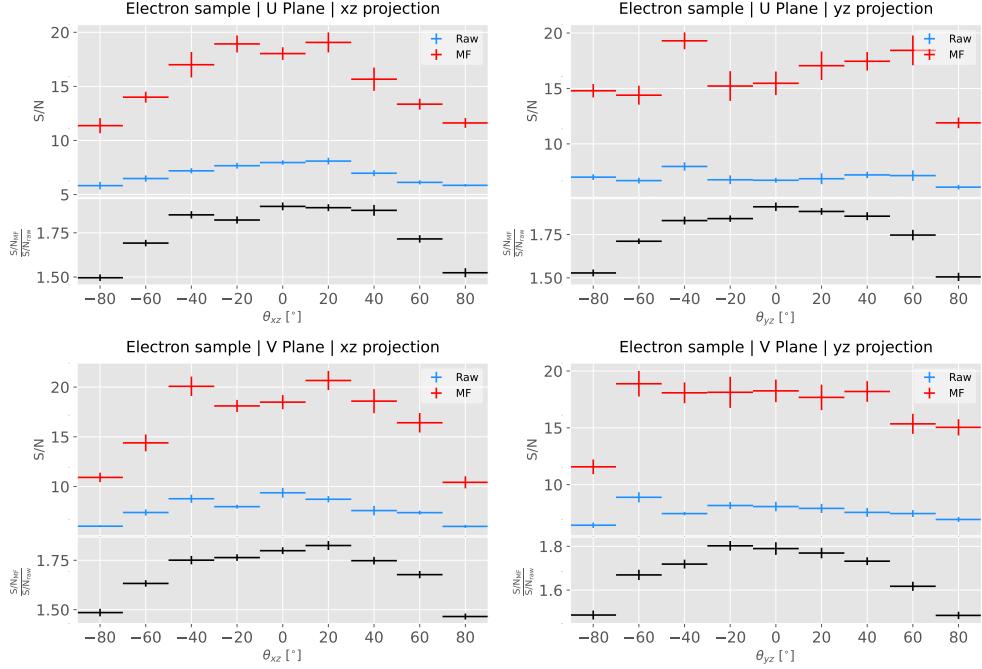


Figure 4.14: Angular dependence of the mean S/N and the S/N improvement for the monoenergetic electron sample. The top and bottom rows correspond to the U and V planes, respectively. The top subplots show the mean S/N for raw (blue) and filtered (red) waveforms whereas the bottom subplots depict the averaged S/N improvement (black).

corresponding waveforms are severely distorted. Therefore, I am interested in seeing how the matched filter behaves in different cases and how the S/N change for those compare to the average.

Figure 4.13 shows the angular dependence of the S/N for the monoenergetic $E_k = 100$ MeV isotropic muons, for the different induction planes and projections. The angles for each event are given by the components of the initial value of the momentum of the particles, taking the angles of the projections on the xz and yz planes with respect to the z axis (more accurately, one needs to compute these angles twice for each event, a pair for the $xy'z'$ coordinate system and the other for the $xy''z''$, as explained previously). The top row shows the dependence on the angles corresponding to the U plane, i.e. $\theta_{xz'}$ and $\theta_{y'z'}$, whereas the bottom row shows the angular dependence viewed from the V plane, $\theta_{xz''}$ and $\theta_{y''z''}$. In each panel, 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

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

averaged S/N change (black). The horizontal lines show the most probable value for the corresponding angular bin, obtained from a fit to a Landau distribution. The vertical lines represent the error in the parameter estimation.

Both for the raw and matched filtered samples, the S/N is lower for tracks that are normal to the APA ($\theta_{xz} \sim \pm 90^\circ$). Similarly, tracks parallel to the wires ($\theta_{yz} \sim \pm 90^\circ$) tend to have higher S/N than those perpendicular to these ($\theta_{yz} \sim 0$). The S/N improvement seems to follow similar trends for both projections in the two planes. In the xz plane there is a slight preference for tracks with $\theta_{xz} \sim \pm 45^\circ$ (particularly in the U plane), whereas in yz the S/N change plateaus around the central region.

Figure 4.14 shows the corresponding angular dependence results for the $E_k = 100$ MeV electrons sample. Although the S/N behaviour in this case is similar to what I observed for the muons, some differences are evident. A possible explanation can be that, because a significant fraction of the hits in these events are produced by the secondary particles generated in the EM shower, some of the S/N ratios do not correspond to the directional information of the primary electron. Even so, the S/N change distribution exhibits a consistent pattern and it is clear that the matched filter enhances the signal regardless of the electron direction.

4.5.2 Distortion and peak asymmetry

As a case study, I select two of the simulated $E_k = 100$ MeV monoenergetic muon events. With respect to the U induction plane, one is parallel to the APA (low $\theta_{xz'}$) and to the wires (high $\theta_{y'z'}$) and the other is normal to the APA plane (high $\theta_{xz'}$) and perpendicular to the wires (low $\theta_{y'z'}$). As expected from the results on the angular dependence discussed above, the former has a higher S/N (both before and after the filtering) when compared to the latter. An interesting thing to notice about these two samples is that, even though one has a much larger S/N than the other, it is the one with the smallest S/N the one that gets a more significant averaged S/N improvement. In Tab. 4.1 I include all the relevant parameters of these two $E_k = 100$ MeV muon events, namely the angles with respect to the $xy'z'$ reference frame, the values of the

4.5. MONTE CARLO STUDIES

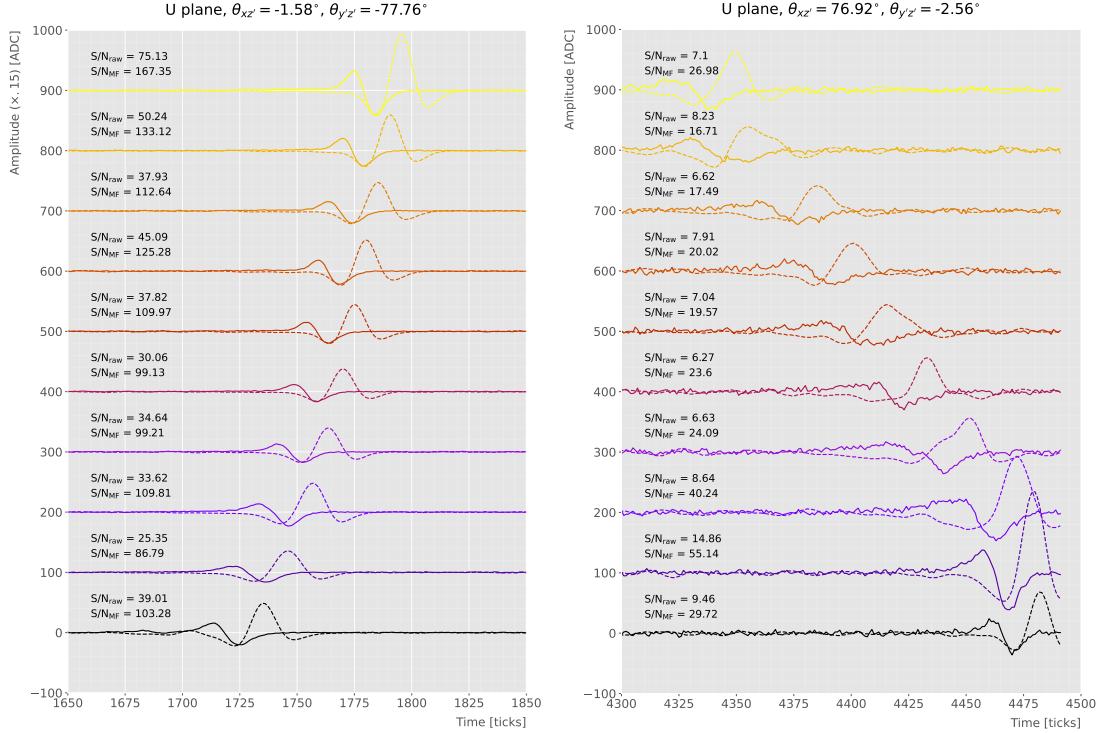


Figure 4.15: Selected consecutive waveforms corresponding to two monoenergetic $E_k = 100$ MeV muon events, one is parallel to the APA and to the wires in the U plane (left panel) and the other is normal to the APA plane and perpendicular to the U plane wires (right panel). The solid lines represent 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 amplitudes to the ones on the right panel.

1388 S/N, the S/N change and also the so-called peak asymmetry Δ_{peak} , that I will define
 1389 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.

	$\theta_{xz'} (\circ)$	$\theta_{yz'} (\circ)$	S/N _{raw}	S/N _{MF}	$\frac{S/N_{MF}}{S/N_{raw}}$	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

1390 One can try to understand better the nature of these two events by looking at the
 1391 raw and filtered data from some of their active channels. Figure 4.15 shows a selection of
 1392 consecutive raw and filtered U plane waveforms from the event with high S/N (left panel)

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

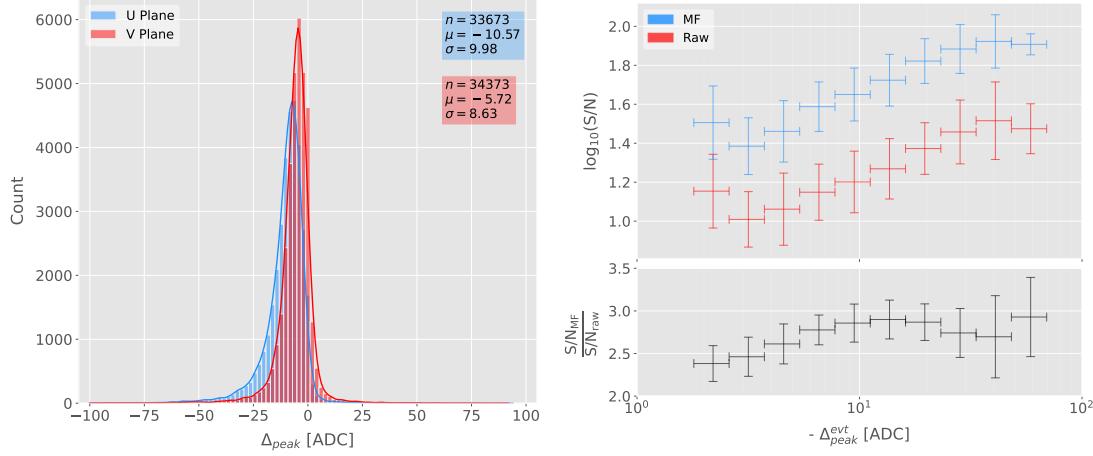


Figure 4.16: Left panel: peak asymmetry distribution for the case of the monoenergetic $E_k = 100$ MeV muon sample. Each value corresponds to a single bipolar signal peak from a channel in any event. The blue distribution represents the peaks on U plane channels, whereas the red corresponds to signal 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.

1393 and the one with low S/N (right panel). To show both collections of waveforms at a
 1394 similar scale I had to apply a factor of 0.15 to the waveforms with high S/N. Additionally,
 1395 next to each waveform I include the values of the raw and matched filtered S/N for the
 1396 corresponding channel. The first thing to notice is that the amplitude of the signal peaks
 1397 from the normal track have a much smaller amplitude, and also appear quite distorted
 1398 when compared to the others. On the other hand, although the matched filtered S/N for
 1399 each channel are still smaller, the relative improvements are larger than in the parallel
 1400 case.

1401 A way to quantify the difference between the shape of the waveforms of these two
 1402 events is using their peak asymmetry. I define the peak asymmetry as the (signed)
 1403 difference between the positive and the negative peaks of the bipolar shape, i.e.:

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

1404 where both heights h_+ and h_- are positive. Figure 4.16 (left panel) shows the distribution

4.5. MONTE CARLO STUDIES

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 means $\mu_{\Delta}^U = -10.57$ ADC and $\mu_{\Delta}^V = -5.72$ ADC, respectively. Notice how the peak asymmetry value of the selected event with the high S/N sits at the left tail of the distribution, whereas the corresponding value of the sample with the low S/N lies around the mean.

It is possible to correlate the peak asymmetry with the S/N and the S/N change per event. Figure 4.16 (right panel) shows the result of comparing 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 how there is an approximate linear relation between the peak asymmetry and the S/N, 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 change. In this case, one can see that there is a clear 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.

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

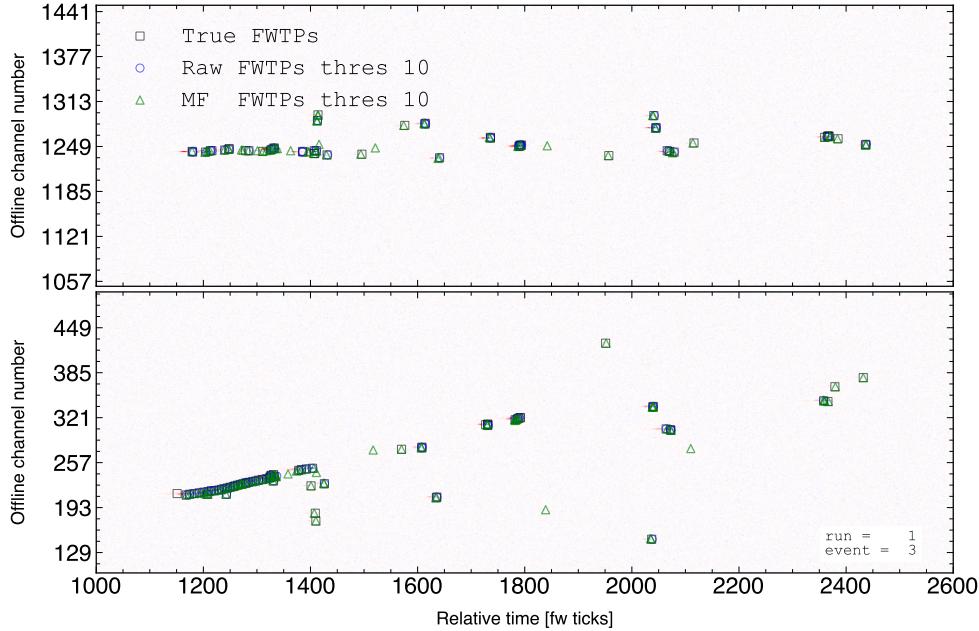


Figure 4.17: Raw event display showing the time (in firmware ticks) versus offline channel number for a $E_k = 100$ MeV electron event. The produced true hits are superimposed (black boxes) as well as the hits coming from the standard hit finder chain (blue circles) and the hit finder using the matched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

4.5.3 Hit sensitivity

One of the advantages of the matched filter, directly related to increasing the S/N, is the capability of forming TPs that before fell below the threshold. For instance, Fig. 4.17 shows the raw ADC data from an example electron event 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 run with a threshold of 10 ADC. Notice that the standard hits match well the true ones in the initial part of the event, where we have a track-like object. However, it misses 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 the matched filter produces more hits as a results of the enhancement

4.5. MONTE CARLO STUDIES

1444 of the signal peaks relative to the noise level, it is also true that it may pick up some
1445 spurious hits not related to any real activity if one lowers the thresholds too much.
1446 Therefore, some optimisation of the threshold is needed, as there is a trade-off between
1447 precision and sensitivity.

1448 Having this in mind, I compare the produced hits from both the standard and the
1449 matched filter hit finders to the true hits. By running the hit finders on the samples
1450 with different values of the threshold I can understand how low these can be pushed,
1451 and then evaluate the gains obtained from this.

1452 To study how the hit formation depends on the energy, I prepared new isotropic
1453 samples with the same types of particles as previously (muons, electrons, protons and
1454 neutral pions) but with a flat kinetic energy distribution ranging from 5 to 100 MeV.

1455 To estimate the hit sensitivity for a certain sample, one needs to recover the set of
1456 true hits to be able to compare these with the ones produced. To do so, I modify the
1457 procedure I use to extract the raw waveforms. For this kind of study, I run the detector
1458 simulation in two steps, first I produce the waveforms without noise and extract them
1459 in the same format I used for the raw data. Then, the noise is added and the noisy
1460 waveforms are similarly written to a file.

1461 To have a better comparison between the true hits and the ones produced from the
1462 raw waveforms after applying the two filters, I apply the FIR filter and the matched
1463 filters to the noiseless waveforms as well. I run the hit finder with a minimal threshold
1464 (in this case I use 1 ADC) on the filtered noiseless waveforms, generating two sets of true
1465 hits. I will refer to these as the standard true hits (with the default FIR filter) and the
1466 matched filter true hits, respectively. This allows for a more precise matching between
1467 the different groups of hits produced, as it will account for any delays and distortions
1468 introduced by the filters.

1469 In the case of the raw waveforms (with noise added), I run the hit finder on them
1470 with different values of the threshold, after applying either the FIR or the matched
1471 filters. I name these simply standard and matched filter hits, respectively. Then, I
1472 match the generated hits to the true hits, the standard hits to the standard true hits

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

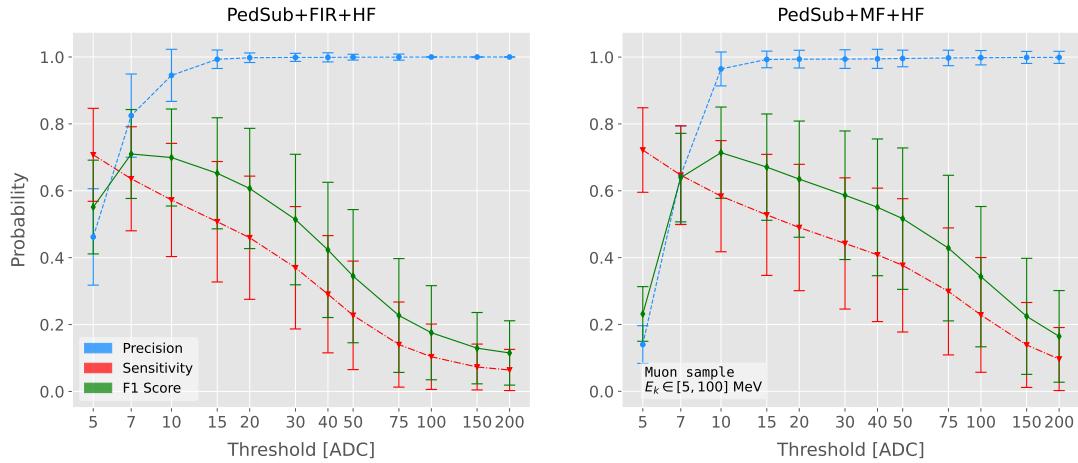


Figure 4.18: Dependence of the precision (blue), sensitivity (red), and F_1 (green) scores on the threshold values used in the hit finder, for the FIR (left panel) and matched filter (right panel) cases. The results were obtained after matching the hits to the true hits in the case of the isotropic muon 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.

and the matched filter hits to the matched filter true hits. The matching is performed by comparing the channel number and the timestamp of the hits. To count as a match, I require that all hits with the same channel number and timestamp have overlapping hit windows, i.e. the time windows between their hit end and hit start times need to overlap. If more than one hit in one of the groups have hit overlap with the same hit in the other group, I only count the match with the closest hit peak time value.

To quantify the performance of the two hit finder approaches, I use a classical method from statistical classification known as confusion matrix [114]. It divides the outputs in four categories: true positive (TP, both truth and predicted values are true), false negative (FN, truth value is true but predicted is false), false positive (FP, truth value is false but predicted is true) and true negative (TN, both truth and predicted values are false).

The contents of the confusion matrix allow us to compute other derived scores to assess the performance of our classification. In this study, I make use of three of these

4.5. MONTE CARLO STUDIES

1487 metrics, namely the precision or positive predictive value:

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

1488 the sensitivity or true positive rate:

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

1489 and the F_1 score [115]:

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

1490 which is the harmonic mean of the precision and the sensitivity.

1491 For this specific case I am not going to make use of the true negative category, as its
1492 definition in this context can be ambiguous because one does not have clear instances in
1493 the classification process. This way, I only count the number of true positives as the
1494 total amount of hits I can match between true and raw populations, the number of false
1495 negatives will be the number of missing true hits, and the false positives the number of
1496 hits which do not match any true hit.

1497 In Fig. 4.18 I show the precision (blue), sensitivity (red) and F_1 -score (green) I
1498 obtain as a function of the threshold used in the hit finder for the muon sample. Because
1499 the matched filters are only applied to induction channels, I consider exclusively the hits
1500 coming from the U and V planes. The panel on the left corresponds to the results I
1501 get when running the hit finder on the FIR filtered waveforms, whereas the right panel
1502 contains the scores for the matched filter case. The points are centered at the threshold
1503 value used and represent the mean value obtained for each score using all the generated
1504 events, while the error bars indicate one standard deviation around the mean value.

1505 One can see that the precision for the matched filter case is lower when the thresholds
1506 are very low, as the noise baseline is slightly amplified, but then rises to high values
1507 quicker than for the FIR case. The other difference one can spot is that the sensitivity
1508 in the FIR case starts dropping faster at around the same threshold values where the

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

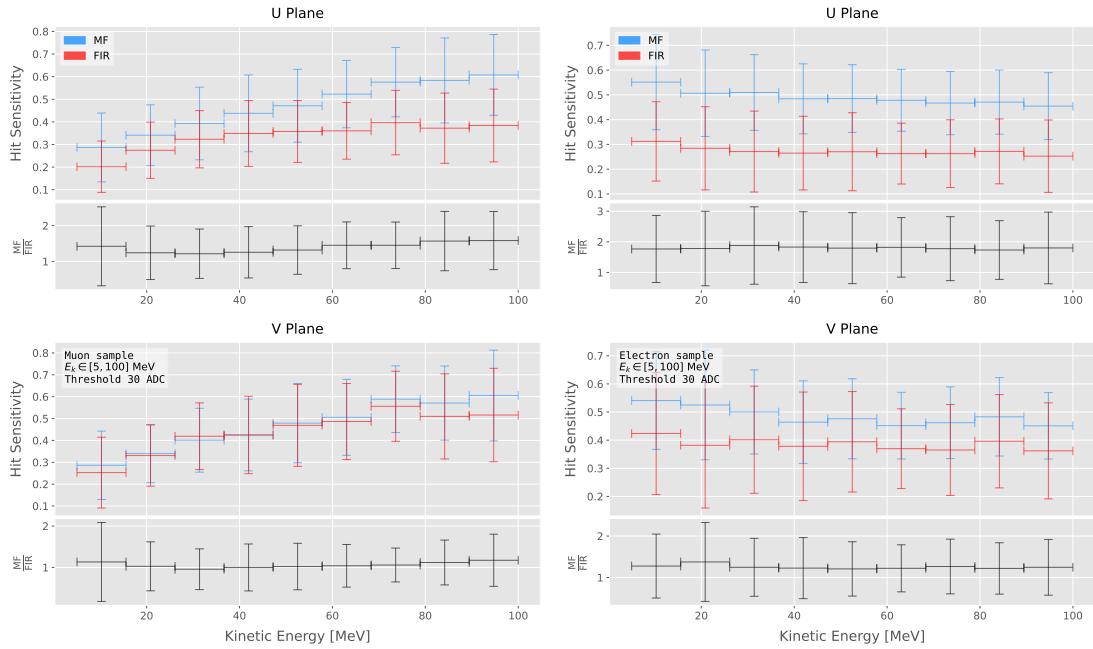


Figure 4.19: Dependence of the averaged hit sensitivity on the kinetic energy of the events for the matched filter (blue) and standard (red) hits, for the case of the muon (left panel) and electron (right panel) samples, separated between U (top plots) and V (bottom plots) induction wire planes. The 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.

precision stabilises around 1, while in contrast for the matched filter this rapid decrease starts at higher threshold values. A similar scan for the same thresholds was performed for the electron sample in the same energy range, yielding similar results.

In Fig. 4.19 I show the average hit sensitivity versus the kinetic energy of the events, both for the matched filter hits (blue) and the standard hits (red). The left panel corresponds to the muon sample, whereas the one on the right corresponds to the electron sample, both with kinetic energies between 5 and 100 MeV. In each panel, the top plot corresponds to hits in the U plane, while the bottom plot contains the same information for the V plane. Each plot contains two subplots, the one on the top shows the hit sensitivity values for the matched filter and standard hits separate, while the bottom subplot depicts the ratio between the matched filter and standard sensitivities. The horizontal lines are placed at the mean value obtained in the fit and represent the

4.5. MONTE CARLO STUDIES

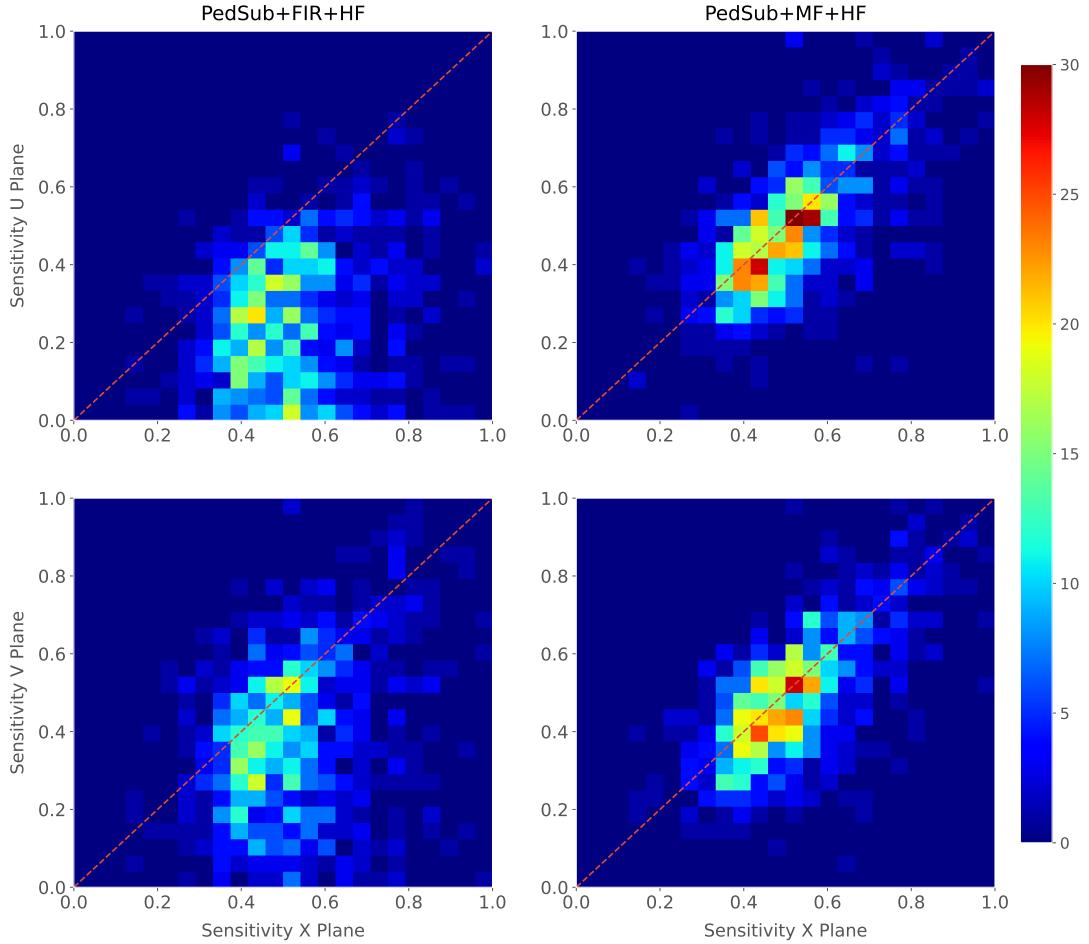


Figure 4.20: Distributions of the hit sensitivity in the U (top panels) and V (bottom panels) planes versus the hit sensitivity in the X plane, both for the standard hits (left panels) and the matched filter hits (right panels), in the case of the electron sample and a threshold of 30 ADC.

width of the E_k bins used, while the vertical error bars indicate one standard deviation around that mean. In both cases, the threshold used was 30 ADC, as I require the precision to be higher than 0.99 for both matched filter and standard cases.

In general, the improvements are better for the U than for the V plane. While for the U channels I achieve a mean improvement of 50% and 80% for muons and electrons, respectively, the improvement in the V plane is stalled at 10% and 25%. Nevertheless, looking at the sensitivities for the matched filter hits in both planes, one can see these have similar mean values for each energy bin. On the other hand, for the standard hits the sensitivity remains higher for the V plane. This way, it looks there is a less

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

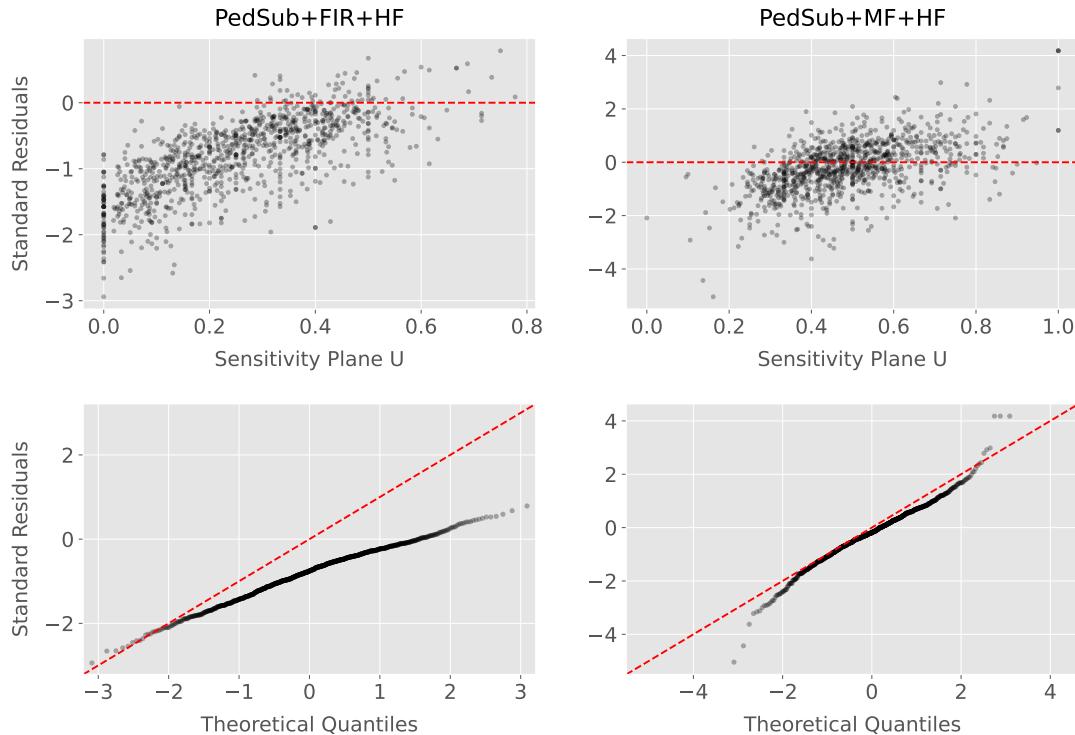


Figure 4.21: Top panels: standard residual plots of the hit sensitivities between the X and U planes. Bottom panels: quantile-quantile plots of the hit sensitivity standard residuals between the X and U planes. In all cases, the left panel corresponds to the standard hits while the right panel represents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

1530 significant gain because the hit sensitivity was already high.

1531 Another interesting observation is the different behaviors for muons and electrons.

1532 While hit sensitivity for muons grows significantly with energy, in the case of electrons it

1533 slightly decreases the higher the kinetic energy of the event is. However, when it comes

1534 to the improvement on the sensitivities, this remains almost constant in all cases.

1535 Furthermore, we can look at how the concurrence of hits between the different wire

1536 planes has changed. For any given event, I expect to have a similar number of hits in

1537 the three planes. As the ionisation electrons need to cross the U and V planes prior

1538 to reach the collection plane X, they will induce current in those wire planes. A way

1539 to check the concurrence of hits across planes is comparing the hit sensitivities in the

1540 different planes for each individual event. Although the sensitivities will not be exactly

1541 equal across planes, ideally they should be normally distributed around the diagonal.

4.6. VD COLDBOX DATA TAKING

1542 Figure 4.20 shows the hit sensitivity in the U (top panels) and V (bottom panels)
1543 planes versus the hit sensitivity in the X plane, for the case of the standard hits (left
1544 panels) and the matched filter hits (right panels). All plots were generated for the
1545 electron sample and a threshold of 30 ADC. From these, one can see a clear trend.
1546 The standard hit finder chain produces hit sensitivities in the induction planes that are
1547 systematically lower than the sensitivity in the X plane, i.e. most of the points sit below
1548 the diagonal (red dashed line). In contrast, when the matched filters are applied, the
1549 majority of the events are distributed around the diagonal. This points out that the
1550 concurrence of hits across planes has improved.

1551 To exemplify the improvement I obtain, I take the residuals of the hit sensitivities
1552 for the X and U planes. Assuming the diagonal hypothesis, i.e. given a dataset of the
1553 form (x, y) for any x I take the predicted y value to be equal to the value of x , I can
1554 compute the standard residuals for the hit sensitivities in U given the sensitivities for
1555 X. In Fig. 4.21 (top panels) I show these standard residuals against the corresponding
1556 values of the hit sensitivity in the U plane, for the electron sample with kinetic energy
1557 between 5 and 100 MeV. Comparing the scatter points in the case of the standard hits
1558 (left panel) and the matched filter hits (right panel), it can be seen that the residuals for
1559 the standard hit finder follow a certain pattern and their mean deviates from 0.

1560 To see clearly if the residuals are normally distributed, in Fig. 4.21 (bottom panels)
1561 I plot the corresponding quantile-quantile plot for both the standard (left panel) and
1562 matched filter (right panel) residuals. One can clearly see that the points for the standard
1563 hit finder case follow a strongly non-linear pattern, suggesting that the residuals do not
1564 follow a normal distribution. In contrast, for the matched filter hits the points conform
1565 to a roughly linear path, implying that in this case the normality condition is fulfilled.

1566 All these results hint at the fact that the concurrence of hits across the wire planes
1567 can be strengthened by applying the matched filters.

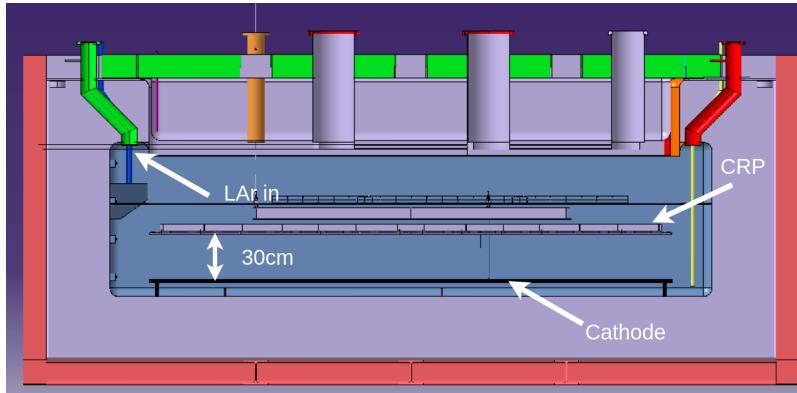


Figure 4.22: Schematic diagram of the vertical drift ColdBox setup at CERN.

1568 4.6 VD ColdBox data taking

1569 Between February and April 2023 the vertical drift (VD) ColdBox setup at CERN,
 1570 shown in Fig. 4.22, was recommissioned for cold electronics testing with CRP5. That
 1571 provided an opportunity for testing the firmware TP generation in a real LArTPC.
 1572 However, during the two run periods new software-related complications that were not
 1573 observed in previous running conditions arose.

1574 These prevented us from taking data with the whole system. As a palliative measure,
 1575 new configurations were developed that allowed to run with TP generation enabled for a
 1576 subset of the ADC links. With these workarounds, we managed to run with up to three
 1577 out of twelve ADC links and the horizontal muon trigger algorithm (HMA).

1578 Additionally, an alternative firmware version was prepared featuring the matched
 1579 filter coefficients optimised for the induction plane hit finding. The version of the filter
 1580 we used for the data taking is slightly different from the one of the previous studies, as
 1581 in this case we needed to apply the same filter coefficients to all channels irrespective
 1582 of the readout plane they come from. With this, we also managed to run with three
 1583 ADC links and the HMA trigger. Figure 4.23 shows an example event display from the
 1584 longest run we recorded with the matched filter firmware.

1585 We used the recorded data, together with our standalone TPG simulation tool, to
 1586 perform comparisons between the firmware and simulated TPs. One such comparison

4.6. VD COLDBOX DATA TAKING

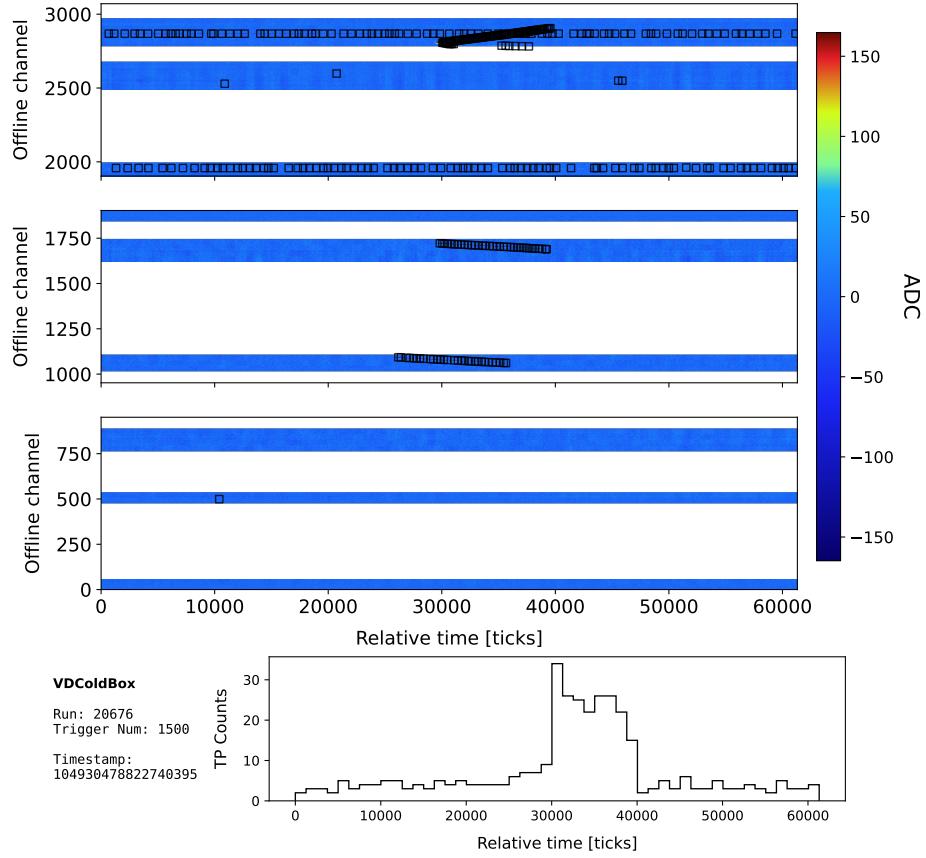


Figure 4.23: Event display of the data taken with the matched filter and HMA trigger at the VD ColdBox. The display shows the data from 3 ADC links for the full trigger window, with the black squares representing the produced TPs. The bottom panel represents the TP counts as a function of time in the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

for a matched filter run can be seen in Fig. 4.24. The agreement achieved is within the expectation, from what we have seen in previous samples.

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

CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

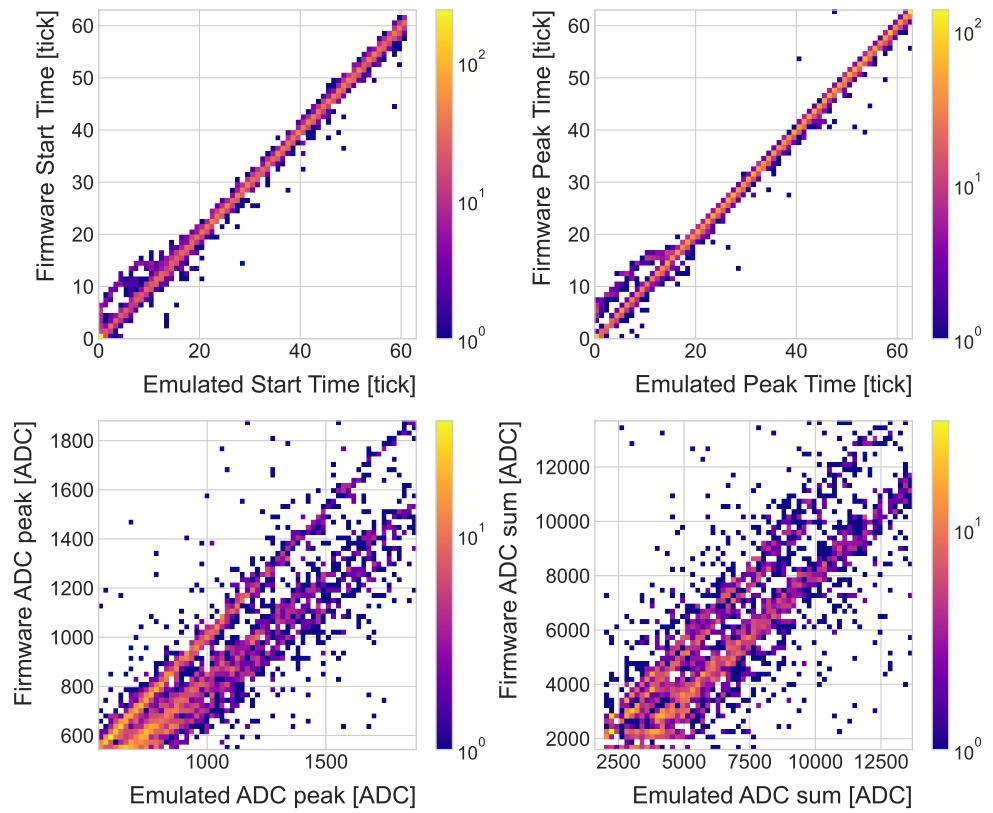


Figure 4.24: Comparison between firmware-produced and simulated TP quantities for a matched filter run at the VD ColdBox.

1594

1595

1596

Dark Matter searches

with neutrinos from the Sun

1597 *He stepped down, trying not to look long at her, as if she were the Sun, yet he
1598 saw her, like the Sun, even without looking.*

1599 – Leo Tolstoy, *Anna Karenina*

1600 The idea of detecting neutrino signals coming from the core of the Sun to probe DM
1601 is not new. The main focus of these searches has usually been high-energy neutrinos
1602 originated from DM annihilations into heavy particles [116–119]. However, recent studies
1603 have proposed to look at the low-energy neutrino flux arising from the decay of light
1604 mesons at rest in the Sun [120–123], previously thought undetectable.

1605 In this Chapter, I try to demonstrate the capability of DUNE to constrain different
1606 DM scenarios. I use the neutrino fluxes arising from DM annihilations in the core of the
1607 Sun to compute the projected limits that DUNE would be able to set on the annihilation
1608 rates of DM particles in the Sun and the DM scattering cross sections.

1609 **5.1 Gravitational capture of DM by the Sun**

1610 The Sun and the centre of the Earth are possible sources of DM annihilations, specially
1611 interesting because of their proximity. Their gravitational attraction ensures the capture
1612 of DM from the local halo through repeated scatterings of DM particles crossing them.
1613 Only neutrinos produced from DM annihilations can escape the dense interior of these

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

1614 objects. Therefore, neutrino telescopes are the most useful experimental layouts to
1615 pursue DM searches from their cores.

1616 The neutrino flux from DM annihilations inside the Sun depends on the DM capture
1617 rate, which is proportional to the DM scattering cross section, and the annihilation rate,
1618 which is proportional to the velocity-averaged DM annihilation cross section. The total
1619 number of DM particles inside the Sun follows the Boltzmann equation [120]:

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

1620 where C_{\odot} and A_{\odot} are the total Sun DM capture and annihilation rates respectively.
1621 In this expression I neglected the evaporation term, proportional to N_{DM} , which only
1622 contribute for $m_{DM} \lesssim 4$ GeV [124]. As the current threshold of neutrino telescopes
1623 is a few GeV, this region falls below the probed range but can be important in future
1624 low-energy projects like DUNE.

1625 This equation has an equilibrium solution:

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

1626 which represents the amount of DM inside the Sun if the capture and annihilation have
1627 reached equilibrium. As the Sun is approximately 4.6 Gyr old [125], it is usually assumed
1628 that equilibrium has been achieved. Therefore, the anomalous neutrino flux from the
1629 Sun would only depend on the DM scattering cross section, enabling us to set limits
1630 on this quantity. If one does not assume equilibrium, some assumptions on the DM
1631 annihilation cross section are necessary to extract predictions from neutrino signals.

1632 Here, I am going to consider three possible scenarios for the DM interactions: DM
1633 scattering off electrons, spin-dependent (SD) and spin-independent (SI) interactions
1634 with nuclei. For these last two, the cross sections will be given in terms of the SD and
1635 SI elastic scattering DM cross section off protons (assuming that the DM interactions

5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

1636 with protons and neutrons are identical), σ_p^{SD} and σ_p^{SI} , as [120, 126]:

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

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

1637 where $\tilde{\mu}_{A_i}$ is the reduced mass of the DM-nucleus i system, $\tilde{\mu}_p$ is the reduced mass
 1638 of the DM-proton system, A_i and J_i the mass number and total angular momentum
 1639 of nucleus i , and $\langle S_{p,i} \rangle$ and $\langle S_{n,i} \rangle$ the expectation value of the spins of protons and
 1640 neutrons averaged over all nucleons, respectively (see Ref. [127] for a review on spin
 1641 expectation values).

1642 Since the Sun is mainly composed of hydrogen, the capture of DM from the halo is
 1643 expected to occur mainly through SD scattering. However, since the SI cross section is
 1644 proportional to the square of the atomic mass, heavy elements can contribute to the
 1645 capture rate (even though they constitute less than 2% of the mass of the Sun). Heavy
 1646 elements can also contribute to the SD cross section if the DM also has momentum-
 1647 dependent interactions [128].

1648 DM particles can get captured by the Sun if after repeated scatterings off solar
 1649 targets their final velocity is lower than the escape velocity of the Sun. In the limit of
 1650 weak cross sections, this capture rate can be approximately written as [129]:

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

1651 where the summation extends over all possible solar targets. In this expression, R_{\odot}
 1652 is the radius of the Sun, ρ_{χ} is the local DM density, m_{χ} the mass of the DM particle,
 1653 $f_{v_{\odot}}(u_{\chi})$ the DM velocity distribution seen from the Sun's reference frame, $R_i^-(\omega \rightarrow v)$
 1654 is the differential rate at which a DM particle with velocity v scatters a solar target of
 1655 mass m_i to end up with a velocity ω and $|F_i(q)|$ is the nuclear form factor of target i .

1656 The differential scattering rate takes a rather simple form when considering velocity-

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

1657 independent and isotropic cross sections. In that case, this quantity is given by [126, 129]:

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

1658 where μ_i is the ratio between the DM mass and the mass of target i , $\mu_{i,\pm}$ is defined as:

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

1659 $n_i(r)$ is the density profile of target i in the solar medium, $u_i(r)$ is the most probable

1660 velocity of target i given by:

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

1661 where $T_\odot(r)$ is the temperature of the Sun, the quantities α_\pm and β_\pm are defined as:

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

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

1662 and the function $\chi(a, b)$ is a Gaussian integral of the form:

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

1663 Finally, if one assumes the DM halo velocity distribution in the galactic rest frame

1664 to be a Maxwell-Boltzmann distribution, one can write the halo velocity distribution for

1665 an observer moving at the speed of the Sun with respect to the DM rest frame as:

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

1666 where:

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

1667 is the DM velocity squared, v_\odot the relative velocity of the Sun from the DM rest frame

5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

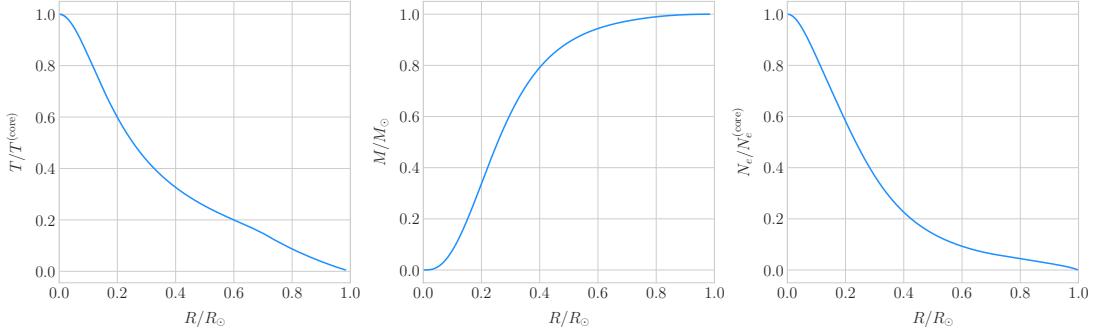


Figure 5.1: Input solar parameters used in the capture rate computation as a function of the solar radius, from left to right: temperature (with respect to the temperature at the core), mass (in solar masses) and electron number density (with respect to the electron density at the core). All quantities shown correspond to the standard solar model BS2005-OP [20].

1668 and $v_d \simeq \sqrt{3/2}v_\odot$ the velocity dispersion.

1669 For the case of strong scattering cross sections, Eq. (5.5) ceases to be valid, as it
 1670 escalates indefinitely with the cross section. In that limit, the capture rate saturates to
 1671 the case where the probability of interaction is equal to one, which can be written as
 1672 [130]:

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

1673 where $\langle v \rangle = \sqrt{8/3\pi}v_d$ is the mean velocity in the DM rest frame and the factor $\xi(v_\odot, v_d)$
 1674 accounts for the suppression due to the motion of the Sun:

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

1675 Having these into account, one can write the total capture rate as a combination of
 1676 both contributions, allowing a smooth transition between the two, as [120]:

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

1677 I computed the capture rate from Eq. (5.16) in the case of interactions with
 1678 electrons. To do so, I used the standard solar model BS2005-OP [20]. Fig. 5.1 shows the

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

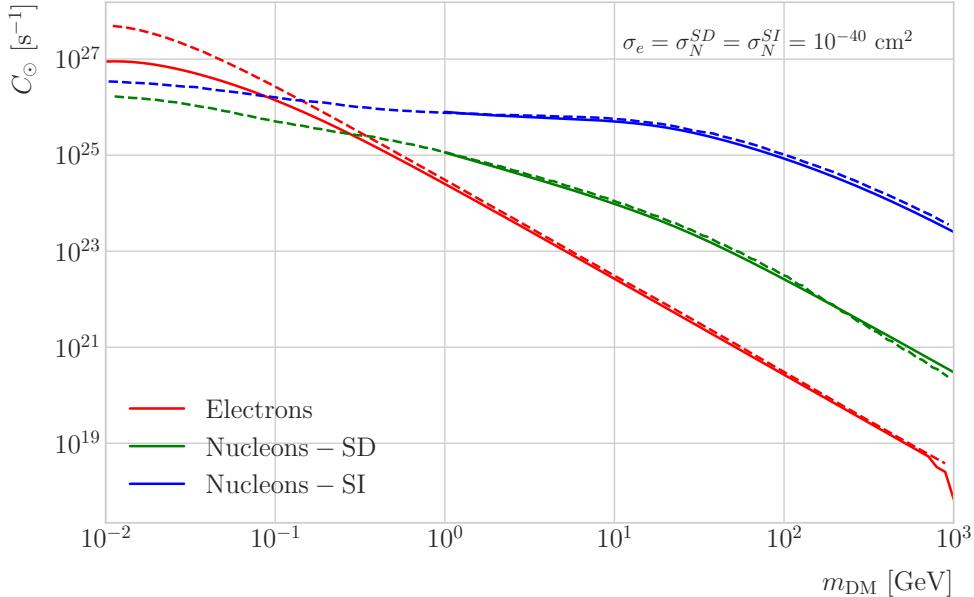


Figure 5.2: Capture rates as a function of the DM mass for the DM-electron interactions (red lines), SD DM-nucleons interactions (green lines), and SI DM-nucleons interactions (blue lines). Solid lines represent the values computed in this work while the dashed lines are the one given in Ref. [126]. All the rates are shown for a choice of scattering cross section of $\sigma_i = 10^{-40} \text{ cm}^2$.

1679 three parameters from the solar model that are needed for the computation, the solar
 1680 temperature (left panel), mass (central panel) and electron density (right panel) profiles.

1681 For the case of the interactions off nuclei, the computations are more convoluted
 1682 as one needs to add up the contributions of the different most abundant nuclei in
 1683 the Sun. Also, in contrast to the electron scenario where the form factor is trivially
 1684 $|F_e(q)|^2 = 1$, for any nucleus i one would need to consider some appropriate nuclear
 1685 density distribution (either a Gaussian approximation, a Woods-Saxon distribution, etc)
 1686 which would complicate the calculations even further.

1687 That is the reason why, at this stage of the study, I decided to take an alternative
 1688 approach to the computation of the DM-nucleus capture rates. I used the **DarkSUSY**
 1689 software, that allows us to compute these quantities performing a full numerical
 1690 integration over the momentum transfer of the form factors [131]. The default standard
 1691 solar model used by **DarkSUSY** is BP2000¹ [132].

¹This is what they say in their manual, but I fear it is somewhat outdated. It appears to me

5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

1692 In Fig. 5.2 I show the results I obtain for the capture rates, for the case of interactions
 1693 off electrons (red solid line), SD (green solid line) and SI (blue solid line) interactions of
 1694 nucleons. In all cases I use a value of the scattering cross sections of $\sigma_i = 10^{-40} \text{ cm}^2$.
 1695 Note here one of the limitations of the **DarkSUSY** approach, one can not extend the
 1696 computation below $m_{\text{DM}} = 1 \text{ GeV}$. Nevertheless, this is not something to worry about in
 1697 this case, as I will discuss next. As a comparison, I added also the values computed in Ref.
 1698 [126] (same color scheme, dashed lines). One can see there is good agreement between
 1699 these and the **DarkSUSY** computation of the SD and SI interactions for $m_{\text{DM}} \geq 1 \text{ GeV}$.
 1700 In this regime their computations also matches quite well the results for the electron
 1701 capture rate. However, these start to differ significantly below $m_{\text{DM}} = 1 \text{ GeV}$, being
 1702 their estimate up to a factor of 5 bigger than ours for low masses. This could be due to
 1703 the use of a different solar model in the calculation.

1704 Let me comment briefly about the assumption I made before about not including
 1705 an evaporation term in the Boltzmann equation. If I include this term in the equation,
 1706 which is proportional to the number of DM particles, the equilibrium solution takes the
 1707 form:

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

1708 where E_{\odot} is the total evaporation rate, τ_{eq} is the equilibrium time in the absence of
 1709 evaporation:

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

1710 and κ is defined as:

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

1711 Now, it is easy to proof that in the case evaporation dominates $\kappa \gg 1$ and therefore:

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

this model is relatively old and do not see why they are not using others like [20]. Maybe one can double-check in the code to make sure.

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

- 1712 In contrast, if evaporation is irrelevant $\kappa \simeq 1$ and one recovers Eq. (5.2).
1713 In this way, one can define the evaporation mass as the mass for which the number
1714 of DM particles in equilibrium approaches Eq. (5.20) at 10% level:

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

1715 This can be regarded as the minimum testable mass one can reach using the annihilation
1716 products of the DM in the Sun.

1717 It was reported in Ref. [126] that, in the case of both SD and SI DM interactions
1718 off nuclei, this value ranges from 2 to 4 GeV depending on the specific scattering
1719 cross section value, compatible with the usual assumptions in the literature. What is
1720 interesting is the case of the electron capture. It was found that, when one applies a
1721 cutoff in the velocity distribution of the DM trapped in the Sun slightly below the escape
1722 velocity, the evaporation mass for the DM-electron interaction decreases remarkably. For
1723 a moderate choice of $v_c(r) = 0.9v_e(r)$ one gets an evaporation mass of around 200 to
1724 600 MeV. This possibility opens a region of the parameter space that could be tested
1725 with the next generation of neutrino detectors.

1726 5.2 Neutrino flux from DM annihilations

1727 When WIMPs annihilate inside the Sun a flux of high-energy neutrinos is expected
1728 from heavy quarks, gauge bosons and $\tau^+\tau^-$ final states, which decay before losing
1729 energy in the dense solar medium [121]. These produce a continuous neutrino spectra
1730 up to $E_{\nu} \sim m_{\chi}$. In the case of direct annihilation into neutrinos, one would have a
1731 monochromatic flux with $E_{\nu} = m_{\chi}$. This kind of signal has been extensively studied in
1732 the literature, allowing to put strong limits on the SD WIMP-proton cross section for
1733 large m_{χ} . However, the number of high-energy neutrinos per WIMP annihilation is small
1734 and the spectrum depends on the unknown final state. Moreover, although background
1735 rejection is easier for large m_{χ} , neutrinos with $E_{\nu} \gtrsim 100$ GeV are significantly attenuated
1736 by interactions in the Sun.

5.3. COMPUTING LIMITS FROM SOLAR NEUTRINO FLUXES

Nevertheless, most WIMP annihilation final states eventually produce a low-energy neutrino spectrum. In this case one does not just consider the more massive final states but also annihilations into e^+e^- , $\mu^+\mu^-$ and light quarks [120]. In particular, light 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 ν_μ flux with $E_\nu = 236$ MeV, while in the case of pions one would have $E_\nu = 29.8$ MeV. In practice, only the K^+ and π^+ contribute to these signals, as the 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. These 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

The first step to use these fluxes to search for DM in the Sun is to determine the expected number of atmospheric background events. For a given exposure, after directionality selection has been applied, this can be written 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

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

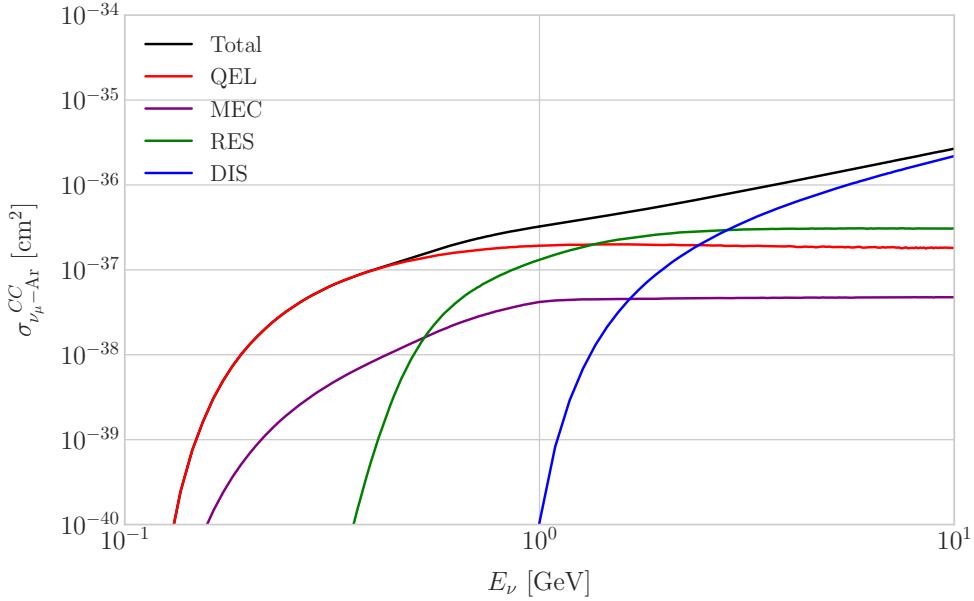


Figure 5.3: NuWro computed $\nu_\mu - {}^{40}\text{Ar}$ charged-current scattering cross section as a function of the neutrino energy. The black line shows to the total cross section, whereas the others correspond to the different contributions (in red quasi-elastic scattering, in green resonant pion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

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, for DUNE we have:

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

where $\sigma_{\nu-\text{Ar}}^{(\mu)}$ is the $\nu_\mu - {}^{40}\text{Ar}$ charged-current scattering cross section. In Fig. 5.3 I show the computed value of the cross section as a function of the neutrino energy E_ν , in the range of interest both for the atmospheric background and signal events. It was computed using the NuWro Monte Carlo neutrino event generator [133], including the CC contributions of the quasi-elastic scattering (red line), resonant pion exchange (green line), deep inelastic scattering (blue line) and meson exchange current (purple line).

The background rejection will depend on the resolution of the detector and the selection one applies on the events. A geometry argument can be used to estimate

5.3. COMPUTING LIMITS FROM SOLAR NEUTRINO FLUXES

the maximum background rejection one can achieve in this case, considering one can efficiently discriminate all events coming from a direction different from that of the Sun. In that case, the optimal background efficiency will simply be the relative angular coverage of the Sun. Taking the angular diameter of the Sun as seen from the Earth to be 0.5° , I have:

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

This value gives an optimistic estimate of the number of background events. However, it can be regarded as an upper limit, as it represents the best case scenario.

In Fig. 5.4 I show the fluxes of atmospheric neutrinos at the Homestake mine during solar minimum, taken from Ref. [134]. The values are averaged over the two angular directions. In blue I have the flux of muon neutrinos while in red I indicate the flux of electron neutrinos. Additionally, the dashed lines correspond to both antineutrino species.

Using these values for the muon neutrino and the corresponding total CC cross section, one can compute the total number of expected background events by integrating over the given energy range. For this I choose the range for DUNE specified in [85], $E_{min} = 10^{-1}$ GeV and $E_{max} = 10$ GeV. Taking all these into account, I find the total number of background events to be:

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

To estimate the sensitivity of DUNE to this kind of signals, one can consider a hypothetical data set where the number of observed neutrinos is taken to be the expected number of background events rounded to the nearest integer, $N_{obs} = \text{round}(N_B)$ [135]. Now, if I assume that the number of signal and background events seen by DUNE are given by Poisson distributions with means equal to the expected number of signal and background events, N_S and N_B , one can denote by N_S^{90} to the number of expected signal events such that the probability of having an experimental run with a number of events greater than N_{obs} is 90%. This number can be obtained as the numerical solution

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

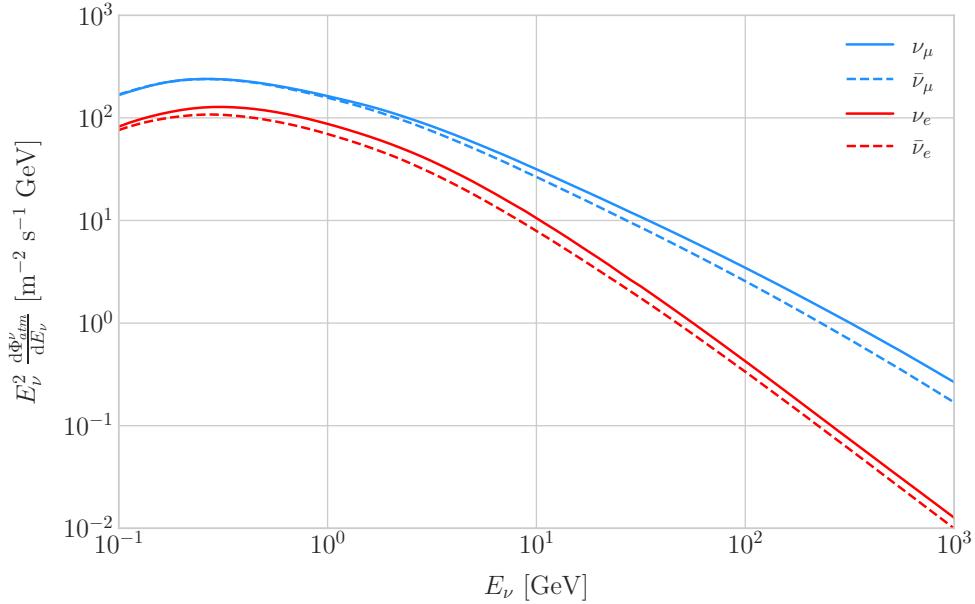


Figure 5.4: Expected atmospheric neutrino flux as a function of the neutrino energy E_ν at Homestake at solar minimum, taken from Ref. [134]. The blue solid (dashed) line correspond to muon neutrinos (antineutrinos) and the red solid (dashed) line correspond to electron neutrinos (antineutrinos).

1798 to the equation:

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

1799 where $\Gamma(x, y)$ is the upper incomplete gamma function.

1800 The number of signal events is related to the neutrino flux from DM annihilations in
1801 a similar way as the background events to the atmospheric neutrino flux. In this case I
1802 have:

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

1803 where η_S is the signal efficiency, Γ_A^{eq} is the total annihilation rate of DM particles at
1804 equilibrium, $\Gamma_A^{eq} = A_\odot (N_{DM}^{eq})^2$, z_{min} and z_{max} the minimum and maximum relative
1805 energies to integrate over (given by $z_{min,max} = E_{min,max}/m_{DM}$ for each m_{DM}) and
1806 $dN_\nu/dAdN_A dz$ the muon neutrino flux per DM annihilation in the Sun.

1807 Having obtained N_S^{90} one can use the relation in Eq. (5.27) to compute $\Gamma_A^{eq,90}$ for
1808 different values of the DM mass. Then, I can directly translate those values into the

5.4. EXAMPLE: KALUZA-KLEIN DARK MATTER

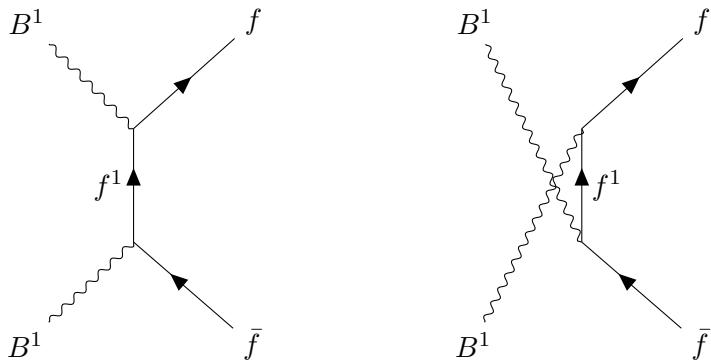


Figure 5.5: Feynman diagrams for $B^1 B^1$ annihilation into SM fermions.

1809 projected sensitivities for DUNE to the DM scattering cross sections, for a given exposure.
 1810 The relation between the annihilation rate and the DM-nucleon cross section comes from
 1811 the equilibrium condition through the solar DM capture rate, discussed above.

1812 5.4 Example: Kaluza-Klein Dark Matter

1813 Even though there are plenty of BSM theories which provide viable DM candidates, the
 1814 Kaluza-Klein type of models [136, 137] within the universal extra dimensions (UED)
 1815 paradigm naturally predict the existence of a massive, stable particle that can play the
 1816 role of the DM. In the UED scenario all the SM fields can propagate in one or more
 1817 compact extra dimensions [138], as opposed to the idea of brane worlds [139, 140], where
 1818 just gravity can propagate in the bulk while SM particles live at fixed points.

1819 Furthermore, in UED there is no violation of the translational invariance along the
 1820 extra dimensions, thus leading to degenerate KK modes masses and also the conservation
 1821 of the KK number in the effective four dimensional theory. At loop level, radiative
 1822 corrections and boundary terms shift the masses of the KK modes and break KK
 1823 number conservation into a KK parity. As a result, this theory only contains interactions
 1824 between an even number of odd KK modes, and therefore the lightest among the first
 1825 KK excitations will be stable. This particle is usually denoted as the lightest Kaluza-
 1826 Klein particle (LKP) and its mass is proportional to $1/R$, being R the size of the extra
 1827 dimension.

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

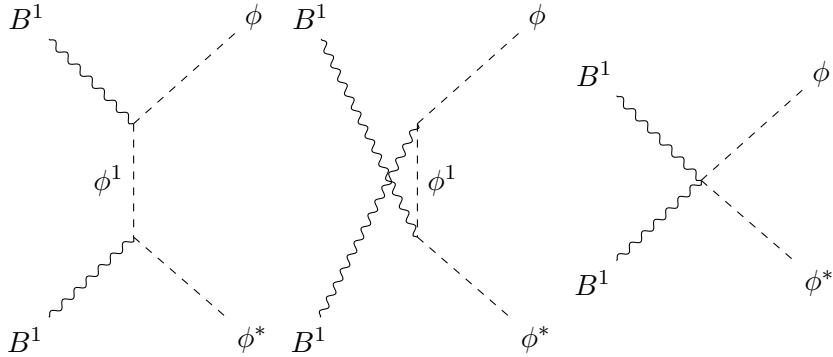


Figure 5.6: Feynman diagrams for $B^1 B^1$ annihilation into a Higgs boson pair.

1828 A viable DM candidate needs to be electrically neutral and non-baryonic, therefore
 1829 good candidates among the first Kaluza-Klein excitations would be the KK neutral
 1830 gauge bosons and the KK neutrinos [141]. Another possible candidate is the first KK
 1831 excitation of the graviton, which receives negligible radiative contributions and therefore
 1832 has a mass almost equal to $1/R$, but it has been shown that the lightest eigenstate
 1833 from the mixing of the gauge mass states (B^1, W_3^1) would be lighter, as these receive
 1834 negative radiative corrections [142]. It is also understood that, when these corrections
 1835 become sizeable, the eigenstates become approximately pure B^1 and W_3^1 states, as the
 1836 Weinberg mixing angle grows small with the KK number [142]. In that case, the LKP
 1837 can be well-approximated as being entirely B^1 .

1838 To estimate the sensitivity of DUNE to this particular DM model, I first need to
 1839 compute the neutrino flux produced by the annihilations of the LKP in the core of the
 1840 Sun, taking into account their propagation in the solar medium, as well as neutrino
 1841 oscillations. To this end I use `WimpSim` [143, 144] to generate 10^6 annihilation events in
 1842 the Sun over a time span of four years, and propagate them to the DUNE FD location
 1843 ($44^\circ 20' \text{ N}, 103^\circ 45' \text{ W}$), for different values of M_{LKP} . The different Feynman diagrams
 1844 for the annihilation of the B^1 into a pair of SM fermions and scalars are shown in Figs.
 1845 5.6 and 5.6.

1846 In Fig. 5.7 I show the obtained muon neutrino spectra arriving to the detector from
 1847 LKP annihilations in the Sun, per unit area and per annihilation, plotted in relative

5.4. EXAMPLE: KALUZA-KLEIN DARK MATTER

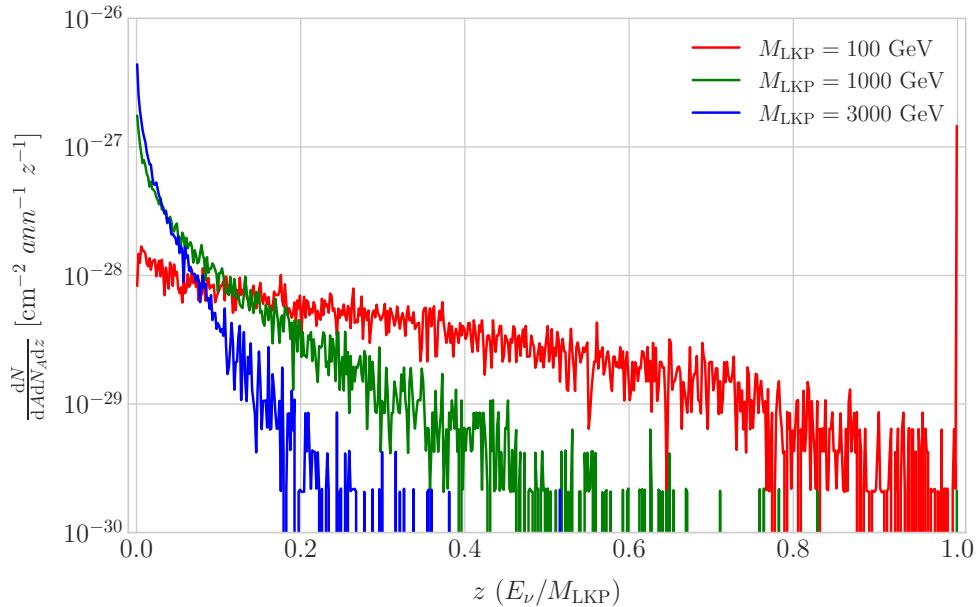


Figure 5.7: Computed spectra of muon neutrinos at the DUNE FD site from B^1 annihilations in the Sun for three different values of M_{LKP} , plotted in relative energy units for legibility.

1848 energy units for different values of the mass. As one could expect the spectra get
 1849 steeper the higher is the mass, due to the absorption of high-energy neutrinos in the
 1850 solar medium. Also, one can see the peak at $z = 1$ due to the direct annihilation into
 1851 neutrinos $\chi\chi \rightarrow \nu\bar{\nu}$.

1852 Now, one can estimate the sensitivity of DUNE to this particular model by using
 1853 the methods I previously discussed. To begin with, I will use the optimistic estimation
 1854 of the background efficiency in Eq. (5.24) to get an upper bound. Using it, one can
 1855 directly compute the number of expected background events to be $N_B = 0.11$ for an
 1856 exposure of 400 kT yr. Then, Eq. (5.26) give us a value of $N_S^{90} = 2.20$ for the 90%
 1857 exclusion number of expected signal events. By using the NuWro generated cross sections
 1858 and the computed neutrino fluxes from B^1 annihilations in the Sun I can estimate the
 1859 limits on the SD and SI DM-nucleus cross section using the relation in Eq. (5.2) and
 1860 the capture rates I computed with DarkSUSY.

1861 In Fig. 5.8 I show the projected sensitivity for DUNE on the spin-dependent B^1 -
 1862 proton scattering cross section versus the mass of the DM particle, for a exposure of

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

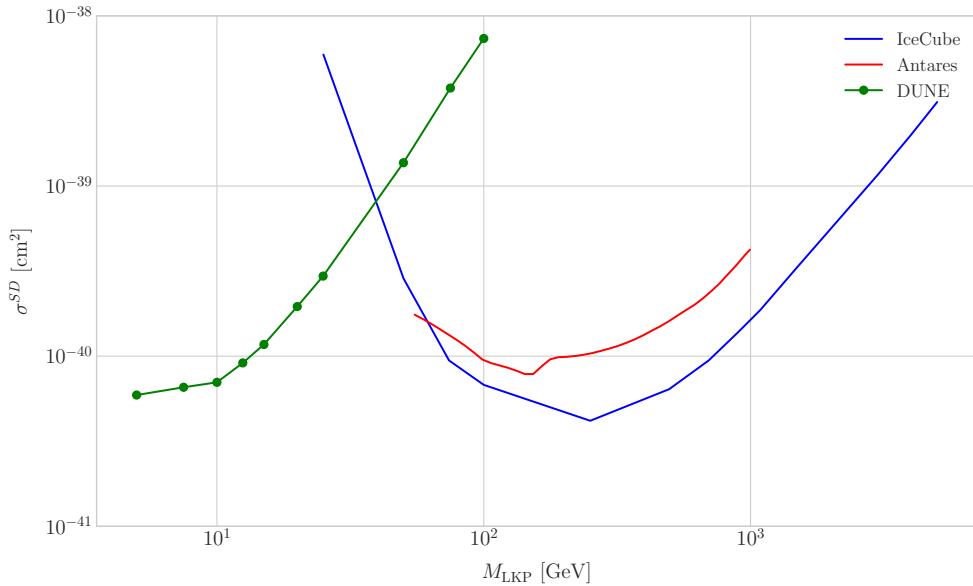


Figure 5.8: Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent B^1 -proton scattering cross section as a function of M_{LKP} (green dots). I also show the previous limits from IceCube [145] (blue line) and Antares [146] (red line) on the LKP cross section. The shaded area represents the disfavoured region (at 95% confidence level) on the mass of the LKP from LHC data [147].

1863 400 kT yr (green dots). I also include the previous results from IceCube [145] (blue line)
1864 and Antares [146] (red line). The shaded area represents the disfavoured region from
1865 combined searches for UED by ATLAS and CMS [147].

1866 From the experimental point of view, this estimation lacked a detailed simulation of
1867 the detector response and thus this must be consider as a mere optimistic sensitivity
1868 computation. However, it shows the potential of DUNE to constrain this kind of exotic
1869 scenarios, showing the region where it will be in a position to compete with other neutrino
1870 telescopes. A more detailed analysis is needed if I am to make a realistic estimation.
1871 Even though the region of the parameter space where DUNE would be sensitive to this
1872 particular model is quite constrained by collider searches [147] and other rare decay
1873 measurements [148, 149], it still constitutes an alternative indirect probe.

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

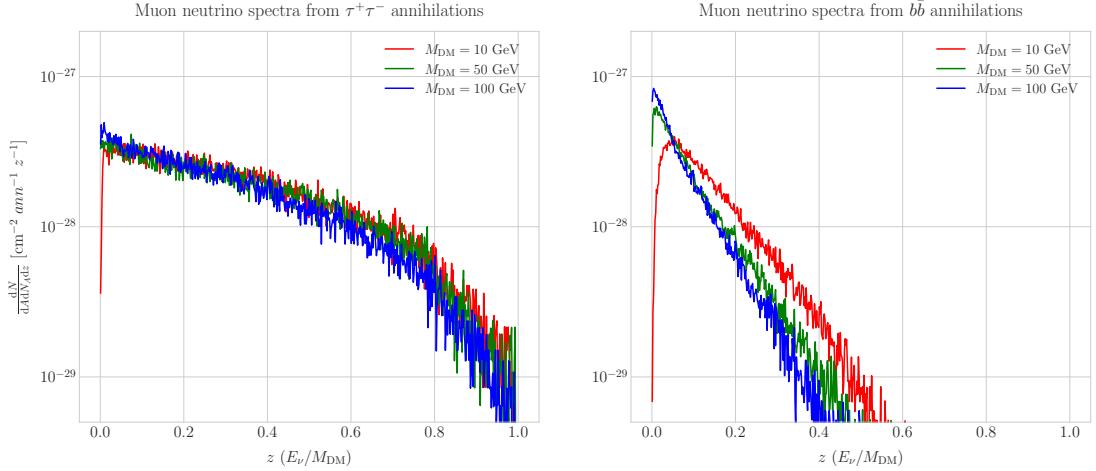


Figure 5.9: Computed spectra of muon neutrinos at the DUNE FD site from $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel) annihilations in the Sun for the DM masses $m_{\text{DM}} = 10 \text{ GeV}$ (red line), 50 GeV (green line) and 100 GeV (blue line), plotted in relative energy units.

1874 5.5 High energy DM neutrino signals

1875 To have better estimates on the capability of the DUNE FD to constrain the parameter
 1876 space of DM using solar neutrino fluxes, I need to start accounting for the detector
 1877 resolution effects and the topologies of the different signatures. As a starting point, I
 1878 will focus on specific annihilation channels. For the case of DUNE, the relevant ones
 1879 are mainly the hard channels $\tau^+\tau^-$ and $\nu\bar{\nu}$ and the soft channel $b\bar{b}$. These are the open
 1880 annihilation channels for relatively low mass WIMPs that will actually give neutrino
 1881 fluxes. Other channels, like W^+W^- and ZZ , are open for more massive WIMPs, but
 1882 those will produce usually a higher energy neutrino flux that will be out of reach for
 1883 DUNE (usually the maximum neutrino energy is taken to be $E_{\text{max}} = 10 \text{ GeV}$).

1884 In Fig. 5.9 I show the `WimpSim` [143, 144] generated muon neutrino spectra at the
 1885 DUNE FD location ($44^\circ 20' \text{ N}, 103^\circ 45' \text{ W}$) from $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel)
 1886 annihilations in the core of the Sun, for different DM masses. Here, one can clearly see
 1887 the meaning of the previous distinction between hard and soft channels. For the same
 1888 DM mass value, the muon neutrino spectrum from the $\tau^+\tau^-$ channel is more flat and
 1889 reaches higher energies than the one from the $b\bar{b}$ channel, which drops faster.

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

1890 In this case, I prepared two sets of files, one for $\tau^+\tau^-$ and the other for $b\bar{b}$, for DM
1891 masses in the range from 5 to 100 GeV (for $b\bar{b}$ the first mass point I take is 7.5 GeV, as
1892 this annihilation channel is not kinematically allowed for a WIMP with $m_{\text{DM}} = 5 \text{ GeV}$).
1893 Then, I prepared the `WimpSim` output fluxes in a specific way to use them as inputs to
1894 `NuWro`, which simulates the neutrino interaction with the argon.

1895 Because `WimpSim` outputs an event list together with the fluxes, I can use the former
1896 to generate the events. The direction of these is given in terms of the azimuth and
1897 altitude angles viewed from the specified location, so first I need to convert these into the
1898 DUNE FD coordinates. Once I have done it, each event can be processed with `NuWro`.
1899 To increase the number of samples and optimise the computation time, I generate 100
1900 interactions (i.e. `NuWro` events) for each `WimpSim` event². I restrict the event generation
1901 to charged current interactions, but I allow all the different contributions to the CC
1902 cross section, i.e. quasielastic scattering (QE), meson exchange current process (MEC),
1903 resonant pion production (RES) and deep inelastic scattering (DIS). I just take into
1904 account the CC contribution because I am only interested in final states with charged
1905 leptons, as we have better chances of reconstructing the kinematics of CC events.

1906 For the atmospheric fluxes I follow a similar procedure, only that this time I do not
1907 have a set of events but the fluxes binned in azimuth and altitude angles. This way, I
1908 transform these to DUNE coordinates and process the fluxes for each bin separated with
1909 `NuWro`.

1910 At this point, I have two sets of events with different energies and final states.
1911 In Fig. 5.10 one can see the distribution of the muon neutrino energies for the case
1912 $m_{\text{DM}} = 10 \text{ GeV}$, both for the $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel) channels, separated
1913 by interaction. One can clearly see that there are different energy regimes where the
1914 primary interaction type is different. This leads to a plurality of event topologies,
1915 therefore making it difficult to implement a general approach to the selection of events
1916 in detriment of the background. As a way to proceed, I decided to focus on a subset of

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

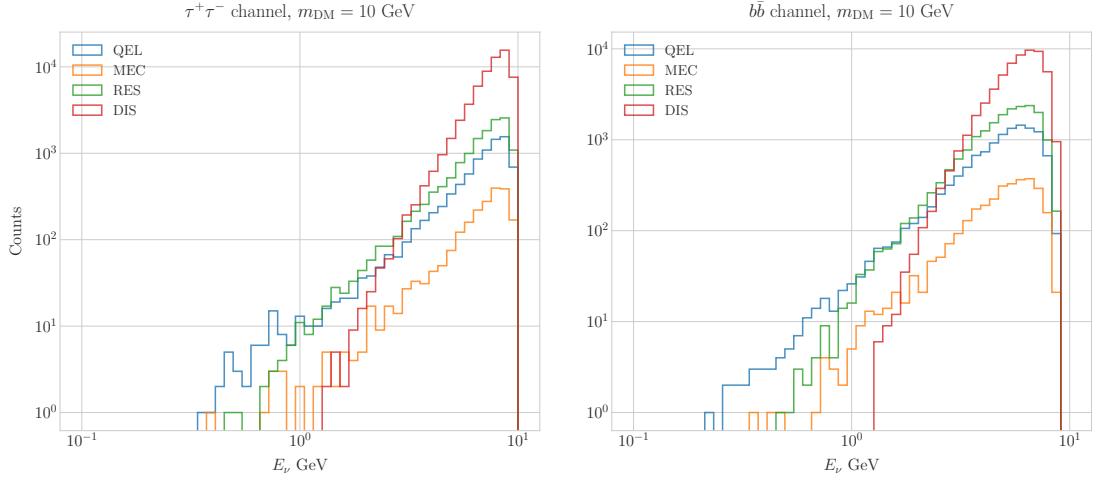


Figure 5.10: Distribution of the muon neutrino energies from the $\tau^+\tau^-$ (left panel) and $b\bar{b}$ (right panel) annihilation channels, for $m_{\text{DM}} = 10 \text{ GeV}$, separated by CC interaction type: QE (blue), MEC (orange), RES (green) and DIS (red).

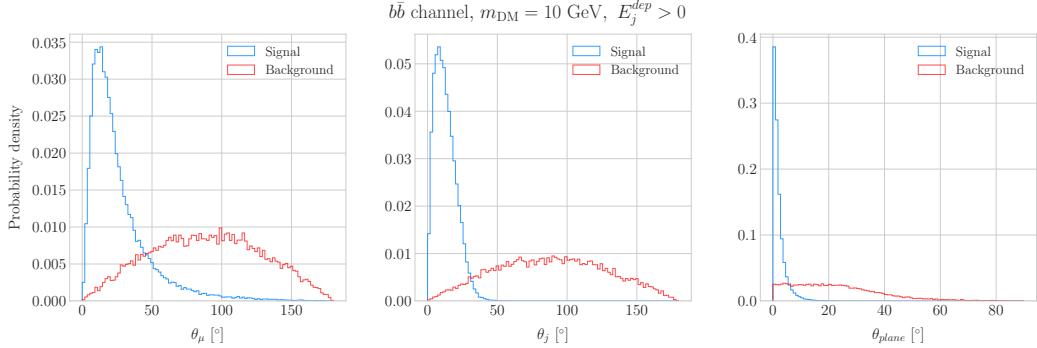


Figure 5.11: Distributions of θ_μ (left panel), θ_j (central panel) and θ_{plane} (right panel) for the $b\bar{b}$ sample with $m_{\text{DM}} = 10 \text{ GeV}$ (blue) and the atmospheric background (red).

1917 the samples, based on the different interaction modes and contents of the final state.

1918 Thus, I consider a CC DIS sample and a single proton CC QE sample.

1919 5.5.1 DIS-like events

1920 To begin with, I consider the high energy part of the spectrum. In this region DIS events
 1921 dominate, i.e. interactions of the form $\nu_\mu + q_d(\bar{q}_u) \rightarrow \mu^- + q_u(\bar{q}_d)$. Therefore, our final
 1922 estates will contain a muon and a hadronic jet from the fragmentation of the outgoing
 1923 quark. As all these events have $E_\nu \gtrsim 1 \text{ GeV}$ the momentum transfer to the remnant
 1924 nucleus is negligible, for this reason the neutrino energy can be effectively reconstructed

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

1925 just taking into account the momenta of the muon and the jet. This technique was
 1926 successfully used in Ref. [150] to select monoenergetic DM solar neutrino events from
 1927 $\nu\bar{\nu}$ annihilation channels.

1928 Using momentum conservation one sees that the plane generated by the momenta
 1929 of the muon and the jet needs to also contain the momentum of the neutrino. As we
 1930 are interested in neutrinos coming from the Sun, the momentum of the neutrino can be
 1931 regarded as known beforehand. This will allow us to define the angle of the outgoing
 1932 muon and jet with respect to the incoming neutrino. Moreover, one can also use that
 1933 information to reject poorly reconstructed jets, checking for deviations of these from the
 1934 momentum conservation plane.

1935 To account for the limited angular resolution of the detector, I smeared the momenta
 1936 of the muons and hadrons. In a LArTPC muons are expected to be tracked with high
 1937 precision, therefore I take the associated angular resolution to be 1° . In the case of
 1938 jets, it is expected that for the hadrons dominating the cascade a detector like DUNE
 1939 has an angular resolution between 1° to 5° [85], so I take the latter, more conservative,
 1940 estimate.

1941 As a first selection step, I will just take into account particles with kinetic energies
 1942 above the detection threshold of DUNE. For muons and photons the specified threshold
 1943 energy is 30 MeV, for charged pions 100 MeV and for other hadrons 50 MeV [85]. This
 1944 way, if the outgoing muon in a certain event has an energy lower than the required
 1945 threshold I will drop such event. For the case of hadrons and photons, I will only require
 1946 to have at least one particle above the energy threshold, so then one can compute the
 1947 jet momentum using the (smeared) momenta of the N particles above threshold as:

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

1948 Additionally, I will also define an estimation of the deposited hadronic energy as:

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

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

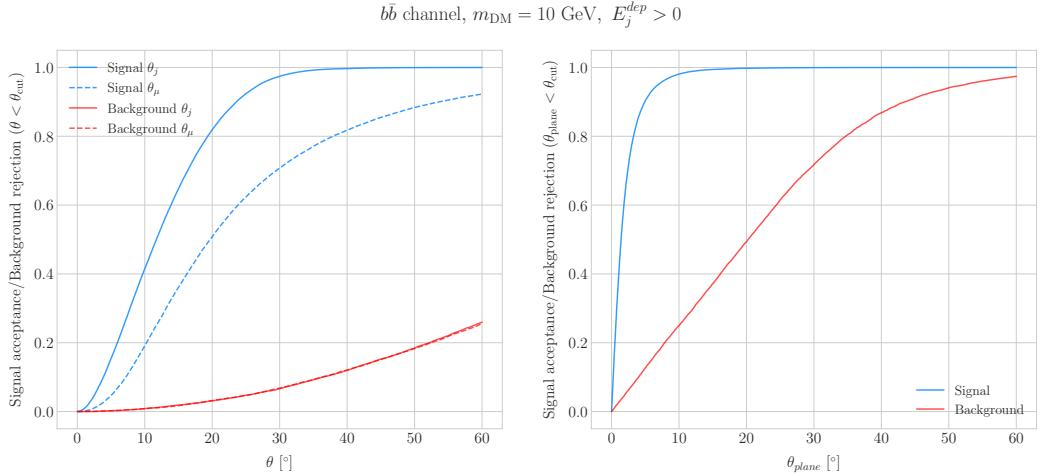


Figure 5.12: Left panel: signal efficiencies (blue lines) and background rejections (red lines) for events passing the cuts $\theta < \theta_{cut}$ for the jet (solid lines) and muon (dashed lines) angles. Right panel: signal efficiency (blue line) and background rejection (red line) for events passing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

1949 This quantity is useful to select events with enough hadronic visible energy in the
 1950 detector. For events where most of the hadronic energy is scattered across plenty of
 1951 hadrons with individual energies below the detection threshold, this estimation will
 1952 give $E_j^{dep} \leq 0$. In these cases it could be expected that the jet momentum is poorly
 1953 reconstructed, and therefore I require events to pass the cut $E_j^{dep} > 0$.

1954 For the events I can compute the angles for the muon and jet with respect to the
 1955 incoming neutrino as:

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

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

1956 and the deviation from the momentum conservation plane as:

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

1957 In Fig. 5.11 I show some distributions of these quantities for the case of the $b\bar{b}$ sample
 1958 with $m_{DM} = 10$ GeV (blue histograms) and for the atmospheric backgrounds (red).

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

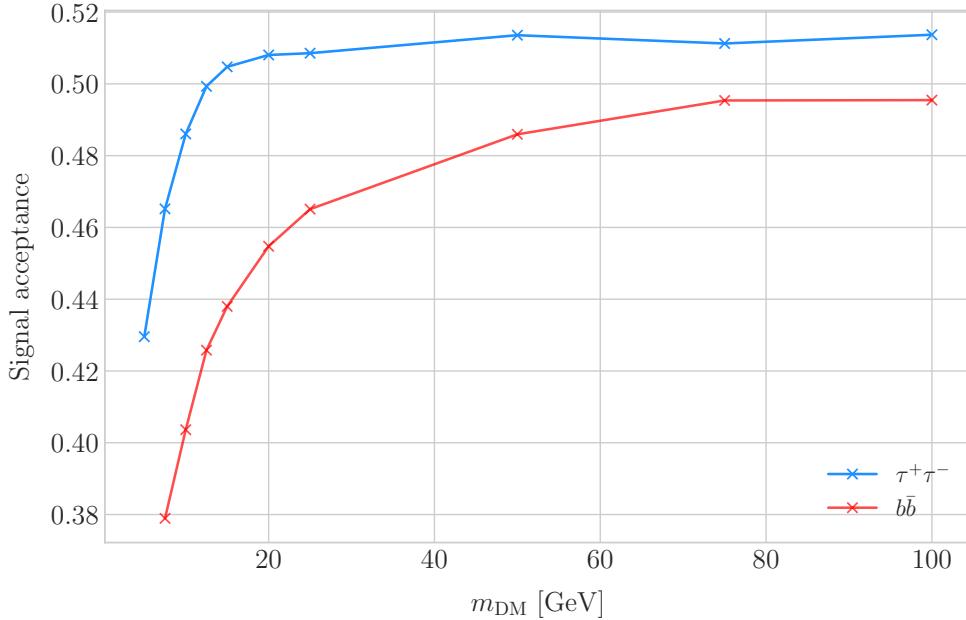


Figure 5.13: Signal efficiencies for the $\tau^+\tau^-$ (blue line) and $b\bar{b}$ (red line) DIS samples as functions of the DM mass, m_{DM} , obtained by applying the optimal angular cuts $\theta_\mu < 27^\circ$, $4^\circ < \theta_j < 26^\circ$ and $\theta_{\text{plane}} < 3.5^\circ$.

1959 In order to select the atmospheric events I followed the same criteria as for the signal
 1960 events. However, because in the signal case I used the true direction of the neutrino
 1961 as input, as it should be that of the Sun at that time and therefore known, in the
 1962 atmospheric case I used a set of solar positions as our ansatz for the neutrino direction.
 1963 From the distributions, one can see that the muon and the jet for the signal events are
 1964 predominantly forward and also that the deviations from the momentum conservation
 1965 plane are peaked at zero, as one should expect.

1966 Now, I can start applying cuts to maximise our signal selection efficiency while at
 1967 the same time I try to minimise the amount of atmospheric background events passing
 1968 the selection. To this end, I will need to find some lower and upper cuts for θ_j and
 1969 θ_μ and an upper bound for θ_{plane} . In Fig. 5.12 I show how upper bound cuts in the
 1970 different angular variables affect the signal efficiency (blue lines) and the background
 1971 rejection (red lines). Notice that the signal efficiency behaves in a quite different way
 1972 when I apply cuts in the jet and the muon angles. On the contrary, the cuts on both
 1973 variables have a similar effect on the background rejection.

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

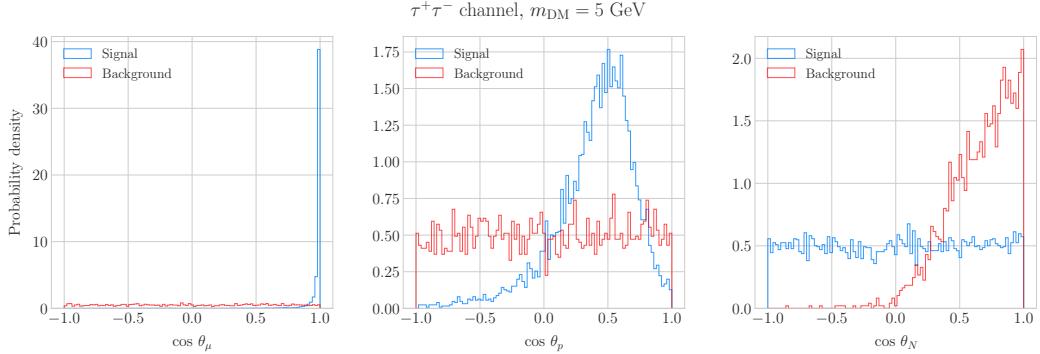


Figure 5.14: Distributions of $\cos \theta_\mu$ (left panel), $\cos \theta_p$ (central panel) and $\cos \theta_N$ (right panel) for the $\tau^+\tau^-$ QE sample with $m_{\text{DM}} = 5$ GeV (blue) and the atmospheric background (red).

1974 In order to obtain the optimal set of cuts, I perform a multidimensional scan. I
 1975 do this separately for the $\tau^+\tau^-$ and the $b\bar{b}$ samples. For each case, I scan the possible
 1976 cuts for each mass point and then I take the mean value of the signal efficiency for
 1977 each configuration, to get the mean efficiency for each set of cuts. I do a similar scan
 1978 for the atmospheric sample independently. Then, I take the sets of cuts such that
 1979 the background rejection achieved is greater than 99.8% and search for the one which
 1980 maximises the $\tau^+\tau^-$ and $b\bar{b}$ sample mean efficiencies. I found that with the cuts $\theta_\mu < 27^\circ$,
 1981 $4^\circ < \theta_j < 26^\circ$ and $\theta_{\text{plane}} < 3.5^\circ$ I get a background rejection of 99.80% while achieving
 1982 a 49.40% and 44.92% mean signal efficiencies for the $\tau^+\tau^-$ and $b\bar{b}$ signals respectively.

1983 In Fig. 5.13 I show the signal efficiencies as a function of the DM mass for the $\tau^+\tau^-$
 1984 (blue line) and the $b\bar{b}$ (red line) DIS events, after applying the cuts discussed above, as
 1985 well as the energy threshold and hadronic visible energy selections. One can see that
 1986 the efficiency grows with the mass, as annihilations of more massive DM particles will
 1987 produce a neutrino spectrum centered at higher energies, where DIS events dominate.
 1988 Notice also that the efficiency is higher for the $\tau^+\tau^-$ case at every mass point, as in
 1989 general this channel produces neutrinos at higher energies than the corresponding $b\bar{b}$
 1990 channel.

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

5.5.2 Single proton QE-like events

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

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

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

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

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

As in the previous case, I need to drop the events where the muon or the proton fall below the kinetic energy detection threshold [85]. Also, I again apply a smearing to the momenta of the particles, a 1% for muons and 5% for protons.

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

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

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

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

Fig. 5.14 shows the distributions of these angular variables for the $\tau^+\tau^-$ QE sample with $m_{\text{DM}} = 5$ GeV (blue) and the atmospheric background (red). Again, for the atmospheric events I used a random solar position as the ansatz for the incoming

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

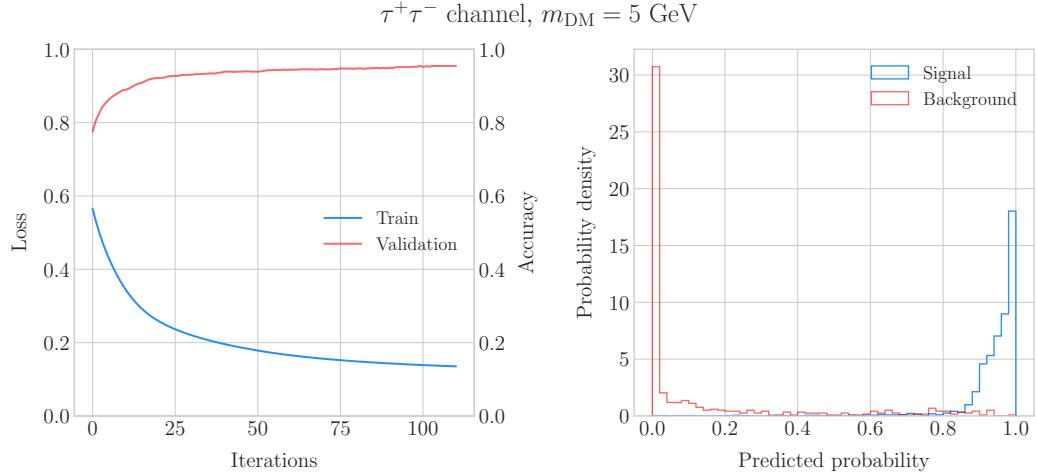


Figure 5.15: Left panel: value of the loss function for the training sample (blue line) and accuracy for the validation sample (red line) versus the number of iterations for the MLP classifier training. Right panel: distributions of the predicted probabilities assigned by the MLP classifier to the test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

2010 neutrino direction. Notice that now, opposed to the DIS case where the signal had very
 2011 sharp distributions for the variables considered, the shapes of the angular distributions
 2012 for signal and background are not that much different.

2013 This effectively means that the usual approach of applying simple angular cuts would
 2014 not work as well as in the previous situation. Therefore, as a possible solution, I tried to
 2015 use a multilayer perceptron (MLP) classifier to separate between signal and background
 2016 events. Thus, the power of the hypothesis test will serve as an estimate of the signal
 2017 efficiency, and in the same way one can take the size of the test to be our background
 2018 rejection.

2019 For each DM mass value and channel, as well as for the background sample, I divide
 2020 our events into training, validation and test samples. The input variables for the classifier
 2021 were the reconstructed neutrino energy from Eq. (5.33) and the angular variables defined
 2022 in Eqs. (5.35 - 5.37). I used the MLP classifier implemented in `scikit-learn` [151], with
 2023 a total of five hidden layers, the rectified linear unit activation function and adaptive
 2024 learning rate. In order to account for fluctuations due to artifacts in the training process I
 2025 repeated the training a thousand times for each sample, redefining each time the training,

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

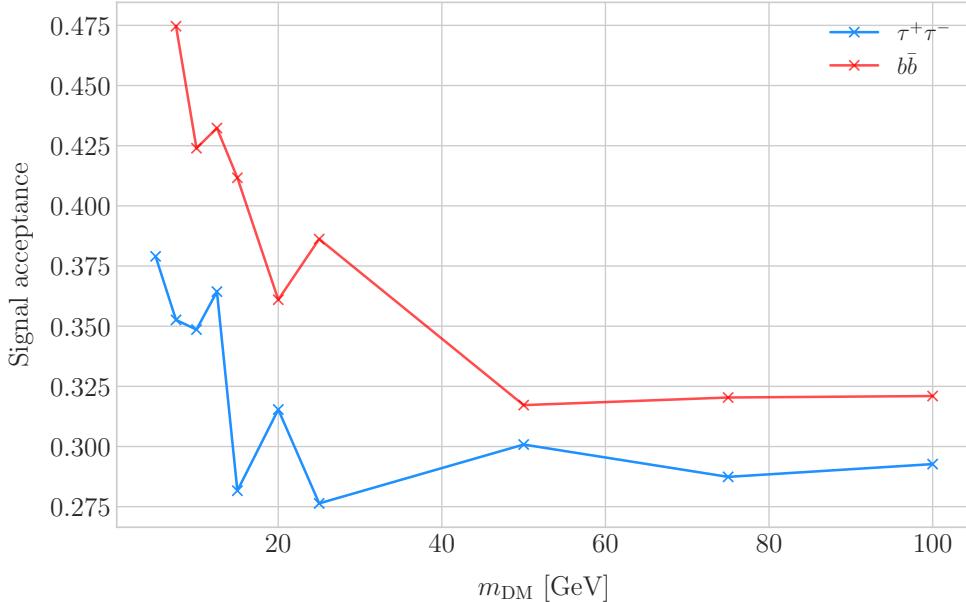


Figure 5.16: Signal efficiencies for the $\tau^+\tau^-$ (blue line) and $b\bar{b}$ (red line) single proton QE samples as functions of the DM mass, m_{DM} , obtained by requiring a minimum predicted probability from the MLP classifier of 0.97 in order to achieve a background rejection greater than 99.8%.

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

Applying this criteria for each sample, I obtain the mean signal efficiencies shown in Fig. 5.16. Notice that the efficiencies for the channel $\tau^+\tau^-$ (blue line) are consistently

5.5. HIGH ENERGY DM NEUTRINO SIGNALS

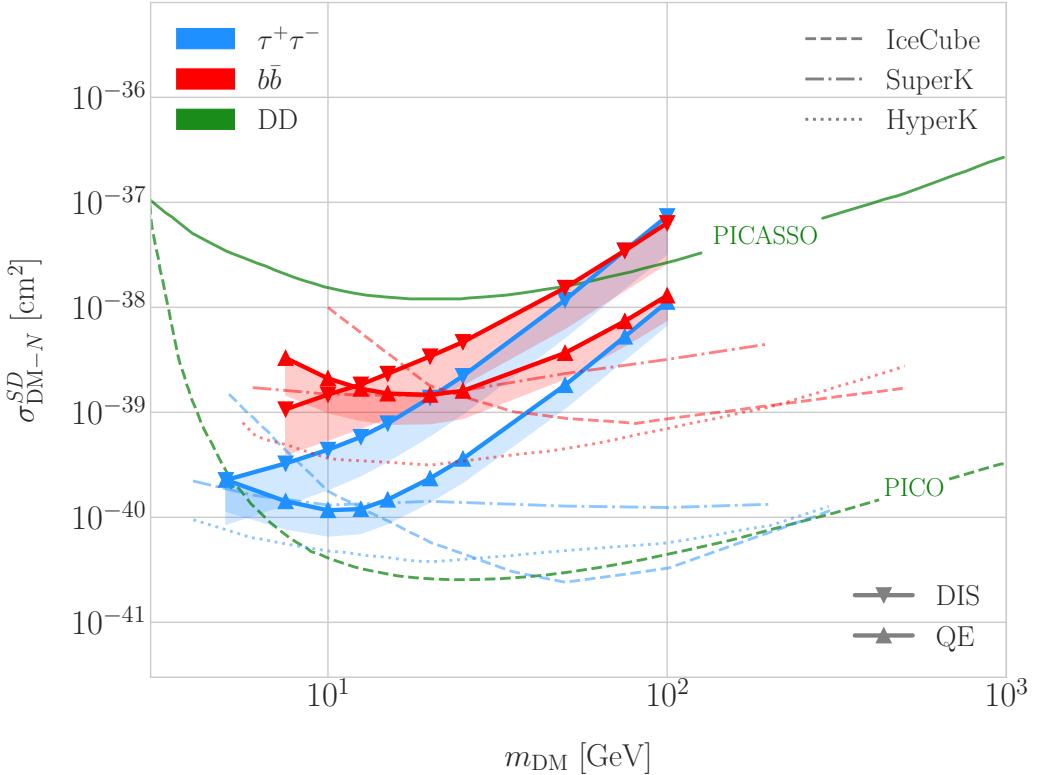


Figure 5.17: Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent DM-nucleon scattering cross section as a function of m_{DM} , for the annihilation channels $\tau^+\tau^-$ (blue) and $b\bar{b}$ (red) separated by interaction mode (up triangles denote DIS interactions whereas down triangles represent QE interactions). I also show the previous limits from IceCube [152] (solid lines) and Super-Kamiokande (dash-dotted), and the projected sensitivities for Hyper-Kamiokande [153] (dotted lines), as well as the direct detection limits from PICASSO [154] (solid green line) and PICO-60 C₃F₈ [155] (dashed green line).

2040 lower than the ones for the $b\bar{b}$ channel (red line). This can be due to the fact that, for
 2041 each DM mass point, the neutrino spectrum coming from the $b\bar{b}$ annihilation channel is
 2042 centered at lower energies when compared to the $\tau^+\tau^-$ spectrum. This directly translates
 2043 into more low energy neutrinos undergoing QE interactions, which give signals that can
 2044 be easily separated from the atmospheric background. This explanation also help us
 2045 understand why in both cases the signal acceptance drops when the DM mass increases.
 2046 In all cases, the background rejection took values between 99.8% to 99.9%. I will assume
 2047 a 99.8% background rejection value in all cases to keep our estimation conservative.

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

2048 5.5.3 Results

2049 In order to estimate the DM-nucleon cross section sensitivities in the present case I need
 2050 again to compute the expected number of background events. As I am now separating
 2051 events by interaction type Eq. (5.25) does not hold anymore, as in that case I integrated
 2052 over the total neutrino-argon cross section. In this instance, the expected background
 2053 events for DIS events is approximately given by:

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

2054 whereas for QE events we have:

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

2055 Now, using these together with Eqs. (5.26) and (5.27) one can obtain the 90% C.L.
 2056 upper limit on the total annihilation rate at equilibrium for both kind of events. Then,
 2057 applying the computed DM-nucleons capture rates I can translate these into limits on
 2058 the DM-nucleon cross section by means of Eqs. (5.2), (5.5) and (5.6).

2059 Fig. 5.17 shows the obtained limits on the SD DM-nucleon cross section for DUNE,
 2060 using the DIS (up triangles) and QE (down triangles) events both for the $\tau^+\tau^-$ (blue)
 2061 and the $b\bar{b}$ (red) samples, for an exposure of 400 kT yr. The coloured bands represent
 2062 the difference between the realistic efficiencies obtained and the limit of perfect signal
 2063 efficiency and optimistic background rejection given by Eq. (5.24). I also include the
 2064 corresponding current limits from IceCube [152] (solid lines) and Super-Kamiokande
 2065 (dash-dotted lines), as well as the projected sensitivities for Hyper-Kamiokande [153]
 2066 (dotted lines). For comparison, I also show the reported direct detection limits from
 2067 PICASSO [154] (solid green line) and PICO-60 C_3F_8 [155] (dashed green line).

2068 Notice that, for most of the mass range, the limits one can set by using the DIS
 2069 events are stronger than those of the QE interactions, except for the low mass part
 2070 of both the $\tau^+\tau^-$ and the $b\bar{b}$ curves where the QE events dominate. In general, the

5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

2071 expected sensitivity of DUNE for DM masses $\lesssim 25$ GeV surpasses the stronger current
 2072 indirect limits. However, experiments like Hyper-Kamiokande are foreseen to have an
 2073 overall better sensitivity in this kind of searches, as they have a bigger active volume
 2074 and accept a broader energy range.

2075 A pending question is what happens when we add the RES and MEC charged-current
 2076 interaction contributions. In that case it would probably be more convenient to split
 2077 the samples by final state interaction topologies. Also, another necessary improvement
 2078 would be adding a full detector simulation and reconstructions. This will also require
 2079 considering the effect of poorly reconstructed events or final states containing neutral
 2080 particles such that they mimic the desired topology at the reconstruction level.

2081 5.6 Example: Leptophilic Dark Matter

2082 In general, the capture rate of DM particles by the Sun via interactions with electrons is
 2083 several orders of magnitude smaller than the capture via DM-nucleus scattering. Thus,
 2084 it would be sub-leading even when nucleon capture is loop suppressed. As I showed in
 2085 Fig. 5.2, the capture rate via scattering off electrons only surpasses the capture rates
 2086 via DM-nucleons interactions for DM masses $\lesssim 100 - 500$ MeV.

2087 However, if one considers a model where DM-nucleon interactions are forbidden even
 2088 at loop level, then electron interactions will be the sole contributor to DM capture in
 2089 the Sun. One can describe such scenario where the DM particles couple to leptons but
 2090 not to the quark sector using effective operators.

2091 In general, assuming that the DM particle is a Dirac fermion, the dimension six
 2092 operators describing the interaction between two DM particles and two leptons can be
 2093 written as:

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

2094 where $G = 1/\Lambda^2$ is the effective coupling strength, Λ the cut-off of the effective field
 2095 theory and ℓ denotes any lepton. In principle, one should consider all the possible
 2096 Lorentz structures Γ_f^i in order to have a complete set of effective operators.

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

2097 However, some combinations will induce interactions with nucleons at loop level.
 2098 As we are specifically interested in interactions which forbid any communication with
 2099 the quark sector, I will not consider those [156]. In addition, some of the effective
 2100 operators give rise to velocity-suppressed scattering cross sections between DM particles
 2101 and leptons. I will also neglect those, as the suppression goes with the square of the DM
 2102 halo velocity which in units of the speed of light is $\sim 10^{-6}$.

2103 This way, the only Lorentz tensor structure that do not induce interactions with
 2104 quarks at loop level and gives a contribution to the scattering cross section that is not
 2105 velocity suppress is the axial-axial interaction. The effective Lagrangian is then given
 2106 by:

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

2107 where c_A^χ and c_A^ℓ are the couplings for the different species. As the DM coupling appears
 2108 as a common factor for any lepton choice, I will redefine the corresponding coupling c_A^ℓ
 2109 to absorb c_A^χ . Also, for simplicity, I will assume that the couplings between the DM
 2110 particles and the leptons are flavour independent, i.e. I have just two couplings, c_A^e for
 2111 charged leptons and c_A^ν for neutrinos.

2112 In the case of a scalar DM particle, the lowest order effective interaction with
 2113 leptons happens through a dimension five operator, generating scalar and pseudoscalar
 2114 interactions. However, the former induces interactions with quarks at two loop level
 2115 whereas the latter gives a velocity suppressed scattering cross section.

2116 From the effective Lagrangian in Eq. (5.41) it can be shown that the axial-axial
 2117 contribution to the scattering cross section for the fermionic DM and a charged lepton
 2118 is given by:

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

2119 If the DM interacts exclusively with fermions, then the only annihilation channels
 2120 that will give us a measurable neutrino flux coming out of the Sun are $\tau^+ \tau^-$ and $\nu \bar{\nu}$. The
 2121 former channel, already explored previously in the more mainstream scenario of the DM
 2122 capture via scattering off nucleons, is open only for $m_{DM} > m_\tau \simeq 1776.86 \pm 0.12$ MeV

5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

[157], 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. [126], 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 [158]:

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

where the sum includes all the possible lepton final states with mass m_{ℓ} .

Solving the Boltzmann equation for the evolution of the DM density gives as a solution a relic density of:

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

where $x_F = m_{\chi}/T_F$ being T_F the freeze-out temperature, g_* the number of relativistic degrees of freedom at freeze-out and a and b the terms in the annihilation cross section

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

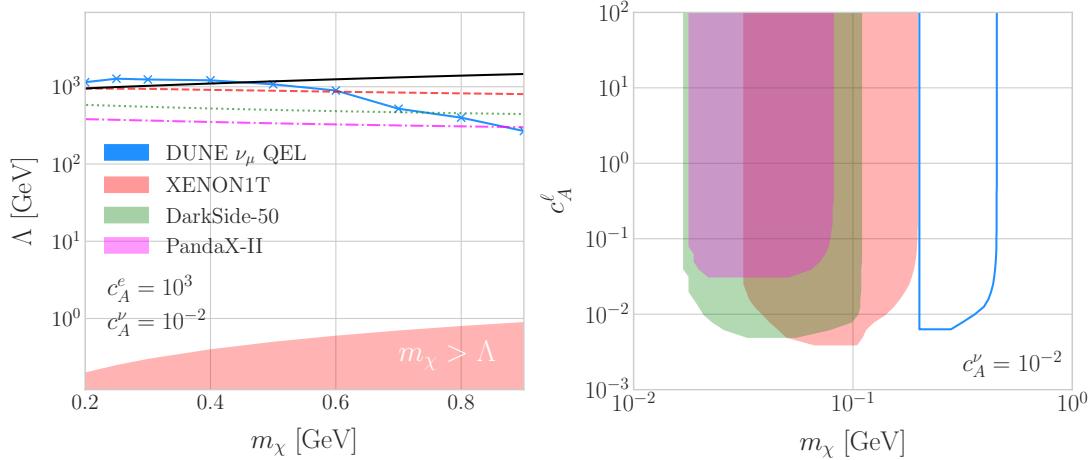


Figure 5.18: Left panel: Projected 90% confidence level sensitivity of DUNE (400 kT yr) to the scale Λ of an EFT containing only leptophilic DM axial-axial interactions (blue line), for the coupling values $c_A^e = 10^3$ and $c_A^\nu = 10^{-2}$. The black line represents the values for which the correct 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 [160] (red), DarkSide-50 [161] (green) and PandaX-II [162] (magenta) are also shown.

expansion $\sigma_{ann}|v| \approx a + bv^2 + \mathcal{O}(v^4)$. Using the current best fit for the relic DM density $\Omega_\chi h^2 = 0.1198 \pm 0.0012$ [159] one can use these relations to compute the required effective theory scale Λ at which the correct density is achieved for any combinations of m_χ and c_A^ℓ .

As discussed before, in the low DM mass region QE interactions dominate. Moreover, if I focus on direct annihilation to neutrinos, the energy of the muon neutrino flux is known as it must be equal to the mass of the DM particle, $E_\nu = m_\chi$. That way, now I do not need to use Eq. (5.33) in order to estimate the momentum transfer to the remnant nucleus, I can simply take:

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

To estimate the signal efficiency and background rejection for this case I used again the MLP classifier from `scikit-learn`, using the same specifications as before. The only difference now is that I add also the reconstructed neutrino energy as one of the

5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

2158 features to train the classifier with, because the characteristic monoenergetic flux for
2159 each m_χ value will help to distinguish between signal and background events.

2160 In this case, for masses below ~ 500 MeV I obtain a signal efficiency close to unity
2161 while keeping a background rejection of 99.9%. For bigger values of the mass, the signal
2162 efficiency drops significantly if I require to keep the background acceptance under 0.01%.
2163 However, because this kind of search is dominated by the background, sacrificing the
2164 signal acceptance to keep the background rejection to a minimum enhances the reach
2165 of the analysis. This way, for DM masses of the order of $m_\chi \sim 1$ GeV I end up with
2166 efficiencies as low as 1%.

2167 Now, estimating the number of background events using Eq. (5.39) one can go on
2168 and apply Eqs. (5.26) and (5.27) together with Eq. (5.42) to derive the sensitivity of
2169 DUNE to this kind of model. Fig. 5.18 (left panel) shows the potential reach of DUNE
2170 to constrain the EFT scale Λ this model containing only leptophilic DM axial-axial
2171 interactions (blue line), for a choice of couplings $c_A^e = 10^3$ and $c_A^\nu = 10^{-2}$. I also included
2172 the current limits on the DM-electron scattering cross section from XENON1T [160]
2173 (dashed red line), DarkSide-50 [161] (dotted green line) and PandaX-II [162] (dash-dotted
2174 magenta line), reworked with Eq. (5.42) to show their implications for the EFT scale.
2175 The values of Λ for which the correct DM relic density value is achieved for each mass are
2176 also shown (black line). This tells us that, for that specific choice of couplings, DUNE
2177 would be sensitive to DM configurations allowed by the relic density constraint up to a
2178 mass of $m_\chi \sim 400$ MeV.

2179 In Fig. 5.18 (right panel) I show similar limits for the excluded values of c_A^ℓ as a
2180 function of the DM mass, for a fixed $c_A^\nu = 10^{-2}$. I do not show the limits for other values
2181 of c_A^ν , as this parameter has little effect on the phenomenology at hand. From this view
2182 one can see that DUNE would be able to offer complementary information to the low
2183 energy DM-electron interaction searches performed by direct detection experiments, in a
2184 slightly higher mass range.

2185 With the present example, although it focuses on a very specific realisation of the DM
2186 interactions, I show the potential of DUNE to constrain exotic DM scenarios. Thanks

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

Table 5.1: Systematic uncertainties for the solar WIMP signal events. Table adapted from Ref. [163].

Systematic	Value
Form factor	Does not apply to SD [164]
Solar model	3% [164]
Local DM density	Not relevant for relative interpretations [164, 165]
Dynamics of solar system	Negligible [166]
Velocity distributions	20% at 20 GeV [164, 165]
Oscillation parameters	8% for $\tau^+\tau^-$, 5% for $b\bar{b}$ [167]
Neutrino interactions in the Sun	10%
Matter effects in the Earth	10%

2187 to its low backgrounds and superb angular resolution DUNE will be able to help with
2188 the systematic searches for dark sectors physics.

2189 5.7 Systematic uncertainties

2190 The estimation of the DM cross sections using neutrinos from WIMP annihilations
2191 inside the Sun is affected by systematic uncertainties from different sources. Surely, the
2192 atmospheric background estimation is also affected by systematic uncertainties. There
2193 are uncertainties common to both types of events, as well as others specific to each. In
2194 this section, I try to provide a comprehensive summary of the main sources of uncertainty
2195 for this analysis, which should be taken into account in any future extensions of the
2196 same.

2197 5.7.1 Systematic uncertainties in the solar WIMP signal

2198 The systematic uncertainties affecting the solar WIMP neutrino signal can be divided in
2199 two categories. On the one hand, we have those affecting the solar WIMP annihilation
2200 rate. On the other hand, there are the ones which modify the neutrino flux resulting
2201 from the annihilations reaching our detector.

2202 • **Uncertainties on the annihilation rate.** These include the astrophysical effects

5.7. SYSTEMATIC UNCERTAINTIES

Table 5.2: Systematic uncertainties for the solar WIMP atmospheric background events. Table adapted from Ref. [49].

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

that affect the normalisation of the solar DM neutrino flux. The main contributions are the solar model choice, the form factor uncertainties (only for SI searches), the gravitational effect of other planets, the local DM density (not relevant for relative comparisons, as it affects direct detection experiments in the same way), and the DM halo and dispersion velocities.

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

Table 5.1 summarises the contributions of the different sources of uncertainty for the signal events. These are the signal systematic uncertainties that have been taken into account in previous solar DM searches with neutrinos [163, 165, 167].

5.7.2 Systematic uncertainties in the atmospheric background

For the atmospheric background events, one needs to take into account the systematic uncertainties affecting the atmospheric ν_μ flux. These have been extensively studied in the context of atmospheric neutrino oscillation measurements. Among these, the energy-dependent flux normalisation uncertainty is the in the low energy regime. Other important contributions to the uncertainty come from the ratios between the muon to electron neutrino and the muon to anti-muon neutrino components of the flux. Additional

CHAPTER 5. DARK MATTER SEARCHES WITH NEUTRINOS FROM THE SUN

2221 uncertainty is introduced by the errors in the pion and kaon production rates calculated
2222 for the hadronic interactions of cosmic rays in the atmosphere [168].

2223 Table 5.2 shows a summary of the leading contributions to the uncertainty on the
2224 atmospheric muon neutrino flux, in the energy range relevant for this analysis.

2225 5.7.3 Common systematic uncertainties

2226 Finally, there are sources of uncertainty common to both signal and backgrounds. These
2227 have two different origins:

2228 • **Uncertainties on the neutrino cross section.** These are introduced by the
2229 modelling of the neutrino-nucleus interactions. In the context of the solar WIMP
2230 analysis, these have been estimated to be 10% for DM masses around 10 GeV
2231 [167].

2232 • **Uncertainties related to the detector.** They affect the measurement of the
2233 neutrino interaction and the final state particles produced. The main detector
2234 uncertainties relevant to this analysis are those of the energy and angular resolutions
2235 of the DUNE FD. Other effects, like the timing and triggering efficiencies, will
2236 also contribute to the uncertainties. The particular values these will take for this
2237 analysis need to be worked out in the context of DUNE.

Particle identification in ND-GAr

2240 *I am no bird; and no net ensnares me; I am a free human being with an
2241 independent will.*

²²⁴² — Charlotte Brontë, *Jane Eyre*

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

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

2254 6.1 GArSoft

2255 GArSoft is a software package developed for the simulation and reconstruction of events
 2256 in ND-GAr. It is inspired by the LArSoft toolkit used for the simulation of LArTPC
 2257 experiments, like the DUNE FD modules. It is based on `art`, the framework for event
 2258 processing in particle physics experiments [169]. Other of its main dependencies are

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2259 ROOT, NuTools, GENIE and Geant4. It allows the user to run all the steps of a generation-
2260 simulation-reconstruction workflow using FHiCL configuration files.

2261 6.1.1 Event generation

2262 The standard generator FHiCLs in GArSoft run the event generation and particle
2263 propagation simulation (i.e. Geant4) in the same job by default. However, it is possible
2264 to split them up if needed. The current version of GArSoft provides five different event
2265 generators, each of them producing `simb::MCTruth` products defined in NuTools. The
2266 available modules are:

2267 • **SingleGen**: particle gun generator. It produces the specified particles with a given
2268 distribution of momenta, initial positions and angles.

2269 • **TextGen**: text file generator. The input file must follow the `hepevt` format¹, the
2270 module simply copies this to `simb::MCTruth` data products.

2271 • **GENIEGen**: GENIE neutrino event generator. The module runs the neutrino-nucleus
2272 interaction generator using the options specified in the driver FHiCL file (flux file,
2273 flavour composition, number of interactions per event, t_0 distribution, ...). Current
2274 default version is v3_04_00.

2275 • **RadioGen**: radiological generator. It produces a set list of particles to model
2276 radiological decays. Not tested.

2277 • **CRYGen**: cosmic ray generator. The module runs the CRY event generator with a
2278 configuration specified in the FHiCL file (latitude and altitude of detector, energy
2279 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

2280 The module **GArG4** searches for all the generated **simb::MCTruth** data products, using
2281 them as inputs to the Geant4 simulation with the specified detector geometry. A constant
2282 0.5 T magnetic field along the drift coordinate is assumed. The main outputs of this step
2283 are **simb::MCParticle** objects for the generated Geant4 particles, **gar::EnergyDeposit**
2284 data products for the energy deposits in the HPgTPC and **gar::CaloDeposit** data
2285 products for the energy deposits in the ECal and muon system.

2286 6.1.2 Detector simulation

2287 The standard detector simulation step in GArSoft is all run with a single FHiCL, but
2288 the different modules can be run independently as well. First the **IonizationReadout**
2289 module simulates the charge readout of the HPgTPC, and later the **SiPMReadout** module
2290 runs twice, once for the ECal and then for the muon system, with different configurations.

2291 The **IonizationAndScintillation** module collects all the **gar::EnergyDeposit**
2292 data products, to compute the equivalent number of ionization electrons for each energy
2293 deposit. The **ElectronDriftAlg** module simulates the electron diffusion numerically
2294 both in the longitudinal and transverse directions and applies an electron lifetime
2295 correction factor. The induced charge on the nearest and neighbouring readout pads
2296 is modeled using the provided pad response functions. The digitisation of the data is
2297 then simulated with the **TPCReadoutSimAlg** module. By default, the ADC sampling
2298 rate used is 50.505 MHz. The resulting raw waveforms for each channel are stored with
2299 zero-suppression, in order to save memory and CPU time. The algorithms keep blocks
2300 of ADC values above a certain threshold, plus some adjustable additional early and late
2301 tick counts. The results of these three steps are **gar::raw::RawDigit** data products.

2302 For the ECal and the muon system the **SiPMReadout** module calls either the
2303 **ECALReadoutSimStandardAlg** or **MuIDReadoutSimStandardAlg** modules. These take
2304 all the **gar::CaloDeposit** data products in the corresponding detector and do the
2305 digitisation depending on whether the hit was in a tile or strip layer. They include single
2306 photon statistics, electronic noise, SiPM saturation and time smearing. The resulting
2307 objects are **gar::raw::CaloRawDigit** data products.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2308 6.1.3 Reconstruction

2309 The reconstruction in GArSoft is also run as a single job by default. It first runs the hit
2310 finding, clustering, track fitting and vertex identification in the HPgTPC, followed by
2311 the hit finding and clustering in the ECal and muon system. After those it produces the
2312 associations between the associations between the tracks and the ECal clusters.

2313 Focusing first on the HPgTPC reconstruction, the `CompressedHitFinder` module
2314 takes the zero-suppressed ADCs from the `gar::raw::RawDigit` data products. The
2315 reconstructed hits largely correspond to the above threshold blocks, however the hit
2316 finder identifies waveforms with more than one maximum, diving them in multiple hits
2317 if they dip below a certain threshold. The data products produced are of the form
2318 `gar::rec::Hit`. These are the inputs to the clustering of hits in the `TPCHitCluster`
2319 module. Hits close in space and time are merged, and the resulting centroids are found.
2320 This module outputs `gar::rec::TPCClusters` objects and associations to the input
2321 hits.

2322 The following step prior to the track fitting is pattern recognition. The module
2323 called `tpcvechitfinder2` uses the `gar::rec::TPCClusters` data products to find track
2324 segments, typically called vector hits. They are identified by performing linear 2D fits
2325 to the positions of the clusters in a 10 cm radius, one fit for each coordinate pair. A
2326 3D fit defines the line segment of the vector hit, using as independent variable the one
2327 whose sum of (absolute value) slopes in the 2D fits is the smallest. The clusters are
2328 merged to a given vector hit if they are less than 2 cm away from the line segment. The
2329 outputs are `gar::rec::VecHit` data products, as well as associations to the clusters. The
2330 `tpcpatrec2` module takes the `gar::rec::VecHit` objects to form the track candidates.
2331 The vector hits are merged together if their direction matches, their centers are within
2332 60 cm and their direction vectors point roughly to their respective centers. Once
2333 the clusters of vector hits are formed they are used to make a first estimation of the
2334 track parameters, simply taking three clusters along the track. The module produces
2335 `gar::rec::Track` data products and associations between these tracks and the clusters

6.1. GArSOFT

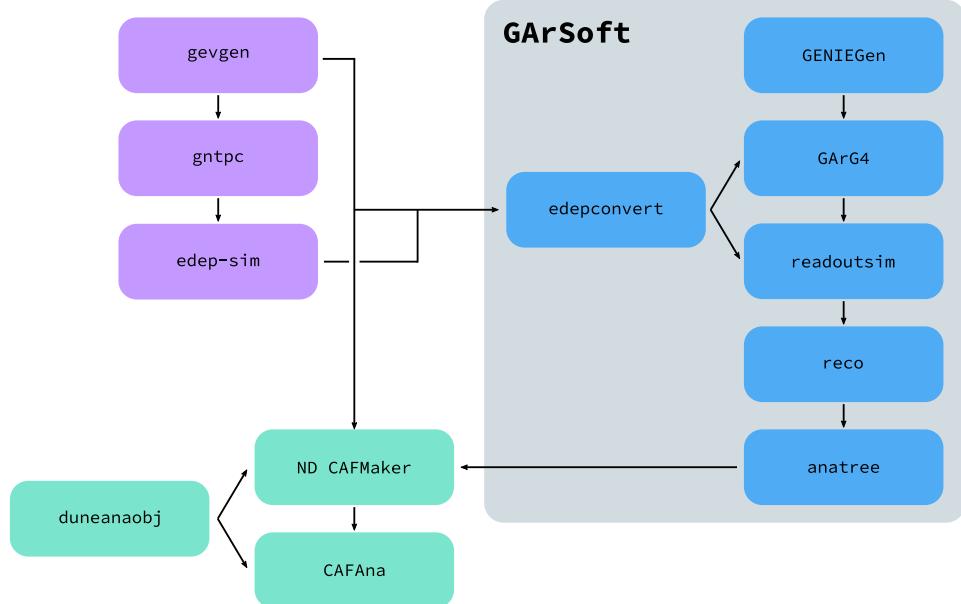


Figure 6.1: Schematic diagram showing the different modules involved in the ND-GAr production.

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

2351 and associations to the tracks and corresponding track ends.

2352 For the ECal and muon tagger, the `SiPMHitFinder` module runs twice with different
2353 configurations, adapted to the particular capabilities of both. The module simply takes
2354 the `gar::raw::CaloRawDigit` products, applies a calibration factor to convert the ADC
2355 counts to MeV and for the strip layer hits it calculates the position along the strip using
2356 the times recorded of both SiPMs. This module produces `gar::rec::CaloHit` data
2357 products. Next, these objects are used as inputs to the `CaloClustering` module. It
2358 merges the hits based on a simple nearest neighbours (NN) algorithm. For the resulting
2359 clusters it also computes the total energy and position of the centroid. The results are
2360 stored as `gar::rec::Cluster` data products, with associations to the hits.

2361 The last step in the reconstruction is associating the reconstructed tracks in the
2362 HPgTPC to the clusters formed in the ECal and muon system. The `TPCECALAssociation`
2363 module checks first the position of the track end points, considering only the points
2364 that are at least 215 cm away from the cathode or have a radial distance to the center
2365 greater than 230 cm. The candidates are propagated up to the radial position, in the
2366 case of clusters in the barrel, or the drift coordinate position, for the end cap cluster, of
2367 the different clusters in the collection using the track parameters computed at the end
2368 point. The end point is associated to the cluster if certain proximity criteria are met.
2369 This module creates associations between the tracks, the end points and the clusters.
2370 The criteria for the associations are slightly different for the ECal and the muon tagger.

2371 6.2 dE/dx measurement in the TPC

2372 Among the parameters extracted from the track fitting, ionisation is particularly useful
2373 for particle identification, as it is a function of the particle velocity. Although for the
2374 case of relativistic particles this dependence is not very strong, measuring the track on
2375 a large number of points may allow us to estimate the amount of ionisation accuratel.
2376 This, paired with a measurement of the momentum, may allow us to identify the particle
2377 type.

6.2. dE/dx MEASUREMENT IN THE TPC

2378 The first calculation of the energy loss per unit length of relativistic particles using a
 2379 quantum-mechanical treatment is due to Bethe [170]. Using this approach, the mean
 2380 ionisation rate of a charged particle traveling through a material medium is (using
 2381 natural units $G = \hbar = c = 1$):

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

2382 where N is the number density of electrons in the medium, e the elementary charge, m_e
 2383 is the electron mass, z the charge of the particle in units of e , β is the velocity of the
 2384 particle, $\gamma = (1 - \beta^2)^{-1}$ and I denotes the effective ionisation potential averaged over
 2385 all electrons. This relation is known as the Bethe-Bloch formula.

2386 From Eq. (6.1) one can see that the ionisation loss does not depend explicitly on
 2387 the mass of the charged particle, that for non-relativistic velocities it falls as β^{-2} , then
 2388 goes through a minimum and increases as the logarithm of γ . This behaviour at high
 2389 velocities is commonly known as the relativistic rise. The physical origin of this effect
 2390 is partly due to the fact that the transverse electromagnetic field of the particle is
 2391 proportional to γ , therefore as it increases so does the cross section.

2392 It was later understood that the relativistic rise could not grow indefinitely with γ .
 2393 A way to add this feature in the Bethe-Bloch formula is by introducing the so-called
 2394 density effect term. It accounts for the polarisation effect of the atoms in the medium,
 2395 which effectively shield the electromagnetic field of the charged particle halting any
 2396 further increase of the energy loss [171]. Denoting the correction as $\delta(\beta)$, one can rewrite
 2397 Eq. (6.1) as:

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

2398 In general, the form of $\delta(\beta)$ depends on the medium and its state of aggregation,
 2399 involving the usage of tabulated parameters and implicit relations [172].

2400 Another standard method to compute the amount of ionisation a charged particle
 2401 produces is the so-called photo-absorption ionisation (PAI) model proposed by Allison
 2402 and Cobb [173]. Within their approach, the mean ionisation is evaluated using a

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

2403 semiclassical calculation in which one characterises the continuum material medium by
2404 means of a complex dielectric constant $\varepsilon(k, \omega)$. However, in order to model the dielectric
2405 constant they rely on the quantum-mechanical picture of photon absorption and collision.
2406 Therefore, in the PAI model the computation of the ionisation loss involves a numerical
2407 integration of the measured photo-absorption cross-section for the relevant material.

2408 In a particle physics experiment, the typical way of determining the energy loss
2409 per unit length as a function of the particle velocity is studying identified particles
2410 over a range of momenta. Once we have established this relation we can use it for
2411 other, unknown particles. In this sense, it makes sense to have a regular mathematical
2412 expression for this relation that one can use.

2413 It happens that neither the Bethe-Bloch theory nor the PAI model from Allison and
2414 Cobb offer a close mathematical form for the ionisation curve. This is the reason why a
2415 full parametrisation of the ionisation curves can be useful. A parametrisation originally
2416 proposed for the ALEPH TPC [174] and later used by the ALICE TPC [175] group that
2417 manages to capture the features of the ionisation energy loss is:

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

2418 where P_i are five free parameters. Hereafter, we will refer to Eq. (6.3) as the ALEPH
2419 dE/dx parametrisation.

2420 6.2.1 Energy calibration

2421 In order to obtain the amount of energy loss by a charged particle due to ionisation
2422 in our TPC we need to determine the conversion between the charge deposited in our
2423 readout planes and the actual energy depositions. This procedure is known as energy
2424 calibration.

2425 In a general, the first step of the calibration involves a non-uniformity correction,
2426 to make sure that the detector response is uniform throughout the TPC. These are
2427 typically divided into three categories, non-uniformities in the transverse YZ plane,

6.2. dE/dx MEASUREMENT IN THE TPC

2428 non-uniformities along the drift direction X and variations of the detector response
2429 over time (would not apply to us as the detector is not built yet). These would correct
2430 for effects such as electron diffusion and attenuation, space charge effects or channel
2431 misconfiguration. However, because at the moment I am only interested in making sure
2432 we recover a sensible result from our simulation, I will not apply uniformity corrections
2433 to our charge deposits.

2434 Other effects, like electron-ion recombination or ADC saturation, lead to a non-linear
2435 relation between the observed charge and the deposited energy in the detector, with the
2436 observed readout charge saturating at high ionisation energies. In this case, because we
2437 are dealing with gaseous argon and therefore recombination is not as important as in
2438 liquid, we do not simulate recombination effects in the TPC. Even so, the simulation of
2439 the electronic response will still introduce charge saturation, and one needs to correct
2440 for it in order to obtain the exact amount of energy loss due to ionisation.

2441 By default, the track fitting algorithm in GArSoft provides a `TrackIonization`
2442 object associated to each reconstructed track. It contains two collections of charge
2443 deposits, one for each fitting direction, consisting on pairs of charge values (dQ , in ADC)
2444 and step sizes (dx , in cm).

2445 In order to estimate the ionisation loss in the ND-GAr TPC, I have used an MC
2446 sample consisting of single, isotropic protons propagating in the TPC. The starting points
2447 of the protons were sampled inside a $50 \times 50 \times 25$ cm box centered at $(100, -150, 1250)$,
2448 and their momenta are uniformly distributed in the range $0.25 - 1.75$ GeV. I ran the
2449 simulated sample through GArSoft's default detector simulation and reconstruction, and
2450 then a custom analyser module that extracts the ionisation data together with other
2451 reconstructed track information from the Kalman fit.

2452 For studying the energy loss of the protons I select the reconstructed tracks that
2453 range out (i.e. slow down to rest) inside the TPC. A characteristic feature of the energy
2454 loss profile of any stopping ionising particle is the so-called Bragg peak, a pronounced
2455 peak that occurs immediately before the particle comes to rest. From Eq. (6.1) we can
2456 see that this behaviour is expected, as the energy loss for non-relativistic particles is

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

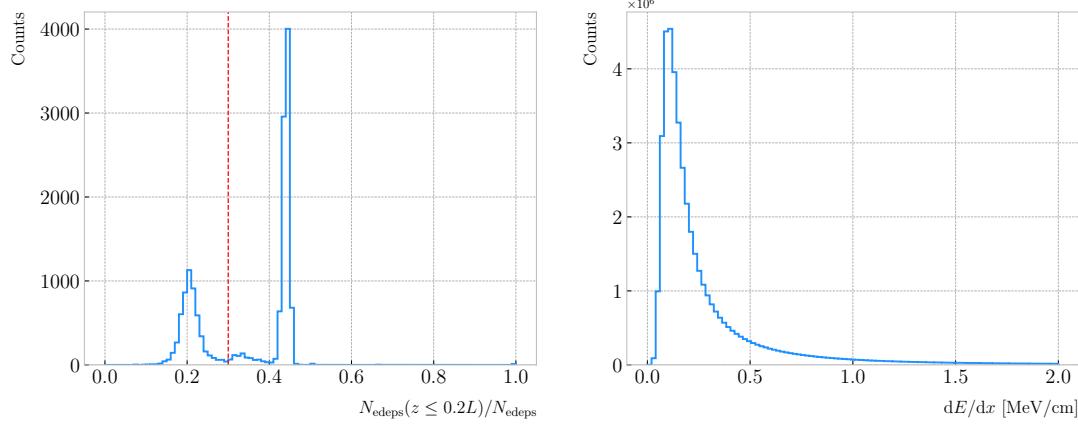


Figure 6.2: Left panel: distribution of the fraction of Geant4-level energy deposits per track with residual range less than 20% of the total track length, for the isotropic proton sample. Right panel: distribution of the ionisation per unit length of the energy deposits in the proton sample after removing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

2457 inversely proportional to β^2 . In data, a way of identifying the Bragg peak, and thus
 2458 select the stopping particles, is checking the number of energy deposits towards the
 2459 end of the track. In this case, I count the fraction of the Geant4 simulated energy
 2460 deposits with a residual range value (the distance from a given energy deposit to the
 2461 last deposit in the track trajectory) less than a 20% of the corresponding track length².
 2462 The distribution of this fraction of energy deposits for our proton sample is shown in
 2463 Fig. 6.2 (left panel). We can clearly see two well separated peaks in this distribution,
 2464 one centered at 0.2 and another, narrower, one centered at a higher value. The first
 2465 one corresponds to non-stopping protons, as in that case the number of energy deposits
 2466 towards the end of the track is uniformly distributed due to the absence of the Bragg
 2467 peak. In that way, I apply a cut in this distribution, requiring that at least 30% of the
 2468 simulated energy deposits sit in the last 20% of the tracks, to ensure that the Bragg
 2469 peak is present.

2470 Figure 6.2 (right panel) shows the distribution of the energy loss per unit length for
 2471 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

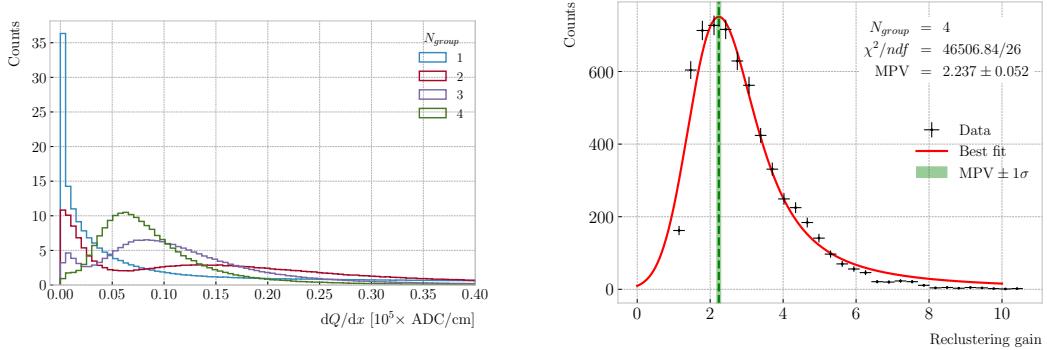


Figure 6.3: Left panel: distribution of the reconstructed ionisation charge per unit length for our MC stopping proton sample. The different colors indicate how many consecutive dQ/dx pairs were grouped together. Right panel: distribution of the median change in dQ/dx per track after $N_{group} = 4$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

it follows the expected shape of a Landau distribution, which describes the fluctuations of the ionisation energy losses [176]. 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

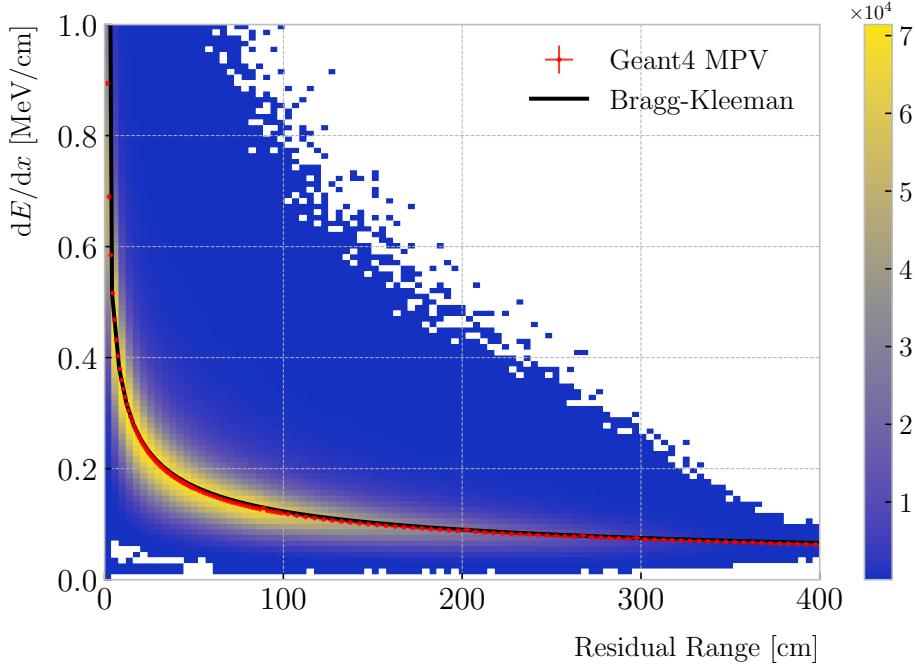


Figure 6.4: Distribution of the Geant4-simulated energy losses per unit length versus residual range for the stopping proton sample. The overlaid points represent the fitted most probable value of the dE/dx distribution in each residual range bin, whereas the curve is their best fit to the Bragg-Kleeman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

2491 dQ/dx per track changes. To do so, we can look at the ratio between the median dQ/dx
 2492 after and before the reclustering. Figure 6.3 (right panel) shows the median enhancement
 2493 in dQ/dx per track for the stopping proton sample in the case $N_{group} = 4$. Fitting a
 2494 Landau distribution convolved with a Gaussian³, I estimate the most probable value of
 2495 this ratio to be $G_{group} = 2.24 \pm 0.05$.

2496 At this point, I am left with determining the conversion between the charge deposits
 2497 per unit length dQ/dx and the energy deposits per unit length dE/dx . To this end, we
 2498 need a way of comparing the two. I can use the residual range z to get a prediction of
 2499 the most probable dE/dx by using the following empirical parametrisation [177]:

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

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

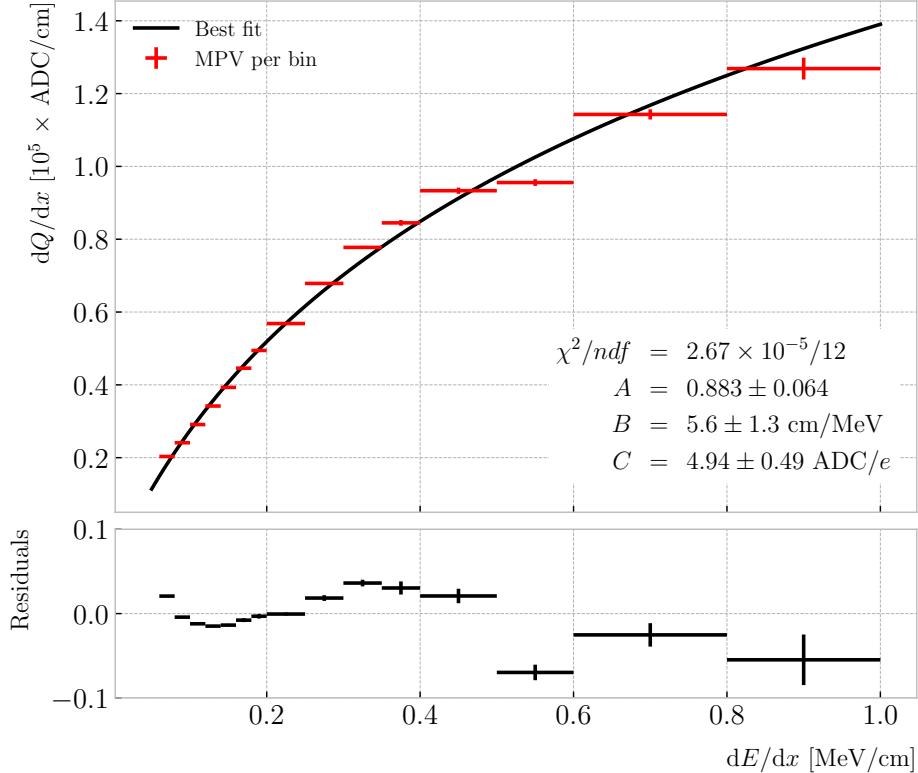


Figure 6.5: Fitted most probable dQ/dx values for each dE/dx bin (red points), obtained from the stopping proton sample. The overlaid curve (black line) represents the best fit to the logarithmic calibration function from Eq. (6.5).

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

2512 Bragg-Kleeman relation to those values (black line). The best fit is obtained for the
 2513 parameter values $p = 1.8192 \pm 0.0005$ and $\Lambda = 0.3497 \pm 0.0008 \text{ cm}/\text{MeV}^p$ ⁴.

2514 Having an analytical expression that relates the residual range to dE/dx , I can take
 2515 our reconstruction-level residual ranges from the stopping proton sample and compute
 2516 the most probable energy loss associated.

2517 In order to parametrise the charge saturation, we can use the following logarithmic
 2518 function inspired by the modified box model for recombination:

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

2519 where A and B are the calibration parameters we need to determine, W_{ion} is the average
 2520 energy to produce an electron-ion pair, G_{group} is the gain from the reclustering discussed
 2521 above and C is the calibration constant to convert number of electrons to ADC counts,
 2522 commonly refer to as gain (also to be obtained in the fit). In this case, I use a value
 2523 for the electron-ion production energy of $W_{ion} = 26.4 \text{ eV}$ [178]. This value, used in our
 2524 simulation as well, was measured for gaseous argon in normal conditions, and therefore
 2525 should be checked in the future to describe correctly the high-pressure argon-CH₄ mixture
 2526 of ND-GAr.

2527 For the calibration fit I follow a procedure similar to the previous one for Eq. (6.4).
 2528 Binning the dE/dx range, I fit a LanGauss distribution to the corresponding dQ/dx
 2529 distribution to obtain the most probable value. The resulting data points (red bars) are
 2530 shown in Fig. 6.5 (top panel), the horizontal error bars depict the width of the dE/dx
 2531 bin whereas the vertical bars represent the error associated to the most probable value
 2532 estimation. A fit to the logarithmic function in Eq. (6.5) is also shown (black line).
 2533 For this I weighted the data points using the inverse of their relative error, obtaining
 2534 a reduced chi-square value of $\chi^2/ndf = 2.22 \times 10^{-6}$. The best fit parameters I found
 2535 from this fit are $A = 0.883 \pm 0.064$, $B = 5.6 \pm 1.3 \text{ cm}/\text{MeV}$ and $C = 4.94 \pm 0.49 \text{ ADC}/e$.
 2536 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

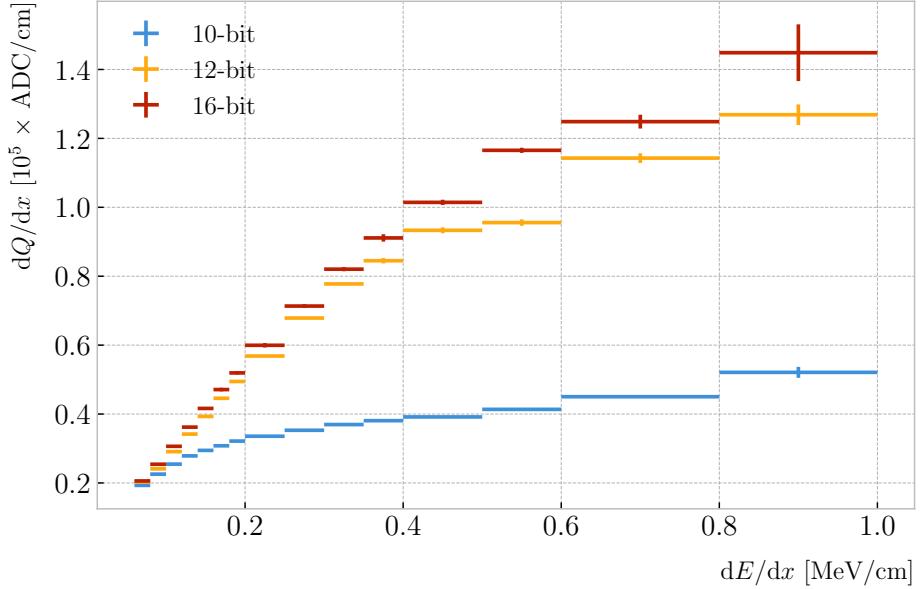


Figure 6.6: Fitted most probable dQ/dx values for each dE/dx bin for three different ADC bit limits, 10 (blue points), 12 (default, yellow points) and 16-bit (red points).

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

2537 The value for the gain I obtained from the fit is in reasonable agreement with our
 2538 expectation. This value is set in GArSoft to 5 ADC/e by default.

2539 One interesting thing to check is what induces this non-linear relation between charge
 2540 and energy. The only effects that modify the amount of electrons reaching the readout
 2541 planes in the simulation are the transverse diffusion and the finite electron lifetime.
 2542 Once the electrons reach the readout chambers, the pad response functions are applied,
 2543 together with an electrons-to-ADC conversion and the ADC saturation limit.

2544 By default, GArSoft applies a 12-bit ADC limit, which can be changed in the
 2545 simulation configuration. However, it can only be increased up to 16-bit, as we represent

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

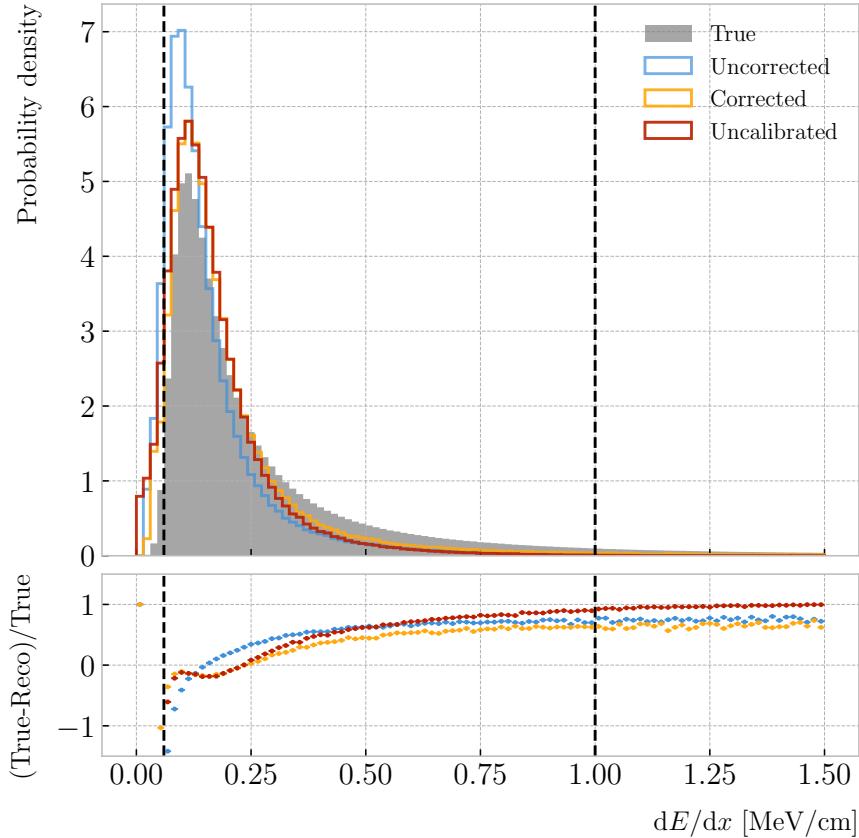


Figure 6.7: Top panel: area normalised dE/dx distributions for the true (solid grey) and the reconstructed energy deposits in the stopping proton sample, both after applying the calibration (blue) and the calibration and the normalisation correction (yellow). Also shown is the distribution obtained by 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.

2546 the ADC collection as a `std::vector<short>`. This way, I tried to change the saturation
 2547 parameter to see how it affects the relation between reconstructed charge and energy.
 2548 Figure 6.6 shows a comparison between the most probable dQ/dx for 10, 12 and 16-
 2549 bit ADC limits. As expected, the lower the limit is the sooner the charge saturates.
 2550 For higher ADC limits the relation between energy and charge remains linear up to
 2551 higher dE/dx values, but even for the 16-bit limit the saturation is noticeable for values
 2552 $\gtrsim 0.5$ MeV/cm.

2553 Table 6.1 shows the results of fitting the samples with 10 and 16-bits ADC limits to
 2554 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

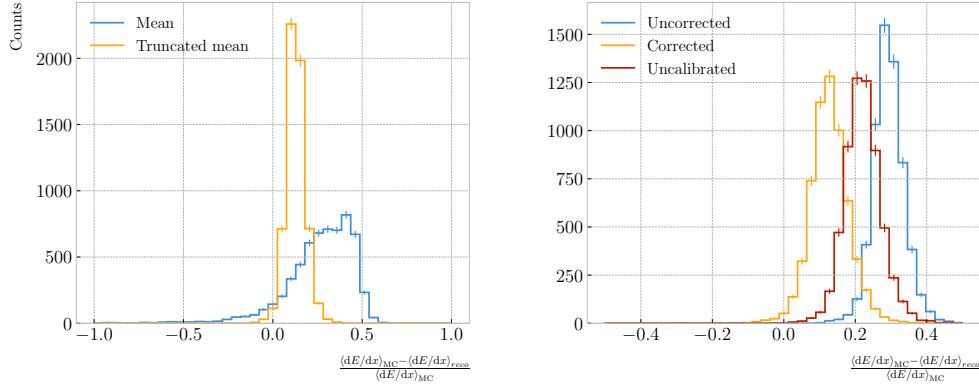


Figure 6.8: Left panel: fractional residuals between the true and the corrected dE/dx means (blue) and the 60% truncated means (yellow), for each event in the stopping proton sample. Right panel: fractional residuals between the true and the uncorrected (blue), corrected (yellow) and uncalibrated (red)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

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

6.2.2 Truncated dE/dx mean

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

However, estimating the most probable dE/dx value for each reconstructed track 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

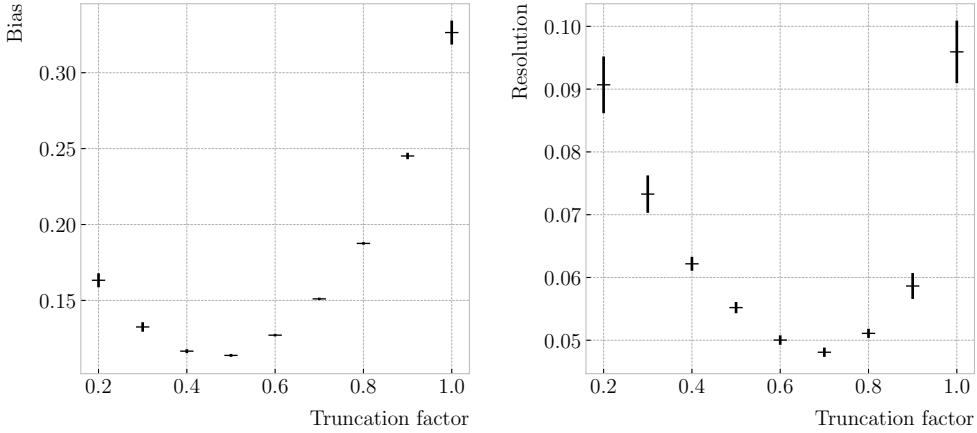


Figure 6.9: Estimated values of the mean dE/dx bias (left panel) and resolution (right panel) obtained using the corrected data from the stopping proton sample, for different values of the truncation factor.

2600 distributions. Therefore, one should perform e.g. a LanGauss fit to correctly estimate
 2601 the most probable values. Automating this kind of fits is often problematic, as they
 2602 usually incur in convergence problems. Moreover, the reconstructed dE/dx distributions
 2603 we obtain tend to have relatively small statistics, which may also produce poor fits. In
 2604 practice, doing these unsupervised fits may degrade our performance, and a more robust
 2605 method is preferred.

2606 A possibility could be taking the mean of the reconstructed dE/dx distribution for
 2607 each particle. The problem with this approach is that the high energy Landau tail,
 2608 combined with our limited statistics, can induce large fluctuations in the computation
 2609 of the mean. Imagine you have two protons with the same kinetic energy, but due to
 2610 reconstruction problems in one case you did not get as many charge deposits reconstructed
 2611 in its high ionisation loss region. If you do not remove the tails the computed dE/dx
 2612 means will be significantly different.

2613 In order to avoid those fluctuations, one can compute the mean of a truncated dE/dx
 2614 distribution instead. By keeping only a given fraction of the lowest energy deposits
 2615 we obtain an estimate of the mean energy loss that is more resilient to reconstruction
 2616 inefficiencies and statistical effects. Figure 6.8 (left panel) shows a comparison between
 2617 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

2618 60% lowest energy clusters (yellow line), for the stopping proton sample. The fractional
 2619 residuals are computed for each proton, taking the corresponding means using their
 2620 collections of true and reconstructed energy deposits. One can see that using the simple
 2621 mean translates into a high bias and uncertainty in the $\langle dE/dx \rangle$ estimation, whereas
 2622 applying the truncation reduces both significantly.

2623 Additionally, I performed a comparison between the 60% truncated mean dE/dx
 2624 obtained using the different calibration methods discussed earlier, namely the uncorrected
 2625 (blue), corrected (yellow) and uncalibrated (red) distributions. The results are shown
 2626 in Fig. 6.8 (right panel). While the widths of these distributions are similar, the bias
 2627 obtained for the corrected sample, i.e. calibration function and correction factor applied,
 2628 is a factor of ~ 2 lower than in the uncalibrated case and almost three times smaller
 2629 than for the uncorrected sample.

2630 The next step is to optimise the level of truncation we are going to apply to our
 2631 data. To do so, I used different truncation factors, i.e. the percentage of energy-ordered
 2632 reconstructed energy deposits we keep to compute the mean, on the corrected dE/dx
 2633 sample of the stopping protons. Then, following the same procedure of computing the
 2634 fractional residuals as before, I fitted the resulting histograms using a double Gaussian
 2635 function. This is simply the sum of two Gaussian functions of the type:

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

2636 I do not add the classical normalisation factor of the Gaussian, $1/\sqrt{2\pi}\sigma$, therefore
 2637 the amplitude A simply represents the maximum of the function. One of the two
 2638 Gaussian functions describes the core part of the distribution, while the other captures
 2639 the behaviour of the tails.

2640 For each truncation factor, I look at the bias and the resolution I obtain. I define
 2641 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

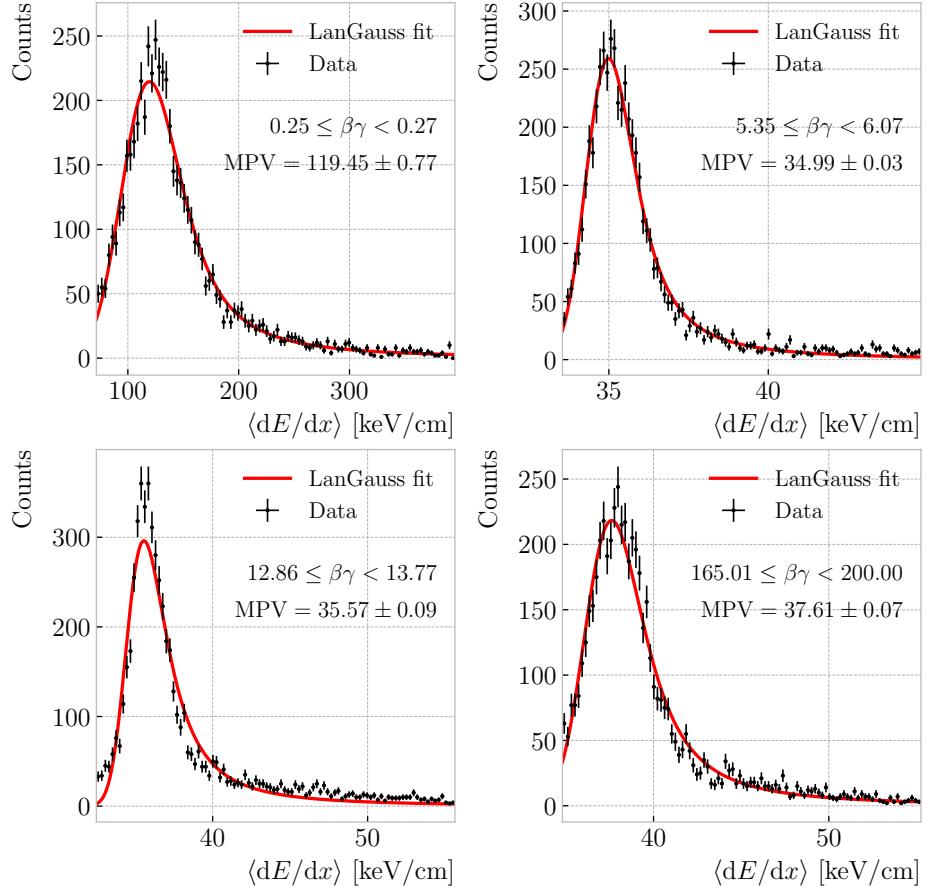


Figure 6.10: Examples of the truncated mean dE/dx LanGauss fits for various $\beta\gamma$ bins, from a simulated FHC neutrino sample.

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

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

6.2.3 Mean dE/dx parametrisation

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

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2652 For this, I used a sample of 10^5 reconstructed FHC neutrino events inside ND-GAr. In
2653 this case I cannot use the stopping proton sample, as we need to cover the full kinematic
2654 range of interest for the neutrino interactions in our detector.

2655 The original data does not contain an estimation of the velocity of the tracks, instead
2656 the tracks have a value for the reconstructed momentum and the associated PDG code
2657 of the Geant4-level particle that created the track. Therefore, one can select some of the
2658 particles in the data, in this case I selected electrons, muons, pions and protons, and
2659 compute β and γ using the reconstructed momentum and their mass. In terms of $\beta\gamma$
2660 the mean dE/dx does not depend on the particle species, so one can consider all the
2661 dataset as a whole. For this fit, I will express β in terms of the $\beta\gamma$ product as:

$$\beta = \frac{\beta\gamma}{\sqrt{1 + (\beta\gamma)^2}}, \quad (6.8)$$

2662 which can be easily proven from the definition of γ .

2663 Next, I bin the data in $\beta\gamma$. I chose a fine binning so as to capture the different
2664 features of the ionisation curve. Instead of fixing the bin width, I select them so each one
2665 has approximately the same statistics. Then, for each $\beta\gamma$ slice, I compute the median
2666 and the interquartile range (IQR) of the $\langle dE/dx \rangle$ distribution. Using these, I make a
2667 histogram in the range [median - IQR, median + 5 IQR], which I fit to a LanGauss
2668 function in order to extract the MPV. Using this range accounts for the asymmetric
2669 nature of the distributions, while also helps avoiding a second, lower maximum present
2670 at low $\beta\gamma$, probably a result of reconstruction failures.

2671 A few examples of these fits are shown in Fig. 6.10. The chosen values of $\beta\gamma$ sit in
2672 very distinct points along the $\langle dE/dx \rangle$ curve, going from the high ionisation region at
2673 low velocities (top left panel), to the minimum point (top right panel), the beginning of
2674 the relativistic rise (bottom left panel), and the plateau produced by the density effect
2675 (bottom right panel).

2676 I used the resulting most probable $\langle dE/dx \rangle$ values and the centres of the $\beta\gamma$ bins as
2677 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

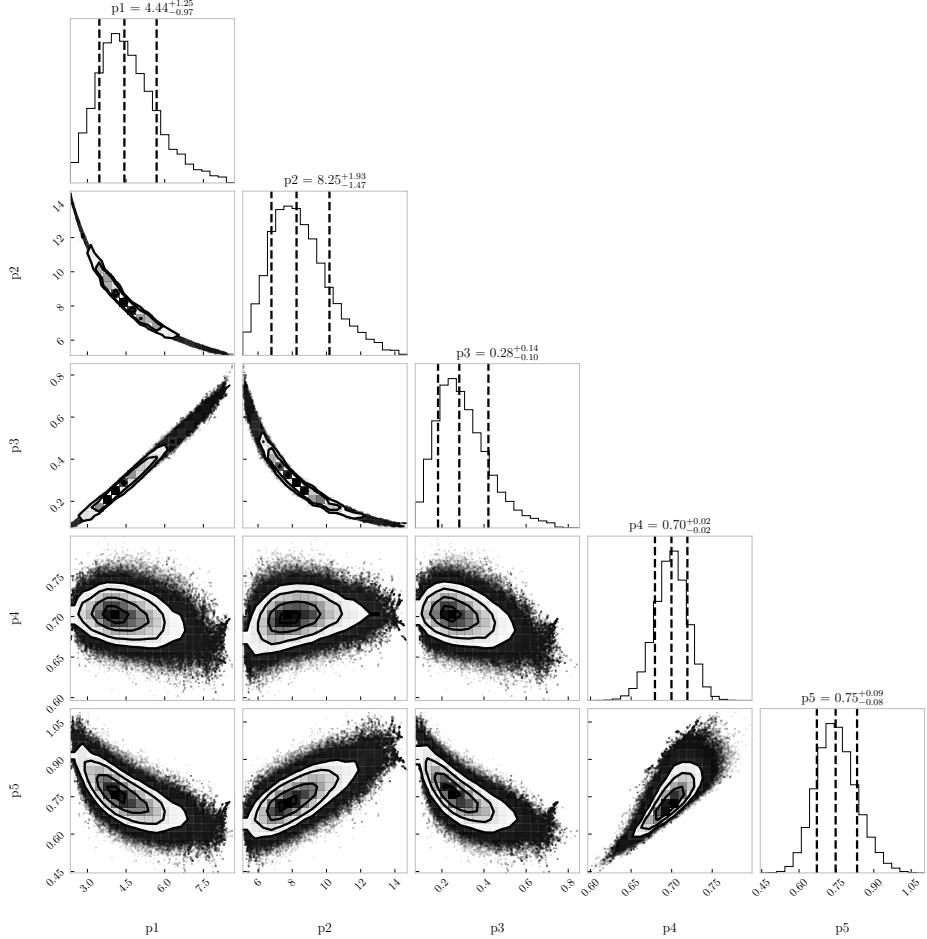


Figure 6.11: Resulting one and two dimensional projections of the posterior probability distributions of the ALEPH $\langle dE/dx \rangle$ parameters obtained by fitting the 60% truncated mean dE/dx values from a FHC neutrino sample in ND-GAr. The vertical dashed lines in the 1D distributions represent 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

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

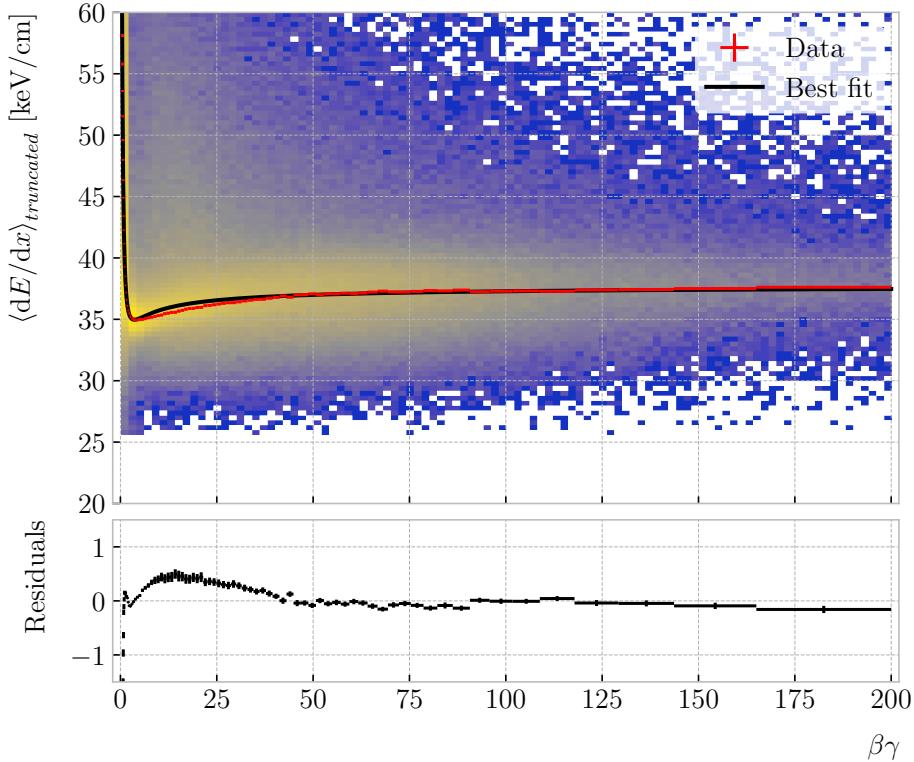


Figure 6.12: Truncated mean dE/dx obtained for the FHC neutrino sample as a function of the $\beta\gamma$ product (upper panel). Also shown are the fitted most probable values for each $\beta\gamma$ bin (red points) and the best fit obtained using the ALEPH parametrisation (black line). The residuals 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).

It is interesting to look at the results of the fit in momentum space, for the different particle species. Figure 6.13 shows the truncated mean dE/dx values versus the reconstructed momentum for the neutrino sample. Using a logarithmic scale for the momentum helps visualising the curves corresponding to the various particles. The resulting fits for electrons, muons, pions and protons are also shown (solid black lines). Notice that each curve stops at different momentum values, as the fits only extend up to $\beta\gamma = 200$ and translating this limit into momentum depends on the particle.

6.2. dE/dx MEASUREMENT IN THE TPC

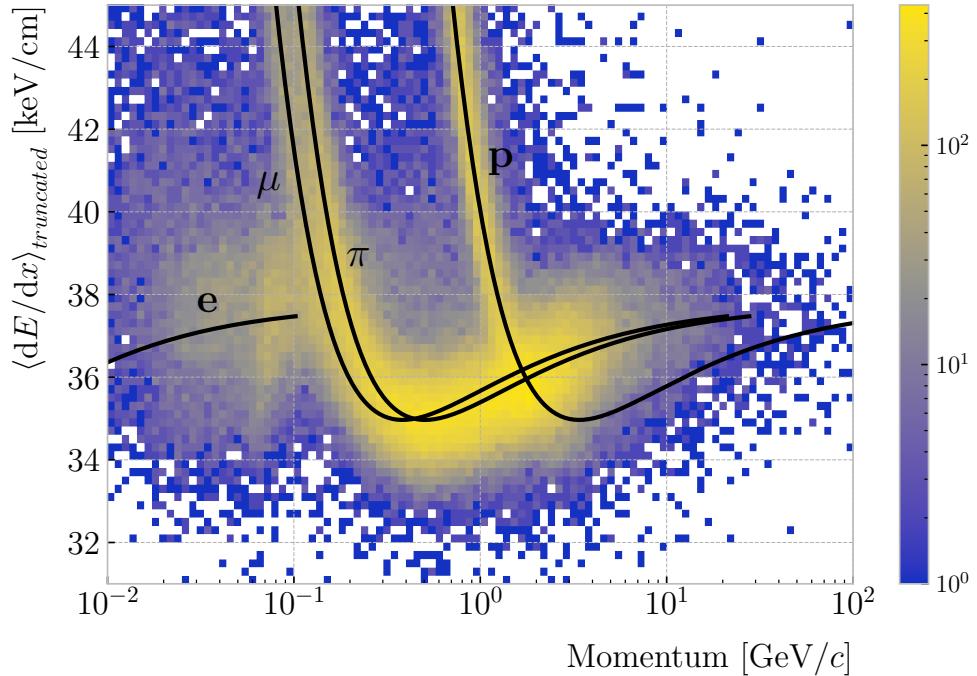


Figure 6.13: Distribution of the 60% truncated mean $\langle dE/dx \rangle$ versus reconstructed momentum for the FHC neutrino sample. The black lines indicate the predictions of the ALEPH parametrisation for electrons, muons, charged pions and protons.

From this plot, the particle separation power of the $\langle dE/dx \rangle$ measurement is evident.
 In the low momentum regime separating electrons, muons and pions is possible, while
 protons can be reliably identified up to $1.5 \text{ GeV}/c$.

Relevant to the separating power is the $\langle dE/dx \rangle$ resolution. This can be obtained from the fit, by taking the ratio of the difference between the expected energy loss for a given particle type and momentum and the measured value over the expectation. Then, performing a double Gaussian fit we can extract the bias and the resolution by means of Eq. (6.7). Figure 6.14 presents the values of the $\langle dE/dx \rangle$ bias (left panel) and resolution (right panel) as a function of the momentum for the true protons in the neutrino sample.

When compared to the values for the resolution obtained for the stopping proton sample (see e.g. Fig. 6.9), it appears that the performance now is much lower. For that low energy sample the resolution obtained was 5%, whereas now we only achieve those numbers for momenta $\geq 0.75 \text{ GeV}/c$. However, there are several differences between these two cases. The former was obtained for a single proton sample, with tracks are fully

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

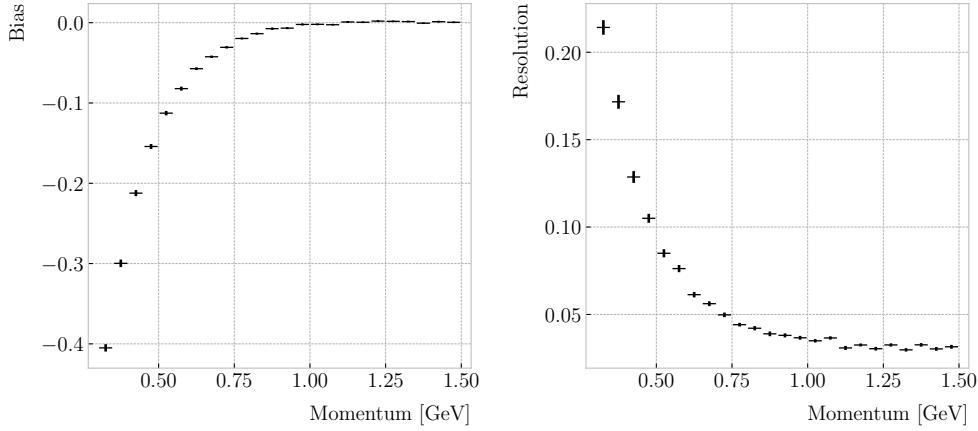


Figure 6.14: Estimated values of the mean dE/dx bias (left panel) and resolution (right panel) obtained for the true protons in the FHC neutrino sample.

2712 contained in the detector volume. On top of that, I refined the selection requiring a single
 2713 reconstructed track per event, which eliminates any misreconstruction effects. In this
 2714 case, we are dealing with tracks that may have fragmented, or even have contributions
 2715 from different true particles. Also, note that at low energies the $\langle dE/dx \rangle$ for protons is
 2716 much higher than it is for other particles. Therefore, having a poor resolution in that
 2717 range does not have an impact on the proton separation.

2718 6.3 Muon and pion separation in the ECal and MuID

2719 As it could be seen from Fig. 6.13, it is not possible to separate muons and charged pions
 2720 in the HPgTPC using dE/dx for momenta $\gtrsim 300$ MeV/ c . In ND-GAr, approximately
 2721 70% of the interactions in FHC mode will be ν_μ CC (compared to the 47% of $\bar{\nu}_\mu$ CC
 2722 interactions when operating in RHC mode), while 24% are neutral currents. Out of
 2723 these, around 53% and 47% of them will produce at least one charged pion in the final
 2724 state, respectively. Figure 6.15 shows a comparison between the spectra of the primary
 2725 muons and the charged pions for ν_μ CC interactions in ND-GAr producing one or more
 2726 charged pions. From this, one can see that (i) the majority of muons and charged pions
 2727 are not going to be distinguishable with a $\langle dE/dx \rangle$ measurement, and that (ii) particle
 2728 identification is necessary both to classify correctly the ν_μ CC events and identify the

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

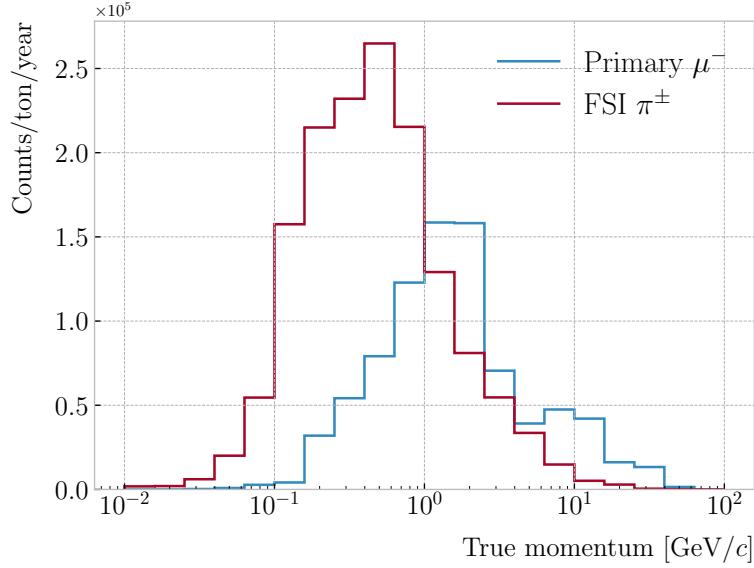


Figure 6.15: True momentum distribution for the primary muon in ν_μ CC $N\pi^\pm$ interactions inside the fiducial volume of ND-GAr (blue line), compared to the post FSI charged pion spectrum (red line).

2729 primary muon within them.

2730 ND-GAr features two other subdetectors which can provide additional information
 2731 for this task, namely the ECal and MuID. The current ECal design, described in (ref
 2732 section), consists of 42 layers, made of 5 mm of Pb, 7 mm of plastic scintillator and a
 2733 1 mm PCB board. The total thickness of this calorimeter is 1.66 nuclear interaction
 2734 lengths or 1.39 pion interaction lengths. The MuID design is in a more conceptual
 2735 stage, however it is envisioned to feature layers with 10 cm of Fe and 2 cm of plastic
 2736 scintillator⁶. With its three layers, it will have a thickness of 1.87 or 1.53 nuclear or pion
 2737 interaction lengths, respectively.

2738 Because pion showers are dominated by inelastic nuclear interactions, the signatures
 2739 of these particles in the calorimeter will look significantly different from those of muons.
 2740 Although our ECal is not thick enough to fully contain the hadronic showers of the
 2741 charged pions at their typical energies in FHC neutrino interactions, they can still be
 2742 used to understand whether the original particle was more hadron-like or MIP-like. In

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.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

2743 Fig. 6.16 I show two examples of energy distributions created by a muon (left panel)
2744 and a charged pion (right panel) of similar momenta interacting in the ECal. These
2745 figures represent the transverse development of the interactions. For each of them, I
2746 computed the principal component and centre of mass of the interaction, projecting
2747 the position of the hits onto the plane perpendicular to that direction, and taking the
2748 distances relative to the centre. It can be seen that the muon follows an almost MIP-like
2749 behaviour, being the central bin in the histogram the one with the highest deposited
2750 energy. On the other hand, the pion not only deposits more energy overall, but also this
2751 energy is more spread-out among the different hits. It is this kind of information that
2752 would allow us to tell apart muons from pions.

2753 This way, I identify three main action points that need to be addressed if one wants
2754 to use these detectors to distinguish between muons and charged pions. These are:

- 2755 1. the way we make the associations between tracks in the HPgTPC to the activities
2756 (what in GArSoft we call clusters) in the ECal and the MuID,
- 2757 2. what variables or features one can extract from the calorimeters that encapsulate
2758 the information we are interested about,
- 2759 3. and how to carry out the classification problem.

2760 6.3.1 Track-ECal matching

2761 One of the main players in the muon and pion separation is the way we associate clusters
2762 in the ECal to reconstructed tracks in the TPC. Missing some associations or making
2763 wrong ones can bias the ECal quantities that we can use for classifying particles. The
2764 current algorithm in GArSoft provides precise associations, i.e. most of the associations
2765 that it produces are correct, but it appears to miss an important number of associations
2766 (at least when using the default configuration).

2767 The current TPC track-ECal cluster association algorithm is divided in four parts.
2768 It first checks whether the track end point fulfils certain conditions to be extrapolated.
2769 There are two cut values in this step, one for the drift direction and other radial.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

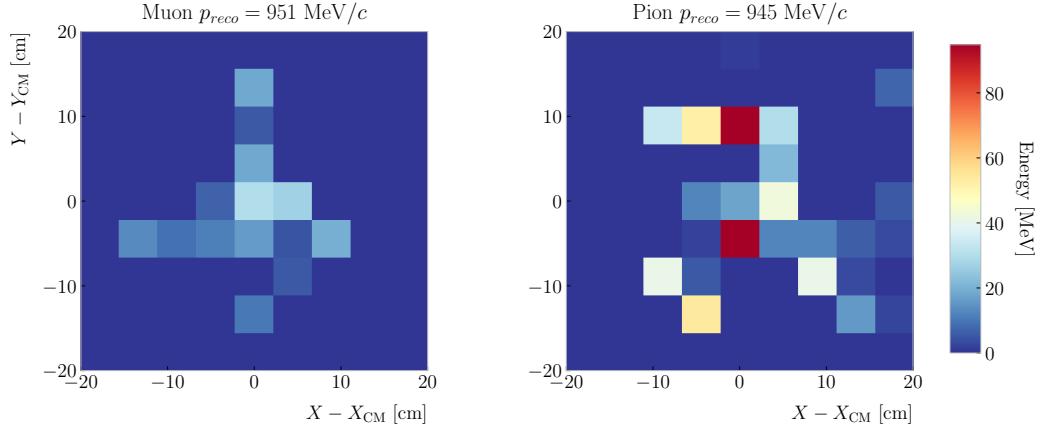


Figure 6.16: Distributions of energy deposits in the ECal for a muon (left panel) and a charged pion (right panel) with similar momenta. The energy is projected onto the plane perpendicular to the principal component of the hits, and the positions are relative to the center of the interaction.

If the point can be extrapolated, the code computes the coordinates of the centre of curvature using the Kalman fit estimates at the track end (y , z , $1/R$, ϕ , $\tan\lambda$). It then compares the distance between this and the cluster in the (z, y) plane with R . This introduces another cut in the perpendicular direction.

The next step is different for clusters in the barrel or in one of the end caps. If it is a barrel cluster the algorithm extrapolates the track up to the radial distance of the cluster. There are three possible outcomes, the extrapolated helix can cut the cylinder of radius r_{clus} two, one or zero times. I get the cut point that is closer to the cluster and check that it is either in the barrel or the end caps. Computing the difference between the x coordinates of the cluster and the extrapolated point, the module checks that this is not greater than a certain cut. If the cluster is in an end cap, I propagate the track up to the x position of the cluster. Then, the algorithm computes the angle in the (z, y) plane between the centre of curvature and the cluster, α , and the centre of curvature and the propagated point, α' . A cut is applied to the quantity $(\alpha - \alpha')R$.

If the cluster contains more than a certain number N of hits, I apply an extra cut to the dot product of the direction of the track at the propagated x value and the cluster direction.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

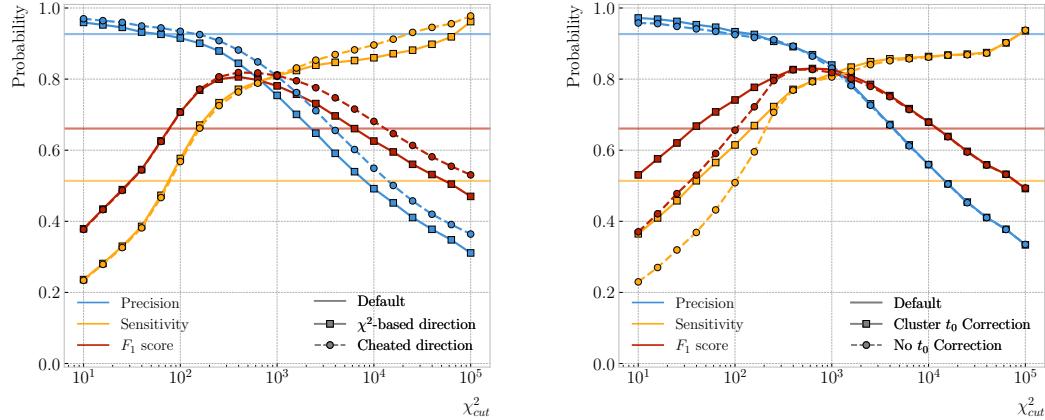


Figure 6.17: Left panel: comparison between the precision (blue), sensitivity (yellow) and F_1 score (red) obtained for the default (horizontal lines) and new algorithms, both with the χ^2 -based direction estimator (squares) and cheating the directions (circles), for different values of the χ^2 cut. Right panel: comparison of the performance of the new algorithm when applying the cluster t_0 correction (squares) and when (circles).

2787 The code makes sure to only associate one end of the track (if any) to a cluster.
2788 However, it can associate more than one track to the same cluster. This makes sense,
2789 as different particles can contribute to the same cluster in the ECal, but it makes it
2790 difficult to quantify the relative contributions of the tracks to a certain cluster.

2791 As a way of comparing the performance of this algorithm, a new, simpler association
2792 module was written. The goal was to have a simple and robust algorithm, which depends
2793 on as few parameters as possible and that can produce a one-to-one matching between
2794 tracks and ECal clusters.

2795 For each reconstructed track, the new algorithms applies the same procedure to the
2796 forward and the backward fits irrespective of their end point positions. It first gets the
2797 Kalman fit parameters at the corresponding end point together with the X position, x_0 ,
2798 $(y_0, z_0, 1/R, \phi_0, \tan\lambda)$.

2799 For each ECal cluster, I compute the radial distance to the centre of the TPC and
2800 find the ϕ value in the range $[\phi_0, \phi_0 + \text{sign}(R)\phi_{max}]$ that makes the propagated helix
2801 intersect with the circle defined with such radius. The (x, y, z) position of the helix for
2802 the ϕ value found (if any) is then computed. In case there are two intersections, I keep
2803 the one that minimises the distance between (y, z) and (y_c, z_c) .

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

2804 I then calculate χ^2 value based on the Euclidean distance between the propagated
2805 point and the cluster:

$$\chi^2/ndf = \frac{\sum_{n=0}^2 (x^{(n)} - x_c^{(n)})^2}{3}. \quad (6.9)$$

2806 If there was no intersection I store a -1 instead. In the end, for each reconstructed track
2807 in the event one ends up with two collections of χ^2 values, one for each ECal cluster
2808 and fit directions.

2809 The current code only supports having ECal clusters associated to one end of each
2810 track. We have two options to decide what track end to keep. The first one tries to
2811 cheat the selection, looking at the distance between the two track ends and the true
2812 start position of the associated MC particle. The second one keeps the track end with
2813 more χ^2 entries below the cut.

2814 This feature of only considering one track end limits the algorithm, making it not
2815 suitable for reconstructing events with particles originating outside the TPC. However,
2816 as for the moment the main concern of the group is the study of neutrino interactions
2817 off the gaseous argon, this is an acceptable assumption.

2818 In order to associate a cluster to a track, I take all clusters with a χ^2 value in the
2819 range $[0, \chi_{cut}^2]$. If a cluster has been assigned to more than one track we leave it with
2820 the one with the lowest χ^2 .

2821 This default behaviour of the algorithm can be modified to associate more than one
2822 track to each cluster. Not only that, but the χ^2 values can be used to assign relative
2823 weights to the different contributions.

2824 To evaluate the performance of the association method, I use a binary classification
2825 approach. In this case, I check the leading MC Track IDs associated to the reconstructed
2826 tracks and ECal clusters. I count an association as true positive (TP) if both Track
2827 IDs coincide. An association is considered false positive (FP) when the Track IDs are
2828 different. If a cluster has not been associated to any track but it shares the Track ID
2829 with a reconstructed track it is counted as a false negative (FN).

2830 For the testing, I used a sample of 10000 FHC neutrino events inside the HPgTPC.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

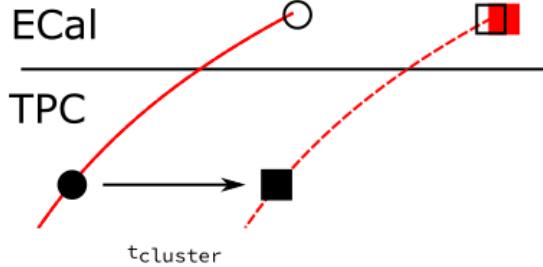


Figure 6.18: Schematics of a possible option to deal with track-ECal associations in non-zero t_0 neutrino interaction events, trying to correct for the drift direction uncertainty in a cluster-by-cluster basis using the cluster time, $t_{cluster}$.

2831 Figure 6.17 (left panel) shows the precision (blue line), sensitivity (yellow line), and F_1
 2832 score (red line) I obtained for different values of χ^2_{cut} . For comparison, the same metrics
 2833 computed for the default algorithm with the current configuration are also shown (dashed
 2834 lines). In the case of the new algorithm, I used both the χ^2 -based method to estimate
 2835 the track direction described earlier (square markers) and the cheated direction from the
 2836 Geant-level information (circle markers). For either of these we achieve similar values of
 2837 the precision compared to the old code, while having a considerably higher sensitivity.
 2838 It can be seen that cheating the direction of the tracks only makes a difference at high
 2839 χ^2_{cut} , past the optimal value of the cut around the F_1 score maximum. Therefore, I set
 2840 the χ^2 method as the default.

2841 One of the possible weak points of this approach is that it relies on the position along
 2842 the drift direction to make the decisions. Within the current ND-GAr design implemented
 2843 in GArSoft, the timing information is provided by the ECal. That effectively means
 2844 that prior to make the track-ECal associations the reconstructed x positions of the track
 2845 trajectories differ from the simulated ones by an amount:

$$x_{reco}^{(n)} - x_{sim}^{(n)} = v_{drift} t_0, \quad (6.10)$$

2846 where v_{drift} is the mean drift velocity in our medium and the initial time is in the range
 2847 $t_0 \in [0, t_{spill}]$ where t_{spill} is the spill length. For a 10 μs spill this translates into a
 2848 maximum 30 cm uncertainty on the drift direction position.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

2849 The current default in GArSoft sets $t_0 = 0$, but the functionality to randomly sample
2850 this within the spill time is in place. Therefore, we need to understand what is the impact
2851 of a non-zero t_0 on the associations algorithm and foresee possible ways of minimising a
2852 loss in performance.

2853 Figure 6.18 represents a possible option to tackle the association problem when
2854 having events with a non-zero initial time t_0 . The black and white circles represent the
2855 original points, whereas the squares indicate the corrected positions. The end points of
2856 the track and the propagated points up to the cluster radius are indicated using filled
2857 and unfilled markers respectively. The red square represents the position of the cluster.

2858 Here I try to correct for the drift coordinate position using the time associated to the
2859 cluster. Assuming that the drift time is much larger than the propagation time, $t_{cluster}$
2860 could be used as a good estimation of the t_0 . An alternative can be using the earliest
2861 time associated to a hit in said cluster. Doing this for each cluster before computing
2862 the χ^2 value could be used as an alternative to knowing the specific value of the t_0 , as
2863 when the association is correct this will provide the right correction but its impact is
2864 small enough to not change the position significantly in the case the cluster does not
2865 correspond to a given track.

2866 I tested the effect of this correction again using a sample of 10000 FHC neutrino
2867 events. Figure 6.17 (right panel) shows the precision (blue line), sensitivity (yellow line),
2868 and F_1 score (red line) for the case the cluster t_0 correction is applied (square markers)
2869 and for the no correction case (circle markers), as a function of χ^2_{cut} . In this case, the
2870 differences are particularly notorious at low values of the cut. It makes sense, as the t_0
2871 effect becomes subdominant when the distance we consider grows large. Overall, the
2872 correction increases the sensitivity while keeping the precision almost unchanged. As a
2873 result, I apply the t_0 correction to the generated samples as the default.

2874 6.3.2 Classification strategy

2875 The problem of the muon and charged pion separation has to be viewed in the broader
2876 context of the particle identification in our detector. Focusing on the beam neutrino

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

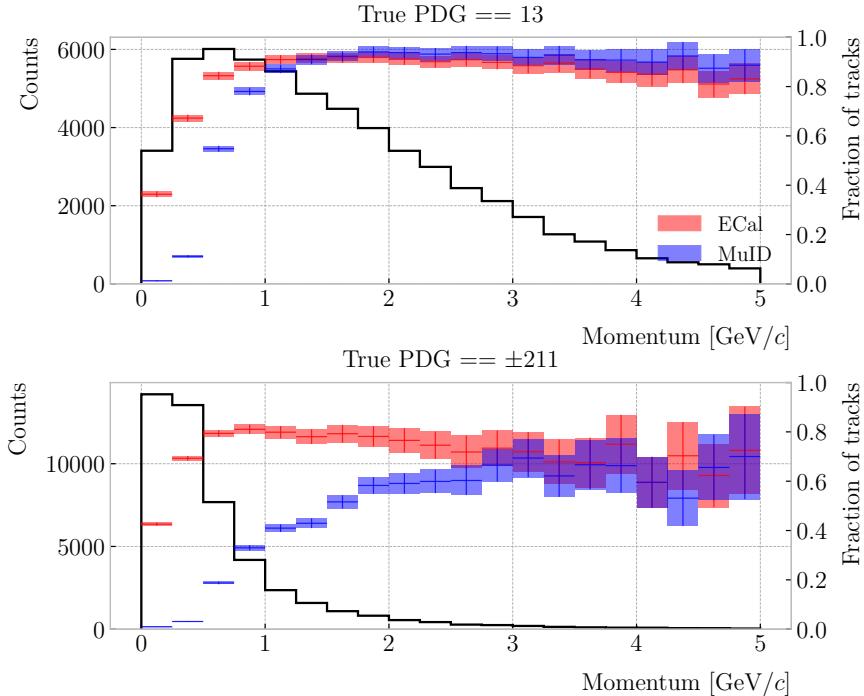


Figure 6.19: Momentum distribution for the reconstructed muons (top panel) and charged pions (bottom panel) in a FHC neutrino sample, together with the fraction of them reaching the ECal (red) and MuID (blue). Each entry corresponds to a reconstructed track, backtracked to a true muon or pion which has not produced any other reconstructed track.

interactions, it is clear that we are going to have muons and pions spanning a broad momentum range. Not only that, but we will also have other particles with similar characteristics that will make the classification even more challenging. Therefore, we are presented with a task that will depend heavily on the kinematic range we are looking at each time, as both the available information and the possible impurities of other particle species vary.

For instance, distinguishing muons from pions could be difficult at low momenta, as a great number of them do not reach the ECal. Therefore, we could think of tailoring a version of the classification for that particular case, which could be complemented with a dE/dx measurement. Likewise, for momenta $\gtrsim 1$ GeV muons and pions reach the calorimeters efficiently, but so do protons. Because of this, one can try to train another classifier for this energy range, and rely on other methods to remove as many of the

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

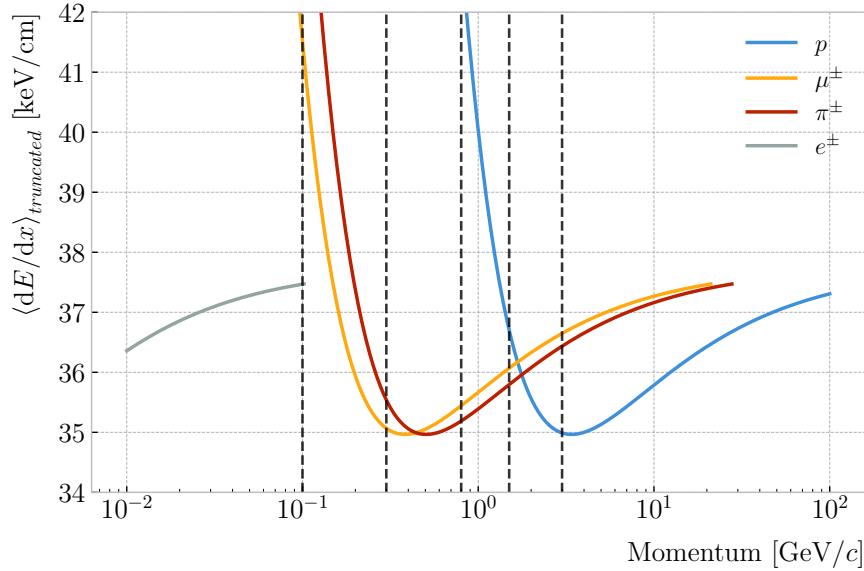


Figure 6.20: Predicted truncated mean dE/dx versus momentum, for electrons, muons, charged pions and protons, obtained using the ALEPH parametrisation. The vertical dashed lines represent the boundaries of the six regions used for the muon and pion classification training.

2889 protons as possible.

2890 Figure 6.19 shows the momentum distribution of the reconstructed muons (top) and
 2891 pions (bottom) in a FHC sample. It also contains the fraction of particles reaching the
 2892 ECal (red) and MuID (blue), for the different momentum bins. In Fig. 6.20 I show the
 2893 mean dE/dx of different particles as a function of the momentum, computed using the
 2894 ALEPH parametrisation with the best fit parameters found in Subsec. 6.2.2.

2895 Using these two figures as references, I decided to approach the classification by
 2896 dividing the problem into six different momentum regions. A summary of these can be
 2897 found in Tab. 6.2. The basic idea is to exploit all the information that is available in
 2898 each region and . For the problem at hand, I prepared separated samples of isotropic
 2899 single muons and pions, with momenta uniformly distributed along the corresponding
 2900 momentum range. Each sample contains 50000 events of the corresponding particle
 2901 species. I did not generate samples for the first region, as it is assumed that the separation
 2902 can be achieved using dE/dx only. For the last region, I generated particles up to a
 2903 momentum of 10 GeV/c , as that is well above the typical energies of muons and pions

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

Table 6.2: Momentum ranges and description of the PID approach assumed for the muon and pion classification task.

Momentum range	Description
< 0.1 GeV/c	All tracks can be separated with dE/dx
[0.1, 0.3) GeV/c	Use ECal for reaching muons and pions, dE/dx for the rest
[0.3, 0.8) GeV/c	Use ECal for muons and pions, dE/dx for protons
[0.8, 1.5) GeV/c	Use ECal and MuID for muons and pions, dE/dx for protons
[1.5, 3.0) GeV/c	Use ECal and MuID for muons and pions, ToF for protons
≥ 3.0 GeV/c	Use ECal and MuID for muons and pions, dE/dx and ToF for protons

2904 from FHC neutrino interactions in ND-GAr.

2905 Additionally, I prepared another sample of 100000 FHC neutrino events. For each
 2906 interaction, I select the reconstructed particles which were backtracked to true muons or
 2907 charged pions. I use this dataset to perform validation checks, to see how the models
 2908 trained with the single particle data generalise to a more realistic scenario.

2909 To tackle this classification problem, I make use of Boosted Decision Trees (BDT). A
 2910 decision tree uses a flowchart-like structure to make decisions based on some input data.
 2911 It starts from a root node, which represents the complete dataset, and then it splits
 2912 this based on the variable or feature which gives the best separation between classes,
 2913 creating two new nodes. The process repeats for each node until it reaches a certain
 2914 limit, like a maximum number of splits or some tolerance criteria. The last set of nodes
 2915 are often called leave nodes, and represent the final prediction of the classifier.

2916 Boosting refers to a family of methods to combine the predictions from multiple
 2917 classifiers, following a sequential approach where each new model learns from the errors
 2918 of the previous one. The process starts with a simple decision tree, which is used to
 2919 make predictions on the training data. Then, the data points misclassified by the first
 2920 model are assigned higher weights, and another decision tree is trained on the data with
 2921 adjusted weights. The predictions of the two trees are then combined, and the cycle
 2922 repeats for a predefined number of iterations. Gradient boosting uses the direction of
 2923 the steepest error descent to guide the learning process and improve the accuracy with

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

2924 each iteration.

2925 6.3.3 Feature selection and importance

2926 Using the reconstructed tracks as a starting point, I compute a number of ECal and
2927 MuID variables for each of them. As there can be more than one cluster associated to a
2928 track, what I do is collect all associated clusters and compute these variables from the
2929 complete collection of associated hits. For the MuID, because it only features three layers
2930 and typically there will be less hits, I also allow single hits to be associated with tracks⁷.
2931 I can roughly divide the variables in three types: energy-related, geometry-related and
2932 statistical. In the following, I briefly describe the variables related exclusively to the
2933 ECal:

2934 • Energy-related ECal

- 2935 – ECal total energy (ClusterTotalEnergy): sum of the energy of all the ECal
2936 hits.
- 2937 – Mean ECal hit energy (HitMeanEnergy): mean of the hit energy distribution.
- 2938 – Standard deviation ECal hit energy (HitStdEnergy): standard deviation of
2939 the hit energy distribution.
- 2940 – Maximum ECal hit energy (HitMaxEnergy): maximum of the hit energy
2941 distribution.

2942 • Geometry-related ECal

- 2943 – Mean distance hit-to-cluster (DistHitClusterMean): mean of the distance
2944 distribution between the hits and the corresponding cluster's main axis.
- 2945 – RMS distance hit-to-cluster (DistHitClusterRMS): root mean square of the
2946 distance distribution between the hits and the corresponding cluster's main
2947 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.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

- 2948 – Maximum distance hit-to-centre (DistHitCenterMax): maximum of the
2949 distance distribution between the hits and the centre of the TPC.
- 2950 – Time-of-Flight velocity (TOFVelocity): slope obtained when fitting a straight
2951 line to the hit time versus hit distance to the centre (i.e. $d = v \times t$).

2952 • Energy and geometry ECal

- 2953 – Radius 90% energy (Radius90E): distance in the hit-to-cluster distribution
2954 for which 90% of the total energy is contained in the hits that are closer to
2955 the axis (i.e. radius that contains 90% of the energy).

2956 • Statistical ECal

- 2957 – Number of hits (NHits): total number of hits associated to the track.
- 2958 – Number of layers with hits (NLayers): not really a count of all layers with
2959 hits but the difference between the last and the first layer with hits.

2960 Figure 6.21 shows the distributions of three different ECal variables, separating true
2961 muons (blue) and charged pions (red), for the five momentum ranges considered. I chose
2962 to show one feature from each category, namely the mean energy per hit (left column),
2963 the mean distance between the hits and the centre of the cluster (middle column), and
2964 the number of ECal layers with hits (right column). These give an idea of the separating
2965 power of the different features, and how it changes considerably with the energy. In
2966 the number of layers with hits distributions, the peak at 6 is due to the fact that the
2967 first six ECal layers sit inside the pressure vessel⁸. Therefore, some of the particles get
2968 stopped crossing it, never making it to the seventh layer.

2969 In the case of the MuID, because at low momenta a significant fraction of the particles
2970 do not make it past the ECal, I only consider the information coming from this detector
2971 for momenta ≥ 0.8 GeV/c, i.e. for the last three momentum regions. The variables I
2972 extract from it are the following:

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.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

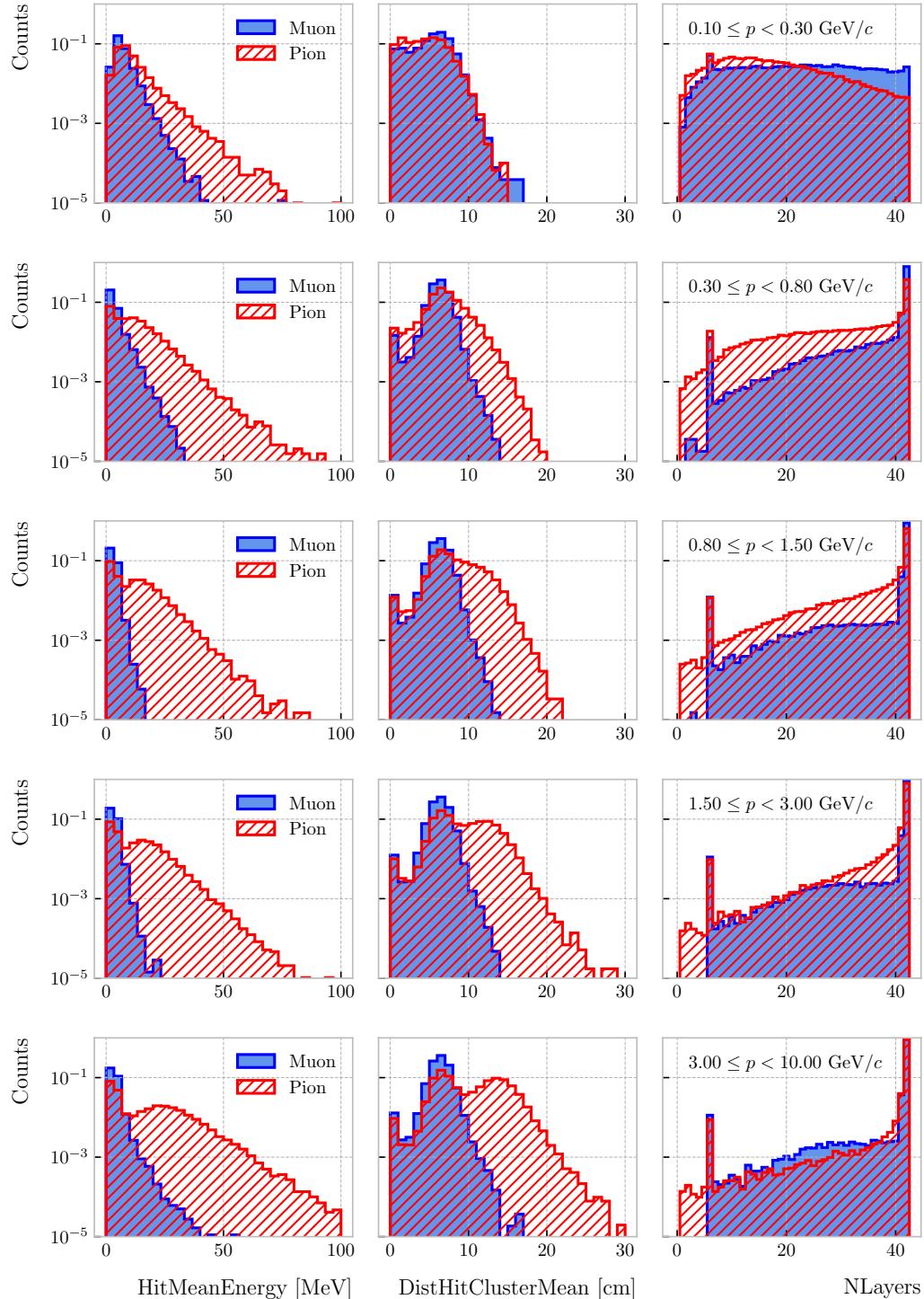


Figure 6.21: Example ECal feature distributions for muons (blue) and charged pions (red) in the five different momentum ranges considered (from top to bottom, in ascending momentum order). From left to right: mean hit energy, mean distance hit-to-cluster, and number of layers with hits.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

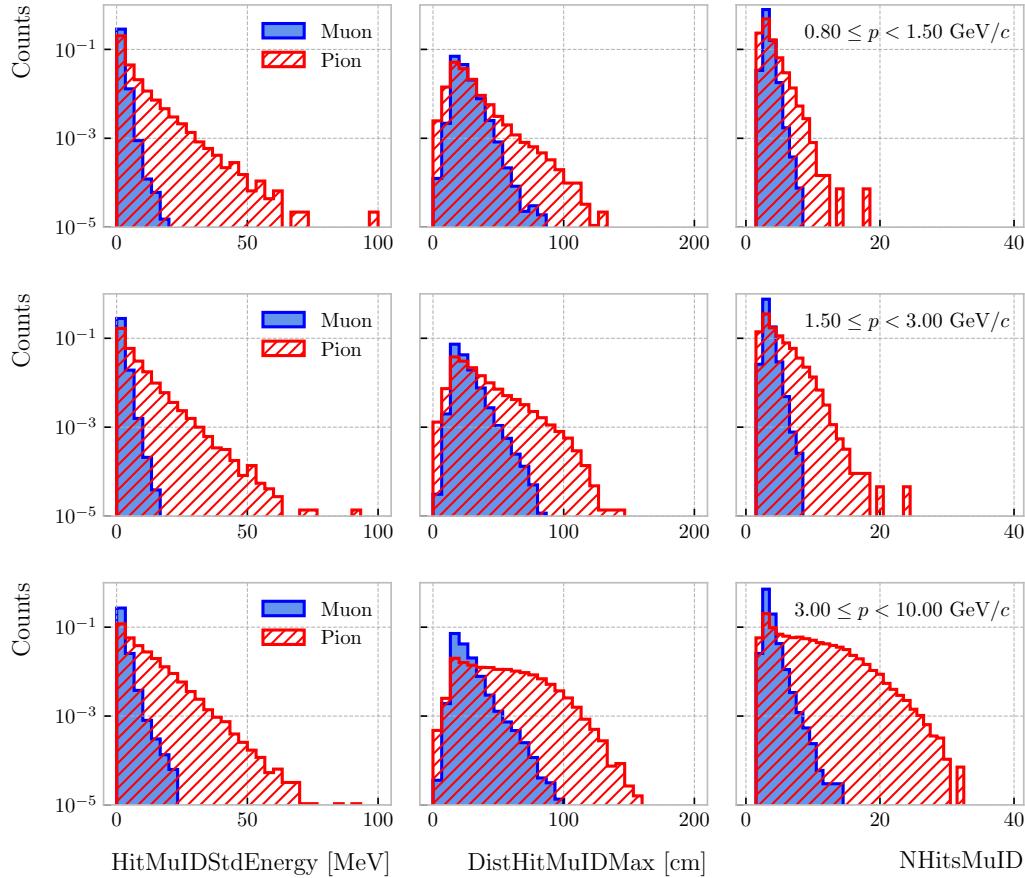


Figure 6.22: Example MuID feature distributions for muons (blue) and charged pions (red) in the three different momentum ranges considered (from top to bottom, in ascending momentum order). From left to right: standard deviation hit energy, maximum distance hit-to-hit, and number of hits.

• Energy-related MuID

- MuID total energy (ClusterMuIDTotalEnergy): sum of the energy of all the MuID hits.
- Mean MuID hit energy (HitMuIDMeanEnergy): mean of the MuID hit energy distribution.
- Standard deviation MuID hit energy (HitMuIDStdEnergy): standard deviation of the MuID hit energy distribution.
- Maximum MuID hit energy (HitMuIDMaxEnergy): maximum of the MuID hit energy distribution.

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

• Geometry-related MuID

- 2983 – Maximum distance MuID hit-to-hit (DistHitMuIDMax): maximum distance
- 2984 between pairs of MuID hits (not sure this is a good variable, distribution
- 2985 looks nuts).
- 2986 – Maximum distance MuID hit-to-centre (DistHitCenterMuIDMax): maximum
- 2987 of the distance distribution between the MuID hits and the centre of the
- 2988 TPC.

• Statistical MuID

- 2990 – Number of hits (NHitsMuID): total number of MuID hits associated to the
- 2991 track.
- 2992 – Number of layers with hits (NLayersMuID): not really a count of all layers
- 2993 with MuID hits but the difference between the last and the first layer with
- 2994 MuIDhits.

2995 Figure 6.22 shows the distributions of three different MuID variables, separating true
2996 muons (blue) and charged pions (red), for the three momentum ranges which use the
2997 muon tagger information. In this case I decided to standard deviation of the MuID hit
2998 energy distribution (left column), the maximum distance between the MuID hit pairs
2999 (middle column), and the number of MuID hits (right column). These variables are used
3000 together with the ECal features at high momenta, providing additional disambiguation
3001 power.

3002 Once our features have been defined, one can do some exploratory analysis to
3003 understand how well the variables describe the target class, and avoid the black-box
3004 approach by what features are most relevant for the learning process. This way, I
3005 performed a feature analysis for each of the momentum ranges I divided this classification
3006 problem into. It follows three steps: first a principal component analysis (PCA), followed
3007 by a feature importance study using Gini and Shapley values, and finally a feature
3008 permutation importance analysis.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

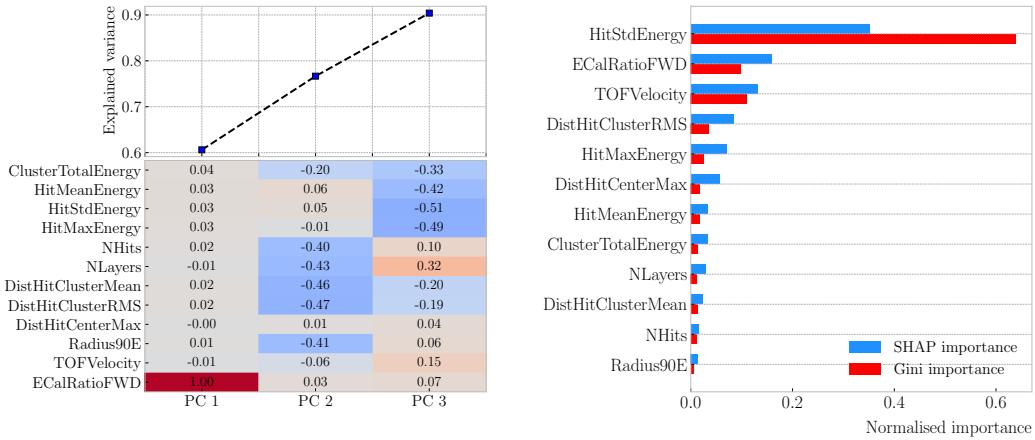


Figure 6.23: Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$.

3009 The PCA is useful to understand the variance of the feature space. It is an
 3010 unsupervised machine learning technique that allows the user to perform a dimensionality
 3011 reduction. It uses a singular value decomposition of the input features to project them
 3012 into a lower dimensional space. The idea is to find the matrix \mathbf{C}_m , whose columns are
 3013 the first m orthonormal eigenvectors of the input covariance matrix. Consider the $n \times p$
 3014 real matrix of input data \mathbf{X} , where n is the number of samples and p the number of
 3015 features. If \mathbf{X} is centred, i.e. the means of its columns are equal to zero, we can write the
 3016 covariance matrix of \mathbf{X} as $\mathbf{C} = \mathbf{X}^\top \mathbf{X} / (n - 1)$. This matrix can be diagonalised, yielding:

$$\mathbf{C} = \mathbf{V} \mathbf{L} \mathbf{V}^\top, \quad (6.11)$$

3017 where \mathbf{V} is a matrix of eigenvectors and \mathbf{L} a diagonal matrix with eigenvalues λ_i . Then,
 3018 performing SVD on \mathbf{X} gives us:

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{W}^\top, \quad (6.12)$$

3019 where \mathbf{U} is a unitary matrix, whose columns are called left singular vectors, \mathbf{S} is a
 3020 diagonal matrix of single values s_i , and \mathbf{W} is another unitary matrix, its columns known

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

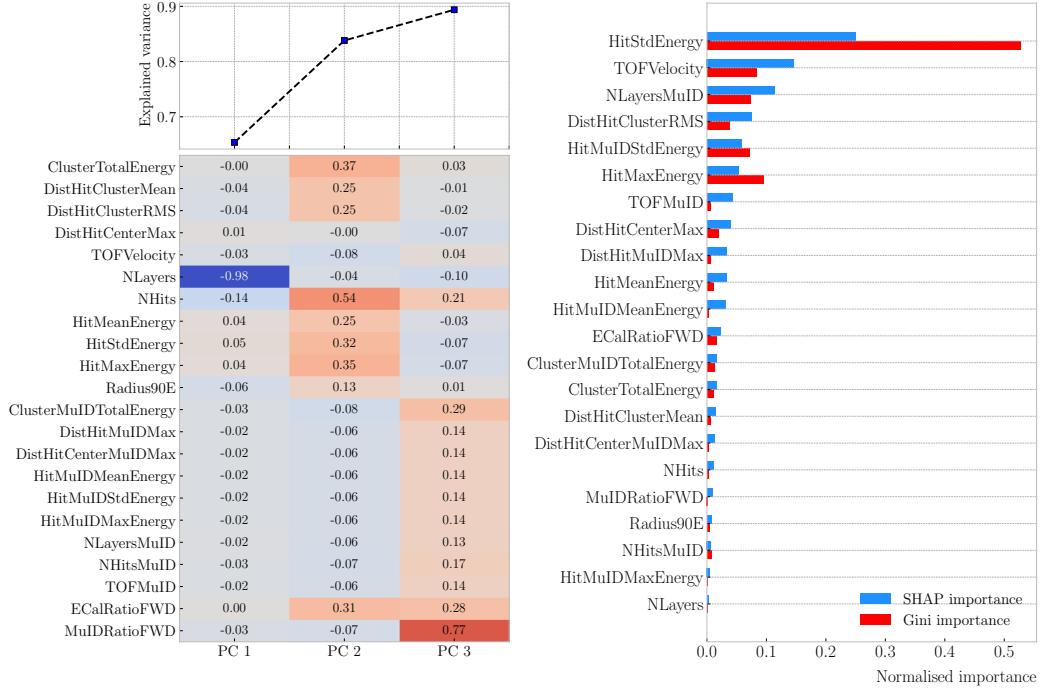


Figure 6.24: Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range $0.8 \leq p < 1.5$ GeV/c.

as right singular vectors. This way, we can write:

$$\mathbf{C} = \mathbf{WSU}^\top \mathbf{USW}^\top / (n - 1) = \mathbf{W} \frac{\mathbf{S}^2}{n - 1} \mathbf{W}^\top. \quad (6.13)$$

meaning that the right singular vectors are also the eigenvectors of the covariance matrix.

The SVD can be computed numerically following an iterative approach.

This way, taking an input data vector $X \in \mathbb{R}^n$, the resulting feature vector $Y \in \mathbb{R}^m$

is given by:

$$Y = \mathbf{C}_m^\top X. \quad (6.14)$$

The new features capture most of the variance of the original sample, while being lower dimensional, as $m < n$.

Before applying the PCA reduction one needs to centre and scale the input data.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3029 Centring is necessary when using SVD to obtain the eigenvectors of the covariance
3030 matrix, as only in that case we can do the identification with the right singular vectors
3031 from the input data. Scaling is needed when variables are on different scales, as some
3032 can then dominate the PCA procedure.

3033 I used the PCA module of `scikit-learn`, together with the `RobustScaler`, which
3034 centres the data and scales it based on the interquartile range. In Fig. 6.23 (left panel)
3035 and Fig. 6.24 (left panel) I show the results I obtained from the PCA for the momentum
3036 ranges $0.3 \leq p < 0.8 \text{ GeV}/c$ and $0.8 \leq p < 1.5 \text{ GeV}/c$, respectively. Notice that in
3037 the second case the number of features increases considerably, as this is the first region
3038 which uses the MuID variables. I found that, in all the cases, adding a fourth PC does
3039 not add additional information. As it can be seen in the top panels of the figures, the
3040 cumulative explained variance is already over 80% with three PCs.

3041 The bottom panels show the contribution of the variables to the principal axes. For
3042 the two first momentum regions, I observe a tendency of the energy-related and the
3043 geometry-related ECal variables to be clustered together. For the other ranges, when
3044 I include the MuID variables, there seems to be a division between ECal and MuID
3045 variables. For these, it seems like the number of ECal layers with hits also plays an
3046 important role.

3047 The next step in the analysis is to quantify the importance of the features based on
3048 two additional metrics, namely the Gini and the Shapley values. The Gini importance,
3049 often called mean decrease impurity, is based on how much a feature contributes to the
3050 purity improvement at the splits in each decision tree. The purity is measured in terms
3051 of the Gini impurity index, defined as:

$$I_G = 1 - \sum_i f_i, \quad (6.15)$$

3052 where f_i is the fractional abundance of the i -th class. Then, for each split one can

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

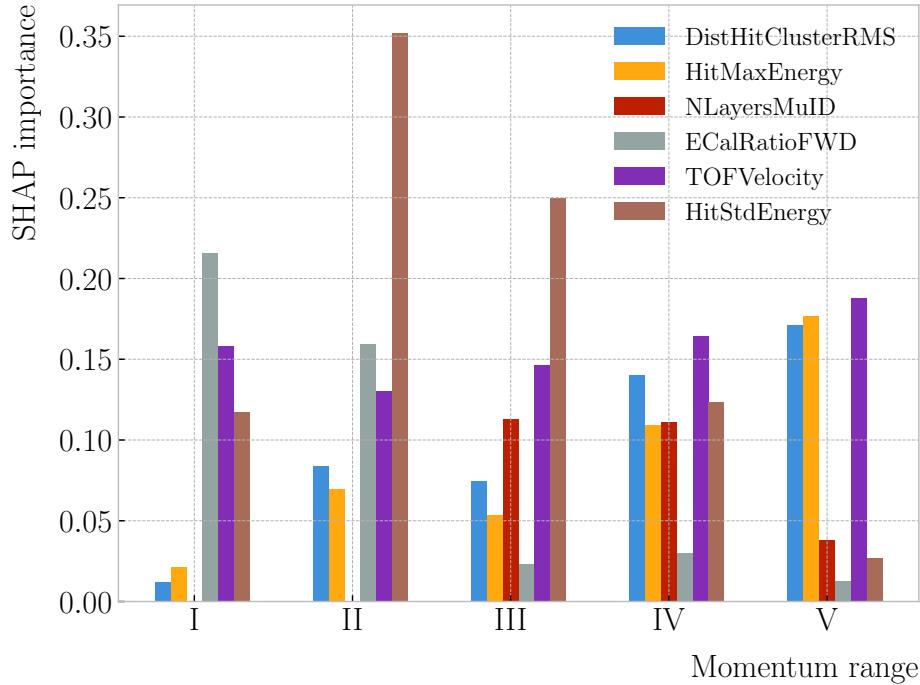


Figure 6.25: Evolution of the SHAP importance for the top six most important features across all five momentum ranges.

3053 compute the weighted decrease in impurity as:

$$\Delta_G = \frac{N_t}{N} \left(I_G - \frac{N_t^R}{N_t} I_G^R - \frac{N_t^L}{N_t} I_G^L \right), \quad (6.16)$$

3054 where N represents the total number of samples, N_t the number of samples at the current
 3055 node, N_t^R and N_t^L the number of samples in the right and left children respectively,
 3056 I_G is the Gini impurity at the current node, and I_G^R and I_G^L the Gini impurities of the
 3057 resulting right and left children.

3058 For each decision tree, one will have a normalised vector with the accumulated
 3059 decrease in Gini impurity for each feature. In the case of a BDT, the feature importances
 3060 are simply the mean for all the estimators in the ensemble⁹.

3061 The concept of Shapley values originated in the context of game theory, and it
 3062 measures the marginal contribution of a feature in enhancing the accuracy of a classifier.

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.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3063 Take F to be the set of all features in a problem, and $S \subseteq F$ the subset of features. To
 3064 compute the Shapley value of the i -th feature, one has to train a model with that feature
 3065 present, $f_{S \cup \{i\}}$, and another model trained without it, f_S . This has to be repeated for
 3066 all possible combinations of subsets $S \subset F \setminus \{i\}$, and evaluating the models predictions
 3067 on the appropriate sets of data x_S . This way, the Shapley value results:

$$\varphi_i = \sum_{S \subset F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]. \quad (6.17)$$

3068 I trained the `GradientBoostingClassifier` from `scikit-learn` with the default
 3069 configuration in order to evaluate both the Gini and Shapley importances. The Gini
 3070 scores are automatically computed by `scikit-learn`, using the training data. For the
 3071 Shapley importance, I used the implementation from the `SHAP` package, computing
 3072 it using the test sample. The results can be seen in Fig. 6.23 (right panel) and
 3073 Fig. 6.24 (right panel), again for the momentum ranges $0.3 \leq p < 0.8 \text{ GeV}/c$ and
 3074 $0.8 \leq p < 1.5 \text{ GeV}/c$. The length of the bars denote either the SHAP (blue) or the Gini
 3075 (red) importance of the feature. One interesting thing to notice is that, when looking at
 3076 the Gini importance, there is always one feature that dominates over the rest. This is
 3077 not the case for the SHAP importance, where importances tend to be more balanced.

3078 Across all momentum ranges, I observe that the most important features are. For
 3079 the five momentum ranges considered, only six variables sit in the top five at least once.
 3080 Figure 6.25 shows the evolution of the SHAP importance of these six features. It is
 3081 interesting to see that the time-of-flight variable keeps its importance almost unchanged
 3082 for all momenta. Also, it looks like the ECal energy ratio gets less relevant the higher the
 3083 momentum is, but the RMS of the hit-to-cluster distance distribution and the maximum
 3084 ECal hit energy become more important in the last momentum ranges.

3085 The last step in the feature selection analysis is the feature permutation. This
 3086 technique measures the contribution of each feature to the performance of a model by
 3087 randomly shuffling its values and checking how some scores degrade. For the present
 3088 case, I am interested in the precision or purity, and the sensitivity or efficiency, as these

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

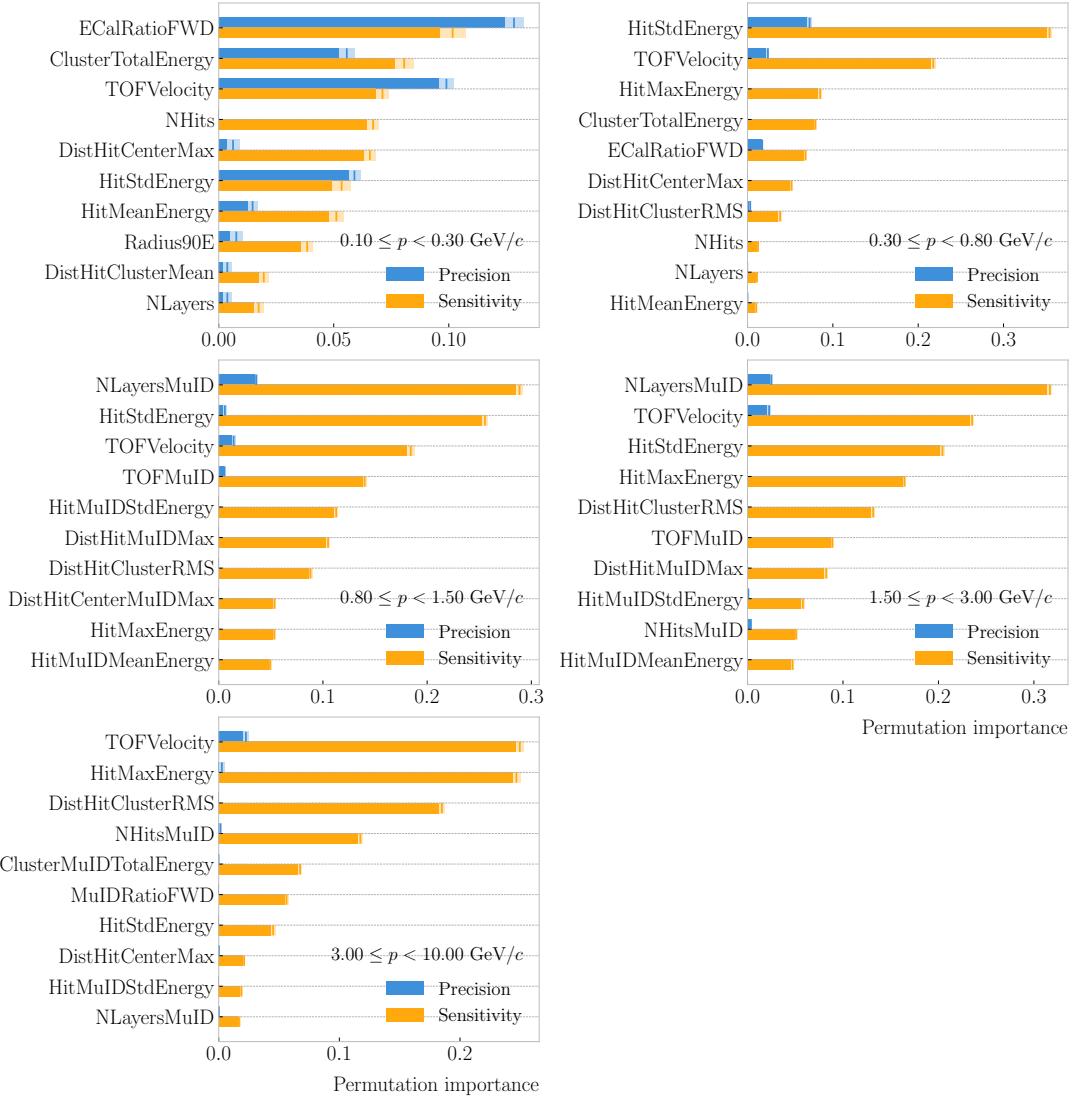


Figure 6.26: Permutation importances for the ten most important features in the different momentum ranges (from left to right, top to bottom, in increasing momentum order). The bars indicate the effect that permutations of each feature have on the purity (blue) and the sensitivity (yellow), the translucent regions representing one standard deviation around the central value.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3089 two are the most relevant metrics from a physics point of view. The `scikit-learn`
3090 module provides the user with a method to perform the permutation scans.

3091 The results of these are shown in Fig. 6.26. For the different momentum ranges
3092 I show the permutation importances for the ten most important features. For each
3093 of the variables I report the effect the permutations have on the precision (blue) and
3094 sensitivity (yellow) of the models. The bars indicate the importance value, with the
3095 lighter part representing one standard deviation around the mean (hinted as an additional
3096 vertical line). Something to notice is that, in the first momentum region, the feature
3097 permutations have an effect on both the precision and the sensitivity. However, for the
3098 rest the precision is almost unaffected, while the sensitivity changes are considerably
3099 larger.

3100 It is also interesting to see that most of the variables identified as important here
3101 are the same I found when looking at the Shapley values. The behaviour of these across
3102 the momentum ranges is also similar, with the same patterns of some features being
3103 important at low momenta and then dropping in importance for the high momentum
3104 ranges.

3105 Wit this, I conclude the study of the features. I have prepared the training and
3106 testing datasets and understood what features are likely to have the largest impact on
3107 the performance of the classifiers.

3108 6.3.4 Hyperparameter optimisation

3109 Any BDT requires the user to specify a number of parameters that will dictate its
3110 behaviour. They can be divided into two categories: (i) tree-specific parameters, which
3111 affect each individual tree in the model, and (ii) boosting parameters, which control the
3112 boosting operation in the model. The value of these so-called hyperparameters affect the
3113 performance and predictive power of the models. Therefore, one needs to carefully select
3114 their optimal values in order to extract as much information as possible from the data.

3115 From all the parameters used to define a tree in the `scikit-learn` implementation
3116 of the BDT classifier, I only consider a subset of them. This is due to the fact that some

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

3117 are mutually exclusive, but also because I noticed that others have little effect on the
3118 problem at hand. Therefore, the parameters I investigate are the following:

- 3119 • `min_samples_split`: defines the minimum number of samples required in a node
3120 to be considered for splitting. Higher values prevent a model from learning relations
3121 which might be highly specific to the particular sample, but may lead to under-fitting
3122 if the value is too low.
- 3123 • `min_samples_leaf`: defines the minimum samples required in a leaf node. For
3124 imbalanced problems it should take a low value, as there will not be many cases
3125 where the minority class dominates.
- 3126 • `max_depth`: maximum depth of a tree. Useful to prevent over-fitting, as higher
3127 depth will allow a model to learn relations specific to the training sample.

3128 In the case of the boosting parameters, the ones I look at are:

- 3129 • `learning_rate`: determines the impact of each tree on the final outcome. Low
3130 values make the model robust to the specific characteristics of a tree, and thus
3131 allow it to generalise well. However, that usually requires a large number of trees
3132 to model the data properly.
- 3133 • `n_estimators`: number of sequential trees to be trained. In general, BDTs are
3134 fairly robust at higher number of trees but it can still overfit at a point.
- 3135 • `subsample`: fraction of observations to be selected for each tree. Values slightly
3136 less than 1 make the model robust by reducing the variance.

3137 In general, hyperparameters depend on each other. Thus, it is not possible to
3138 optimise them independently. In the literature, we find two main strategies to explore
3139 the hyperparameter space. We could use a grid search, in which one discretises a
3140 portion of the space of hyperparameters and evaluates the model at each point. Another
3141 approach is the randomised search, where a certain number of random configurations of
3142 hyperparameters are explored.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

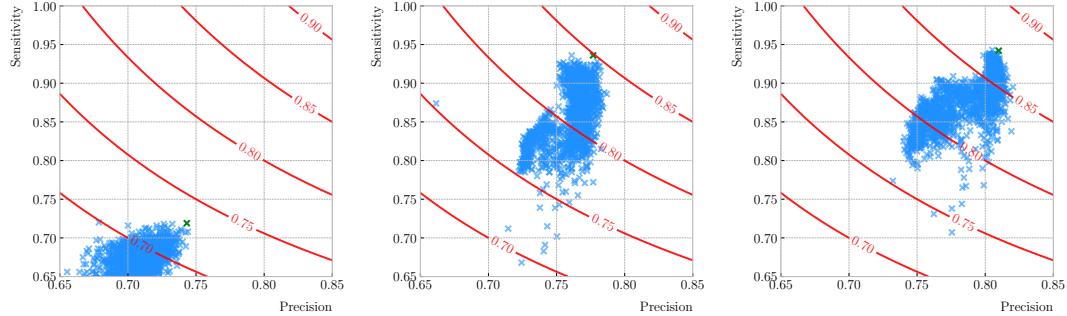


Figure 6.27: Values of the precision and sensitivity obtained for 10000 BDT hyperparameter configurations, for the momentum regions I, III and V. The red contours indicate the curves of equal F_1 -score, while the green crosses are the selected configurations.

3143 In this case, I used the random search to scan the hyperparameter space. Also,
 3144 because it is not guaranteed that a set of hyperparameters can be efficiently applied
 3145 across different datasets, I perform the optimisation for each of the momentum ranges
 3146 considered. Table 6.3 shows the list of hyperparameters considered, and the range within
 3147 which I let them vary. I decided to fix the number of estimators to 400 in all cases, as
 3148 its value is correlated with that of the learning rate.

3149 I evaluate 10000 different hyperparameter configurations for each momentum range.
 3150 For the hyperparameter tuning, I used subsamples containing 10% of the full datasets,
 3151 keeping the original proportions between classes, in order to reduce the computational
 3152 load. The performance of the models was assessed using a stratified 3-fold cross-validation
 3153 with replacement. Cross-validation involves dividing the data in a number of subsets,
 3154 training the model using some of them, and testing it with the rest. In our case, I
 3155 divide the data in 3 equal-sized subsets, maintaining the class proportions of the original
 3156 dataset. Then, for 3 consecutive iterations, I train the models using 2 of the subsets
 3157 while I compute the precision and sensitivity scores with the other. This approach
 3158 provides a more robust estimate of the performance on unseen data.

3159 Figure 6.27 shows the results in the precision versus sensitivity plane, for the
 3160 momentum regions I, III and V (from left to right). The contours represent the curves
 3161 of equal F_1 -score, i.e. the harmonic mean of the precision and the sensitivity. In order

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

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

3162 to select the optimal configurations (indicated in the plots with a green cross), I chose
 3163 the point with the highest F_1 -score.

3164 The results for the different momentum ranges are summarised in Tab. 6.3. One
 3165 can see some consistency in hyperparameter choices, with models generally preferring
 3166 small values for the tree-specific parameters, small learning rate, and relatively large
 3167 subsample sizes.

3168 Now that I have obtained the optimal values of the hyperparameters, I can train
 3169 the different BDTs. In this case I use the complete datasets, keeping 20% of the data
 3170 for testing. Table 6.4 shows the values of the different performance metrics obtained
 3171 using the selected hyperparameters and 5-fold cross-validation. The last row indicates
 3172 the value of the area under the receiver operating characteristic (ROC) curve. This
 3173 represents the sensitivity of a model as a function of the false positive rate. I have

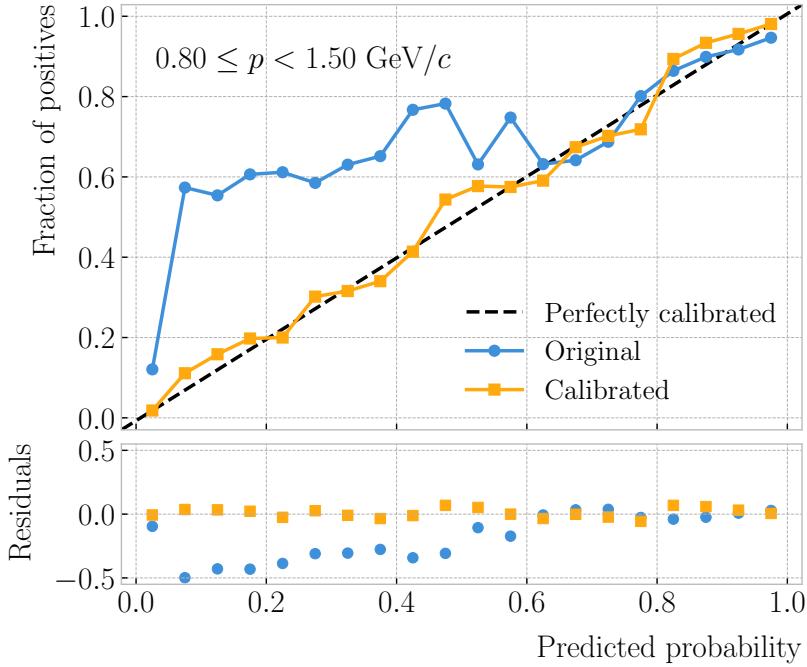


Figure 6.28: Reliability diagrams for the BDT classifier used in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$, both for the original (blue circles) and calibrated (yellow squares) responses. For reference, the response of a perfectly calibrated classifier is also shown (black dashed line).

3174 included it here as it is a classic model metric used in the machine learning community.

3175 Overall, there is a clear trend of models performing better at higher momentum.

3176 6.3.5 Probability calibration

3177 So far, the trained BDTs are able to provide predictions of the class labels. Ideally,
3178 one would like the output of a classifier to give a confidence level about the prediction.
3179 However, it is not straightforward to interpret the outputs of our BDTs in terms of
3180 probabilities.

3181 A way to visualise how well the predictions of a classifier are calibrated is using
3182 reliability diagrams [179]. They represent the probability of the positive label versus the
3183 probability predicted by the classifier. These can be obtained by binning the predicted
3184 probabilities, and then compute the conditional probability $P(y_{true} = 1 | y_i \leq y_{pred} <$
3185 $y_{i+1})$ by checking the fraction of true positive instances in each bin. The reliability

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

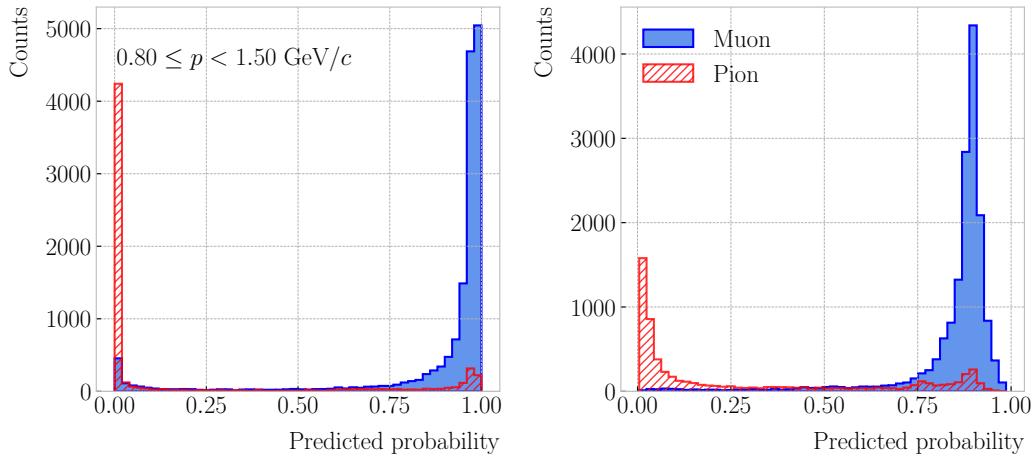


Figure 6.29: Uncalibrated (left panel) and calibrated (right panel) predicted probabilities assigned by the BDT classifiers for true muons (blue) and charged pions (red) in the momentum range $0.3 \leq p < 0.8$ GeV/ c .

3186 diagram of a perfectly calibrated classifier would be a diagonal line.

3187 In this case, I try to correct the raw response of the classifiers by applying a sigmoid
3188 function:

$$\sigma(x; A, B) = \frac{1}{1 + e^{Ax+B}}, \quad (6.18)$$

3189 where the parameters A and B are real numbers determined using the method of least
3190 squares.

3191 For each classifier, I perform a grid search to obtain the optimal values of A and B .
3192 For any pair, I compute the predicted probabilities as $y_{pred} = \sigma(y_{raw}; A, B)$, where y_{raw}
3193 are the raw predictions of the classifier¹⁰. Then, I calculate the corresponding reliability
3194 curve, and take the sum of the squared residuals between it and the response of the
3195 perfectly calibrated classifier.

3196 Figure 6.28 shows the reliability diagrams for the original (blue) and calibrated
3197 (yellow) probability predictions of the classifier for the III momentum range, $0.3 \leq p <$
3198 0.8 GeV/ c . The original response of the classifier is given by $y_{pred} = \sigma(y_{raw}; -2, 0)$,
3199 which is the transformation applied by `scikit-learn` to produce the probability estimate.
3200 Notice how the calibrated prediction matches the ideal response much better than the

¹⁰In `scikit-learn` these correspond to the outputs of the `decision_function` method.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

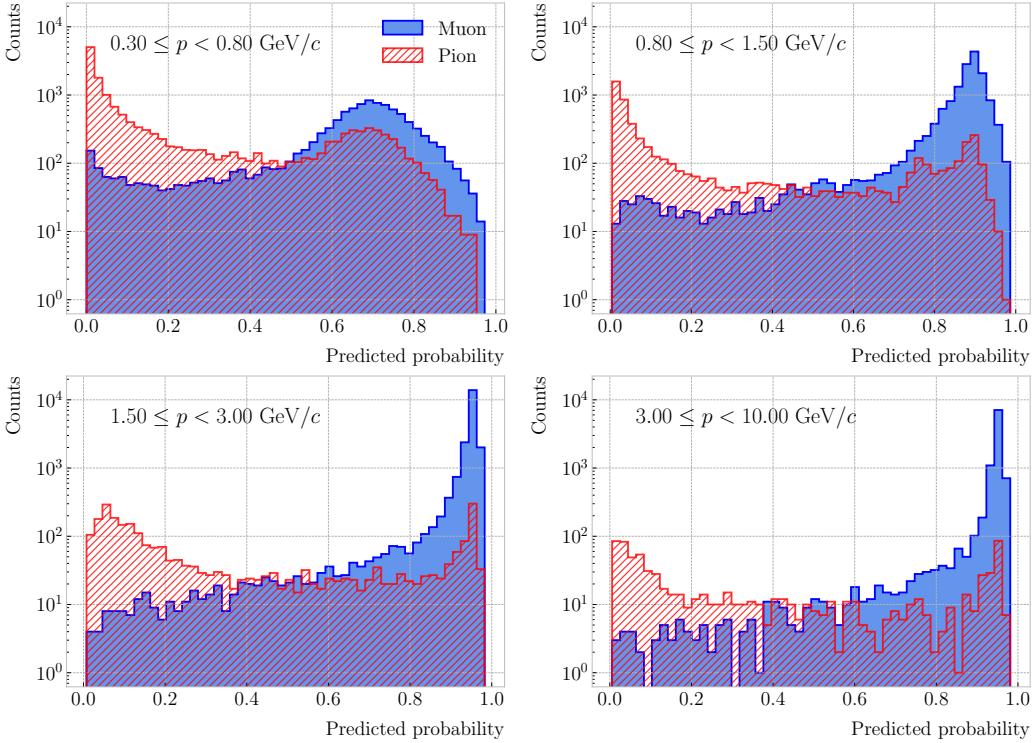


Figure 6.30: Calibrated predicted probabilities assigned by the BDT classifiers for the true muons (blue) and charged pions (red) in a FHC neutrino sample.

3201 original, across all the probability range.

3202 One can also compare the responses of the uncalibrated and calibrated classifiers
 3203 broken down by true particle type, as shown in Fig. 6.29. It can be seen that the
 3204 distributions for both muons (blue) and charged pions (red) smoothen after calibration,
 3205 but still the separating power of the classifier remains unchanged.

3206 6.3.6 Performance

3207 At this point, having the trained classifiers and the probability calibration parameters, I
 3208 am able to assess the performance of the classification strategy in a physics-relevant case.
 3209 For this, I prepared a sample of 10^5 FHC neutrino interaction events in the HPgTPC.
 3210 Using the truth matching information, I select all true muons and pions, and apply the
 3211 corresponding BDT classifier based on their momentum.

3212 Figure 6.31 shows the resulting calibrated output of the classifiers for the different

6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

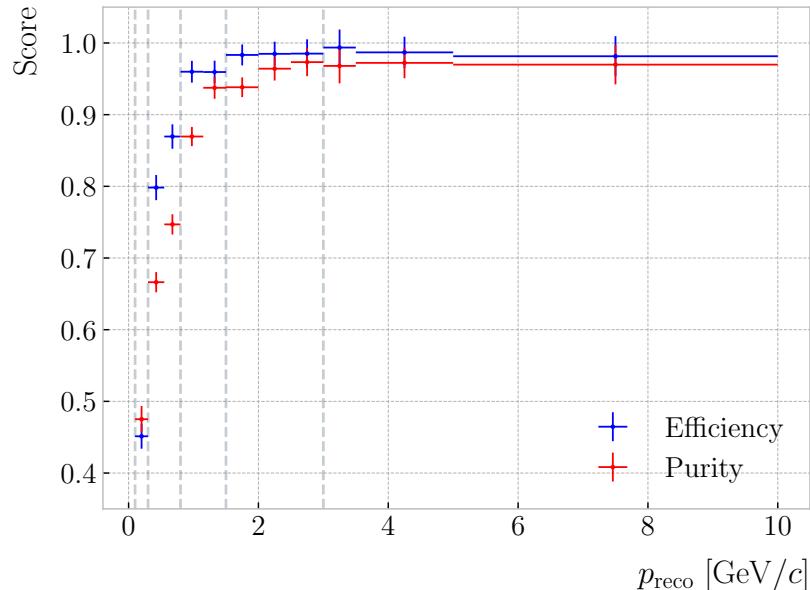


Figure 6.31: Efficiency (blue) and purity (red) of the muon selection as a function of the reconstructed momentum for the FHC neutrino sample.

3213 momentum regions. I do not include the first region, $0.10 \leq p < 0.30$ GeV/ c , as it only
 3214 contains a small fraction of signal events. The distributions obtained for this validation
 3215 sample

3216 I also studied the performance of the muon identification in a track-by-track selection.
 3217 To do so, I apply a simple cut on the output of the BDT classifiers. Every particle
 3218 with a predicted probability higher than the cut is considered a muon, while the ones
 3219 not passing the cut are taken to be pions. The results obtained for a cut of 0.50 are
 3220 shown in Fig. 6.31. Both the efficiency (blue) and the purity (red) of the selection are
 3221 displayed as a function of the momentum. The binning was chosen so that there were no
 3222 bins in between different momentum ranges and each had roughly the same number of
 3223 events. Even without optimising the value of the cut, the performance of the selection
 3224 is excellent. The only issues the first momentum range, where efficiency and purity sit
 3225 slightly below 0.50. However, a dE/dx measurement could help enhance the selection
 3226 there.

3227 This shows that the method behaves as expected when using unseen data, generalising
 3228 without problems from single particle to full neutrino events. In the coming Chapter, I

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

3229 will study how to use the outputs of the BDT, hereafter referred to as muon scores, to
3230 perform realistic event selections in ND-GAr.

3231 6.4 ECal time-of-flight

3232 Looking at Fig. 6.20, it is clear that for momentum values in the range $1.0 - 3.0 \text{ GeV}/c$
3233 it is not possible to separate pions and protons using a $\langle dE/dx \rangle$ measurement in the
3234 HPgTPC. However, in the previous section I assumed that protons at those energies
3235 could be identified by other means, and therefore were not an issue for the muon and
3236 pion discrimination.

3237 Some detectors, like ALICE [180] or the ILD concept [181], complement the PID
3238 capabilities of their gaseous trackers with time-of-flight measurements. The use of
3239 fast timing silicon sensors, with hit time resolutions under 100 ps, would allow for the
3240 identification of charged hadrons via a ToF measurement up to $5.0 \text{ GeV}/c$. In the case
3241 of ND-GAr, one could think of using the inner layers of the ECal, the ones consisting of
3242 high-granularity tiles, to obtain a ToF-based PID, with some inputs from the TPC.

3243 Measuring the momentum and the velocity of a charged particle allows for a
3244 determination of the mass through the relativistic momentum formula:

$$m = \frac{p}{\beta} \sqrt{1 - \beta^2}. \quad (6.19)$$

3245 In our case, the momentum is measured in the TPC, using the curvature and the dip
3246 angle of the helix inside the magnetic field. The velocity of the particle can be written
3247 as:

$$\beta = \frac{\ell_{track}}{c\tau}, \quad (6.20)$$

3248 where ℓ_{track} is the length of the track, and τ the arrival time to the ECal.

3249 In GArSoft, the track length is computed at the Kalman filter stage. It is simply the
3250 sum of the line segments along the track, either in the forward or backward fit. In this
3251 case, because we are only interested in the particles that make it to the ECal, I choose

6.4. ECAL TIME-OF-FLIGHT

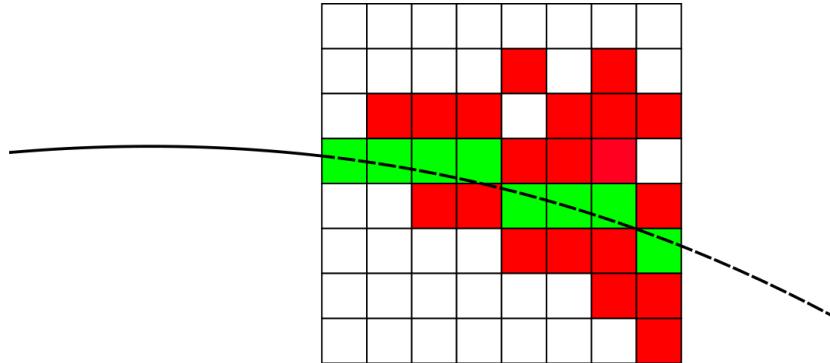


Figure 6.32: Schematic of the hit selection used for the ToF measurement. The grid represents the layers of the inner ECal, with coloured squares indicating the tiles with hits. Green squares indicate the selected hits.

3252 the fit direction based on the results of the track-cluster associations.

3253 Additionally, because the last 30 cm of the TPC radius are uninstrumented¹¹, I need
 3254 to correct for the length of the tracks. Using the track fit parameters to propagate the
 3255 helix to its entry point in the ECal, one can write the total track length as:

$$\ell_{track} = \ell + \left| \frac{\phi_{EP} - \phi}{R^{-1}} \right| \sqrt{1 + \tan^2 \lambda}, \quad (6.21)$$

3256 where ϕ_{EP} is the angle of rotation at the entry point to the calorimeter, and ℓ , ϕ , R and
 3257 λ are the track length, angle of rotation, radius of curvature and dip angle at the last
 3258 point in the fit, respectively.

3259 To test the idea of performing a ToF measurement with the inner ECal, I generated
 3260 two data samples. Each consists of 10000 single particle events, either charged pions or
 3261 protons. Their momenta are uniformly distributed in the range $0.5 - 5.0$ GeV/ c , and
 3262 their directions are isotropic. I process each sample using different values of the time
 3263 resolution, from $\Delta\tau = 0$, the perfect time resolution case for comparison, to the current
 3264 nominal value of $\Delta\tau = 0.7$ ns, and the worse scenario of $\Delta\tau = 1.0$ ns.

¹¹Note to self: check this number.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

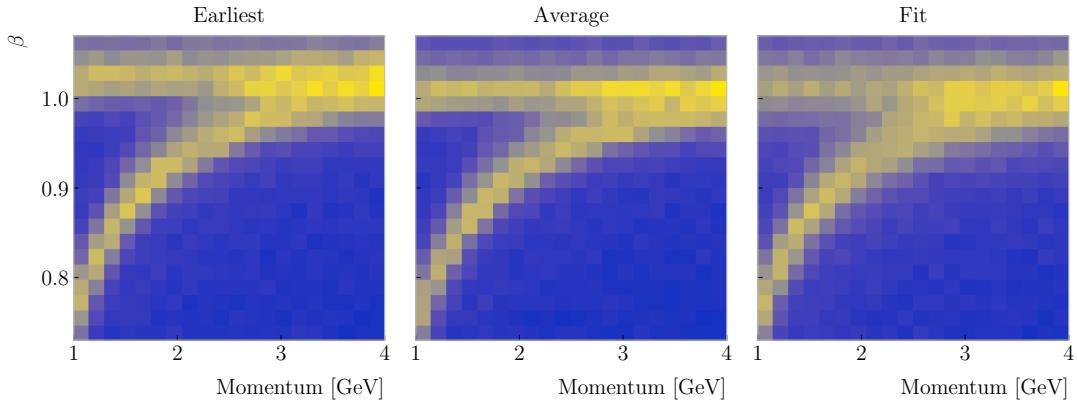


Figure 6.33: Particle velocity versus momentum measured with different ECal arrival time estimations. From left to right: earliest hit time, average hit time, and fitted hit time. In all cases the time resolution is $\Delta\tau = 0.1$ ns.

3265 6.4.1 Arrival time estimations

3266 In the simulation, the limited time resolution of the ECal is taken into account by
 3267 applying a Gaussian smearing to the true hit times. Other effects, like the digitisation
 3268 of the signals, are not taken into account and fall beyond the scope of this study. After
 3269 the track-cluster, one ends up with a collection of ECal hits associated to each particle.
 3270 From these, the arrival time of the particle to the ECal can be extracted.

3271 The simplest possibilities are to either take the time of the earliest hit or the hit
 3272 closest to the entry point. Because these two coincide, in general, I focused only in
 3273 the earliest hit time. However, this needs to be corrected, to account for the distance
 3274 travelled from the entry point to the position of the hit:

$$\tau_{earliest} = \tau_{hit} - \frac{d_{EP-hit}}{c}, \quad (6.22)$$

3275 where τ_{hit} is the time of the earliest hit, and d_{EP-hit} is the distance between that hit
 3276 and the entry point of the particle to the ECal. This is computed as the arc length
 3277 between the entry point and the point of the extrapolated helix up to the layer of the
 3278 hit. This way of correcting the time assumes c for the propagation of the particle, which
 3279 may lead to biased estimates.

6.4. ECAL TIME-OF-FLIGHT

3280 I also tried to estimate the arrival times using information from the rest of the hits.
3281 In order to do this, as a simplifying assumption, I approximate the hadronic shower
3282 considering only its MIP component. For each layer, I keep only the hit in the tile closest
3283 to the point of the extrapolated track up to that layer. Figure 6.32 shows an example of
3284 how this hit selection works. The dashed line represents the extrapolated track, while
3285 the coloured squares are the tiles containing hits. Green indicates the tiles closer to the
3286 track in each layer (in the sketch they correspond to the grid columns).

3287 Now, I can use these collections of hits to estimate the arrival times. A possibility
3288 is to take the average of the times of the selected hits, denoted $\tau_{average}$. For that to
3289 work, one needs to correct these times, in a similar way as in Eq. (6.22), before taking
3290 the average. However, as before, this correction assumes that the particle travels at the
3291 speed of light inside the ECal. Another option is to perform a linear fit to the hit times
3292 and the distances to the entry point. In that case, the arrival time would be the fitted
3293 value of the intercept, τ_{fit} . This method would not assume a speed of light propagation.

3294 Figure 6.33 shows the velocity estimations as a function of the particle momentum,
3295 for the earliest hit time (left panel), average hit time (middle panel), and fitted hit time
3296 (right panel). The two bands correspond to the π^\pm and the p particles. $\Delta\tau = 0.1$ ns.
3297 Notice how, for the earliest hit time method, the velocities are significantly biased
3298 towards larger values. For the multi-hit methods, the τ_{fit} estimate appears to produce a
3299 larger variance than when using the $\tau_{average}$ method.

3300 6.4.2 Proton and pion separation

3301 Once we have the velocities of the particles, one can estimate their masses through
3302 Eq. (6.19). The resulting mass spectra are shown in Fig. 6.34. I computed the masses
3303 for the three arrival time estimates discussed above, and three different values of the
3304 time resolution: $\Delta\tau = 0.00$ (perfect time resolution), $\Delta\tau = 0.10$ ns, and $\Delta\tau = 0.70$ ns.
3305 Although in all cases we have the same number of events, it appears as if the entries
3306 in the histograms decrease as the time resolution increases. Sometimes, the particles
3307 get unphysical values of $\beta > 1$, and in turn they do not contribute to the mass spectra.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

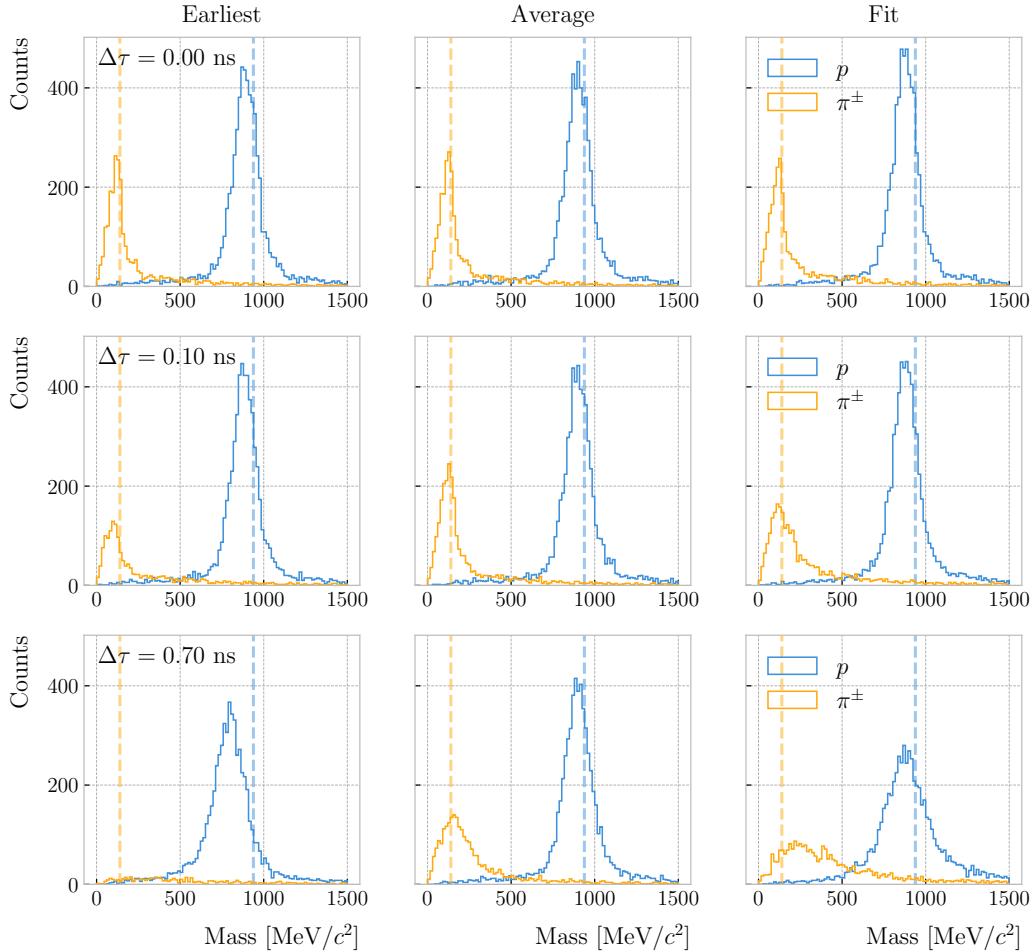


Figure 6.34: Mass spectra for p (blue) and π^\pm (yellow) particles, using different ECal time resolution values (from top to bottom, in ascending order), and arrival time estimates. From left to right: earliest hit time, average hit time, and fitted hit time. The dashed lines indicate the true masses of the particles.

3308 This is more likely to happen for higher values of $\Delta\tau$.

3309 As noted before, the average hit time method produces the most robust estimates
 3310 when increasing $\Delta\tau$. Intuitively this makes sense, as by taking the mean one averages
 3311 out the effect of the Gaussian smearing. Going forward, I will use this arrival time
 3312 estimator, as it appears to be the best performing one.

3313 It is possible to use the velocity estimations to select a sample of protons. In this
 3314 case, I do so by dividing the relevant momentum range in bins of $0.1 \text{ GeV}/c$. For each
 3315 momentum bin, I compute the expected velocity for the protons via the inverse of Eq.

6.4. ECAL TIME-OF-FLIGHT

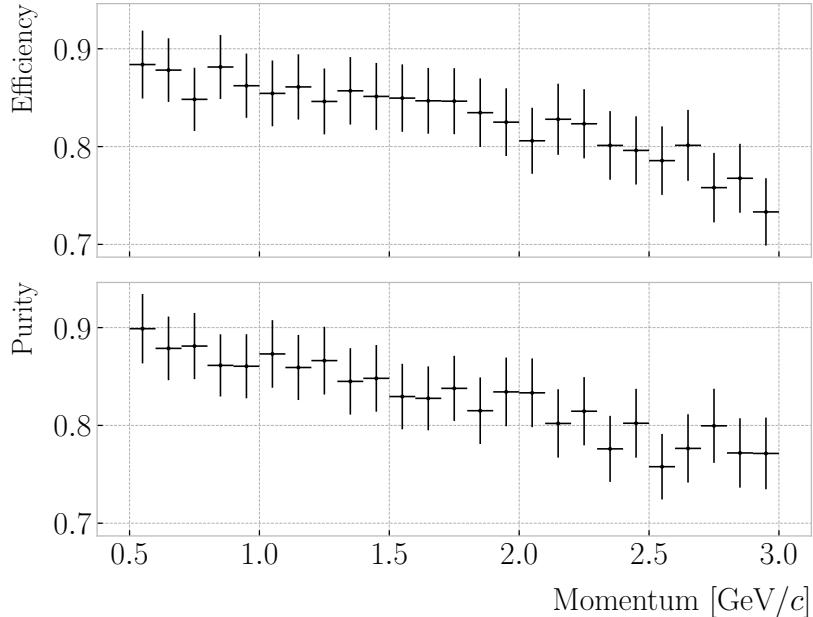


Figure 6.35: Efficiency (top panel) and purity (bottom panel) for the proton selection as a function of the momentum, for $\Delta\tau = 0.10$ ns.

3316 (6.19), and then take the fractional residuals of the measured velocities. Using that
 3317 distribution, I choose the cut that maximises the F_1 -score of the proton selection.

3318 The results can be seen in Fig. 6.35, for the case $\Delta\tau = 0.10$ ns. As expected from
 3319 Fig. 6.33, the performance of the selection degrades rapidly with increasing momentum.
 3320 However, the purity is still around 75% at 3.0 GeV/c. This is likely to be sufficient, as
 3321 we do not expect protons or charged pions with higher energies from the beam neutrino
 3322 interactions.

3323 Figure 6.36 shows a few examples of the ToF velocity estimation in a FHC neutrino
 3324 sample. Here, for the different momentum bins, I have taken the fractional residual of
 3325 the expected value of β for a proton and the measured values (black data points). The
 3326 coloured lines represent Gaussian fits to the distributions of the different true particle,
 3327 with the gray line being the sum of these. It can be seen that, even for momenta close
 3328 to 2.0 GeV/c, a good proton separation can be achieved. This idea will be explored
 3329 further later, in the context of the event selection.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

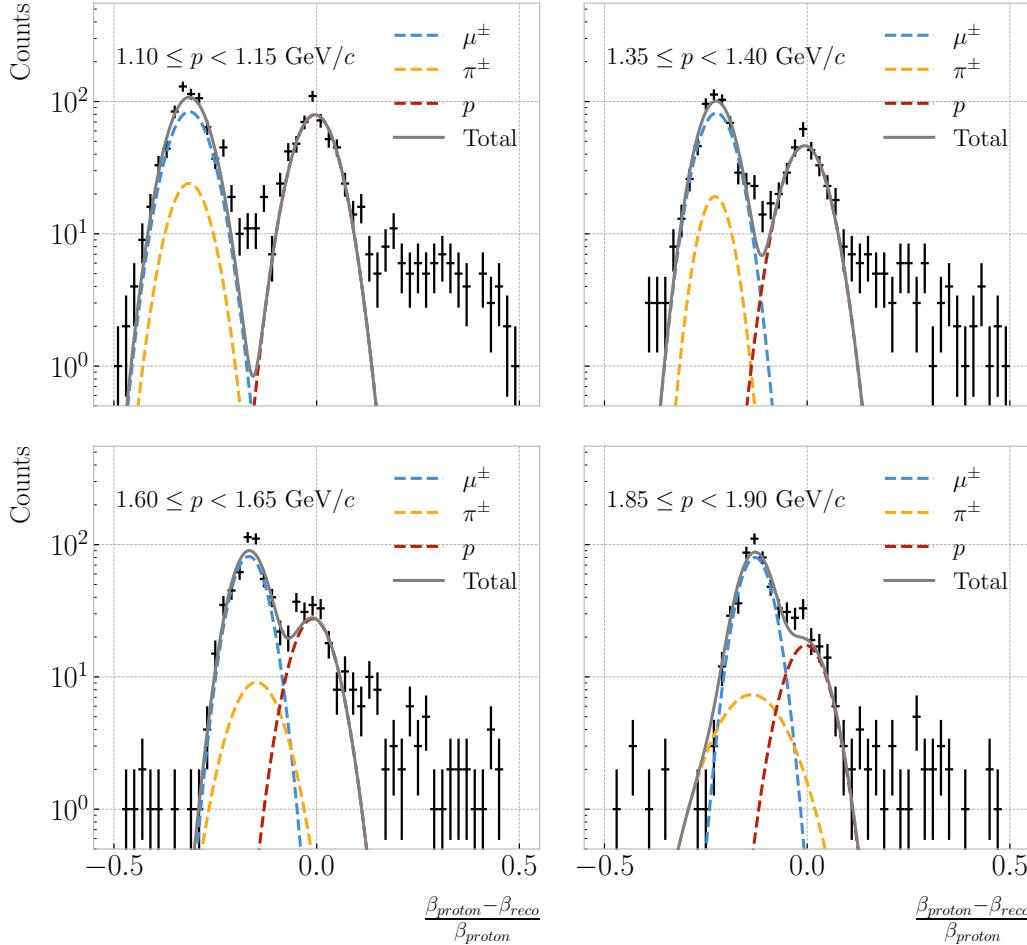


Figure 6.36: Distributions of the velocities measured by ToF with the inner ECal, for different momentum bins, in a FHC neutrino interaction sample. The Gaussian fits are performed around the maxima for each particle species.

3330 6.5 Charged pion decay in flight

3331 As discussed previously, in GArSoft the TPC tracks are formed after a pattern recognition
 3332 algorithm and a Kalman filter are applied to the TPC clusters. These two steps can
 3333 find discontinuities in the track candidates (e.g. due to a particle decay) when these
 3334 so-called breakpoints are large enough. However, for some, more subtle, cases they may
 3335 miss them and form a single reconstructed track. It has been noted in the literature
 3336 that Kalman filters offer, as a by-product, additional information to form test statistics
 3337 to identify these breakpoints [182, 183].

6.5. CHARGED PION DECAY IN FLIGHT

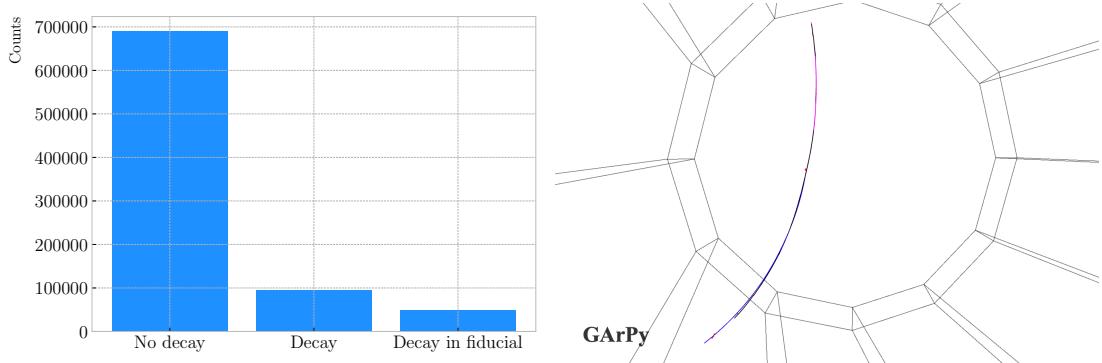


Figure 6.37: Left panel: number of non-decaying, decaying and decaying in the fiducial volume pions for a MC sample of 100000, $p = 500 \text{ MeV}/c$ isotropic positively charged pions inside the TPC. Right panel: event display for a positive pion decaying inside the fiducial volume, with a single reconstructed track for the pion and muon system.

3338 Considering the mean life of the charged pion, $\tau = (2.6033 \pm 0.0005) \times 10^{-8} \text{ s}$, one
 3339 can estimate that about 12% of the pions with momentum $p \sim \mathcal{O}(500 \text{ MeV}/c)$ (roughly
 3340 the peak of the pion momentum distribution in ν_μ CC interactions off argon) decay
 3341 inside the TPC. Figure 6.37 (left panel) shows the amount of charged pions decaying in
 3342 the full TPC and fiducial volumes from an isotropic, monoenergetic sample of 100000
 3343 negatively charged pions with $p = 500 \text{ MeV}/c$. We see that about 10% of those decayed,
 3344 with more than half of them decaying inside the TPC fiducial volume.

3345 Figure 6.37 (right panel) shows an example event display of a charged pion (magenta
 3346 line) decays in flight inside the TPC, but because the angle of the muon (blue line) is
 3347 small both were reconstructed as one single track (black line). In this case, the composite
 3348 track reaches the ECal, where it undergoes a muon-like interaction, thus being classified
 3349 as a muon.

3350 A way to understand what decaying pion tracks were totally or partially reconstructed
 3351 together with the daughter muon is looking at the relative energy contributions to the
 3352 reconstructed track. In order to select a sample of such events, I require that a minimum
 3353 50% of the total energy comes from the pion and at least 20% from the muon.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3354 6.5.1 Track breakpoints

3355 To identify potential decays we can use the information we obtain from the Kalman
 3356 filter at each step of the fitted track. The simplest test we can think about is computing
 3357 the χ^2 of the mismatch between all the parameters in the forward and the backward fits:

$$3358 \quad \chi_k^{2(FB)} = (\hat{x}_k^B - \hat{x}_k^F)^T [V^{(\hat{x}_k, B)} + V^{(\hat{x}_k, F)}]^{-1} (\hat{x}_k^B - \hat{x}_k^F), \quad (6.23)$$

3358 where \hat{x}_k^F , \hat{x}_k^B are the Kalman filter state vector estimates at step k in the forward and
 3359 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.
 3360 Using the values of the χ^2 at measurement k for the forward and backward fits we can
 3361 compute another χ^2 value that characterises the overall track fit:

$$\chi_{track}^2 = \chi_k^{2(F)} + \chi_k^{2(B)} + \chi_k^{2(FB)}, \quad (6.24)$$

3362 which remains approximately constant for all k .

3363 An alternative approach proposed in the context of the NOMAD experiment was
 3364 using a fit with a more elaborate breakpoint hypothesis, so we can perform a comparison
 3365 of the χ^2 with and without breakpoints. This can be achieved by using some alternative
 3366 parametrisation with extra parameters, which allows some of the track parameters to
 3367 be discontinuous at certain points. A decay changes the momentum magnitude and
 3368 direction, so we can use the new state vector:

$$\alpha = (y, z, 1/R_F, 1/R_B, \phi_F, \phi_B, \tan\lambda_F, \tan\lambda_B)^T. \quad (6.25)$$

3369 As we already have the estimates from the standard Kalman filter and their
 3370 covariance matrices at each point, we do not need to repeat the Kalman fit for the new
 3371 parametrisation. Instead, I can compute the values of α at each point k that minimise

6.5. CHARGED PION DECAY IN FLIGHT

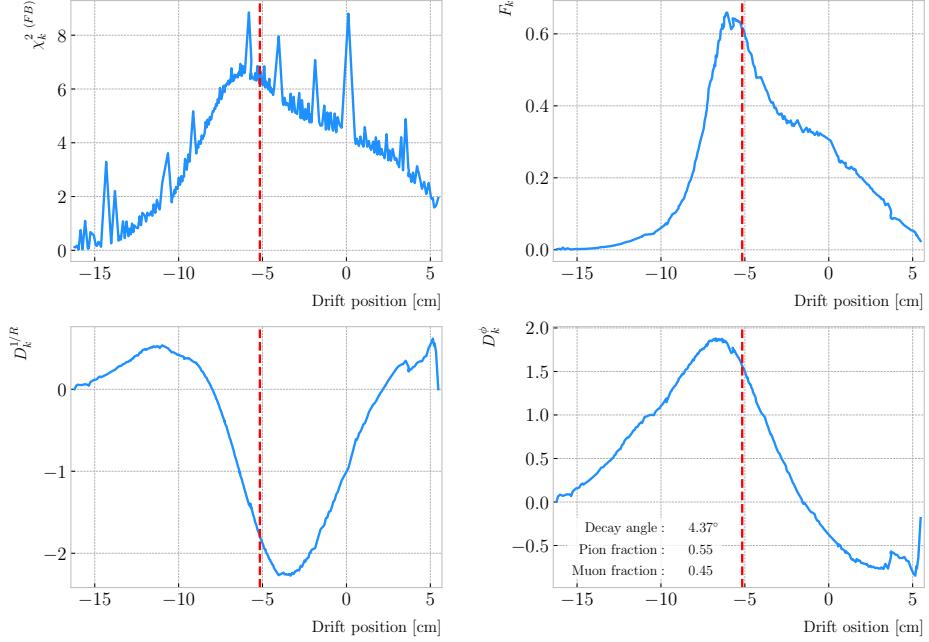


Figure 6.38: Values of $\chi_k^2(FB)$ (top left panel), F_k (top right panel), $D_k^{1/R}$ (bottom left panel) and D_k^ϕ (bottom right panel) versus position along the drift direction for a reconstructed track in a positive pion decay event. The vertical red dashed line indicates the true location of the decay point.

3372 the χ^2 resulting from comparing them to $\{\hat{x}_k^B, \hat{x}_k^F\}$. Introducing the two 5×8 matrices:

$$H^F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad H^B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (6.26)$$

3373 we can write this as:

$$\begin{aligned} \chi_k^2(FB)(\alpha) &= (\hat{x}_k^F - H^F \alpha)^T \left[V^{(\hat{x}_k, F)} \right]^{-1} (\hat{x}_k^F - H^F \alpha) \\ &\quad + (\hat{x}_k^B - H^B \alpha)^T \left[V^{(\hat{x}_k, B)} \right]^{-1} (\hat{x}_k^B - H^B \alpha). \end{aligned} \quad (6.27)$$

3374 The minimum of $\chi_k^2(FB)(\alpha)$ is found when the measured new state vector takes the
3375 value:

$$\hat{\alpha}_k = V^{(\hat{\alpha}_k)} H^T (V^{(\hat{x}_k)})^{-1} \hat{X}, \quad (6.28)$$

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

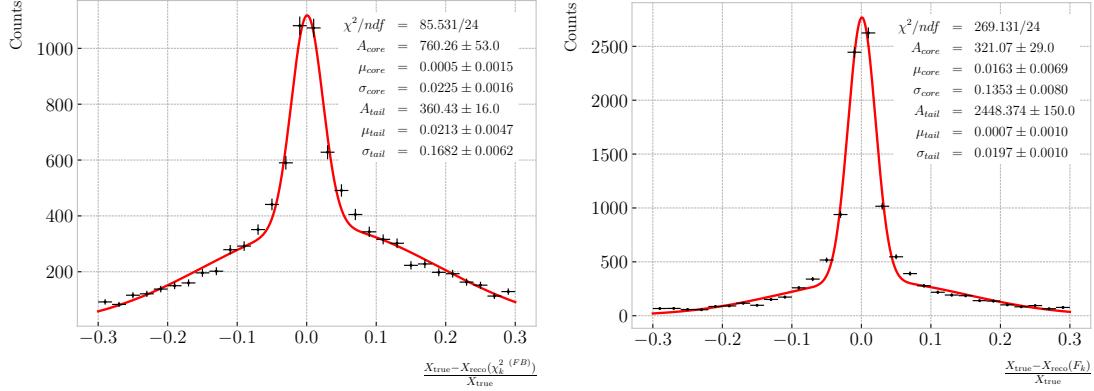


Figure 6.39: Fractional residual distributions of the true and reconstructed decay position along the drift coordinate, using the position of the maximum of $\chi_k^2(FB)$ (left panel) and F_k (right panel) as estimates of the decay position. Also shown are double Gaussian fits to these points (red lines).

3376 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 $V^{(\hat{x}_k, B)}$
3377 and $V^{(\hat{\alpha}_k)}$ is the covariance matrix of $\hat{\alpha}_k$, given by:

$$V^{(\hat{\alpha}_k)} = \left(H^T (V^{(\hat{x}_k)})^{-1} H \right)^{-1}. \quad (6.29)$$

3378 From these new fit estimates we can compute the F statistic, which tells us whether
3379 the model with breakpoint provides a statistically significant better fit:

$$F_k = \left(\frac{\chi_{track,k}^2 - \chi_{full,k}^2}{8 - 5} \right) / \left(\frac{\chi_{full,k}^2}{N - 8} \right). \quad (6.30)$$

3380 One can also compute the signed difference of the duplicated variables divided by
3381 their standard deviation at each point. These represent how significant the discontinuity
3382 in each variable is. For any variable η we can write it as:

$$D_k^\eta = \frac{\hat{\eta}_k^B - \hat{\eta}_k^F}{\sqrt{\text{Var}[\hat{\eta}_k^F] + \text{Var}[\hat{\eta}_k^B] - 2\text{Cov}[\hat{\eta}_k^F, \hat{\eta}_k^B]}}. \quad (6.31)$$

3383 In our case, the relevant ones to look at are $D_k^{1/R}$ and D_k^ϕ .

3384 Figure 6.38 shows the values of $\chi_k^2(FB)$, F_k , $D_k^{1/R}$ and D_k^ϕ as functions of the position
3385 along the drift direction, for an example reconstructed track with 55.5% of the energy

6.5. CHARGED PION DECAY IN FLIGHT

3386 coming from the charged pion and 45.5% from the daughter muon. The true position of
 3387 the decay is indicated (dashed red lines). Notice how $\chi_k^2(FB)$ and F_k , $D_k^{1/R}$ reach their
 3388 maxima near the decay point. In the former case this indicates a large forward-backward
 3389 difference in the track fit. In the later it represents that the extended state vector
 3390 improves the fit particularly around that point.

3391 I can estimate the decay position finding resolution by computing the difference
 3392 between the X position of the maxima of $\chi_k^2(FB)$ and F_k and the X position of the
 3393 true decay. Figure 6.39 represent the the fractional residual distributions for both
 3394 cases, from the sample of tracks containing pion decays. Fitting a double Gaussian to
 3395 the distributions (red lines) I find a resolution of $(3.31 \pm 0.15)\%$ and $(6.94 \pm 0.31)\%$
 3396 respectively.

3397 In principle, the F -statistic should follow a Fisher distribution with $(8 - 5)$ and
 3398 $(N - 8)$ degrees of freedom under the null hypothesis. In most of our cases $N \sim \mathcal{O}(100)$,
 3399 so the probability density functions will look very similar. In this case, it is safe to take
 3400 the limit $N \rightarrow \infty$ in the Fisher PDF:

$$\begin{aligned}\tilde{f}(x; a - b) &= \lim_{N \rightarrow \infty} f(x; a - b, N - a) \\ &= \frac{2^{-\frac{a-b}{2}}}{\Gamma\left(\frac{a-b}{2}\right)} (a - b)^{\frac{a-b}{2}} x^{\frac{a-b}{2}-1} e^{-\frac{a-b}{2}x}.\end{aligned}\tag{6.32}$$

3401 In our case $a - b = 8 - 5 = 3$, so we would obtain a p-value of 0.05 at $x = 2.60$.

3402 Figure 6.40 contains the distributions of the maxima of $\chi_k^2(FB)$, F_k and D_k^ϕ and the
 3403 minima of $D_k^{1/R}$ for a sample of non-decaying pion tracks (blue) and another sample of
 3404 reconstructed tracks containing part of the pion and the daughter muon from a decay
 3405 inside the fiducial volume (red). Notice that, even though the values of $F_k^{(max)}$ for the
 3406 decay sample are typically larger than for the non-decaying one, just a small fraction of
 3407 the events go beyond the aforementioned value of $F = 2.60$. Therefore, from a practical
 3408 point of view, it is not the most efficient variable to use for selecting the decay events.

3409 However, looking at the $D_k^{1/R \text{ (min)}}$ distribution we can see there is a big difference
 3410 between non-decaying and decaying events in this variable. One can use a combination

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

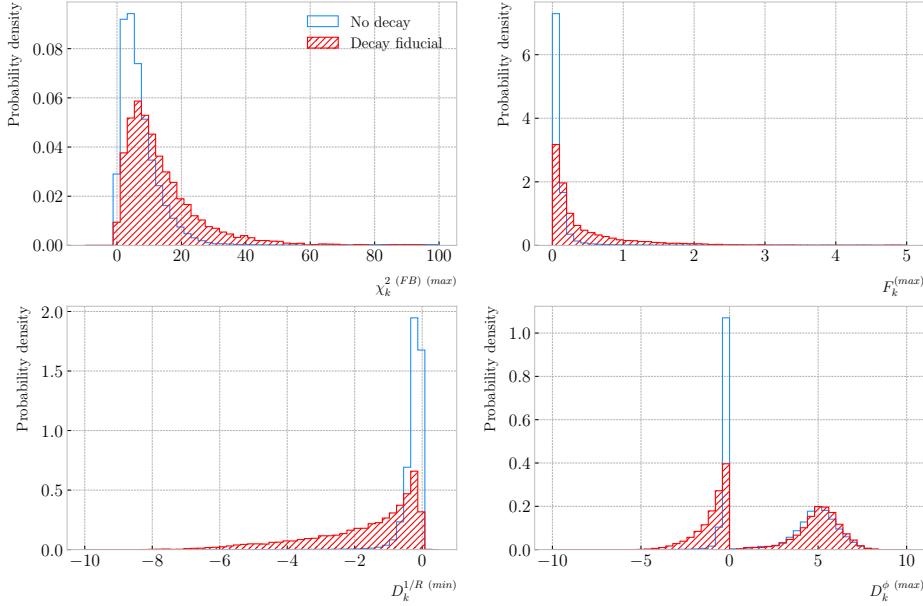


Figure 6.40: Distributions of the extreme values of $\chi_k^2(FB)$ (top left panel), F_k (top right panel), $D_k^{1/R}$ (bottom left panel) and D_k^ϕ (bottom right panel) for non-decaying reconstructed pion tracks (blue) and tracks which include the decay inside the fiducial volume (red).

3411 of these four variables to distinguish between the pion decay events (signal) and the
 3412 non-decaying pions (background).

3413 An approach to this classification could be using a boosted decision tree (BDT). One
 3414 of the advantages of BDTs is that they are easy to interpret and identify the relative
 3415 importance of the different input variables. Training a BDT with 400 estimators and a
 3416 maximum depth of 4 I can obtain an efficient classification without overtraining. Figure
 3417 6.41 (left panel) shows the distribution of probabilities predicted by the BDT for a test
 3418 sample. The signal efficiency as a function of background acceptance, the so-called ROC
 3419 curve, is shown in Fig. 6.41 (right panel). With a relative importance of 0.83, the most
 3420 important variable turned out to be $D_k^{1/R} \text{ (min)}$.

3421 One thing we can check is how the resolution to the decay and the signal efficiency in
 3422 the classification changes with the true decay angle. Using an equal-frequency binning
 3423 for the decay angles, we can repeat the previous steps for each bin.

3424 Figure 6.42 (left panel) shows the dependence on the decay angle of the decay finding

6.5. CHARGED PION DECAY IN FLIGHT

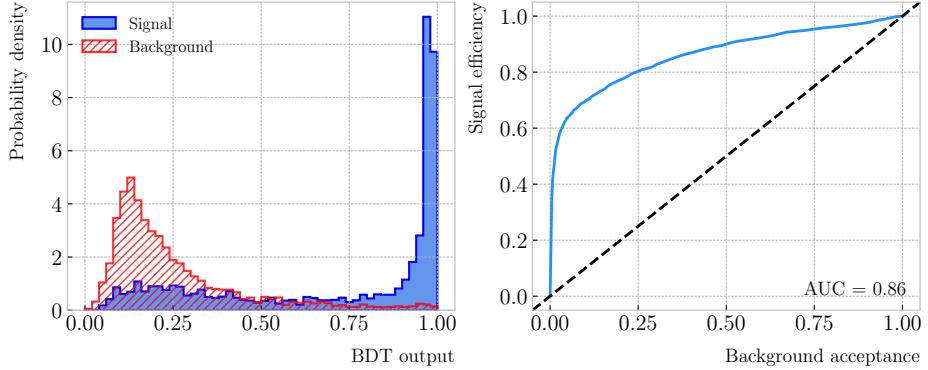


Figure 6.41: Left panel: distributions of the predicted probabilities assigned by the BDT classifier to a test sample of decaying pion+muon tracks (blue) and non-decaying pion tracks (red). Left: signal efficiency versus background acceptance (ROC curve) obtained from the BDT for the test sample.

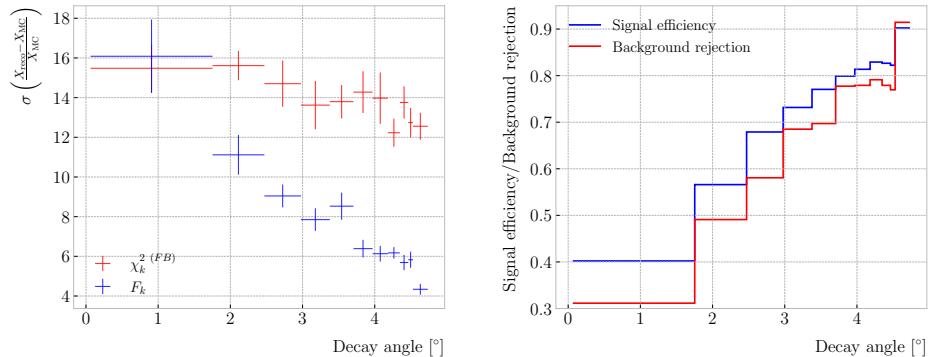


Figure 6.42: Left panel: dependence of the decay position finding resolution on the true value of the decay angle for the $\chi_k^2(FB)$ (red) and F_k (blue) methods. Right panel: signal efficiency (blue line) and background rejection (red line) from the BDT classifier versus true decay angle.

resolution. We can see that for the $\chi_k^2(FB)$ maximum location method the resolution consistently lies between 12 to 16%. However, the $F_k^{(\max)}$ approach gives a significantly better resolution for high angle values, reaching the 4 – 6% range for decay angles $\geq 4^\circ$.

For the classification dependence on the angle, I use the same classifier I trained before but evaluating the test sample for each individual angular bin. I compute the signal efficiency in each bin for a fixed value of the background rejection, in this case 90%. Similarly, for the background rejection estimation I use a fixed signal efficiency value of 90%. Figure 6.42 (right panel) represents the change in signal efficiency (blue)

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

3433 and background rejection (red) with the value of the true decay angles.

3434 6.6 Neutral particle identification

3435 6.6.1 ECal clustering

3436 Another important reconstruction item is the clustering algorithm of ECal hits in
3437 GArSoft. The default module features a NN algorithm that treats all hits in the same
3438 way, independently of the layer each hit comes from. However, the current ECal design
3439 of ND-GAr has two very different types of scintillator layers. The inner layers are made
3440 out of tiles, which provide excellent angular and timing resolutions. On the other hand,
3441 the outer layers are cross scintillator strips. That way, an algorithm that treats hits
3442 from both kinds of layers differently may be able to improve the current performance.

3443 Inspired by the reconstruction of T2K’s ND280 downstream ECal [184], the idea
3444 was to put together a clustering module that first builds clusters for the different ECal
3445 views (tiles, strips segmented in the X direction and strips segmented in Y direction),
3446 and then tries to match them together to form the final clusters.

3447 Working on a module-by-module basis, the algorithm first separates the hits depending
3448 on the layer type they come from. Then, it performs a NN clustering for the 3 sets of
3449 hits separately. For the tile hits it clusters together all the hits which are in nearest-
3450 neighbouring tiles and nearest-neighbouring layers, for strip hits it looks at nearest-
3451 neighbouring strips and next-to-nearest-neighbouring layers (as the layers with strips
3452 along the two directions are alternated). For strip clusters an additional cut in the
3453 direction along the strip length is needed.

3454 After this first clustering I then apply a recursive re-clustering for each collection
3455 of strip clusters based on a PCA method. In each case, we loop over the clusters with
3456 $N_{hits} \geq 2$, computing the centre of mass and three principal components. Propagating
3457 these axes up to the layers of the rest of the clusters, we check if the propagated point
3458 and the centre of mass of the second cluster are within next-to-nearest-neighbouring
3459 strips. An additional cut in the direction along the strip length is also needed. Moreover,

6.6. NEUTRAL PARTICLE IDENTIFICATION

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

3460 I require that the two closest hits across the two clusters are at most in next-to-nearest-
3461 neighbouring strips. I merge the clusters if these three conditions are satisfied. The
3462 re-clustering is repeated until no more cluster pairs pass the cuts.

3463 The clusters in each strip view are combined if their centres of mass are close enough
3464 and they point in the same direction. An alternative approach for the strip cluster
3465 merging could be to compute the overlap between the ellipsoids defined by the principal
3466 axes of the clusters, and then merge the pair if the overlap exceeds some threshold.
3467 Further study is needed to understand if this change would have an impact in the overall
3468 clustering performance.

3469 To merge the tile clusters to the combined strip clusters I propagate the principal
3470 axis of the strip cluster towards the inner layers, up to the centre of mass layer of the
3471 tile cluster. I merge the clusters if the distance between the propagated point and the
3472 centre of mass is below a certain cut.

3473 The last step is to check if clusters in neighbouring modules should be merged
3474 together, both across two barrel modules, across end cap modules and between barrel
3475 end cap modules. I check the distance between the two closest hits in the pair of clusters
3476 and merge them if it passes this and an additional direction cut.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

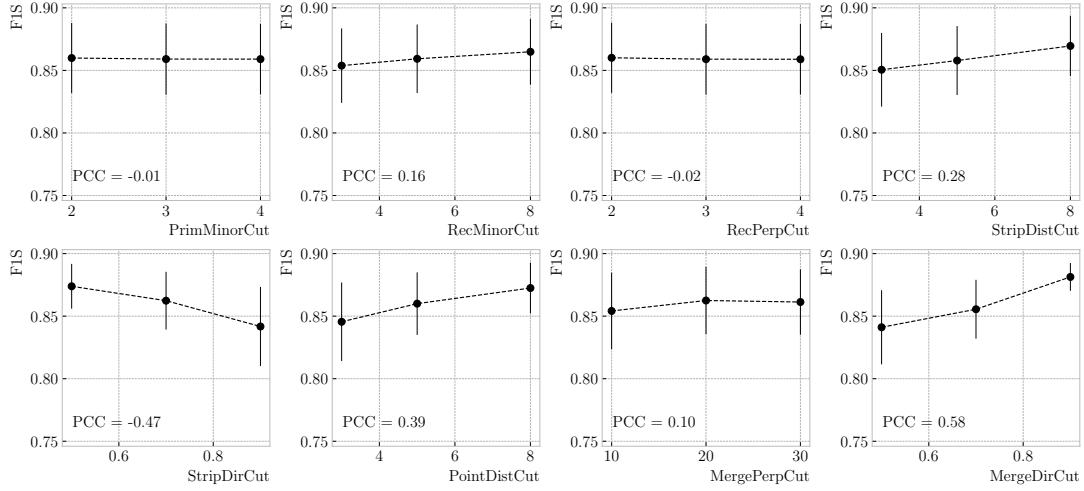


Figure 6.43: Mean values of the F_1 -score marginal distributions for the different free parameters of the new clustering algorithm, with the error bars representing one standard deviation around the mean. The F_1 -score values were computed for the 6561 possible parameter configurations using 1000 ν_μ CC interaction events.

3477 This algorithm has a total number of eight free parameters that need to be optimised.

3478 I used a sample of 1000 ν_μ CC interactions in order to obtain the optimal configuration of
 3479 clustering parameters. This sample was generated up to the default ECal hit clustering
 3480 level, so then I could run the new clustering algorithm each time with a different
 3481 configuration of parameters. As the number of parameters is relatively large, I only
 3482 performed a coarse-grained scan of the parameter space. Sampling each of the eight
 3483 parameters at three different points each I obtain 6561 different configurations. These
 3484 parameters, together with the used values, are summarised in Tab. 6.5.

3485 In order to measure the performance of the clustering, I use a binary classification
 3486 approach. For each formed cluster, I identify the Geant4 Track ID of the matching MC
 3487 particle and the energy fraction of each hit. Then, I assign to each cluster the Track ID
 3488 with the highest total energy fraction. For each of the different Track IDs associated to
 3489 the clusters, I select the cluster with the highest energy (only from the hits with the
 3490 same Track ID). I identify such a cluster as the main cluster for that Track ID. I count
 3491 as true positives (TPs) the hits with the correct Track ID in each main cluster. False
 3492 positives (FPs) are the hits with the incorrect Track ID for the cluster they are in, not

6.6. NEUTRAL PARTICLE IDENTIFICATION

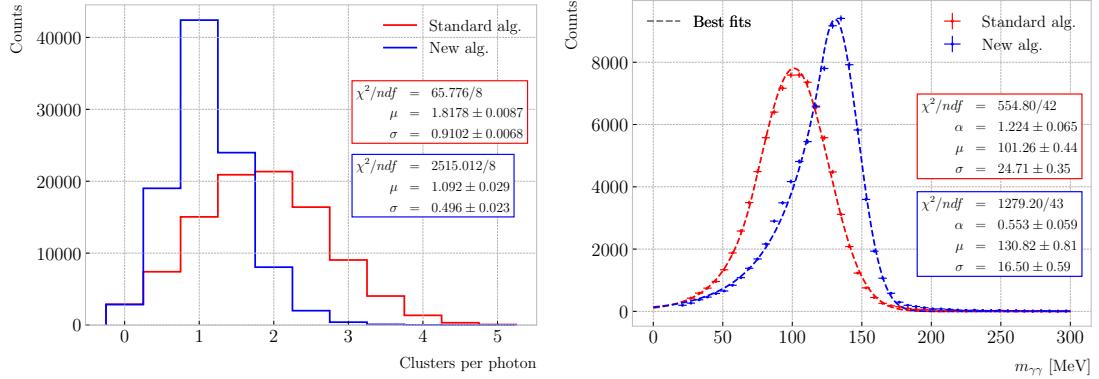


Figure 6.44: Left panel: distributions of the number of ECal clusters per photon from π^0 decays for the standard (red) and new (blue) clustering algorithms. Right panel: reconstructed invariant mass distributions for photon pairs from single π^0 events using the standard (red) and new (blue) ECal clustering algorithms.

only main clusters. The false negatives (FNs) are the hits with the correct Track ID in clusters other than the main.

Figure 6.43 shows the computed F_1 -score values for the different cuts. In each case, the central value represents the mean of the F_1 -score distribution for the specified value of the corresponding variable and the vertical error bar represents one standard deviation around the mean. Also shown are the Pearson correlation coefficients of these central values. We can see that five of the variables have a sizeable effect on the F_1 -score, with an absolute difference between the last and first values as big as 4%.

The working configuration is obtained as follows. I first select all configurations with purity $\geq 90\%$. Among those, I choose the combinations that yield the maximum F_1 -score. If more than one configuration remains I select the one with the highest sensitivity. Doing so, I end up with a parameter configuration with an efficiency of 88% and a 90% purity. Compared with the default algorithm, which gives an efficiency of 76% and a purity of 91% for the same sample, I have managed to improve the efficiency by a factor of 1.16.

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAR

3508 6.6.2 π^0 reconstruction

3509 One of the potential applications of the new ECal hit clustering is the reconstruction of
 3510 neutral particles, in particular pions. Neutral pions decay promptly after being produced,
 3511 through the $\pi^0 \rightarrow \gamma\gamma$ channel (98.823 ± 0.034)% of the time. The photon pair does
 3512 not leave any traces in the HPgTPC (unless one or both of them converts into an
 3513 electron-positron pair), but each of them will produce an electromagnetic shower in
 3514 the ECal.

3515 To test the potential impact of the new algorithm in π^0 reconstruction, I generated
 3516 a MC sample of single, isotropic neutral pions inside the HPgTPC. All pions were
 3517 generated with a momentum $p = 500$ MeV and their initial positions were uniformly
 3518 sampled inside a $2 \times 2 \times 2$ m box aligned with the centre of the TPC. I ran both the
 3519 default and the new clustering algorithms, using for the latter the optimised configuration
 3520 discussed above.

3521 The first thing to notice is that the number of clusters produced per photon has
 3522 decreased. Figure 6.44 (left panel) shows these distributions for the default (red) and
 3523 new (blue) algorithms. Using a simple Gaussian fit, we see that the mean number of
 3524 ECal clusters per photon went from 1.82 ± 0.01 to 1.09 ± 0.03 . This effectively means that
 3525 with the new algorithm the ECal activity of one true particle is typically reconstructed
 3526 as a single object. From the reconstruction point of view this can be an advantage. As
 3527 now most of the photon energy ends up in a single ECal cluster, I can simply use cluster
 3528 pairs to identify the π^0 decay.

3529 In general, one calculates the invariant mass of the photon pair as:

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos \theta)}, \quad (6.33)$$

3530 where E_i are the energies of the photons and θ the opening angle between them. In this
 3531 case I can use the energies deposited in the ECal and their incident directions. This
 3532 quantity is computed for all possible pairs of clusters, using their position together with
 3533 the true decay point. In a more realistic scenario, e.g. ν_μ CC interaction, one could use

6.7. INTEGRATION IN GArSOFT

3534 the position of the reconstructed primary vertex instead. I also tried to use the principal
3535 direction of the clusters, but that approach gave considerably worse results. For each
3536 event I only keep the pair with an invariant mass closer to the true π^0 mass value.

3537 Figure 6.44 (right panel) shows the invariant mass distributions for the photon pairs
3538 we get using the default (red) and the new (blue) ECal clustering algorithms. For the fit
3539 I used a modified version of the Crystal Ball function [185], obtained by taking the limit
3540 where the parameter controlling the power-law tail goes to infinity:

$$f(x; N, \mu, \sigma, \alpha) = N \cdot \begin{cases} e^{\frac{\alpha(2x-2\mu+\alpha\sigma)}{2\sigma}}; & x \leq \mu - \alpha\sigma, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}; & x > \mu - \alpha\sigma. \end{cases} \quad (6.34)$$

3541 Comparing the fitted mean and standard deviation values for the Gaussian cores, we
3542 see that the distribution for the new algorithm is a 67% narrower and also peaks much
3543 closer to the true m_{π^0} value, going from 101.3 ± 0.4 MeV to 130.8 ± 0.6 MeV.

3544 6.7 Integration in GArSoft

3545 All the additions and improvements to the reconstruction discussed in this Chapter
3546 had to be integrated in the GArSoft framework. This is necessary both to allow a
3547 more streamlined path for development, as this makes testing and adding features
3548 straightforward, as well as make the changes usable in future productions of simulated
3549 data. In this section, I outline the current status of the integration in GArSoft of the
3550 reconstruction work presented above.

3551 The new track-cluster association code has been implemented in GArSoft, under
3552 the name of `TPCECALAssociation2`, and has now become the new default in the
3553 reconstruction. The structure of the module is similar to the previous implementation,
3554 and the data products they output are identical in form. Therefore, any existing code
3555 using the association objects does not need to be modified.

3556 The computation of the truncated mean dE/dx of the tracks, the evaluation of
3557 the muon score for muon and pion separation, and the estimation of the velocity from

CHAPTER 6. PARTICLE IDENTIFICATION IN ND-GAr

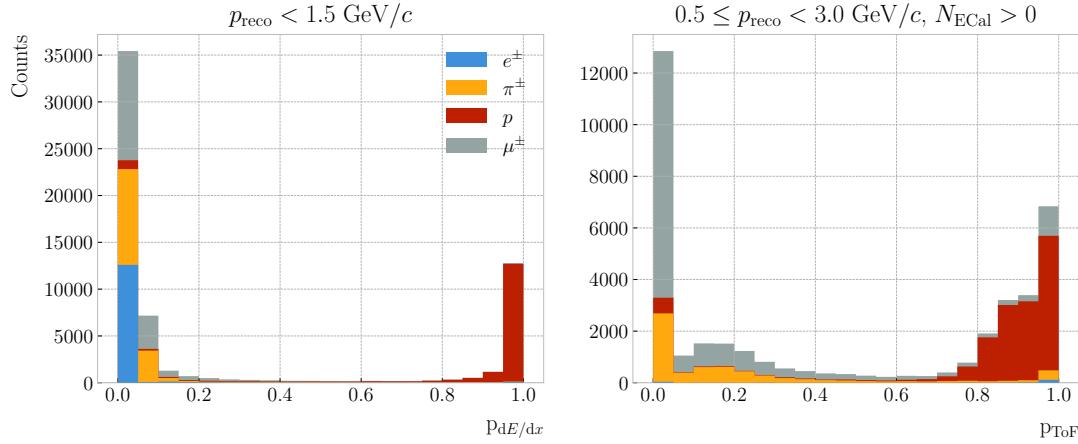


Figure 6.45: Distributions of proton dE/dx (left panel) and ToF (right panel) scores for a sample of 100000 FHC neutrino interactions in the HPgTPC. The distributions are broken down by the true particle type associated to the reconstructed particle.

time-of-flight are all orchestrated by the `CreateRecoParticles` module. Each one of these is implemented as a separate algorithm, which is then called by the parent module. This generates the `gar::rec::RecoParticle` products, a new high-level data object in GArSoft. These combine the information from the HPgTPC, ECal, and μ ID to create an object useful for analysers. At the moment, these data products are only generated for charged particles. However, in the future the module can be extended to incorporate other algorithms used for the identification of neutral particles, like neutral pions and neutrons.

Additionally, analogous to the muon score, the `gar::rec::RecoParticle` objects contain two other scores based on the $\langle dE/dx \rangle$ and ToF estimates which measure the “protoness” of a reconstructed particle. These are obtained in a number of momentum bins, and are a measure of the distance to the point in the corresponding distribution that maximises the F_1 -score for the proton separation. This distance is then transformed applying a sigmoid function, which produces a score in the $0 - 1$ range, with coefficients obtained following a procedure similar to the one used to calibrate the response of the muon score. The dE/dx proton score is defined for all particles with momenta $p_{\text{reco}} < 1.5 \text{ GeV}/c$, whereas the ToF proton score is available for the particles with at least one associated hit in the inner ECal and momentum in the range $0.5 \leq p_{\text{reco}} < 3.0 \text{ GeV}/c$.

6.7. INTEGRATION IN GArSOFT

3576 As an example, Fig. 6.45 shows the distributions of the dE/dx (left panel) and ToF
3577 (right panel) proton scores for the reconstructed particles in a 100000 FHC neutrinos
3578 sample.

3579 The calculation of the track breakpoint variables for pion decay identification is
3580 currently implemented as an analysis module in GArSoft. It would be interesting to add
3581 this information to the `gar::rec::RecoParticle` products, possibly calling the code as
3582 an additional algorithm in the `CreateRecoParticles` module. However, the best way
3583 to propagate the information to the high-level objects is still unclear.

3584 About the new ECal clustering algorithm, it is still in a development phase, and
3585 as such it has not replaced the current clustering module. At the moment, its latest
3586 version is integrated in GArSoft as the `CaloClustering2` module. The algorithm used
3587 is implemented separately, and then invoked in the main code. The module can be
3588 run standalone on the outputs of the reconstruction, creating a second instance of the
3589 `gar::rec::Cluster` collection. In the future it may replace the current algorithm as
3590 the default in the reconstruction chain. However, more work is needed in order to
3591 understand its performance in all the different use cases.

Event selection in ND-GAr

3594 *You have power over your mind, not outside events. Realise this, and you will
3595 find strength.*

3596 – Marcus Aurelius, *Meditations*

3597 As discussed previously, it is necessary to evaluate the capabilities of ND-GAr at
3598 identifying different particles. In the context of the LBL analysis, we want ND-GAr to
3599 provide data samples containing events of specific topologies, like ν_μ CC $1\pi^\pm$, ν_μ CC
3600 $1p1\pi^\pm$, etcetera. Thus, developing a strategy for the event selection using the current
3601 reconstruction is required.

3602 In this Chapter, I present the results of a number of preliminary studies focused on
3603 the event selection in ND-GAr, particularly the ν_μ CC selection and the pion tagging
3604 strategies. I also investigate the neutrino energy reconstruction, as well as the systematic
3605 uncertainties relevant for our detector.

3606 7.1 Data sample

3607 For the event selection studies I used a MC sample consisting of 10^5 FHC neutrino
3608 interaction events inside the HPgTPC volume. The version of **GENIE** used was v3_04_00,
3609 with the G18 tune. This is a preliminary version of the re-tune produced from CCQE,
3610 CC 1π , CC 2π , and CC inclusive bubble chamber cross section data [186]. It uses the local
3611 Fermi gas as a description of the nuclear model. The quasielastic-like events are described
3612 by the Nieves quasielastic [187] and Valencia $2p2h$ [188] models. The Berger-Seghal

CHAPTER 7. EVENT SELECTION IN ND-GAR

3613 model [189, 190] is used for the resonant and coherent pion production. As in all the
3614 GENIE tunes, the Bodek-Yang model [191] describes the DIS interactions. Finally, the
3615 FSI are described using the effective intranuclear transport model in INTRANUKE.

3616 For this sample, I used `GArG4` instead of `edep-sim` for the particle propagation.
3617 Because both `Geant4` wrappers use different configurations for the simulation, the results
3618 obtained are different. The default `edep-sim` configuration used by the DUNE ND
3619 is appropriate for ND-LAr, where thresholds for particle production are higher. In
3620 the case of ND-GAr, these parameters need to be adjusted accordingly. For the time
3621 being, in these first productions of analysis files, we will use our standalone `Geant4`
3622 implementation.

3623 The detector simulation and reconstruction used was GArSoft version v02_21_00. I
3624 made use of the standard routines for the readout simulation and the reconstruction,
3625 which include the additions described in section 6.7. A summary of the GArSoft outputs
3626 is extracted in the form of a plain `ROOT TTree`. These are then used, together with the
3627 GENIE output files, to produce the files known within DUNE as common analysis files
3628 (CAFs). The version of the CAF format used in this analysis is `duneanaobj v3_07_00`.

3629 This sample only includes single interaction events. In the future, we will move
3630 to simulate full neutrino spills. Also, we will need to include neutrino interactions in
3631 the other detector volumes (ECal, magnet, . . .), as well as rock muons making it to
3632 ND-GAr. However, this will require a significant amount of work to go into the so-called
3633 interaction slicer, the part of the reconstruction in charge of splitting the reconstructed
3634 events.

3635 Looking forward, these sort of small samples are useful to prepare for launching a
3636 full production of ND-GAr events. In the original DUNE TDR LBL analysis, the event
3637 rates are calculated with a 1.1×10^{21} POT/year assumption, which assumes a combined
3638 uptime and efficiency of the accelerator complex and the LBNF beamline of 57% [59].
3639 If we have one spill every 1.2 s, that translates into 7.5×10^{13} POT/spill. Therefore,
3640 assuming that the POT/spill scales linearly with beam power, in Phase II we will have
3641 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

3642 using the same efficiency. The event rates per year in ND-GAr computed for these two
 3643 possible values of the POT/year are shown in Tab. 7.1.

3644 The latest PRISM plan requires 1.50 POT · years of data on-axis, followed by
 3645 0.25 POT · years at each off-axis position ($2, 4, 8, 12, 16, 20, 24$, and 28 m), both for
 3646 FHC and RHC mode. This implies that a full on-axis ND-GAr production will require
 3647 a total of 2.85×10^{21} POT for both horn currents. The production of these samples
 3648 is necessary to understand the impact of ND-GAr on the LBL sensitivities, and the
 3649 studies presented here should be considered as a first step towards the realisation of
 3650 such analysis.

3651 7.2 ν_μ CC selection

3652 In a ν_μ CC inclusive selection, the signal topology we look for is a neutrino-induced
 3653 muon with or without other final state particles. Here, I also require the neutrino vertex
 3654 to be located inside the fiducial volume (FV) of ND-GAr.

3655 The FV is defined as a smaller cylinder within the cylindrical volume of the HPgTPC.

CHAPTER 7. EVENT SELECTION IN ND-GAR

3656 The FV has a radius R_{FV} and a half-length L_{FV} . For a particle position to lie within
3657 the FV it must satisfy:

$$\vec{x}_i \in \left\{ \vec{x} \in \mathbb{R}^3 \mid |x_0| \leq L_{\text{FV}} \text{ \& } \sqrt{x_1^2 + x_2^2} \leq R_{\text{FV}} \right\}, \quad (7.1)$$

3658 in the reference frame of the HPgTPC. For convenience, I define:

$$\begin{aligned} \Delta R_{\text{FV}} &= R_{\text{HPgTPC}} - R_{\text{FV}}, \\ \Delta L_{\text{FV}} &= L_{\text{HPgTPC}} - L_{\text{FV}}, \end{aligned} \quad (7.2)$$

3659 where R_{HPgTPC} and L_{HPgTPC} refer to the radius and the half-length of the HPgTPC,
3660 respectively. Figure 7.1 shows the HPgTPC volume with the FV inside of it. In that
3661 representation, the FV is defined as $\Delta L_{\text{FV}} = 30.0$ cm and $\Delta R_{\text{FV}} = 30.0$ cm. Also shown
3662 is the HPgTPC reference frame, with x being the drift direction and z aligned along the
3663 beam direction.

3664 In some cases, it is interesting to divide the signal events in different categories
3665 based on their true interaction mode. In this work, I will distinguish between charged-
3666 current quasi-elastic (CCQE), coherent (CCCOH), resonant (CCRES), and deep-inelastic
3667 (CCDIS) interactions. I also use a separate category for the interactions not included in
3668 any of the other categories (CCOther).

3669 Any other events are considered backgrounds. For this selection, I use the following
3670 categorisation of background events:

- 3671 • Out of FV: if the true neutrino vertex lies outside the defined FV.
- 3672 • NC: if the event is a true neutral-current event.
- 3673 • $\bar{\nu}_\mu$ CC: if the true neutrino candidate is of muon antineutrino flavour.
- 3674 • Other: if the event is not signal nor falls in any of the other background categories.

3675 The key to the CC selection is the identification of a primary muon candidate.
3676 Typically, this is the longest track in the event. However, sometimes protons and pions

7.2. ν_μ CC SELECTION

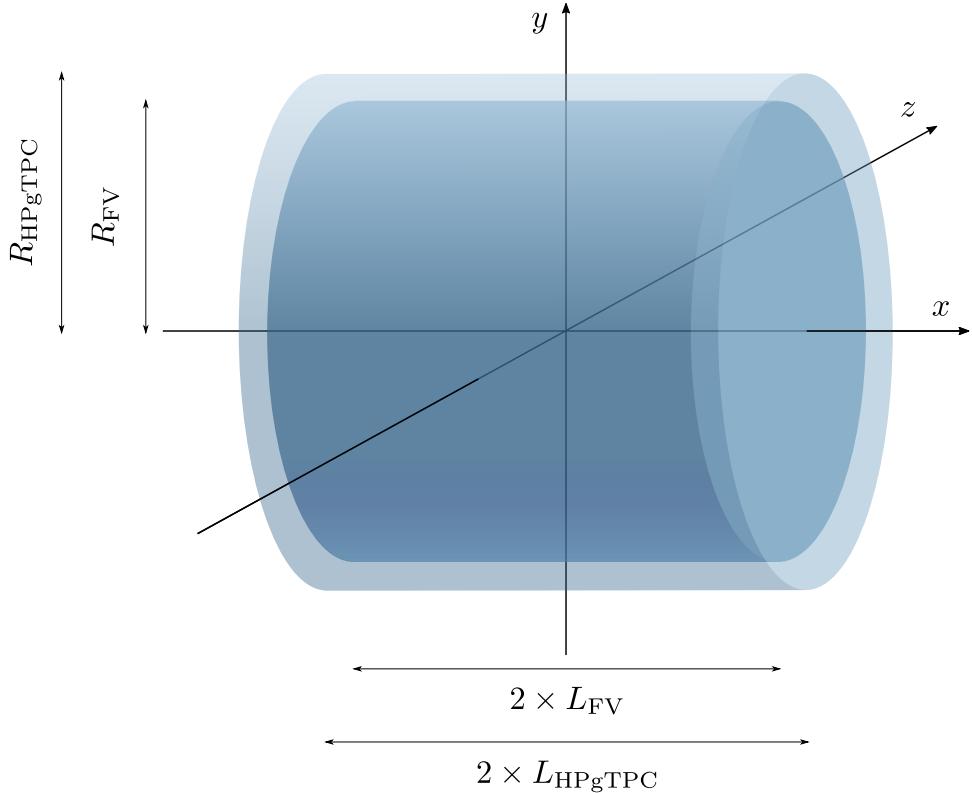


Figure 7.1: Schematic diagram of the HPgTPC including the fiducial volume (FV). In this case the FV is given by $\Delta L_{\text{FV}} = 30.0$ cm and $\Delta R_{\text{FV}} = 30.0$ cm.

3677 leave tracks longer than that of the muon. This is particularly important in the GAr
 3678 medium, considerably less dense than the LAr. For this reason, the muon identification
 3679 in ND-GAr relies heavily on the capabilities of the ECal.

3680 The selection strategy proposed combines the information coming from the three
 3681 main detection systems of ND-GAr: the HPgTPC charge readout, and the ECal and
 3682 μ ID detectors. It consists of five steps:

- 3683 1. Event contains reconstructed particles.
- 3684 2. Select particles with reconstructed negative charge, $q_{\text{reco}} = -1$.
- 3685 3. Select particles passing the muon score cut, $\mu_{\text{score}} \geq \mu_{\text{score}}^{\text{cut}}$.
- 3686 4. Keep reconstructed particle with the highest momentum, $\max [p_{\text{reco}}]$.
- 3687 5. Check that the remaining particle starts within the FV.

CHAPTER 7. EVENT SELECTION IN ND-GAR

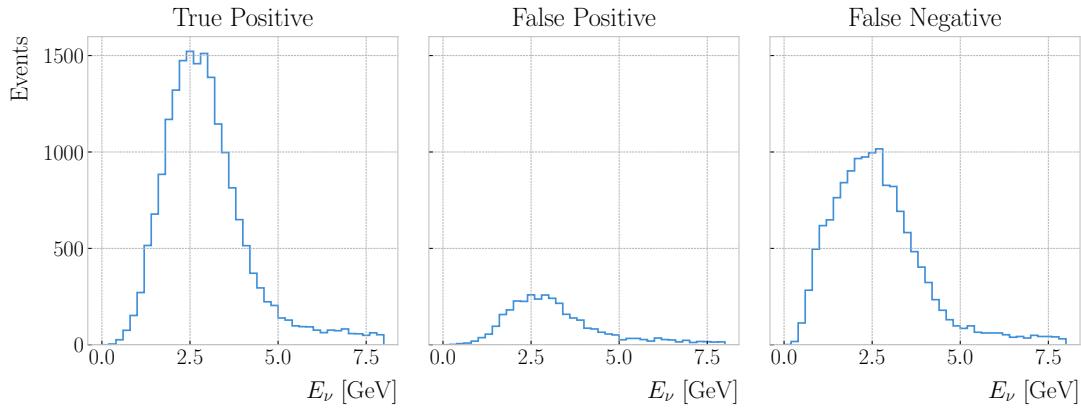


Figure 7.2: True positive (left panel), false positive (middle panel), and false negative (right panel) true neutrino energy distributions for the ν_μ CC selection given by a muon score cut of $\mu_{\text{score}}^{\text{cut}} = 0.75$, and a FV defined as $\Delta L_{\text{FV}} = 30.0$ cm and $\Delta R_{\text{FV}} = 30.0$ cm.

3688 All the events passing these cuts are classified as signal, and the selected particle is
 3689 regarded as the primary muon candidate.

3690 **7.2.1 Selection optimisation**

3691 I performed an optimisation of this selection, comparing the performance of a number of
 3692 configurations. For the muon selection, I varied the value of $\mu_{\text{score}}^{\text{cut}}$ from 0.05 to 0.95,
 3693 using a step size of 0.05. Additionally, to optimise the FV, I systematically explored a
 3694 number of different parameter configurations, moving within the 10.0 – 70.0 cm range for
 3695 ΔL_{FV} and 25.0 – 75.0 cm for ΔR_{FV} , in increments of 10.0 cm and 5.0 cm respectively.

3696 For each parameter configuration, I extract three different true neutrino energy
 3697 distributions. These are built combining the results of the selection described previously,
 3698 which we can refer to as the “reco” selection, and a “true” selection. The later identifies
 3699 the true ν_μ CC events using the GENIE event records, and checks that the true neutrino
 3700 vertices are contained in the FV.

3701 The first distribution consists of the events passing both selections, i.e., these are
 3702 the true ν_μ CC events which pass the “reco” selection. The second distribution contains
 3703 the events passing the “reco” selection but failing the “true” selection. These are
 3704 the background events that the selection misidentifies. Finally, the third distribution

7.2. ν_μ CC SELECTION

3705 corresponds to the events picked by the “true” selection but not by the “reco” one. In
 3706 other words, these are the true ν_μ CC events that our selection misses. In analogy to
 3707 the machine learning jargon, I refer to these distributions as the true positive (TP),
 3708 false positive (FP), and false negative (FN) spectra, respectively. Figure 7.2 shows an
 3709 example of these three distributions for the case $\mu_{\text{score}}^{\text{cut}} = 0.75$, $\Delta L_{\text{FV}} = 30.0$ cm, and
 3710 $\Delta R_{\text{FV}} = 30.0$ cm.

3711 By making different combinations of these distributions one can compute a series of
 3712 performance metrics. Using the full information from the spectra allows to obtain the
 3713 scores as a function of the true neutrino energy, whereas the totals can be obtained by
 3714 integrating the histograms. This way, the efficiency of the selection is given by:

$$\begin{aligned} \text{Efficiency} &= \frac{\text{Selected true } \nu_\mu \text{ CC events}}{\text{Total true } \nu_\mu \text{ CC events}} \\ &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \end{aligned} \quad (7.3)$$

3715 while the purity can be written as:

$$\begin{aligned} \text{Purity} &= \frac{\text{Selected true } \nu_\mu \text{ CC events}}{\text{Total selected events}} \\ &= \frac{\text{TP}}{\text{TP} + \text{FP}}. \end{aligned} \quad (7.4)$$

3716 Another scoring metric typically used when quantifying the performance of a selection
 3717 is the significance. It is defined as:

$$\text{Significance} = \frac{S}{\sqrt{S+B}} = \frac{\text{TP}}{\sqrt{\text{TP}+\text{FP}}}. \quad (7.5)$$

3718 The significance measures the relative size of the true signal within the selection, $S = \text{TP}$
 3719 with respect to one standard deviation of the counting experiment. Assuming Poisson
 3720 statistics, the variance is equal to the number of observations, and therefore the standard
 3721 deviation equals to $\sqrt{N} = \sqrt{S+B} = \sqrt{\text{TP}+\text{FP}}$. I use this metric to

3722 Figure 7.3 shows the change in efficiency (blue) and purity (red) of the ν_μ CC
 3723 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

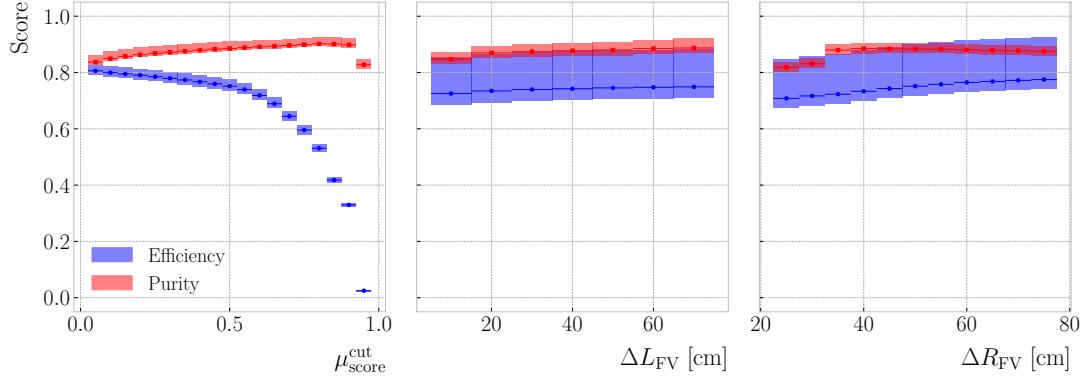


Figure 7.3: Efficiency (blue) and purity (red) for the ν_μ CC selection as a function of the muon score cut (left panel), FV half-length cut (middle panel), and radial cut (right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the horizontal line corresponds to the median.

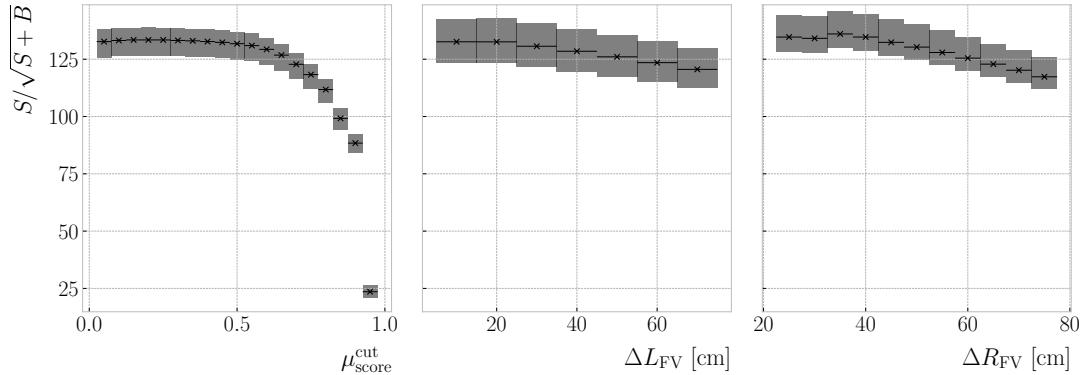


Figure 7.4: Significance for the ν_μ CC selection as a function of the muon score cut (left panel), FV half-length cut (middle panel), and radial cut (right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the horizontal line corresponds to the median.

3724 and ΔR_{FV} . For each value of the cuts, I compute the median and IQR (represented
 3725 by the horizontal lines and the heights of the boxes, respectively) of the corresponding
 3726 conditional distributions of efficiency and purity. This representation is useful to get
 3727 an idea of the general trend the scores follow with the cuts, as well as the spread. It
 3728 is clear that the muon score cut has the biggest impact on the efficiency, which ranges
 3729 between 0.05 to 0.80, whereas the purity remains stable with values around 0.85.

3730 A similar depiction of the significance can be found in Fig. 7.4. In this case, one can
 3731 see that the $S/\sqrt{S+B}$ decreases as the cuts grow tighter. However, there are hints of

7.2. ν_μ CC SELECTION

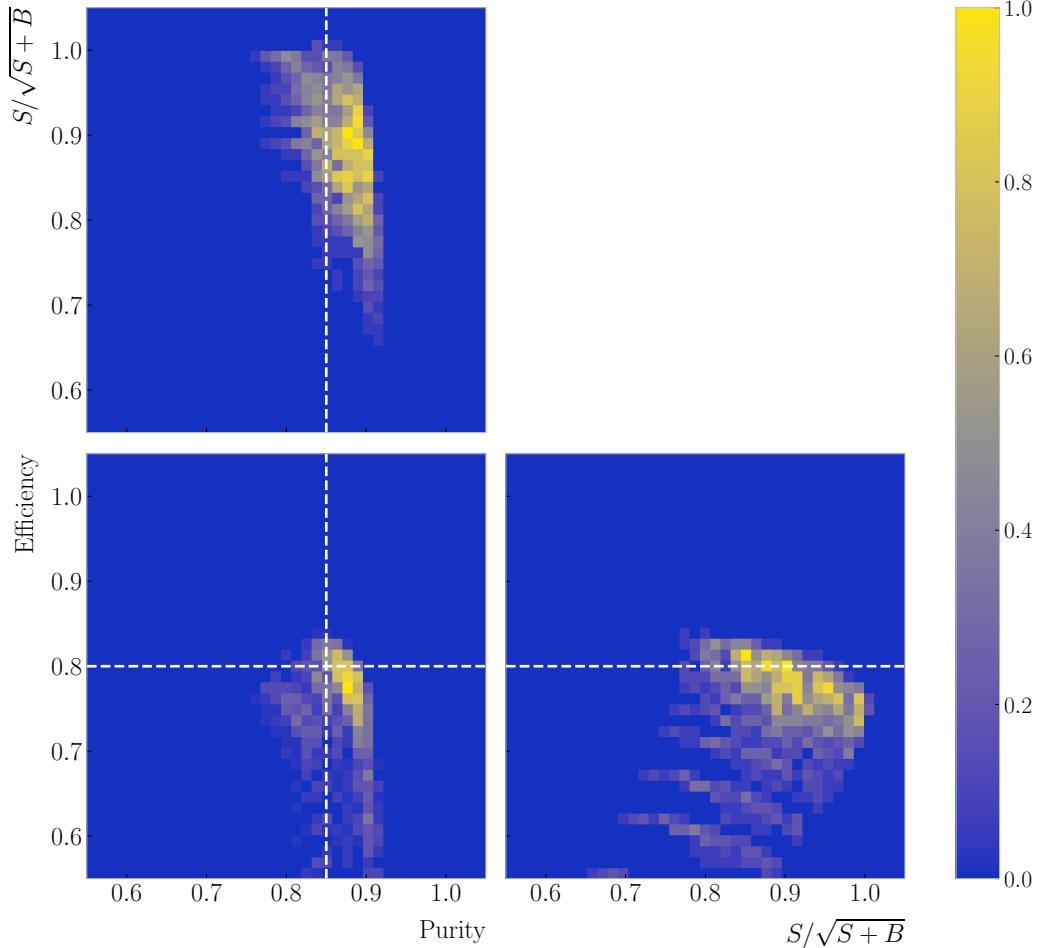


Figure 7.5: Normalised 2D distributions of efficiency, purity and significance for the ν_μ CC selection. The $S/\sqrt{S+B}$ is normalised to the highest value achieved. The vertical dashed line indicates a purity value of 0.85, whereas the horizontal one corresponds to an efficiency of 0.80.

3732 local maxima at intermediate values.

3733 Selecting the cut configuration with the highest significance, 147 ± 11 for the parameter
 3734 values explored here, results in an efficiency and purity of 0.754 ± 0.006 and 0.833 ± 0.007 ,
 3735 respectively. Figure 7.5 shows the 2D distributions resulting when combining pairs of
 3736 efficiency, purity and significance, obtained for the cut configurations explored. The
 3737 significance is normalised to the highest value obtained in the parameter scan. Looking
 3738 at this, it is clear that a selection with highest efficiency and purity can be achieved,
 3739 maintaining a similar significance level.

CHAPTER 7. EVENT SELECTION IN ND-GAR

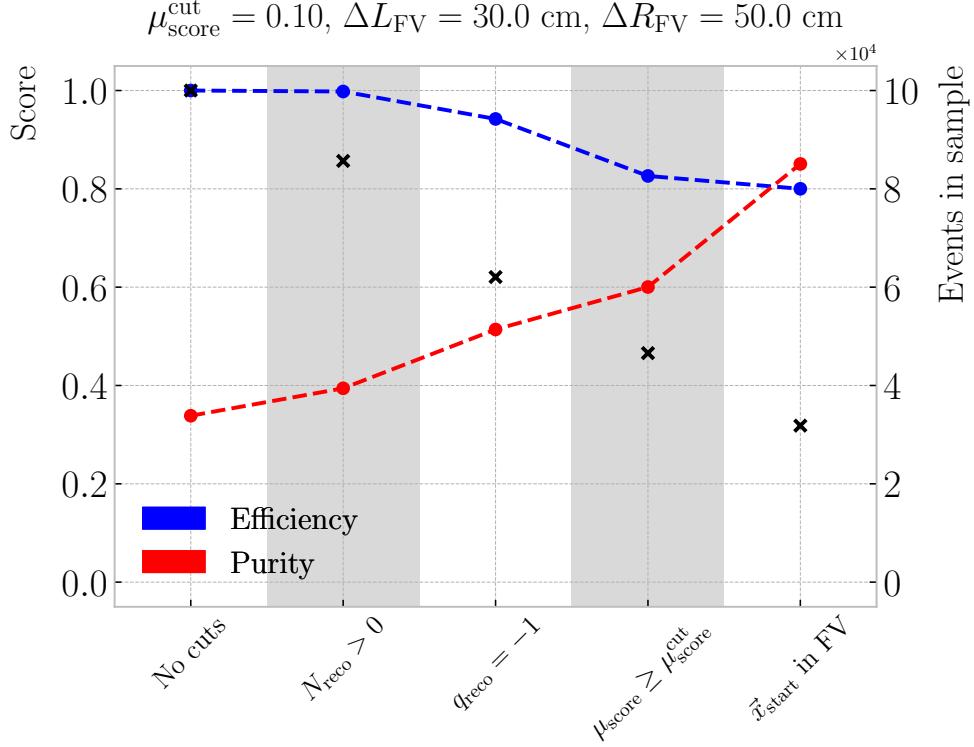


Figure 7.6: Cumulative efficiency (blue) and purity (red) of the ν_μ CC selection. The secondary axis indicates the number of events in the sample after each cut (black crosses).

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%)

Therefore, to get a more refined selection, I first select the configurations with a purity and an efficiency higher than 0.85 and 0.80, respectively. After that, I select the tuple of cuts yielding the highest significance. The resulting value for the muon score cut is $\mu_{\text{score}}^{\text{cut}} = 0.10$, and the FV is given by $\Delta L_{\text{FV}} = 30.0 \text{ cm}$ and $\Delta R_{\text{FV}} = 50.0 \text{ cm}$. 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

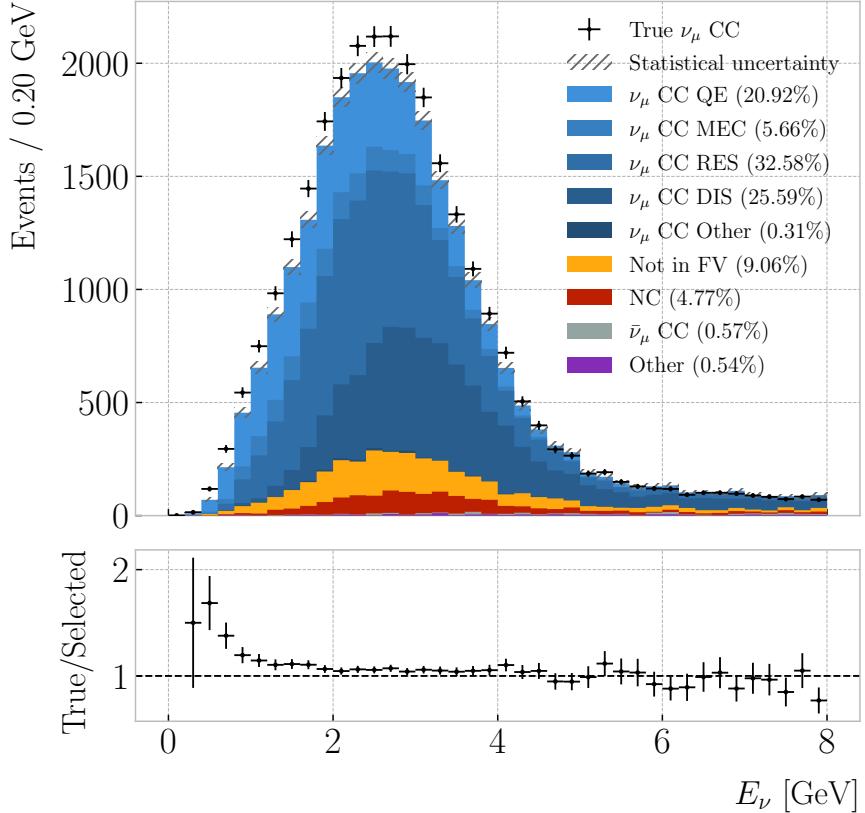


Figure 7.7: True neutrino energy spectra for the ν_μ CC selection. The selected events correspond to the coloured stacked histogram, broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray. The true distribution is also shown with the black data points. The bottom panel shows the ratio between the number of true and selected ν_μ CC events per bin.

3745 with a significance of 138 ± 11 . Hereafter, I use this optimised selection cuts, unless
 3746 specified otherwise.

3747 A summary of the selection can be found in Tab. 7.2. It shows the number of
 3748 events in the selected sample after each selection cut, as well as the absolute and relative
 3749 passing rates. Figure 7.6 shows the overall efficiencies (blue) and purities (red) after
 3750 each cut in the event selection is applied. As expected, the efficiency drops while the
 3751 purity increases with the successive cuts.

3752 Notice how, out of the cuts prior to the FV constraint, the sign selection produces
 3753 the highest increase in purity. This is one of the advantages of having a magnetised
 3754 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

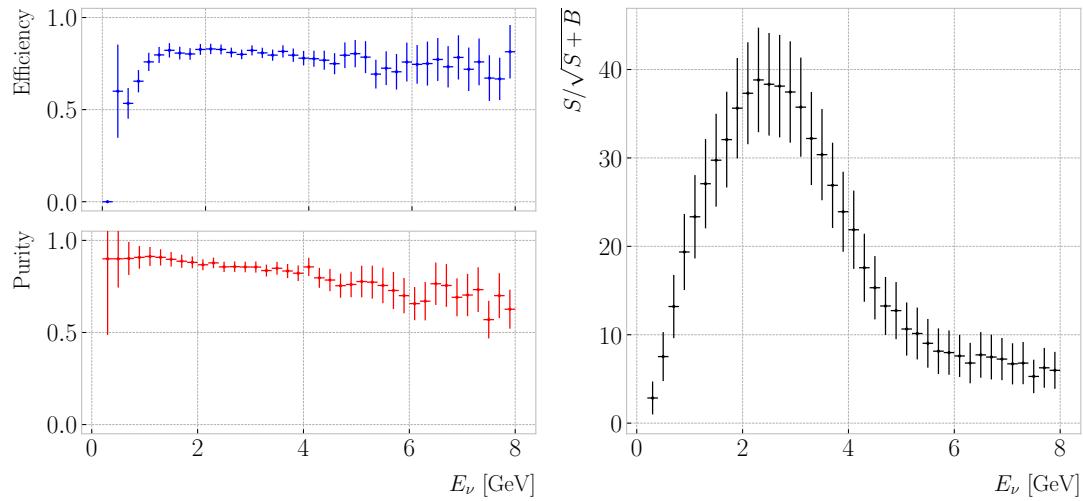


Figure 7.8: Left panel: efficiency (top panel) and purity (bottom panel) for the ν_μ CC selection as a function of the true neutrino energy. Right panel: significance for the ν_μ CC selection as a function of the true neutrino energy

3755 7.2.2 Selection performance

3756 Using the stored spectra discussed above, the true neutrino energy distribution for the
 3757 selected events can be recovered doing TP + FP. Similarly, the combination TP + FN
 3758 gives the true spectrum. Figure 7.7 shows the true (black data points) and selected
 3759 (coloured stacked histogram) E_ν distributions for the optimised ν_μ CC selection. The
 3760 colours in the selected spectrum indicate the different signal categories and backgrounds,
 3761 with the overall statistical uncertainty represented by the gray hatched mess. The ratio
 3762 between the true and selected events is also shown. One can see that it sits around 1 for
 3763 most of the energy range. However, for energies ≤ 1 GeV there is a significant deficit of
 3764 selected events.

3765 These spectra also allow to compute the efficiency and purity of the selection as
 3766 a function of the true neutrino energy, as shown in Fig. 7.8 (left panel). As it could
 3767 be expected from the previous ratio plot, the efficiency is low at low neutrino energies.
 3768 Nonetheless, it raises quickly with the energy, until it stabilises around a value of 0.80.
 3769 Looking at the purity, one may notice that, although it starts at around 0.90, there is a
 3770 significant decrease towards the high end of the spectrum. Figure 7.8 (right panel) also

7.2. ν_μ CC SELECTION

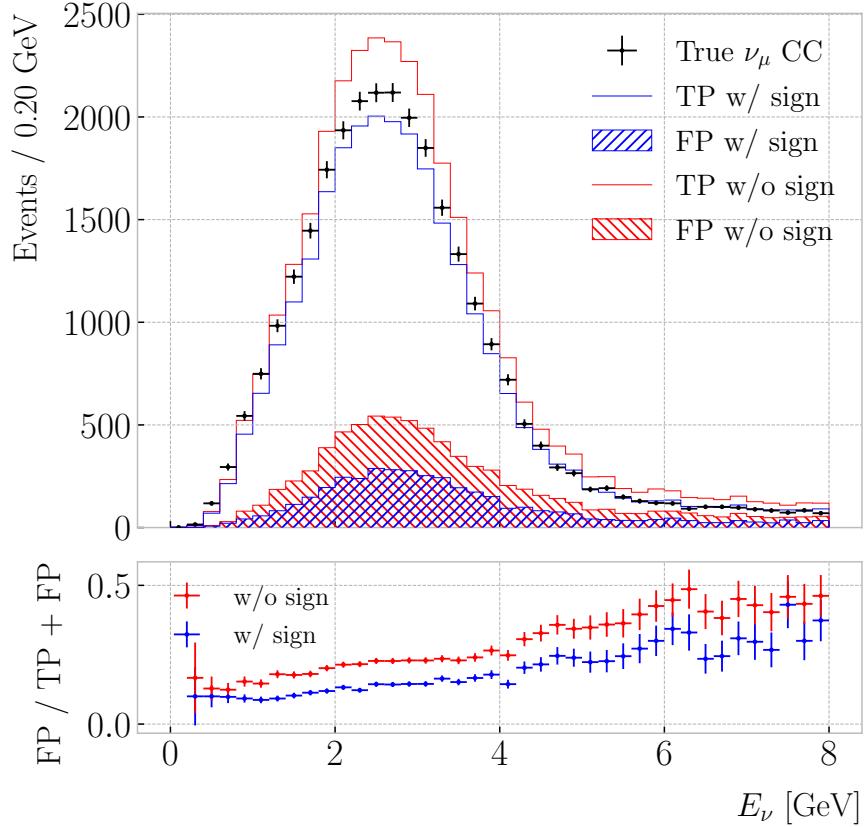


Figure 7.9: True neutrino energy spectra for the ν_μ CC selection with (blue) and without (red) sign selection. The selected events are broken down by true positives (signal) and false positives (background). The true distribution is also shown (black data points). The bottom panel shows the ratios between the number of false positives and total selected events per bin.

3771 shows the significance as a function of the energy. In this case, the highest $S/\sqrt{S+B}$ is
 3772 achieved around the energies where the spectrum peaks.

3773 A variation of the ν_μ CC selection one can try is to apply it without the reconstructed
 3774 charge cut. Figure 7.9 (top panel) shows the E_ν distributions corresponding to the
 3775 selection with (blue stacked histogram) and without (red stacked histogram) the sign
 3776 selection. In the former case, the out of FV contamination amounts to 9.06% of the
 3777 total, while the NC contamination results 4.77% and the wrong-sign contamination
 3778 0.57%. For the later, these backgrounds account for the 10.01%, 10.82%, and 2.18%
 3779 of the selected events, respectively. As expected, removing the positive particles does
 3780 not change the FV-related effects noticeably. However, the sign selection proves its

CHAPTER 7. EVENT SELECTION IN ND-GAR

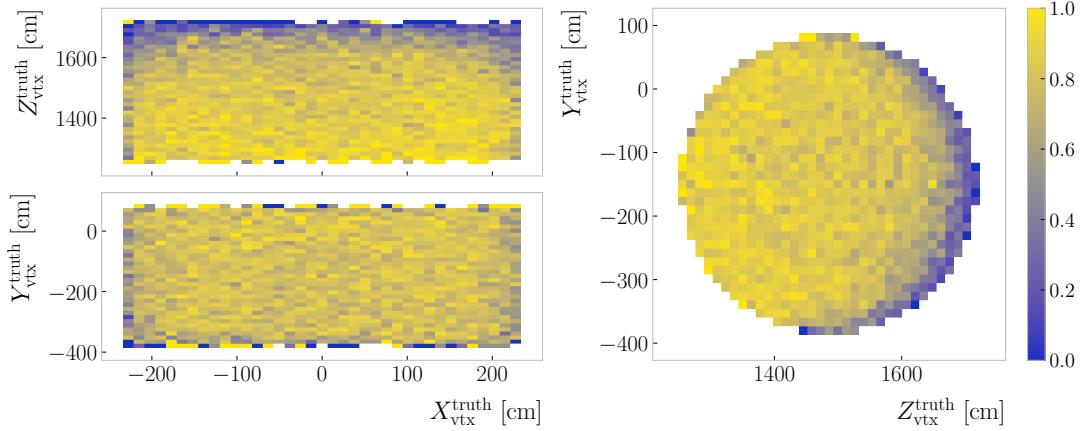


Figure 7.10: Efficiency 2D distributions for the ν_μ CC selection given the true position of the interaction vertex.

worth in the rejection of $\bar{\nu}_\mu$ CC events, which drop almost by one order of magnitude.
 Additionally, the charge selection cuts the NC events in half, as it reduces the chances
 of misidentifying a positively charged hadron for a muon.

As an additional check, I explored how the performance of the ν_μ CC selection
 depends on the position of the neutrino interaction within the HPgTPC. Maps of the
 selection efficiency for the X, Z (top left panel), X, Y (bottom left panel), and Z, Y
 (right panel) true vertex position pairs are given in Fig. 7.10. It can be seen that the
 efficiency remains stable along the drift direction, only slightly degrading close to the
 edges of the FV. Regarding the radial direction, it is clear that an important number of
 events with high $Z_{\text{vtx}}^{\text{truth}}$ are not being selected. Intuitively, the muons arising from these
 interactions will leave short tracks. As their directions are typically aligned with the
 beam direction, they enter the ECal shortly after production. This is likely to affect
 the tracking, and therefore their identification. As a result, the regions with the lowest
 efficiency are the downstream corners of the HPgTPC, i.e. the areas with high $|X_{\text{vtx}}^{\text{truth}}|$
 and $Z_{\text{vtx}}^{\text{truth}}$.

7.2.3 Primary muon kinematics

This ν_μ CC selection relies on the identification of the a primary muon, meaning that
 for each selected event a particle is picked out as the muon candidate. It is because of

7.2. ν_μ CC SELECTION

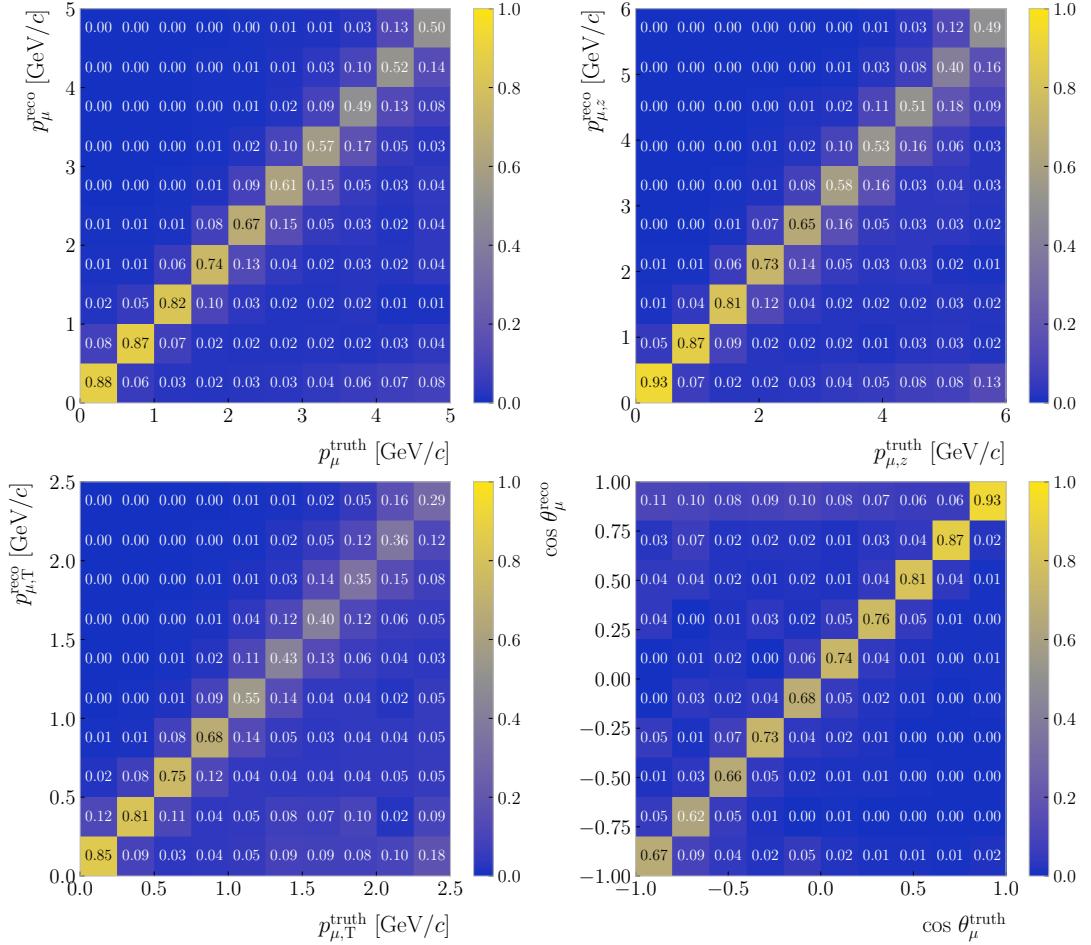


Figure 7.11: Distributions for the reconstructed versus truth generated primary muon momentum (top left panel), longitudinal momentum (top right panel), transverse momentum (bottom left panel), and beam angle (bottom right panel). The reconstructed values correspond to the selected primary muon candidate, whereas the truth values come from the true primary muon in the event.

3799 this that one can study the kinematics of these selected primary muons.

3800 Figure 7.11 shows a comparison between some of the reconstructed and truth primary
 3801 muon kinematic variables. From top to bottom, left to right, we have muon momentum,
 3802 longitudinal momentum, transverse momentum and beam angle. The histograms are
 3803 column-normalised, and so the diagonal entries give an idea of the resolution for the
 3804 different variables. The match between truth and reconstructed values can only be done
 3805 for the selected true ν_μ CC events, as the others do not have a primary muon. However,
 3806 for this comparison I do not require the events to start inside the FV.

CHAPTER 7. EVENT SELECTION IN ND-GAR

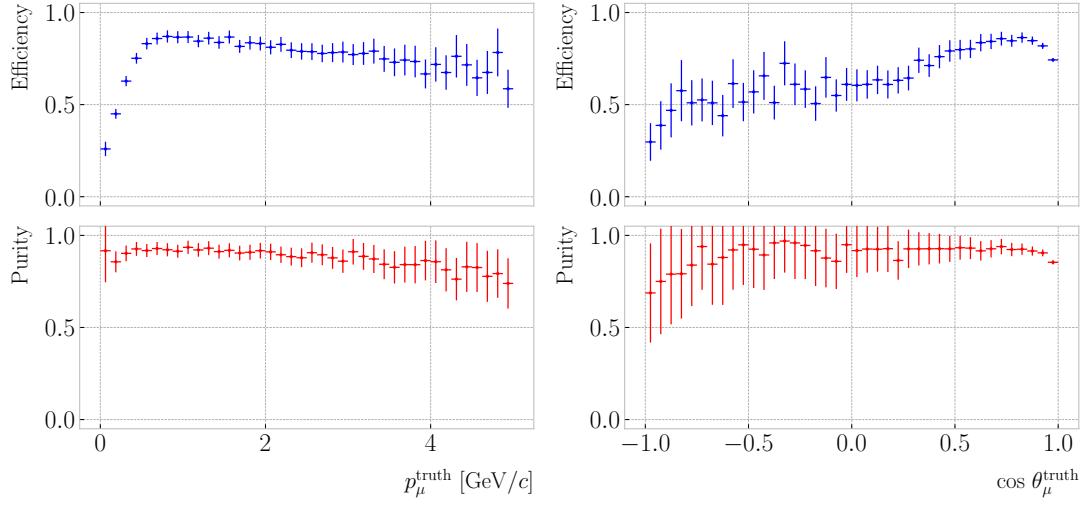


Figure 7.12: Efficiency (blue) and purity (red) of the ν_μ CC selection as a function of the primary muon true momentum (left panel) and beam angle (right panel).

3807 Notice that, for the reconstructed values, the variables do not necessarily come
 3808 from a reconstructed particle that matches the true primary muon. In other words,
 3809 sometimes, even though the event was correctly identified, the primary muon may have
 3810 been confused with another particle. That means that in these distributions include
 3811 both reconstruction and selection deficiencies.

3812 I also studied the performance of the ν_μ CC selection as a function of the kinematic
 3813 variables of the primary muon. As before, these metrics are only possible to compute for
 3814 true ν_μ CC events. The efficiency (top panels) and purity (bottom panels) as a function
 3815 of the truth muon momentum (left) and beam angle (right) are shown in Fig. 7.12. One
 3816 can see that there are some similarities in the behaviour of both metrics between the
 3817 true neutrino energy and the muon momentum cases. This is to be expected, as these
 3818 two variables are highly correlated. For the efficiency, there is a rapid increase at low
 3819 momentum values until it peaks at around 1 GeV/ c , after which it starts decreasing
 3820 slowly. The purity remains relatively constant, with a slight drop towards high p_μ^{truth}
 3821 values. In the case of the muon angle, the decrease in efficiency at high $\theta_\mu^{\text{truth}}$ is more
 3822 noticeable. However, note that the number of events with backward-going muons is
 3823 much smaller than those aimed towards the forward direction, as can be seen from the

7.2. ν_μ CC SELECTION

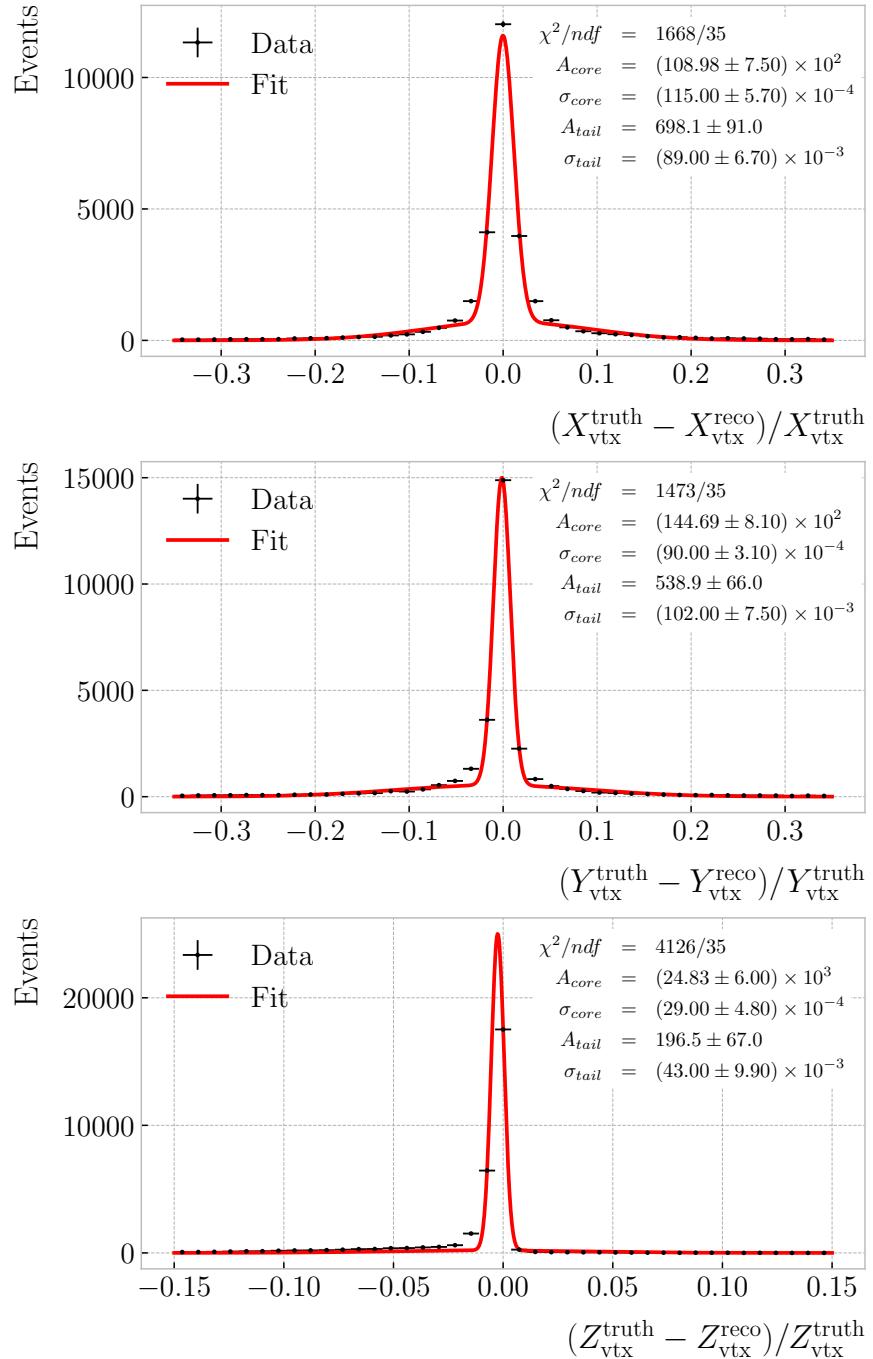


Figure 7.13: Fractional residual distributions for the position of the primary vertex in the ν_μ CC selection. The best fits to a double Gaussian function are also shown (red lines).

CHAPTER 7. EVENT SELECTION IN ND-GAR

3824 size of the vertical error bars. There is also a decline in the purity with the beam angle,
3825 but this effect is much smaller.

3826 A byproduct of selecting the primary lepton in the interaction is the position
3827 of the reconstructed neutrino vertex candidate. Checking how the position of the
3828 selected reconstructed primary vertex and the true vertex position compare is needed to
3829 understand the validity of our method. Figure 7.13 shows the distributions of fractional
3830 residuals between the truth and reconstructed vertex positions in the X (top panel),
3831 Y (middle panel), and Z (bottom panel) directions. Performing a double Gaussian fit
3832 to the distributions (red lines), I estimate the reconstructed vertex resolution achieved
3833 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 ,
3834 and Z directions, respectively. As expected, the resolution along the drift direction.
3835 However, the significant difference in resolution between the two transverse directions is
3836 worth noting. Not only the resolution is better for the Z direction, but the layout of the
3837 residual distribution is highly asymmetrical. This may be related to the variability in
3838 the selection efficiency along that direction.

3839 7.3 Charged pion identification

3840 Now that I have checked the robustness of the proposed ν_μ CC selection, it can be
3841 used as a starting point for other, more convoluted, selections. One of the priorities
3842 of ND-GAr, as mentioned previously, is the identification of pions. With its lower
3843 tracking thresholds, ND-GAr is expected to do better regarding π^\pm identification than
3844 the traditional LArTPCs, like ND-LAr. Moreover, it can make use of the different
3845 detector subcomponents to tag the charged pions.

3846 The ν_μ CC selection provides a starting point for the pion identification. The first
3847 thing one can do is rule out the selected primary muon candidate. Then, by looking at
3848 the properties of the rest of the reconstructed particles, one can start the counting of
3849 the charged pions.

3850 The two proton scores, the one based on the dE/dx in the HPgTPC and the one

7.3. CHARGED PION IDENTIFICATION

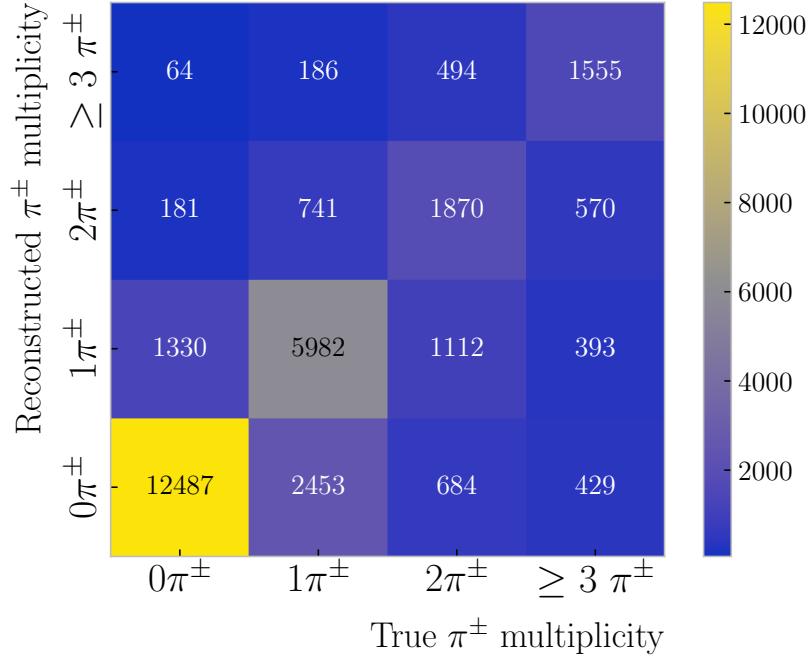


Figure 7.14: Distribution of events given their true and reconstructed π^\pm multiplicity, for the selection given by $p_{dE/dx}^{\text{cut}} = p_{\text{ToF}}^{\text{cut}} = 0.50$, $\Delta_{dE/dx}^{\pi^\pm} = 0.20$, and $d_\mu^{\text{cut}} = 50.0$ cm.

3851 obtained from the ToF measurement in the ECal, can be used to separate the protons
 3852 from the sample of charged pions. By providing appropriate cuts for these, a good
 3853 separation can be achieved.

3854 Another source of information available is the dE/dx of the track associated to the
 3855 reconstructed particle. To select the charged pions, we can require that the measured
 3856 mean dE/dx is compatible with the expectation for a true π^\pm , in other words:

$$\left\langle \frac{dE}{dx} \right\rangle_{\pi^\pm} \left(1 - \Delta_{dE/dx}^{\pi^\pm} \right) \leq \left\langle \frac{dE}{dx} \right\rangle_{\text{meas.}} < \left\langle \frac{dE}{dx} \right\rangle_{\pi^\pm} \left(1 + \Delta_{dE/dx}^{\pi^\pm} \right), \quad (7.6)$$

3857 where the parameter $\Delta_{dE/dx}^{\pi^\pm}$ measures the fractional variation one allows around the
 3858 theoretical expectation. To obtain the expected mean dE/dx of a charged pion with a
 3859 given momentum, I use the ALEPH parametrisation with the parameter values obtained
 3860 previously.

3861 Also, as we are only interested in the primary pions, and because these are by
 3862 definition close to the interaction vertex, one can apply an additional distance cut. Using

CHAPTER 7. EVENT SELECTION IN ND-GAR

3863 the start position of the muon candidate, we can restrict the starting point of pions to a
3864 certain volume around the vertex.

3865 Combining all these ideas, I propose the following procedure to identify the charged
3866 pions in an event:

- 3867 1. Apply ν_μ CC selection.
- 3868 2. Disregard particle selected as primary muon.
- 3869 3. Remove particles with momentum below threshold.
- 3870 4. Select particles with proton dE/dx score below threshold.
- 3871 5. Select particles with proton ToF score below threshold.
- 3872 6. Select particles with mean dE/dx around the expected value for a pion.
- 3873 7. Remove particles with a distance between the start of the track and the primary
3874 vertex greater than the cut.

3875 The remaining particles after all these cuts are taken to be charged pion candidates.

3876 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
3877 d_μ^{cut} in order of appearance. The momentum threshold is necessary to compare with
3878 the true multiplicity. For values of the kinetic energy lower than 10 – 20 MeV, we
3879 do not expect to be able to tag individual pions. Such low energy particles just leave
3880 small traces in the TPC which, together with the busy environment of the neutrino
3881 interaction vertex, leaves one with no other option but to only account for their energy
3882 calorimetrically. As such, the true pion counting also features this momentum threshold.

3883 I performed an optimisation of the charged pion counting by scanning the space of
3884 possible cut configurations. For the two proton scores, I let them vary between 0.10 to
3885 0.90, in increments of 0.10. Similarly, the parameter $\Delta_{dE/dx}^{\pi^\pm}$ takes values in the range
3886 0.05 – 0.25, with a step size of 0.05. Finally, the distance cut changes in 10 cm steps,
3887 from 10 to 120 cm.

7.3. CHARGED PION IDENTIFICATION

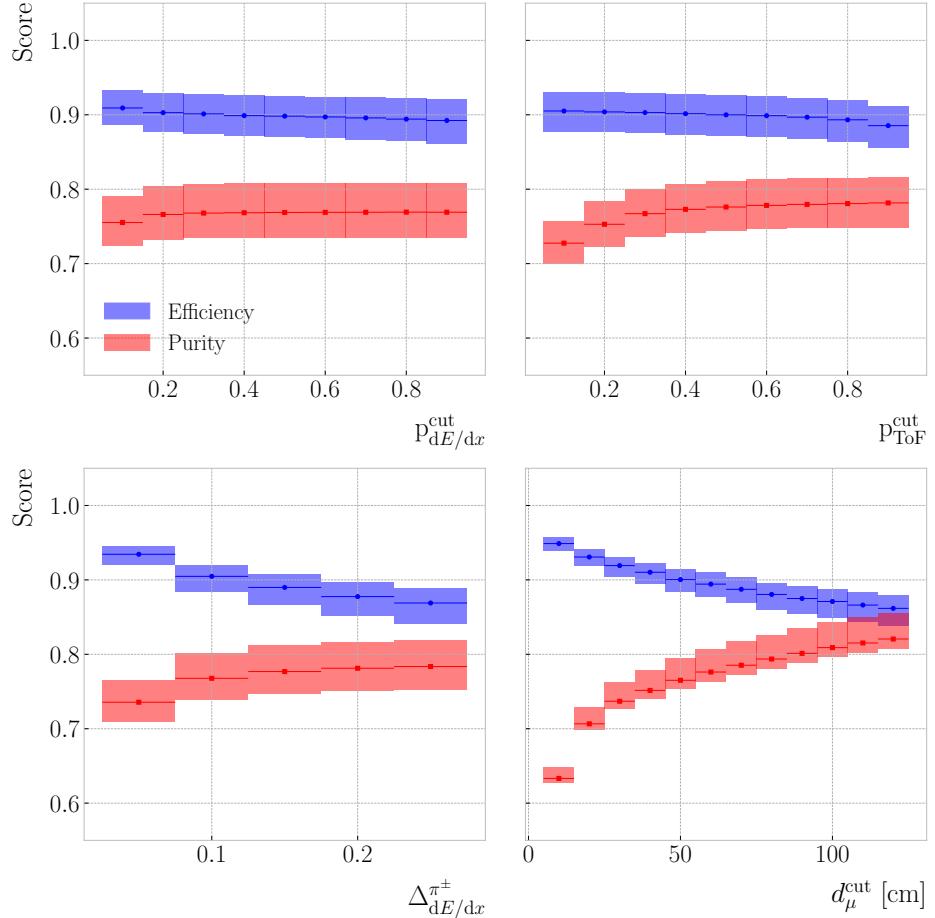


Figure 7.15: Efficiency (blue) and purity (red) for the ν_μ CC $0\pi^\pm$ selection as a function of the proton dE/dx score cut (top left panel), proton ToF score cut (top right panel), pion dE/dx cut (bottom left panel), and distance to muon cut (bottom right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the line corresponds to the median.

3888 For each combination of selection cuts, I compare the true charged pion multiplicity
 3889 given by GENIE with the number of charged pion candidates I count with this method,
 3890 hereafter referred to as the reconstructed π^\pm multiplicity. The result of this comparison
 3891 is a matrix, with columns and rows indicating true and reconstructed charged pion
 3892 multiplicity, respectively. An example of one of these matrices, obtained for a certain
 3893 configuration of cuts, can be seen in Fig. 7.14. From these multiplicity matrices one can
 3894 extract performance metrics, like efficiency, purity, and significance.

3895 Given a multiplicity matrix \mathbf{M} , the efficiency for the i -th multiplicity value can be

CHAPTER 7. EVENT SELECTION IN ND-GAR

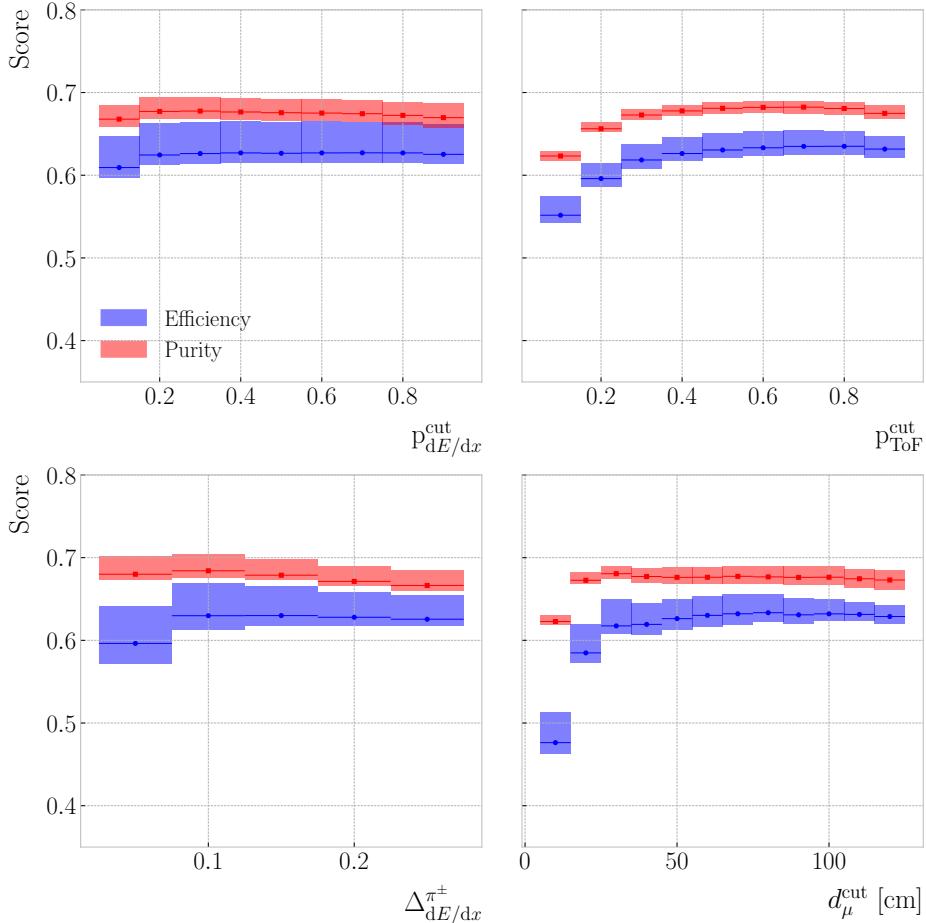


Figure 7.16: Efficiency (blue) and purity (red) for the ν_μ CC $1\pi^\pm$ selection as a function of the proton dE/dx score cut (top left panel), proton ToF score cut (top right panel), pion dE/dx cut (bottom left panel), and distance to muon cut (bottom right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the line corresponds to the median.

3896 computed as:

$$\text{Efficiency}|_i = \frac{M_{ii}}{\sum_j M_{ij}}, \quad (7.7)$$

3897 or in other words, dividing the corresponding diagonal entry by the sum of all the entries

3898 in the same column. On the other hand, the purity is given by:

$$\text{Purity}|_i = \frac{M_{ii}}{\sum_j M_{ji}}, \quad (7.8)$$

3899 which is just the ratio between the diagonal entry and the sum of the entries in the

7.3. CHARGED PION IDENTIFICATION

3900 corresponding row. Similarly, the significance is obtained by taking the square root of
 3901 the denominator in the previous expression:

$$\text{Significance}|_i = \frac{S}{\sqrt{S+B}}|_i = \frac{M_{ii}}{\sqrt{\sum_j M_{ji}}}. \quad (7.9)$$

3902 Figures 7.15 and 7.16 show the efficiency (blue) and the purity (red) for the ν_μ
 3903 CC $0\pi^\pm$ and $1\pi^\pm$ selections, respectively, as a function of the different cut values. In
 3904 the figures, each box represents the IQR of the conditional distribution for the fixed
 3905 value of the corresponding cut, and the horizontal lines correspond to the medians. The
 3906 first thing one notices is that the efficiency is always higher than the purity in the $0\pi^\pm$
 3907 selection, while the opposite is true for the $1\pi^\pm$ selection. Also, it is clear that the range
 3908 within these metrics fluctuate in the $0\pi^\pm$ selection is significantly higher than it is for
 3909 the $1\pi^\pm$ case. This shows that it is easier to assess that no charged pions are present in
 3910 the event than actually tagging them.

3911 For the ν_μ CC $0\pi^\pm$ selection, the performance metrics follow the expected tendency.
 3912 As the purity grows with a cut value, the efficiency decreases. Interestingly, this is not
 3913 the case for the $1\pi^\pm$ selection, where both efficiency and purity follow roughly the same
 3914 trends along the different cuts. This makes sense when one comprehends that this is not
 3915 a traditional cut-based selection, but more of a counting exercise. Some restrictive cut
 3916 configurations will not tag any particles as pions. On the contrary, loose cuts will render
 3917 every particle as a π^\pm . Therefore, when looking at a specific multiplicity, the relation
 3918 between the cut value and the performance metrics is not obvious. Thus, sometimes
 3919 efficiency and purity can both increase, as the cuts refine the definition of a reconstructed
 3920 pion.

3921 To have a working point for our studies, I chose the cut configuration that yields
 3922 the maximum significance for the ν_μ CC $1\pi^\pm$ selection. Of course, other cuts would be
 3923 more appropriate in certain scenarios. However, this provides us with a starting point
 3924 to understand the performance of the selection. A significance of 66 ± 7 for the $1\pi^\pm$
 3925 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

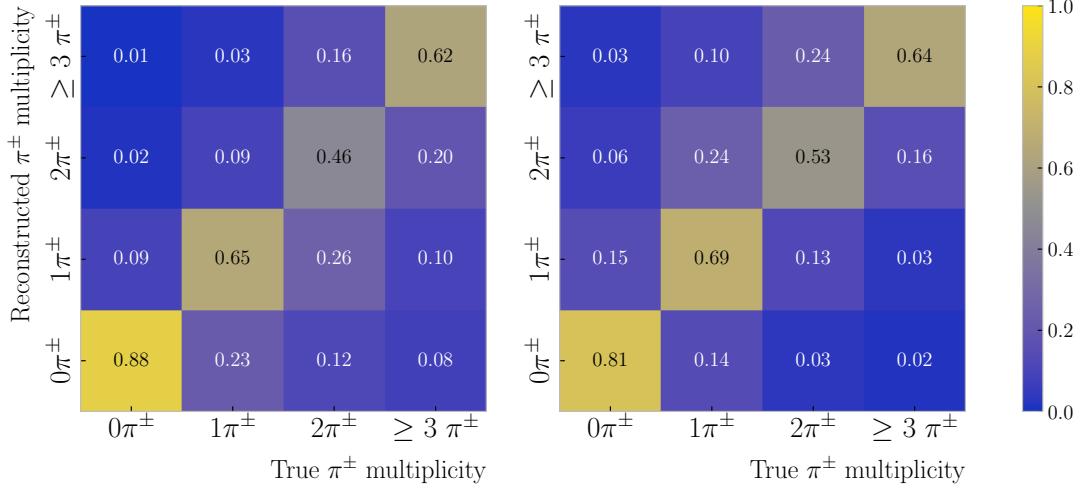


Figure 7.17: Distribution of events given their true and reconstructed π^\pm multiplicity, both column-wise (left panel) and row-wise (right panel) normalised, for the selection that maximises the significance of the ν_μ CC $1\pi^\pm$ selection. The column normalisation yields the efficiency in the diagonal entries, whereas the row normalisation reveals the purity.

3926 $d_\mu^{\text{cut}} = 110.0$ cm.

3927 Figure 7.17 shows the multiplicity matrices resulting from this optimised ν_μ CC $1\pi^\pm$
 3928 selection. Although both matrices are produced with the same selection cuts, one is
 3929 column normalised (left panel), whereas the other is row normalised (right panel). It
 3930 follows from the definitions in Eqs. (7.7) and ((7.8)) that the diagonal entries of these
 3931 matrices correspond to the efficiencies and the purities, respectively, for each of the
 3932 possible charged pion multiplicity selections.

3933 An additional check to make is understand how this configuration performs when
 3934 applied to the other selections, like ν_μ CC $0\pi^\pm$, and how it compares to the other
 3935 possible configurations. A comparison between the different pion multiplicity selections,
 3936 indicated with colours, in the purity versus efficiency space is shown in Fig. 7.18. For
 3937 each of the possible multiplicity choices, the performance obtained for the $1\pi^\pm$ optimised
 3938 selection is indicated by an outlined point. From this, one can see that the selected
 3939 configuration performs reasonably well, within the limits of what can be achieved in
 3940 each case, across the different multiplicities.

3941 At this point, one can study the charged pion selection performance as a function of

7.3. CHARGED PION IDENTIFICATION

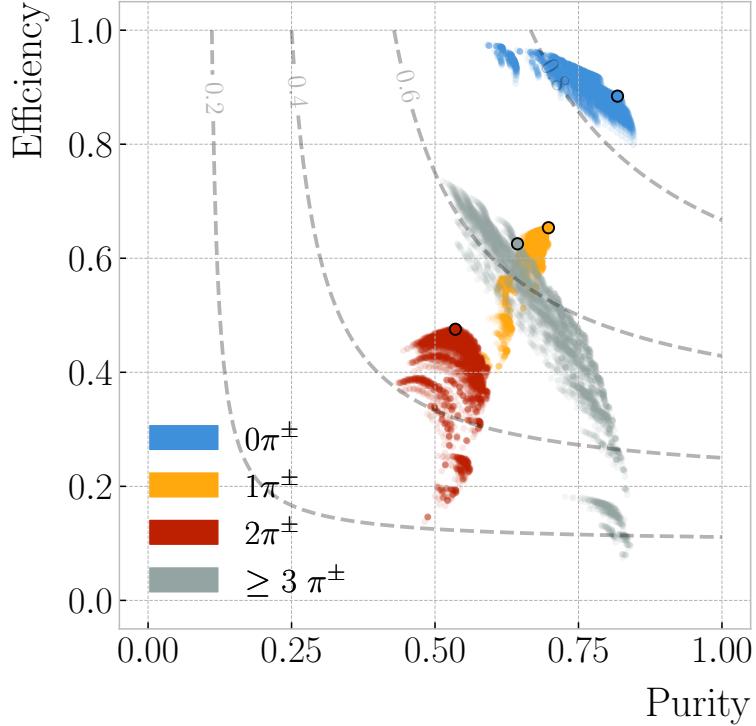


Figure 7.18: Purity versus efficiency achieved for the different cut configurations explored separated by the various ν_μ CC $N\pi^\pm$ selections. The outlined points indicate the state for each possible multiplicity when using the configuration that maximises the significance of the ν_μ CC $1\pi^\pm$ selection. The contours indicate the surfaces of equal F_1 -score.

3942 other quantities of interest. A natural variable to check is the hadronic invariant mass.

3943 It is defined as:

$$W = \sqrt{-Q^2 + 2m_n q_0 + m_n^2} \quad (7.10)$$

3944 where Q^2 is the momentum transfer from the neutrino to the primary muon, q_0 the
 3945 energy transfer, and m_n the mass of the nucleon. This quantity is related to the elasticity
 3946 of the neutrino interaction, and defines the transitions between the QEL, RES and DIS
 3947 regions. An interesting invariant mass range for DUNE is the one that extends between
 3948 the mass of the Δ resonance, even though it is typically extended down to $m_p + m_{\pi^\pm}$,
 3949 and 2.0 GeV. It is estimated that roughly 7 in every 10 events in our ND will take
 3950 place in this region. Although the RES production dominates at these W values, this
 3951 range also includes the transition to the DIS regime. Thus, it is often called the shallow

CHAPTER 7. EVENT SELECTION IN ND-GAR

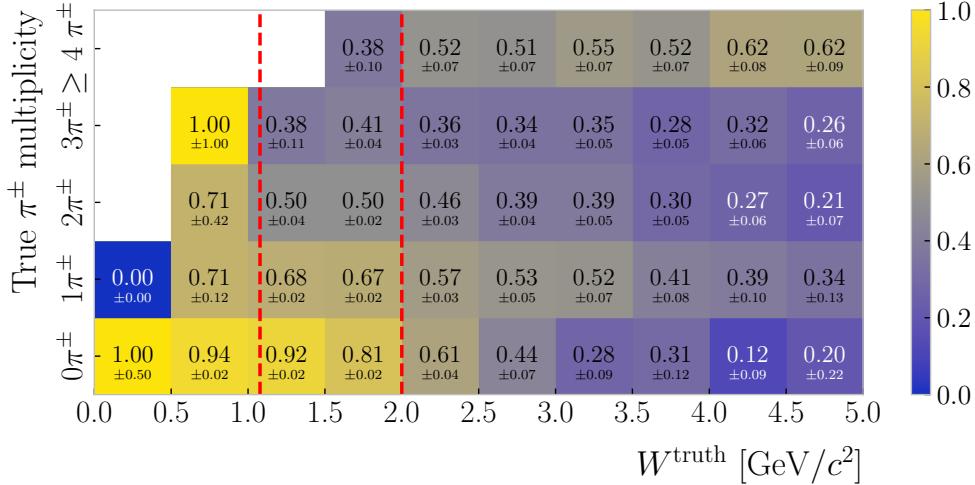


Figure 7.19: Efficiency of the various ν_μ CC $N\pi^\pm$ selections as a function of the true hadronic invariant mass. The dashed vertical lines correspond (from left to right) to the values $m_p + m_{\pi^\pm}$ and $2.0 \text{ GeV}/c^2$, which define the shallow inelastic scattering region.

3952 inelastic scattering (SIS) region.

3953 Within these boundaries, the resonant events produce either 1 or 2 charged pions,
 3954 whereas the multipion events are typically associated to non-resonant production.
 3955 Therefore, our ability of correctly select events with $\geq 2\pi^\pm$ in the SIS region will
 3956 impact our ability of constraining this poorly understood regime. Figure 7.19 shows the
 3957 efficiency of the various charged pion multiplicity selections in a number of hadronic
 3958 invariant mass bins. The two red dashed lines indicate the boundaries of the SIS region.
 3959 One can see that, although not as good as the single pion selection, the efficiency for the
 3960 multipion events is reasonable in the relevant invariant mass range. The total efficiency
 3961 for the ν_μ CC $\geq 2\pi^\pm$ selection in the SIS regime is estimated to be 0.65 ± 0.02 .

3962 7.3.1 ν_μ CC $1\pi^\pm$ selection

3963 By focusing on the $1\pi^\pm$ selection, one can study the kinematics of the selected pion.
 3964 This allows one to understand how well the charged pions are tagged. This is difficult
 3965 to do only using the multiplicity matrices, as with them one can only check that the
 3966 number of charged pions is the same as in the truth. Sometime, even if the estimated
 3967 pion multiplicity is correct, the identified particles may not be true pions.

7.3. CHARGED PION IDENTIFICATION

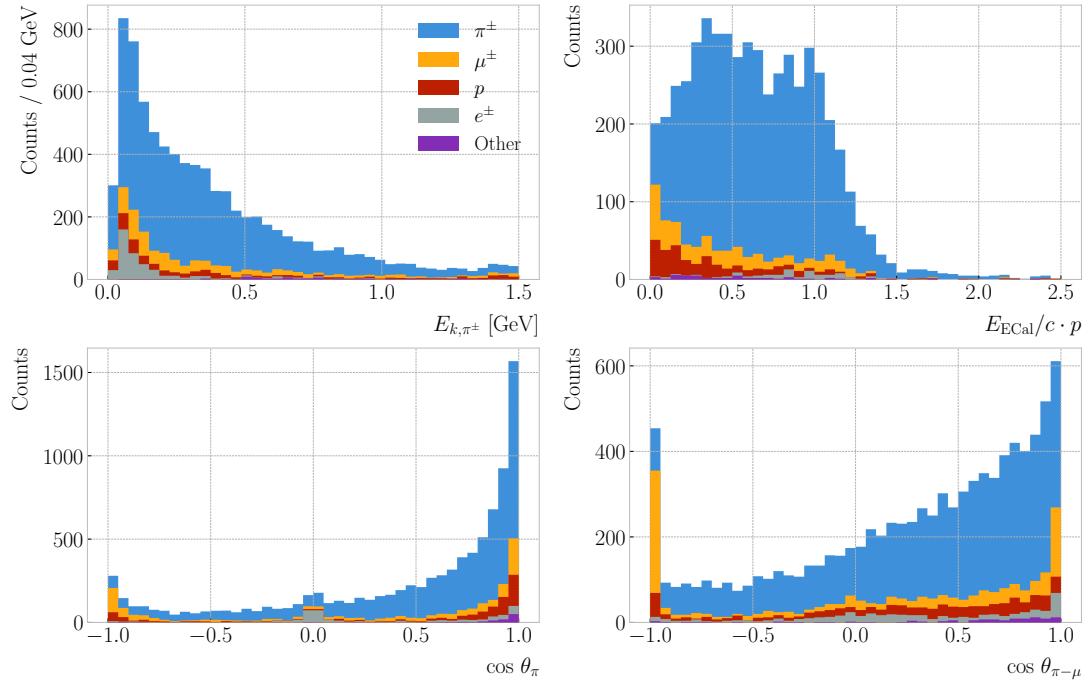


Figure 7.20: Reconstructed kinematic distributions for the pion candidate in the ν_μ CC $1\pi^\pm$ selection, broken down by true ID of the particle. From left to right, top to bottom, we have the kinetic energy given the pion hypothesis, the ratio between the energy deposited in the ECal and the momentum, the pion angle, and the angle between the muon and pion candidates.

3968 Figure 7.20 displays the distributions of various reconstructed kinematic variables
 3969 for the selected pion candidate. The different colours indicate the ID of the true particle
 3970 associated to the reconstructed pion.

3971 First, we have the kinetic energy distribution. For this set of reconstructed particles,
 3972 because they have been tagged as charged pions, the kinetic energy is computed using their
 3973 momentum assuming the pion hypothesis. One can see that most of the contaminants
 3974 sit in low energy range, up to around 0.2 GeV.

3975 The next distribution presents the ratio between the energy deposited in the ECal
 3976 associated to the particle over the momentum measured in the HPgTPC. This variable is
 3977 restricted to particles with at least one associated hit in the ECal. It is interesting to see
 3978 two peak structure in the true pion distribution. The first one presumably corresponds
 3979 to the pions punching-through the ECal, while the latter is probably due to the ones

CHAPTER 7. EVENT SELECTION IN ND-GAR

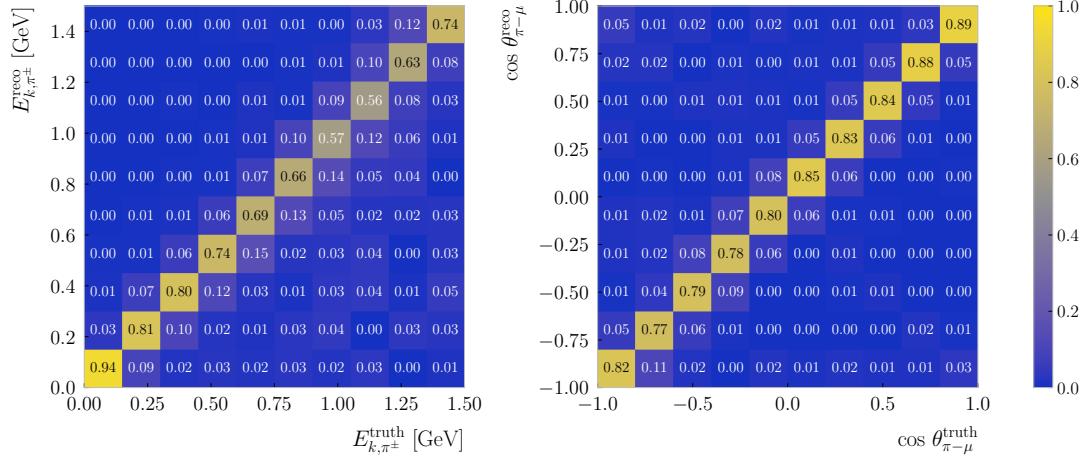


Figure 7.21: Distributions for the reconstructed versus truth generated pion kinetic energy (left panel) and cosine of the angle between the pion and muon (right panel). The reconstructed values correspond to the selected primary muon and pion candidates in the ν_μ CC $1\pi^\pm$ selection, whereas the truth values come from the true primary muon and pion in the events.

3980 stopping in it. On the other hand, the misidentified particles, other than the electrons,
 3981 tend to lower ratios. This is expected for protons, as this could not be higher than 0.5
 3982 for momenta ≤ 1 GeV/ c even if they stopped, but for the muons it may point to a
 3983 misreconstruction.

3984 The following distribution shows the angle of the pion candidates with respect to the
 3985 beam direction. Although most of them are aimed in the forward direction, it can be
 3986 noted that an important number of the misidentified muons seem to be backward-going.
 3987 This is likely a reconstruction artifact, produced by broken tracks that got assigned the
 3988 wrong propagation direction. Also, there is a sizeable number of true electrons with
 3989 directions perpendicular to the beam, probably delta electrons from the primary muon.

3990 Finally, I included the reconstructed pion-muon angular distribution. Even though
 3991 it shares some similarities with the previous distribution, as the primary muon typically
 3992 goes forward, the pion distribution is not as prominently forward-going in this case.
 3993 Also, it may be noted that approximately 25% of the muons misidentified as pions have
 3994 $\cos \theta_{\pi-\mu} \leq -0.95$. Therefore, putting an additional angular cut improves the purity of
 3995 the charged pion selection from 0.74 ± 0.01 to 0.77 ± 0.01 , while not loosing a substantial

7.4. NEUTRAL PION IDENTIFICATION

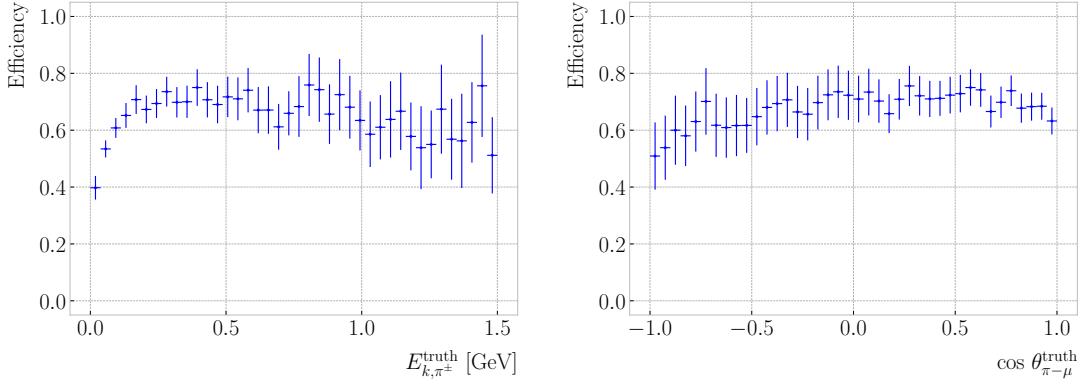


Figure 7.22: Efficiency of the ν_μ CC $1\pi^\pm$ selection as a function of the true pion kinetic energy (left panel) and pion-muon angle (right panel).

3996 amount of true pions.

3997 A comparison between the true and the reconstructed values of the pion kinetic
 3998 energy (left panel) and pion-muon angle (right panel) is shown in Fig. 7.21. The
 3999 distributions are column normalised, which allows to see the fraction of events in the
 4000 correct bins. For this, I selected the events where only one reconstructed pion and
 4001 one true pion were identified, as that is the only case where a pairing of the variables is
 4002 possible. It showcases the excellent agreement between the reconstruction and the truth
 4003 information.

4004 One can also study the performance of the pion selection as a function of the
 4005 truth pion kinematics. Fig. 7.22 shows the selection efficiency versus the true kinetic
 4006 energy (left panel) and the angle between the true primary pion and muon (right panel).
 4007 The efficiency is computed from the events with a single true and reconstructed pion,
 4008 comparing their number to the total of events with one true pion. Notice how the
 4009 efficiency, although it starts with relatively low values, plateaus around 0.70 quickly
 4010 after 0.20 GeV. In terms of the pion-muon angle, the efficiency looks relatively flat, only
 4011 dropping slightly towards the back-to-back case.

CHAPTER 7. EVENT SELECTION IN ND-GAR

4012 7.4 Neutral pion identification

4013 The ν_μ CC selection can also be used as a stepping stone for the identification of
4014 neutral pions. Although these particles do not leave any traces in the HPgTPC, using a
4015 combination of the different detectors within ND-GAr. Being able to tag the neutral
4016 pions is a valuable asset for the estimation of the reconstructed neutrino energy, as both
4017 their kinetic and mass components can then be added in the calculation.

4018 In the case that both photons from the π^0 decay do not undergo pair production
4019 of a e^+e^- pair, they will reach the ECal where they will produce an electromagnetic
4020 shower. This activity inside the ECal will not be associated to any charged particle track
4021 inside the HPgTPC, unless there is a reconstruction failure. Thus, having a neutrino
4022 interaction vertex candidate from the ν_μ CC selection, one can reconstruct the mass of
4023 the π^0 using the energy and position of the photons. I already used this same technique
4024 in section 6.6 for a single π^0 sample. However, here I apply it to neutrino interaction
4025 events, and the vertex position is not cheated but selected from the reconstruction
4026 products.

4027 The idea is to look for all the ECal clusters that were not associated to tracks in
4028 each event. Then, if two or more were identified, compute the invariant mass for all
4029 possible combinations. At this point, I select the pair whose invariant mass is closest to
4030 m_{π^0} , remove the pairs containing any of the two selected clusters from the collection,
4031 and iterate until no more pairs can be formed.

4032 I repeat this procedure for the events with 0, 1, 2 and 3 or more true neutral pions.
4033 For each of them, I extract the invariant mass of the first three cluster pair candidates
4034 (in order of proximity to m_{π^0}), in case they can be formed. If the number of the cluster
4035 pair is lower than the true neutral pion multiplicity of the event, that entry will be
4036 counted as signal. The additional candidates for an event of a given multiplicity are
4037 considered background. The resulting distribution is shown in Fig. 7.23 (black data
4038 points).

7.4. NEUTRAL PION IDENTIFICATION

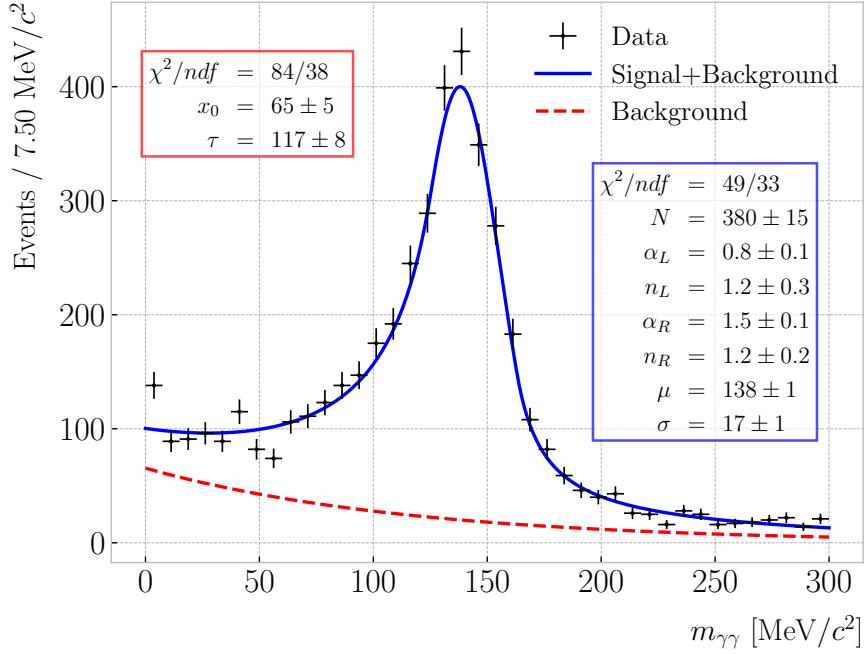


Figure 7.23: Invariant mass distribution obtained for the unassociated ECal cluster pair in the event with the value closest to the true π^0 mass. The best fits for signal plus background (solid blue line) and background only (dashed red line) are also shown.

4039 I fit the signal distribution to a double-sided Crystal Ball function:

$$f_s(x; \alpha_{L,R}, n_{L,R}, \mu, \sigma, N) = N \begin{cases} A_L \left(B_L - \frac{x-\mu}{\sigma} \right)^{-n_L}, & \text{for } \frac{x-\mu}{\sigma} < -\alpha_L, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & \text{for } -\alpha_L \leq \frac{x-\mu}{\sigma} \leq \alpha_R, \\ A_R \left(B_R + \frac{x-\mu}{\sigma} \right)^{-n_R}, & \text{for } \frac{x-\mu}{\sigma} > \alpha_R, \end{cases} \quad (7.11)$$

4040 where $A_{L,R}$ and $B_{L,R}$ are given by:

$$A_{L,R} = \left(\frac{n_{L,R}}{|\alpha_{L,R}|} \right)^{n_{L,R}} e^{-\frac{|\alpha_{L,R}|^2}{2}}, \quad (7.12)$$

$$B_{L,R} = \frac{n_{L,R}}{|\alpha_{L,R}|} - |\alpha_{L,R}|.$$

4041 The tails of this distribution accommodate the asymmetric shape of the misreconstruction
 4042 effects. The values obtained for the best fit parameters are indicated in Fig. 7.23 (blue
 4043 box).

CHAPTER 7. EVENT SELECTION IN ND-GAR

4044 The background is characterised by an exponential function of the form:

$$f_b(x; x_0, \tau) = x_0 e^{-x/\tau}. \quad (7.13)$$

4045 Similarly, the best fit values can be seen in Fig. 7.23 (red box).

4046 Figure 7.23 also shows the results of the fits for the signal plus background (blue line)
4047 and the background only (dashed red line) cases. Using these, I estimate the tagging
4048 efficiency of this method to be 0.90 ± 0.01 with a purity of 0.85 ± 0.01 , when selecting
4049 the candidates with an invariant mass in the range $54.0 - 288.0 \text{ GeV}/c^2$.

4050 This is a robust method to identify the photon pair from the π^0 decay. However,
4051 this approach is not enough to efficiently identify all the events containing neutral pions
4052 in the sample. A quick calculation reveals that only 20% of the ν_μ CC $1\pi^0$ events can
4053 be correctly identified with it.

4054 This approach can be complemented with the identification of the secondary vertices
4055 from the e^+e^- conversions. This will make it possible to cover the cases when either
4056 one or both photons convert in the HPgTPC. In those cases, one can try pairing the
4057 e^+e^- with unassociated activities in the ECal, or matching pairs of secondary vertices.
4058 However, this will require further work on the reconstruction, and thus falls out of the
4059 scope of this analysis.

4060 7.5 Neutrino energy reconstruction

4061 In a neutrino-nucleus CC interaction, where alongside the charged lepton N nucleons
4062 where knocked out and M mesons produced, the reconstructed neutrino energy can be
4063 computed as:

$$E_{\text{rec}} = S_n + E_\ell + \sum_{i=0}^N E_{k,n_i} + \sum_{j=0}^M E_{m_j}, \quad (7.14)$$

4064 where S_n is the average single-nucleon separation energy, E_ℓ the energy of the primary
4065 lepton, E_{k,n_i} is the kinetic energy of the i -th knocked-out nucleon and E_{m_j} the total
4066 energy of the j -th produced meson.

7.5. NEUTRINO ENERGY RECONSTRUCTION

4067 This represents the ideal scenario, where all the kinetic energy of the nucleons is
 4068 visible in the detector and one can identify all mesons produced in the interaction. In a
 4069 real experiment, some of these energy components will not be , and this needs to be
 4070 accounted for in any estimation of the reconstructed energy.

4071 For instance, in ND-GAr neutrons are complicated to account for, as they do not
 4072 produce tracks in the TPC. They may be identified either from scatterings off Ar nuclei
 4073 in the HPgTPC, or performing a ToF measurement in the ECal. However, these methods
 4074 are not fully mature in the current reconstruction, and their development is beyond the
 4075 scope of this study. So, in the following, I will completely ignore the contribution of
 4076 neutrons.

4077 Also, with a real detector we can not expect to tag all the charged pions irrespective
 4078 of their energy. This is why one has to introduce detection thresholds in the energy
 4079 estimation. Thus, in the reconstructed energy calculation I will add only the kinetic
 4080 energy for the charged pions below the threshold, and the total energy for the pions
 4081 above the threshold.

4082 Likewise, the identification of all neutral pions in the sample is challenging. As
 4083 discussed in the previous section, with our ECal we are able to identify the photons
 4084 from the π^0 decays, but that selection still needs to be completed with other methods.
 4085 Therefore, for this first study I do not take into account the energy contribution of the
 4086 neutral pions.

4087 With all this in mind, using the truth information from the events I compute the
 4088 reconstructed neutrino energy as:

$$E_{\text{rec}}^{\text{truth}} = S_n + E_\ell + \sum_{i=0}^{N_p} E_{k,p_i} + \sum_{j=0}^{M_{\pi^\pm}^<} E_{k,\pi_j^\pm} + \sum_{k=0}^{M_{\pi^\pm}^>} E_{\pi_k^\pm}, \quad (7.15)$$

4089 where N_p is the number of protons, and $M_{\pi^\pm}^<$ and $M_{\pi^\pm}^>$ the number of charged pions
 4090 below and above the threshold, respectively. As before, I assume a kinetic energy
 4091 threshold of 20 MeV for the charged pions.

4092 At the reconstruction level, I use the energy of the primary muon candidate, computed

CHAPTER 7. EVENT SELECTION IN ND-GAR

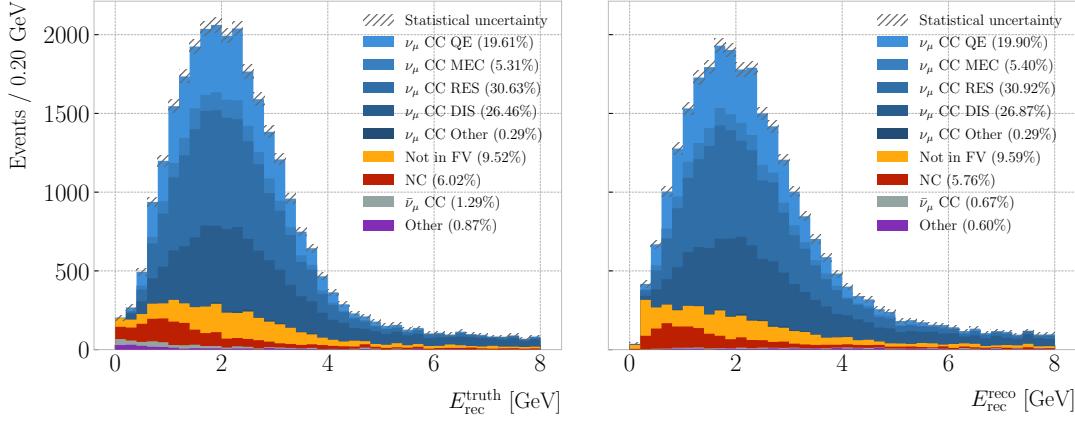


Figure 7.24: Reconstructed neutrino energy spectra for the ν_μ CC selection obtained using the truth (left panel) and reconstructed (right panel) information. The selected events are broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray.

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

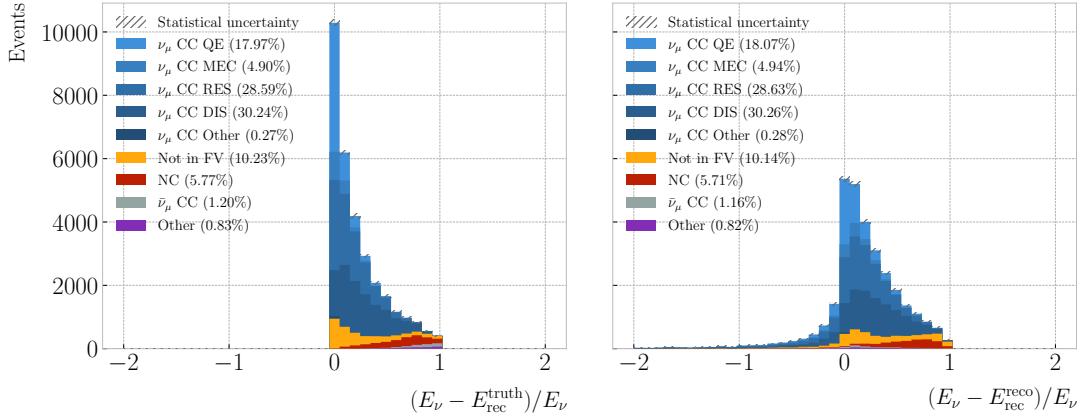


Figure 7.25: Neutrino energy residuals distributions for the ν_μ CC selection obtained using the truth (left panel) and reconstructed (right panel) information. The selected events are broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray.

prone to underestimate the neutrino energy, due to deficiencies in the reconstruction.

7.6 Systematic uncertainties

Although the implementation and study of the systematic uncertainties relevant for ND-GAr is out of the scope of this preliminary analysis, in this section I give an extended overview of the topic. can be classified in three categories: neutrino flux uncertainties, neutrino-nucleus interaction model uncertainties, and detector response uncertainties.

7.6.1 Flux uncertainties

The neutrino flux prediction is affected by systematic uncertainties arising from two sources: the uncertainties in the production of hadrons in the target and the uncertainties in the design parameters of the beamline itself. These fluxes and their uncertainties are generated with the G4LBNF simulation [85], a Geant4 implementation of the LBNF beamline, and the Package to Predict the FluX (PPFX) framework, originally developed for MINERvA [192].

The hadron production uncertainties are associated to the kinematic distributions of the hadrons produced when the protons interact with the carbon target, as well

CHAPTER 7. EVENT SELECTION IN ND-GAR

as the possible interactions of the hadrons with the beamline materials. The PPFX package estimates these uncertainties by performing a number of random throws of the production model parameters [193]. This way, different predictions of the LBNF flux are generated, which can be compared to the nominal prediction to build a matrix of the covariances between neutrino energies, flavours and running modes (either FHC or RHC). The resulting hadron production uncertainties are described by the eigenvectors associated to the largest eigenvalues in this matrix, obtained performing a PCA analysis.

The other set of uncertainties affecting the neutrino flux prediction come from the limited precision with which we know the parameters of the different components in the beamline. These include the specifications of the target, the dimensions of the decay pipe, and the current and alignment of the magnetic horns. The effects on the flux predictions of these uncertainties are estimated using the G4LBNF simulation. For each of the parameters, the simulation runs with said parameter shifted by $\pm 1\sigma$ from the nominal value, and the resulting flux prediction is compared to the nominal one.

7.6.2 Cross section uncertainties

As discussed previously in section 2.6, the neutrino-nucleus interaction model is of great importance for neutrino experiments, as it maps the true neutrino energy to the kinematics of the final state particles. The uncertainties on the cross section model are implemented in three ways: varying the parameters used in the GENIE simulation, using weights that parametrise cross section effects not accounted for in GENIE, and comparing the GENIE predictions to other interaction models.

Within the DUNE TDR LBL analysis, the default interaction model was that implemented in GENIE v2_12_10 [123]. A summary of the cross section systematic parameters present in GENIE used in that analysis is presented in Tab. 7.3. The additional systematic parameters used in the analysis are described in Tab. 7.4.

In this default GENIE configuration, the initial state of the nucleons is described by the Bodek-Ritchie global Fermi gas model [194]. The model is known give a poor agreement whe compared to neutrino-nucleon data [195]. 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. [123].

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. [123].

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

4154 the model, the current versions of **GENIE** use the local Fermi gas approach, which takes
4155 into account the correlation between the momentum of the nucleons and their location
4156 within the nucleus.

4157 For the CCQE events, the dominant model uncertainties arise from the axial form
4158 factor of the nucleon, for which a dipole parametrisation is used, and the nuclear
4159 correlation effects computed using the random phase approximation (RPA). In the
4160 analysis, a parametrisation of the Valencia RPA effect [188] is used. This consists of
4161 a third-order Bernstein polynomial up to $Q^2 = 1.2 \text{ GeV}^2$ followed by an exponential
4162 decay (BeRPA), originally proposed by the T2K collaboration [196].

4163 The $2p2h$ interactions are included using the Valencia model [188], with an additional
4164 correction following the observation of an underprediction of these events in MINERvA
4165 [197]. Additional uncertainties for the energy dependence of the missing strength were
4166 added. Also, the uncertainties in the scaling from carbon to argon are included, based
4167 on measurements of electron scattering off short-range correlated (SRC) nucleon pairs
4168 on multiple targets [198].

4169 In this version of **GENIE**, the Rein-Sehgal model describes the single pion resonant
4170 production events [199]. It includes 16 different resonances, with no interference between
4171 them. Two parameters account for the uncertainties on the axial and vector masses of
4172 the resonances. In subsequent **GENIE** tunes, like the one used in the studies presented in
4173 this Chapter, the Berger-Sehgal model is used [189]. This is an improved version of the

7.6. SYSTEMATIC UNCERTAINTIES

4174 Rein-Sehgal model, which includes the lepton mass effects in the calculations.

4175 The Bodek-Yang parametrisation is used to describe the DIS events [191]. The
4176 parameters A_{HT} and B_{HT} account for higher twist effects in the scaling variable, while
4177 C_{v1u} and C_{v2u} control the form of the valence quark K factors. For the analysis, the
4178 uncertainties on the values of these parameters are taken into consideration. Also, due to
4179 the difficulties of GENIE at describing the transition region between RES and DIS events,
4180 a set of systematic parameters affecting the different non-resonant pion production
4181 channels were developed, following the example of NOvA [200]. There are independent
4182 parameters for the interactions on protons and neutrons, except for the CC DIS 1π case
4183 where they are merged. All start with an uncertainty of 50% for $W \leq 3 \text{ GeV}/c^2$, which
4184 linearly decreases until reaching a 5% at $W = 5 \text{ GeV}/c^2$.

4185 For the TDR analysis, an additional 20% normalisation uncertainty was added to all
4186 NC events in the ND. It was implemented to understand if the NC events passing the
4187 selection cuts affected the results of the analysis [123].

4188 Finally, the effective intranuclear transport model (often denoted as hA) is a part
4189 of GENIE, implemented in the INTRANUKE module. GENIE features a large number of
4190 parameters for the uncertainties on the intranuclear cascade model, which are summarised
4191 in the last portion of Tab. 7.3. In following GENIE releases, updated versions of the
4192 INTRANUKE model are used.

4193 Although part of this cross section systematic treatment is outdated, as the tunes
4194 currently used feature different models, it gives a good idea of what systematic effects
4195 are relevant for the different measurements we may want to perform in the future. At
4196 the moment, a significant effort is channeled to the creation of new tunes specifically
4197 tailored for DUNE, including the development of parametrisations particularly relevant
4198 for ND-GAr.

4199 7.6.3 Detector uncertainties

4200 The DUNE ND CDR [91] presents a number of studies on the performance of ND-GAr.
4201 These were based on the reference design described in section 3.5. Because the detector

CHAPTER 7. EVENT SELECTION IN ND-GAR

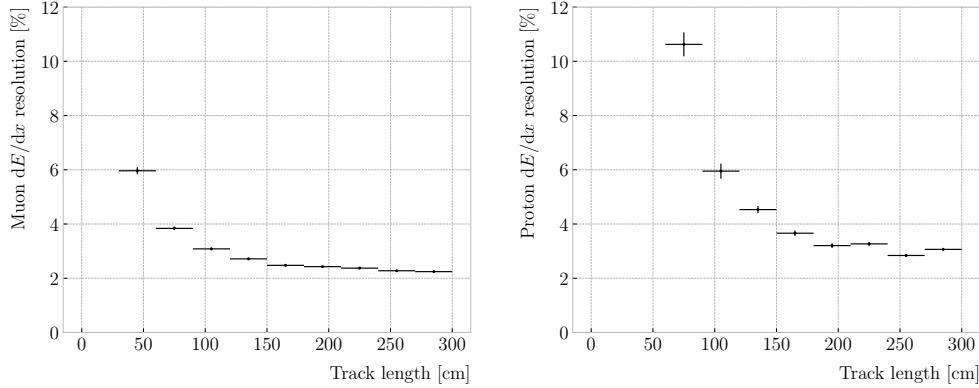


Figure 7.26: Estimated dE/dx resolution as a function of the track length for true muons (left panel) and protons (right panel) in a ν_μ CC sample.

is still in an R&D stage, with the design continually evolving, these performance metrics will need to be revisited in the future. However, they still provide valuable information. These studies help understand what detector requirements are needed to achieve the physics goals of the experiment, and on what design aspects we need to improve.

Since the reference design of ND-GAr repurposes the ALICE MWPCs and other hardware components, the ALICE TPC operation experience point of reference for the spatial resolution performance. They reported a single hit resolution of 0.25 mm and 1.50 mm in the directions perpendicular and parallel to the drift direction, respectively [95]. Nevertheless, the MWPCs are not the leading option for the charge readout anymore. Current efforts focus on the study of the effects of different pixelisation choices. of the GEMs setups.

For other performance metrics, a fairer comparison for the ND-GAr HPgTPC could be the PEP-4. It operated with a 80:20 Ar:CH₄ mixture at 8.5 bar, achieving a two-track separation of 1 cm [201, 202]. This metric is particularly relevant in our case, as the neutrino interaction vertex can be an area of very high track multiplicity. Thus, our track separation capabilities will have a direct impact on the primary vertex identification and resolution. There are several difference between our HPgTPC and PEP-4. The operating pressure of ND-GAr will be higher, and the gas mixture likely to contain a higher fraction of Ar.

7.6. SYSTEMATIC UNCERTAINTIES

4221 In terms of the ionisation measurement, both the experience from ALICE and PEP-4
4222 are relevant, as this depends on the readout and the running conditions (pressure and
4223 gas mixture). They obtained resolutions of 4.5% and 3.0% for typical track lengths of
4224 160 and 75 cm, respectively. According to previous studies on the dE/dx resolution in
4225 gaseous detectors [203], ND-GAr can achieve a 2% resolution for a typical track length of
4226 200 cm. Figure 7.26 shows the values of the resolution I estimate for muons (left panel)
4227 and proton (right panel) tracks with different lengths, using the procedure described in
4228 section 6.2.

4229 The tracking capabilities of ND-GAr were studied in the context of ν_μ CC interactions.
4230 Using a sample of reconstructed tracks from true muons and charged pions, the tracking
4231 efficiency was estimated to be above 90% for momenta $\geq 40 \text{ MeV}/c$, with it steadily
4232 rising with the momentum. As a function of the angle with respect to the beam direction,
4233 the efficiency was almost flat for particles with $p \geq 200 \text{ MeV}/c$. In the case of protons,
4234 the tracking performs for kinetic energies $\geq 20 \text{ MeV}$. A machine learning algorithm is
4235 being developed for low energy proton track identification near the interaction vertex.
4236 Preliminary results show an efficiency of 30% at 5 MeV for this method.

4237 The same samples used for the tracking studies were employed to estimate the
4238 momentum resolution. The momentum is computed from the curvature of the tracks
4239 in the magnetic field, and is therefore limited by the track length in the direction
4240 perpendicular to the field. Focusing on the tracks associated to true muons, a double
4241 Gaussian fit revealed a width of 2.7% and 12% for the core and tails of the momentum
4242 distribution. This same study determined the 3D angular resolution in the HPgTPC to
4243 be 0.80° .

4244 The main source of uncertainty in the momentum measurement is the value of the
4245 magnetic field. The magnetic field simulations indicate that the overall uncertainty on
4246 the central field value is $< 0.05\%$. A preliminary study investigated the use of K_s^0 decays
4247 in the HPgTPC to measure any deviations of the magnetic field from its nominal 0.5 T
4248 value. This showed that even a magnetic field bias of 1% will shift the reconstructed
4249 invariant mass distribution significantly.

CHAPTER 7. EVENT SELECTION IN ND-GAR

4250 The results presented for the ECal in Ref. [91] use an outdated version of the
4251 geometry, where the entire ECal sits outside of the pressure vessel and the layers consist
4252 of 5 mm of scintillator and 2 mm of Cu. The sample used consists of single photons in
4253 a 20° cone aligned with the beam direction. In the simulation, an energy threshold of
4254 200 keV and a time resolution of 0.25 ns are assumed.

4255 The energy resolution of the photons is obtained from a Gaussian fit. The resulting
4256 resolutions are then fitted to a function of the form:

$$\frac{\sigma_E}{E} = \frac{A}{\sqrt{E}} \oplus \frac{B}{E} \oplus C, \quad (7.16)$$

4257 where A is the stochastic term, B the noise term and C the constant term. The best fit
4258 finds the values $A = 6.1\%$, $B = 1.6\%$, and $C = 4.5\%$. The photon angle is estimated
4259 using a PCA analysis of the associated ECal cluster, taken to be the direction of the
4260 first principal component. The angular resolution is computed from a Gaussian fit to
4261 the core of the distribution. As a function of the photon energy, the values obtained are
4262 $\frac{8.17^\circ}{\sqrt{E}} + 4.18^\circ$. Different arrangements of the layers and alternative absorber choices may
4263 improve these results.

Conclusion and outlook

4266 *Our plans miscarry because they have no aim. When a man does not know
4267 what harbour he is making for, no wind is the right wind.*

4268 – Lucio Anneo Seneca, *Epistulae morales ad Lucilium*

4269 This thesis is a compilation of three different projects within DUNE. However,
4270 the idea behind each one of them is the same. The common theme is the prospect
4271 of improving or extending the physics of DUNE. In the first case, by enhancing the
4272 production of TPs in the induction channels what I seek is to provide more useful
4273 information to the FD data selection. The investigations with both data and MC, as
4274 well as the opportunity to run with a live detector, showed that such an enhancement
4275 is possible and should be pursued. Next, the solar DM analysis adds to the already
4276 rich BSM programme of DUNE. With the results of these preliminary studies, I want
4277 to show that DUNE can be complementary to the large-volume neutrino detectors in
4278 this kind of searches. Finally, the goal of the ND-GAr reconstruction improvements
4279 was the development of the PID strategy of the detector. For this, I tried to extract all
4280 the possible information from its different subcomponents. With the PID at hand, it
4281 is possible to form the selections of the different ND-GAr samples, as I have shown in
4282 this work. These will help understand how the detector is going to further constrain the
4283 neutrino interaction uncertainties in DUNE Phase II, which will eventually allow DUNE
4284 to reach its ultimate physics goals.

4285 The DAQ system of the DUNE FD relies on the online identification of hits on

CHAPTER 8. CONCLUSION AND OUTLOOK

4286 channels, the so-called TPs, to form decisions data to store. The goal of Chapter 4
4287 is to motivate a method to enhance the production of TPs in the induction channels
4288 of the detectors. Forming TPs from all the charge readout planes will improve the
4289 redundancy of the trigger algorithms. Not only that, but this may be the key to have
4290 more complex trigger logic that requires directional information. The aspect I focused on
4291 to improve the hit finding is the filtering of the waveforms. In section 4.3 I use a sample
4292 of ProtoDUNE-SP cosmic data to show how different low-pass FIR filters affect the S/N
4293 in the collection and induction planes. Then, I introduce the concept of the matched
4294 filter in section 4.4. Using the same dataset, I demonstrate that the improvement in the
4295 S/N of the induction channels achieved with these filters can be significantly higher than
4296 with the standard filter approach. A series of studies using MC samples are presented in
4297 section 4.5. These allow to study the dependence of the filtering on the orientation and
4298 the energy of the tracks. I also use them to assess the impact of this method on the hit
4299 sensitivity. Finally, in section 4.6 I briefly summarise the results from the VD ColdBox
4300 runs which featured the matched filter.

4301 With these studies, I showed that the matched filter puts the production of TPs in
4302 the induction and collection planes on the same level. The natural next step will be to
4303 understand the impact that this has in the context of the current trigger algorithms.
4304 Then, explore the development of new trigger routines, like triggers based on coincidence
4305 across planes. At the same time, these alternative hit finder chains should be implemented
4306 in the trigger simulations currently under development.

4307 The solar DM analysis is covered in Chapter 5. There I explain how the DUNE
4308 FD can be used to probe DM interactions by measuring the neutrino flux coming from
4309 DM annihilations in the core of the Sun. After introducing the topic of DM capture
4310 and annihilation in a massive object like the Sun, I describe what kind of neutrino
4311 signals one can expect from such events in section 5.2. Later, I comment on how
4312 DUNE could constrain the DM parameter space by performing counting experiments.
4313 In section 5.5 I study the selection efficiency for the $\tau^+\tau^-$ and $b\bar{b}$ channels. I focus
4314 on two different kinematic regimes: the high energy neutrinos where DIS interactions

4315 with argon dominate, and the low energy part of the spectrum where neutrinos mainly
4316 undergo QEL interactions. This allows me to compute the projected generator-level DM
4317 cross section sensitivities, showing how DUNE can be complementary to other indirect
4318 DM searches. Additionally, I explore two specific realisations of the DM interactions,
4319 namely Kaluza-Klein and leptophilic DM.

4320 At this stage, this analysis already shows the potential of DUNE to explore these
4321 scenarios. However, including the full simulation and reconstruction of the events will
4322 be necessary moving forward. At the moment, a significant effort is aimed towards the
4323 reconstruction of atmospheric neutrinos in the DUNE FD, which could be relevant for
4324 the case at hand. Also, following iterations of the analysis should include all the relevant
4325 systematic uncertainties. A summary of these is presented in section 5.7.

4326 Chapter 6 reviews my work on the reconstruction for ND-GAr. In section 6.2 I try
4327 to establish the relation between the measured charge in the readout and the deposited
4328 energy from a stopping proton sample, using the residual range of the tracks. This
4329 calibration allows to compute the mean dE/dx for the particles. I finish the section
4330 providing a parametrisation for how this depends on the momentum. The problem of the
4331 muon and pion separation is the topic of section 6.3. I propose to use the information
4332 from the ECal to achieve this classification. In this section, I describe the features and
4333 the procedure I follow to train the classifier, showing its performance as a function of
4334 the particle momentum. In section 6.4 I explore the possibility of performing a ToF
4335 measurement with the ECal. With this, I achieve a separation between pions and protons
4336 in a momentum range beyond the reach of the HPgTPC alone. Section 6.5 is devoted to
4337 the identification of charged particle decays inside the HPgTPC where the parent plus
4338 (charged) daughter system is reconstructed as a single track. I use the information from
4339 the track fit to construct a series of variables which can identify the tracks containing
4340 decays, as well as locate their position. I finish the Chapter introducing a new clustering
4341 algorithm for the ECal hits in section 6.6. It aims at having a one-to-one correspondence
4342 between particles and clusters, which will facilitate the reconstruction of neutral particles
4343 in the ECal.

CHAPTER 8. CONCLUSION AND OUTLOOK

4344 The goal of these developments was establishing a robust PID strategy for ND-GAr,
4345 that allows to reconstruct the multiplicity of pions and other hadrons in the neutrino
4346 interactions final states. In section 6.7 I describe the status of the integration of the
4347 different additions to the reconstruction chain.

4348 Finally, in Chapter 7 I apply to the event selection in ND-GAr. I start by describing
4349 a method for selecting ν_μ CC events in section 7.2. This is mainly based on the muon
4350 score derived from the muon/pion classification I developed. Additionally, I perform
4351 an optimisation of the FV. As part of this study, I also examined the kinematics of
4352 the selected primary muon and the reconstructed interaction vertex. Next, in section
4353 7.3 I explore the capabilities of ND-GAr and its reconstruction at identifying charged
4354 pions. I optimise a selection based on the reconstructed charged pion multiplicity, for
4355 events with 0, 1, 2, and $\geq 3\pi^\pm$ in the final state. I the performance of the selection as a
4356 function of the truth hadronic invariant mass, as well as the true pion kinematics for the
4357 ν_μ CC $1\pi^\pm$ case. I briefly discuss the possibility of tagging events with neutral pions
4358 by reconstructing the invariant mass of the photon pairs from their decay in section
4359 7.4. Lastly, in section 7.5 I study the neutrino energy reconstruction of the selected ν_μ
4360 CC events using a calorimetric approach. For this, I compare the values obtained using
4361 generator-level and reconstructed information.

4362 These studies constitute the first try at an event selection in ND-GAr using full
4363 simulation and reconstruction. It will serve as a stepping stone for the development
4364 of other selections and analyses. Ultimately, the goal is to quantify the impact of
4365 ND-GAr on the LBL analysis in DUNE. For this, including the effect of the systematic
4366 uncertainties outlined in section 7.6 will be necessary.

4367 In summary, this thesis provides an overview of three novel topics within DUNE. As
4368 a single sentence, in this work I investigate the enhancement of the triggering capabilities
4369 of the FD, study the sensitivity of the FD to solar DM signatures, and develop the
4370 particle identification and event selection strategies for the Phase II ND. Each Chapter
4371 aims to be a comprehensive summary of the status of the different studies. I hope they
4372 can be helpful guides for future work both in the ND and FD.

Bibliography

- 4374 [1] S.L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** 579.
 4375 33, 37
- 4376 [2] S. Weinberg, *A model of leptons*, *Physical Review Letters* **19** 1264. 33, 37
- 4377 [3] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968)
 4378 367. 33, 37
- 4379 [4] J. Erler and M. Schott, *Electroweak Precision Tests of the Standard Model after*
 4380 *the Discovery of the Higgs Boson*, *Prog. Part. Nucl. Phys.* **106** (2019) 68
 4381 [1902.05142]. 33
- 4382 [5] L. Canetti, M. Drewes and M. Shaposhnikov, *Matter and Antimatter in the*
 4383 *Universe*, *New J. Phys.* **14** (2012) 095012 [1204.4186]. 33
- 4384 [6] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates*
 4385 *and constraints*, *Phys. Rept.* **405** (2005) 279 [[hep-ph/0404175](#)]. 33
- 4386 [7] S.F. King, A. Merle, S. Morisi, Y. Shimizu and M. Tanimoto, *Neutrino Mass and*
 4387 *Mixing: from Theory to Experiment*, *New J. Phys.* **16** (2014) 045018 [1402.4271].
 4388 33
- 4389 [8] A.D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry*
 4390 *of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32. 33
- 4391 [9] M.B. Gavela, P. Hernandez, J. Orloff and O. Pene, *Standard model CP violation*
 4392 *and baryon asymmetry*, *Mod. Phys. Lett. A* **9** (1994) 795 [[hep-ph/9312215](#)]. 33

BIBLIOGRAPHY

- 4393 [10] E.K. Akhmedov, V.A. Rubakov and A.Y. Smirnov, *Baryogenesis via neutrino*
4394 *oscillations*, *Phys. Rev. Lett.* **81** (1998) 1359 [[hep-ph/9803255](#)]. 33
- 4395 [11] B.W. Lee and S. Weinberg, *Cosmological Lower Bound on Heavy Neutrino*
4396 *Masses*, *Phys. Rev. Lett.* **39** (1977) 165. 34
- 4397 [12] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys.*
4398 *Rept.* **267** (1996) 195 [[hep-ph/9506380](#)]. 34
- 4399 [13] G. Arcadi, D. Cabo-Almeida, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini
4400 et al., *The Waning of the WIMP: Endgame?*, [2403.15860](#). 34
- 4401 [14] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*
4402 *Detector Technical Design Report, Volume I Introduction to DUNE*, *JINST* **15**
4403 (2020) T08008 [[2002.02967](#)]. 34, 61, 62, 63, 69, 70, 79, 80, 81, 82
- 4404 [15] W. Pauli, *Dear radioactive ladies and gentlemen*, *Phys. Today* **31N9** (1978) 27. 37
- 4405 [16] F. Reines and C.L. Cowan, *Detection of the free neutrino*, *Phys. Rev.* **92** (1953)
4406 830. 37
- 4407 [17] A. Pich, *The Standard Model of Electroweak Interactions*, in *2010 European*
4408 *School of High Energy Physics*, pp. 1–50, 1, 2012 [[1201.0537](#)]. 38
- 4409 [18] ALEPH, DELPHI, L3, OPAL, SLD, LEP ELECTROWEAK WORKING GROUP,
4410 SLD ELECTROWEAK GROUP, SLD HEAVY FLAVOUR GROUP collaboration,
4411 *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006)
4412 257 [[hep-ex/0509008](#)]. 41
- 4413 [19] R. Davis, D.S. Harmer and K.C. Hoffman, *Search for neutrinos from the sun*,
4414 *Physical Review Letters* **20** 1205. 42
- 4415 [20] J.N. Bahcall, A.M. Serenelli and S. Basu, *New solar opacities, abundances,*
4416 *helioseismology, and neutrino fluxes*, *Astrophys. J. Lett.* **621** (2005) L85
4417 [[astro-ph/0412440](#)]. 42, 127, 129

BIBLIOGRAPHY

- 4418 [21] J.N. Bahcall, N.A. Bahcall and G. Shaviv, *Present Status of the Theoretical*
 4419 *Predictions for the ^{37}Cl Solar-Neutrino Experiment*, . 42
- 4420 [22] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,
 4421 *Measurement of the solar electron neutrino flux with the Homestake chlorine*
 4422 *detector, Astrophys. J.* **496** (1998) 505. 42
- 4423 [23] SAGE collaboration, *Measurement of the solar neutrino capture rate with gallium*
 4424 *metal. III: Results for the 2002–2007 data-taking period, Phys. Rev. C* **80** (2009)
 4425 015807 [0901.2200]. 42, 54
- 4426 [24] F. Kaether, W. Hampel, G. Heusser, J. Kiko and T. Kirsten, *Reanalysis of the*
 4427 *GALLEX solar neutrino flux and source experiments, Phys. Lett. B* **685** (2010) 47
 4428 [1001.2731]. 42, 54
- 4429 [25] Q.R. Ahmad, R.C. Allen, T.C. Andersen, J.D. Anglin, G. Bühler, J.C. Barton
 4430 et al., *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by*
 4431 *^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, . 43
- 4432 [26] Q.R. Ahmad, R.C. Allen, T.C. Andersen, J. D. Anglin, J.C. Barton, E.W. Beier
 4433 et al., *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current*
 4434 *Interactions in the Sudbury Neutrino Observatory*, . 43
- 4435 [27] T.K. Gaisser and M. Honda, *Flux of atmospheric neutrinos*, . 43
- 4436 [28] K. Hirata, T. Kajita, M. Koshiba, M. Nakahata, S. Ohara, Y. Oyama et al.,
 4437 *Experimental study of the atmospheric neutrino flux*, . 44
- 4438 [29] D. Casper, R. Becker-Szendy, C.B. Bratton, D.R. Cady, R. Claus, S.T. Dye et al.,
 4439 *Measurement of atmospheric neutrino composition with the imb-3 detector*, . 44
- 4440 [30] M. Ambrosio, R. Antolini, C. Aramo, G. Auriemma, A. Baldini, G. C. Barbarino
 4441 et al., *Measurement of the atmospheric neutrino-induced upgoing muon flux using*
 4442 *macro*, . 44

BIBLIOGRAPHY

- 4443 [31] W. Allison, G. Alner, D. Ayres, W. Barrett, C. Bode, P. Border et al.,
4444 *Measurement of the atmospheric neutrino flavour composition in soudan 2*, . 44
- 4445 [32] SUPER-KAMIOKANDE collaboration, *Evidence for oscillation of atmospheric*
4446 *neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [[hep-ex/9807003](#)]. 44
- 4447 [33] P. Minkowski, $\mu \rightarrow e\gamma$ at a Rate of One Out of 10^9 Muon Decays?, *Phys. Lett. B*
4448 **67** (1977) 421. 47
- 4449 [34] M. Gell-Mann, P. Ramond and R. Slansky, *Complex Spinors and Unified Theories*,
4450 *Conf. Proc. C* **790927** (1979) 315 [[1306.4669](#)]. 47
- 4451 [35] T. Yanagida, *Horizontal gauge symmetry and masses of neutrinos*, *Conf. Proc. C*
4452 **7902131** (1979) 95. 47
- 4453 [36] R.N. Mohapatra and G. Senjanovic, *Neutrino Mass and Spontaneous Parity*
4454 *Nonconservation*, *Phys. Rev. Lett.* **44** (1980) 912. 47
- 4455 [37] J. Schechter and J.W.F. Valle, *Neutrino Masses in $SU(2) \times U(1)$ Theories*, *Phys.*
4456 *Rev. D* **22** (1980) 2227. 47
- 4457 [38] B. Pontecorvo, *Mesonium and anti-mesonium*, *Sov. Phys. JETP* **6** (1957) 429. 47
- 4458 [39] M. Gell-Mann and A. Pais, *Behavior of neutral particles under charge conjugation*,
4459 . 47
- 4460 [40] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of*
4461 *Leptonic Charge*, *Zh. Eksp. Teor. Fiz.* **53** (1967) 1717. 48
- 4462 [41] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, *Zh.*
4463 *Eksp. Teor. Fiz.* **34** (1957) 247. 49
- 4464 [42] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary*
4465 *particles*, *Prog. Theor. Phys.* **28** (1962) 870. 49
- 4466 [43] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *Phys. Rev. D*
4467 **110** (2024) 030001. 51

BIBLIOGRAPHY

- 4468 [44] L. Wolfenstein, *Neutrino Oscillations in Matter*, *Phys. Rev. D* **17** (1978) 2369. 52
- 4469 [45] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,
4470 *Measurement of the solar electron neutrino flux with the Homestake chlorine*
4471 *detector*, *Astrophys. J.* **496** (1998) 505. 54
- 4472 [46] G. Bellini et al., *Precision measurement of the ${}^7\text{Be}$ solar neutrino interaction rate*
4473 *in Borexino*, *Phys. Rev. Lett.* **107** (2011) 141302 [[1104.1816](#)]. 54
- 4474 [47] SUPER-KAMIOKANDE collaboration, *Solar neutrino measurements in*
4475 *super-Kamiokande-I*, *Phys. Rev. D* **73** (2006) 112001 [[hep-ex/0508053](#)]. 54
- 4476 [48] SNO collaboration, *Combined Analysis of all Three Phases of Solar Neutrino*
4477 *Data from the Sudbury Neutrino Observatory*, *Phys. Rev. C* **88** (2013) 025501
4478 [[1109.0763](#)]. 54
- 4479 [49] SUPER-KAMIOKANDE collaboration, *Atmospheric neutrino oscillation analysis*
4480 *with external constraints in Super-Kamiokande I-IV*, *Phys. Rev. D* **97** (2018)
4481 072001 [[1710.09126](#)]. 54, 157
- 4482 [50] ICECUBE collaboration, *Measurement of Atmospheric Neutrino Oscillations at*
4483 *6–56 GeV with IceCube DeepCore*, *Phys. Rev. Lett.* **120** (2018) 071801
4484 [[1707.07081](#)]. 54
- 4485 [51] KAMLAND collaboration, *Reactor On-Off Antineutrino Measurement with*
4486 *KamLAND*, *Phys. Rev. D* **88** (2013) 033001 [[1303.4667](#)]. 55
- 4487 [52] RENO collaboration, *Measurement of Reactor Antineutrino Oscillation Amplitude*
4488 *and Frequency at RENO*, *Phys. Rev. Lett.* **121** (2018) 201801 [[1806.00248](#)]. 55
- 4489 [53] DAYA BAY collaboration, *Measurement of the Electron Antineutrino Oscillation*
4490 *with 1958 Days of Operation at Daya Bay*, *Phys. Rev. Lett.* **121** (2018) 241805
4491 [[1809.02261](#)]. 55

BIBLIOGRAPHY

- 4492 [54] K.J. Kelly, P.A.N. Machado, S.J. Parke, Y.F. Perez-Gonzalez and R.Z. Funchal,
4493 *Neutrino mass ordering in light of recent data*, *Phys. Rev. D* **103** (2021) 013004.
4494 55
- 4495 [55] P. Dunne, “Latest Neutrino Oscillation Results from T2K.” Jul, 2020. 55
- 4496 [56] MINOS collaboration, *Combined analysis of ν_μ disappearance and $\nu_\mu \rightarrow \nu_e$*
4497 *appearance in MINOS using accelerator and atmospheric neutrinos*, *Phys. Rev.*
4498 *Lett.* **112** (2014) 191801 [[1403.0867](#)]. 55
- 4499 [57] OPERA collaboration, *Final Results of the OPERA Experiment on ν_τ*
4500 *Appearance in the CNGS Neutrino Beam*, *Phys. Rev. Lett.* **120** (2018) 211801
4501 [[1804.04912](#)]. 55
- 4502 [58] K2K collaboration, *Measurement of Neutrino Oscillation by the K2K Experiment*,
4503 *Phys. Rev. D* **74** (2006) 072003 [[hep-ex/0606032](#)]. 55
- 4504 [59] DUNE collaboration, *Long-baseline neutrino oscillation physics potential of the*
4505 *DUNE experiment*, *Eur. Phys. J. C* **80** (2020) 978 [[2006.16043](#)]. 55, 238
- 4506 [60] HYPER-KAMIOKANDE collaboration, *Physics potential of Hyper-Kamiokande for*
4507 *neutrino oscillation measurements*, *PoS NuFact2019* (2019) 040. 55
- 4508 [61] P.F. de Salas, D.V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena,
4509 C.A. Ternes et al., *2020 global reassessment of the neutrino oscillation picture*,
4510 *JHEP* **02** (2021) 071 [[2006.11237](#)]. 55, 56, 63
- 4511 [62] T2K collaboration, *Measurements of neutrino oscillation in appearance and*
4512 *disappearance channels by the T2K experiment with 6.6×10^{20} protons on target*,
4513 *Phys. Rev. D* **91** (2015) 072010 [[1502.01550](#)]. 56
- 4514 [63] NOvA collaboration, *First measurement of muon-neutrino disappearance in*
4515 *NOvA*, *Phys. Rev. D* **93** (2016) 051104 [[1601.05037](#)]. 56

BIBLIOGRAPHY

- 4516 [64] SUPERNEMO collaboration, *Probing New Physics Models of Neutrinoless Double*
 4517 *Beta Decay with SuperNEMO*, *Eur. Phys. J. C* **70** (2010) 927 [[1005.1241](#)]. 56
- 4518 [65] SNO+ collaboration, *Current Status and Future Prospects of the SNO+*
 4519 *Experiment*, *Adv. High Energy Phys.* **2016** (2016) 6194250 [[1508.05759](#)]. 56
- 4520 [66] NEXT collaboration, *Sensitivity of a tonne-scale NEXT detector for neutrinoless*
 4521 *double beta decay searches*, *JHEP* **2021** (2021) 164 [[2005.06467](#)]. 56
- 4522 [67] P. Coloma and P. Huber, *Impact of nuclear effects on the extraction of neutrino*
 4523 *oscillation parameters*, *Phys. Rev. Lett.* **111** (2013) 221802 [[1307.1243](#)]. 57
- 4524 [68] P. Coloma, P. Huber, C.-M. Jen and C. Mariani, *Neutrino-nucleus interaction*
 4525 *models and their impact on oscillation analyses*, *Phys. Rev. D* **89** (2014) 073015
 4526 [[1311.4506](#)]. 57
- 4527 [69] U. Mosel, *Neutrino Interactions with Nucleons and Nuclei: Importance for*
 4528 *Long-Baseline Experiments*, *Ann. Rev. Nucl. Part. Sci.* **66** (2016) 171
 4529 [[1602.00696](#)]. 57
- 4530 [70] J.A. Formaggio and G.P. Zeller, *From ev to eev: Neutrino cross sections across*
 4531 *energy scales*, *Rev. Mod. Phys.* **84** (2012) 1307 [[1305.7513](#)]. 58, 59
- 4532 [71] L. Bathe-Peters, S. Gardiner and R. Guenette, *Comparing generator predictions*
 4533 *of transverse kinematic imbalance in neutrino-argon scattering*, [2201.04664](#). 59
- 4534 [72] R.A. Smith and E.J. Moniz, *Neutrino reactions on nuclear targets*, *Nucl. Phys. B*
 4535 **43** (1972) 605. 60
- 4536 [73] H. Nakamura and R. Seki, *Quasi-elastic neutrino-nucleus scattering and spectral*
 4537 *function*, *Nuclear Physics B - Proceedings Supplements* **112** 197. 60
- 4538 [74] V. Pandey, N. Jachowicz, T. Van Cuyck, J. Ryckebusch and M. Martini,
 4539 *Low-energy excitations and quasielastic contribution to electron-nucleus and*

BIBLIOGRAPHY

- 4540 *neutrino-nucleus scattering in the continuum random-phase approximation*, *Phys.*
4541 *Rev. C* **92** (2015) 024606 [[1412.4624](#)]. 60
- 4542 [75] A. Nikolakopoulos, R. González-Jiménez, N. Jachowicz, K. Niewczas, F. Sánchez
4543 and J.M. Udías, *Benchmarking intranuclear cascade models for neutrino*
4544 *scattering with relativistic optical potentials*, *Phys. Rev. C* **105** (2022) 054603
4545 [[2202.01689](#)]. 60
- 4546 [76] MINIBOONE collaboration, *First Measurement of the Muon Neutrino Charged*
4547 *Current Quasielastic Double Differential Cross Section*, *Phys. Rev. D* **81** (2010)
4548 092005 [[1002.2680](#)]. 60
- 4549 [77] MINERVA collaboration, *Measurements of the Inclusive Neutrino and*
4550 *Antineutrino Charged Current Cross Sections in MINERvA Using the Low- ν Flux*
4551 *Method*, *Phys. Rev. D* **94** (2016) 112007 [[1610.04746](#)]. 60
- 4552 [78] MICROBooNE collaboration, *First Measurement of Energy-Dependent Inclusive*
4553 *Muon Neutrino Charged-Current Cross Sections on Argon with the MicroBooNE*
4554 *Detector*, *Phys. Rev. Lett.* **128** (2022) 151801 [[2110.14023](#)]. 60
- 4555 [79] SBND collaboration, *Neutrino cross-section measurement prospects with SBND*,
4556 *PoS NuFact2017* (2018) 067. 60
- 4557 [80] NOvA collaboration, *Measurement of the double-differential muon-neutrino*
4558 *charged-current inclusive cross section in the NOvA near detector*, *Phys. Rev. D*
4559 **107** (2023) 052011 [[2109.12220](#)]. 60
- 4560 [81] T2K collaboration, *Measurement of the ν_μ charged-current cross sections on*
4561 *water, hydrocarbon, iron, and their ratios with the T2K on-axis detectors*, *PTEP*
4562 **2019** (2019) 093C02 [[1904.09611](#)]. 60
- 4563 [82] \$NU\$PRISM collaboration, *ν PRISM: A new way of probing neutrino*
4564 *interactions*, *PoS NuFACT2014* (2015) 046. 60

BIBLIOGRAPHY

- 4565 [83] C. Hasnip, *DUNE-PRISM – A New Method to Measure Neutrino Oscillations*,
 4566 Ph.D. thesis, Oxford U., 2023. 60, 71
- 4567 [84] DUNE collaboration, *Snowmass Neutrino Frontier: DUNE Physics Summary*,
 4568 2203.06100. 63, 64
- 4569 [85] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*
 4570 *Detector Technical Design Report, Volume II: DUNE Physics*, 2002.03005. 63, 65,
 4571 67, 73, 133, 142, 146, 271
- 4572 [86] SUPER-KAMIOKANDE collaboration, *Search for Proton Decay via $p \rightarrow e^+ \pi_0$ and*
 4573 *$p \rightarrow \mu^+ \pi_0$ in a Large Water Cherenkov Detector*, *Phys. Rev. Lett.* **102** (2009)
 4574 141801 [0903.0676]. 64
- 4575 [87] S. Raby, *Grand Unified Theories*, in *2nd World Summit: Physics Beyond the*
 4576 *Standard Model*, 8, 2006 [[hep-ph/0608183](#)]. 64
- 4577 [88] KAMIOKANDE-II collaboration, *Observation of a Neutrino Burst from the*
 4578 *Supernova SN 1987a*, *Phys. Rev. Lett.* **58** (1987) 1490. 65
- 4579 [89] R.M. Bionta et al., *Observation of a Neutrino Burst in Coincidence with*
 4580 *Supernova SN 1987a in the Large Magellanic Cloud*, *Phys. Rev. Lett.* **58** (1987)
 4581 1494. 65
- 4582 [90] J. Strait, E. McCluskey, T. Lundin, J. Willhite, T. Hamernik, V. Papadimitriou
 4583 et al., *Long-baseline neutrino facility (lbnf) and deep underground neutrino*
 4584 *experiment (dune) conceptual design report volume 3: Long-baseline neutrino*
 4585 *facility for dune june 24, 2015*, 1601.05823. 65, 66
- 4586 [91] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE) Near*
 4587 *Detector Conceptual Design Report, Instruments* **5** (2021) 31 [[2103.13910](#)]. 68,
 4588 72, 73, 74, 75, 275, 278
- 4589 [92] J. Asaadi et al., *A New Concept for Kilotonne Scale Liquid Argon Time*
 4590 *Projection Chambers, Instruments* **4** (2020) 6 [[1908.10956](#)]. 69

BIBLIOGRAPHY

- 4591 [93] DUNE collaboration, *DUNE Phase II: Scientific Opportunities, Detector*
4592 *Concepts, Technological Solutions*, 2408.12725. 71, 74, 76
- 4593 [94] DUNE collaboration, *A Gaseous Argon-Based Near Detector to Enhance the*
4594 *Physics Capabilities of DUNE*, 2203.06281. 73
- 4595 [95] ALICE collaboration, *Alice: Physics performance report, volume ii*, . 74, 276
- 4596 [96] F. Sauli, *Gem: A new concept for electron amplification in gas detectors*, *Nuclear*
4597 *Instruments and Methods in Physics Research Section A: Accelerators,*
4598 *Spectrometers, Detectors and Associated Equipment* **386** 531. 74, 78
- 4599 [97] DUNE collaboration, *SPY: a conceptual design study of a magnet system for a*
4600 *high-pressure gaseous TPC neutrino detector*, *JINST* **19** (2024) P06018
4601 [2311.16063]. 76
- 4602 [98] A. Ritchie-Yates et al., *First operation of an ALICE OROC operated in high*
4603 *pressure Ar-CO₂ and Ar-CH₄*, *Eur. Phys. J. C* **83** (2023) 1139 [2305.08822]. 77
- 4604 [99] ALICE TPC collaboration, *The upgrade of the ALICE TPC with GEMs and*
4605 *continuous readout*, *JINST* **16** (2021) P03022 [2012.09518]. 77
- 4606 [100] F. Sauli, *The gas electron multiplier (gem): Operating principles and applications*,
4607 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,*
4608 *Spectrometers, Detectors and Associated Equipment* **805** 2. 78
- 4609 [101] C. Lippmann, *A continuous read-out tpc for the alice upgrade*, *Nuclear*
4610 *Instruments and Methods in Physics Research Section A: Accelerators,*
4611 *Spectrometers, Detectors and Associated Equipment* **824** 543. 78
- 4612 [102] C. Calabria, *Large-size triple gem detectors for the cms forward muon upgrade*,
4613 *Nuclear and Particle Physics Proceedings* **273–275** 1042. 78
- 4614 [103] DUNE collaboration, *The DUNE Far Detector Vertical Drift Technology*,
4615 *Technical Design Report*, 2312.03130. 83, 84

BIBLIOGRAPHY

- ⁴⁶¹⁶ [104] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume IV: Far Detector Single-phase Technology*, JINST **15** (2020) T08010 [2002.03010]. 85
- ⁴⁶¹⁹ [105] DUNE DAQ Project, “Trigger and Data AcQuisition (TDAQ) System Design.” EDMS-2826454, 2022. 87
- ⁴⁶²¹ [106] DUNE DAQ, “dtp-firmware.”
⁴⁶²² <https://gitlab.cern.ch/dune-daq/readout/dtp-firmware>, 2020. 88
- ⁴⁶²³ [107] J. McClellan and T. Parks, *A personal history of the Parks-McClellan algorithm*, IEEE Signal Processing Magazine **22** (2005) 82. 91
- ⁴⁶²⁵ [108] L. Weinberg and P. Slepian, *Takahasi’s results on tchebycheff and butterworth ladder networks*, IRE Transactions on Circuit Theory **7** (1960) 88. 91
- ⁴⁶²⁷ [109] G. Turin, *An introduction to matched filters*, IEEE Transactions on Information Theory **6** (1960) 311. 95
- ⁴⁶²⁹ [110] J.W. Goodman, *Statistical Optics*, Wiley (1985). 96
- ⁴⁶³⁰ [111] B. Dwork, *Detection of a pulse superimposed on fluctuation noise*, Proceedings of the IRE **38** (1950) 771. 97
- ⁴⁶³² [112] L. Wainstein and V. Zubakov, *Extraction of Signals from Noise*, Prentice-Hall (1962). 98
- ⁴⁶³⁴ [113] E.D. Church, *Larsoft: A software package for liquid argon time projection drift chambers*, 1311.6774. 101
- ⁴⁶³⁶ [114] S.V. Stehman, *Selecting and interpreting measures of thematic classification accuracy*, Remote Sensing of Environment **62** (1997) 77. 114
- ⁴⁶³⁸ [115] A.A. Taha and A. Hanbury, *Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool*, BMC Medical Imaging **15** (2015) . 115

BIBLIOGRAPHY

- 4640 [116] J. Silk, K.A. Olive and M. Srednicki, *The Photino, the Sun and High-Energy*
4641 *Neutrinos*, *Phys. Rev. Lett.* **55** (1985) 257. 123
- 4642 [117] M. Srednicki, K.A. Olive and J. Silk, *High-Energy Neutrinos from the Sun and*
4643 *Cold Dark Matter*, *Nucl. Phys. B* **279** (1987) 804. 123
- 4644 [118] J.S. Hagelin, K.W. Ng and K.A. Olive, *A High-energy Neutrino Signature From*
4645 *Supersymmetric Relics*, *Phys. Lett. B* **180** (1986) 375. 123
- 4646 [119] T.K. Gaisser, G. Steigman and S. Tilav, *Limits on Cold Dark Matter Candidates*
4647 *from Deep Underground Detectors*, *Phys. Rev. D* **34** (1986) 2206. 123
- 4648 [120] N. Bernal, J. Martín-Albo and S. Palomares-Ruiz, *A novel way of constraining*
4649 *WIMPs annihilations in the Sun: MeV neutrinos*, *JCAP* **08** (2013) 011
4650 [1208.0834]. 123, 124, 125, 127, 131
- 4651 [121] C. Rott, J. Siegal-Gaskins and J.F. Beacom, *New Sensitivity to Solar WIMP*
4652 *Annihilation using Low-Energy Neutrinos*, *Phys. Rev. D* **88** (2013) 055005
4653 [1208.0827]. 123, 130
- 4654 [122] C. Rott, S. In, J. Kumar and D. Yaylali, *Dark Matter Searches for Monoenergetic*
4655 *Neutrinos Arising from Stopped Meson Decay in the Sun*, *JCAP* **11** (2015) 039
4656 [1510.00170]. 123
- 4657 [123] DUNE collaboration, *Searching for solar KDAR with DUNE*, *JCAP* **10** (2021)
4658 065 [2107.09109]. 123, 272, 273, 274, 275
- 4659 [124] G. Busoni, A. De Simone and W.-C. Huang, *On the Minimum Dark Matter Mass*
4660 *Testable by Neutrinos from the Sun*, *JCAP* **07** (2013) 010 [1305.1817]. 124
- 4661 [125] J.N. Bahcall and M.H. Pinsonneault, *Solar models with helium and heavy element*
4662 *diffusion*, *Rev. Mod. Phys.* **67** (1995) 781 [[hep-ph/9505425](#)]. 124
- 4663 [126] R. Garani and S. Palomares-Ruiz, *Dark matter in the Sun: scattering off electrons*
4664 *vs nucleons*, *JCAP* **05** (2017) 007 [1702.02768]. 125, 126, 128, 129, 130, 153

BIBLIOGRAPHY

- 4665 [127] V.A. Bednyakov and F. Simkovic, *Nuclear spin structure in dark matter search: The Zero momentum transfer limit*, *Phys. Part. Nucl.* **36** (2005) 131
 4666 [hep-ph/0406218]. 125
- 4668 [128] R. Catena and B. Schwabe, *Form factors for dark matter capture by the Sun in effective theories*, *JCAP* **04** (2015) 042 [1501.03729]. 125
- 4670 [129] A. Gould, *WIMP Distribution in and Evaporation From the Sun*, *Astrophys. J.*
 4671 **321** (1987) 560. 125, 126
- 4672 [130] A. Gould, *Resonant Enhancements in WIMP Capture by the Earth*, *Astrophys. J.*
 4673 **321** (1987) 571. 127
- 4674 [131] T. Bringmann, J. Edsjö, P. Gondolo, P. Ullio and L. Bergström, *DarkSUSY 6 : An Advanced Tool to Compute Dark Matter Properties Numerically*, *JCAP* **07**
 4675 (2018) 033 [1802.03399]. 128
- 4677 [132] J.N. Bahcall, M.H. Pinsonneault and S. Basu, *Solar models: Current epoch and time dependences, neutrinos, and helioseismological properties*, *Astrophys. J.* **555**
 4678 (2001) 990 [astro-ph/0010346]. 128
- 4680 [133] T. Golan, J.T. Sobczyk and J. Zmuda, *NuWro: the Wroclaw Monte Carlo Generator of Neutrino Interactions*, *Nucl. Phys. B Proc. Suppl.* **229-232** (2012)
 4681 499. 132
- 4683 [134] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara and S. Midorikawa,
 4684 *Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model*,
 4685 *Phys. Rev. D* **92** (2015) 023004 [1502.03916]. 133, 134
- 4686 [135] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727].
 4687 133
- 4689 [136] T. Kaluza, *Zum Unitätsproblem der Physik*, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1921** (1921) 966 [1803.08616]. 135

BIBLIOGRAPHY

- 4691 [137] O. Klein, *Quantum Theory and Five-Dimensional Theory of Relativity. (In
4692 German and English)*, *Z. Phys.* **37** (1926) 895. 135
- 4693 [138] T. Appelquist, H.-C. Cheng and B.A. Dobrescu, *Bounds on universal extra
4694 dimensions*, *Phys. Rev. D* **64** (2001) 035002 [[hep-ph/0012100](#)]. 135
- 4695 [139] N. Arkani-Hamed, S. Dimopoulos and G.R. Dvali, *The Hierarchy problem and new
4696 dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263 [[hep-ph/9803315](#)]. 135
- 4697 [140] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra
4698 dimension*, *Phys. Rev. Lett.* **83** (1999) 3370 [[hep-ph/9905221](#)]. 135
- 4699 [141] G. Servant and T.M.P. Tait, *Is the lightest Kaluza-Klein particle a viable dark
4700 matter candidate?*, *Nucl. Phys. B* **650** (2003) 391 [[hep-ph/0206071](#)]. 136
- 4701 [142] H.-C. Cheng, K.T. Matchev and M. Schmaltz, *Radiative corrections to
4702 Kaluza-Klein masses*, *Phys. Rev. D* **66** (2002) 036005 [[hep-ph/0204342](#)]. 136
- 4703 [143] M. Blennow, J. Edsjö and T. Ohlsson, *Neutrinos from WIMP annihilations using
4704 a full three-flavor Monte Carlo*, *JCAP* **01** (2008) 021 [[0709.3898](#)]. 136, 139
- 4705 [144] J. Edsjö, J. Elevant and C. Niblaeus, *WimpSim Neutrino Monte Carlo*. 136, 139
- 4706 [145] M. Colom i Bernadich and C. Pérez de los Heros, *Limits on Kaluza-Klein dark
4707 matter annihilation in the Sun from recent IceCube results*, *Eur. Phys. J. C*, **80** 2
4708 (2020) 129 **80** (2019) [[1912.04585](#)]. 138
- 4709 [146] ANTARES collaboration, *Search for Dark Matter in the Sun with the ANTARES
4710 Neutrino Telescope in the CMSSM and mUED frameworks*, *Nucl. Instrum. Meth.
4711 A* **725** (2013) 76 [[1204.5290](#)]. 138
- 4712 [147] N. Deutschmann, T. Flacke and J.S. Kim, *Current LHC Constraints on Minimal
4713 Universal Extra Dimensions*, *Phys. Lett. B* **771** (2017) 515 [[1702.00410](#)]. 138
- 4714 [148] U. Haisch and A. Weiler, *Bound on minimal universal extra dimensions from
4715 anti-B → X(s)gamma*, *Phys. Rev. D* **76** (2007) 034014 [[hep-ph/0703064](#)]. 138

BIBLIOGRAPHY

- 4716 [149] A. Freitas and U. Haisch, *Anti- $B \rightarrow X(s)$ gamma in two universal extra*
4717 *dimensions, Phys. Rev. D* **77** (2008) 093008 [[0801.4346](#)]. 138
- 4718 [150] C. Rott, D. Jeong, J. Kumar and D. Yaylali, *Neutrino Topology Reconstruction at*
4719 *DUNE and Applications to Searches for Dark Matter Annihilation in the Sun,*
4720 *JCAP* **07** (2019) 006 [[1903.04175](#)]. 142
- 4721 [151] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al.,
4722 *Scikit-learn: Machine learning in Python, Journal of Machine Learning Research*
4723 **12** (2011) 2825. 147
- 4724 [152] ICECUBE collaboration, *Search for GeV-scale dark matter annihilation in the Sun*
4725 *with IceCube DeepCore, Phys. Rev. D* **105** (2022) 062004 [[2111.09970](#)]. 149, 150
- 4726 [153] N.F. Bell, M.J. Dolan and S. Robles, *Searching for dark matter in the Sun using*
4727 *Hyper-Kamiokande, JCAP* **11** (2021) 004 [[2107.04216](#)]. 149, 150
- 4728 [154] E. Behnke et al., *Final Results of the PICASSO Dark Matter Search Experiment,*
4729 *Astropart. Phys.* **90** (2017) 85 [[1611.01499](#)]. 149, 150
- 4730 [155] PICO collaboration, *Dark Matter Search Results from the Complete Exposure of*
4731 *the PICO-60 C₃F₈ Bubble Chamber, Phys. Rev. D* **100** (2019) 022001
4732 [[1902.04031](#)]. 149, 150
- 4733 [156] J. Kopp, V. Niro, T. Schwetz and J. Zupan, *DAMA/LIBRA and leptonically*
4734 *interacting Dark Matter, Phys. Rev. D* **80** (2009) 083502 [[0907.3159](#)]. 152
- 4735 [157] PARTICLE DATA GROUP collaboration, *Review of Particle Physics, PTEP* **2020**
4736 (2020) 083C01. 153
- 4737 [158] M. Beltran, D. Hooper, E.W. Kolb and Z.C. Krusberg, *Deducing the nature of*
4738 *dark matter from direct and indirect detection experiments in the absence of*
4739 *collider signatures of new physics, Phys. Rev. D* **80** (2009) 043509 [[0808.3384](#)].
- 4740 153

BIBLIOGRAPHY

- 4741 [159] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters, Astron.*
4742 *Astrophys.* **641** (2020) A6 [[1807.06209](#)]. 154
- 4743 [160] XENON collaboration, *Light Dark Matter Search with Ionization Signals in*
4744 *XENON1T, Phys. Rev. Lett.* **123** (2019) 251801 [[1907.11485](#)]. 154, 155
- 4745 [161] DARKSIDE collaboration, *Search for Dark Matter Particle Interactions with*
4746 *Electron Final States with DarkSide-50, Phys. Rev. Lett.* **130** (2023) 101002
4747 [[2207.11968](#)]. 154, 155
- 4748 [162] PANDAX-II collaboration, *Search for Light Dark Matter-Electron Scatterings in*
4749 *the PandaX-II Experiment, Phys. Rev. Lett.* **126** (2021) 211803 [[2101.07479](#)].
4750 154, 155
- 4751 [163] C. Principato, *An Indirect Search for Weakly Interacting Massive Particles in the*
4752 *Sun Using Upward-going Muons in NOvA*, Ph.D. thesis, Virginia U., 2021.
4753 [10.18130/x5z2-1466](#). 156, 157
- 4754 [164] G. Wikström and J. Edsjö, *Limits on the wimp-nucleon scattering cross-section*
4755 *from neutrino telescopes*, [0903.2986](#). 156
- 4756 [165] SUPER-KAMIOKANDE collaboration, *Search for neutrinos from annihilation of*
4757 *captured low-mass dark matter particles in the Sun by Super-Kamiokande, Phys.*
4758 *Rev. Lett.* **114** (2015) 141301 [[1503.04858](#)]. 156, 157
- 4759 [166] C. Rott, T. Tanaka and Y. Itow, *Enhanced sensitivity to dark matter*
4760 *self-annihilations in the sun using neutrino spectral information*, [1107.3182](#). 156
- 4761 [167] M.M. Boliev, S.V. Demidov, S.P. Mikheyev and O.V. Suvorova, *Search for muon*
4762 *signal from dark matter annihilations in the Sun with the Baksan Underground*
4763 *Scintillator Telescope for 24.12 years, JCAP* **09** (2013) 019 [[1301.1138](#)]. 156, 157,
4764 158

BIBLIOGRAPHY

- 4765 [168] 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](#)]. 158
- 4768 [169] C. Green, J. Kowalkowski, M. Paterno, M. Fischler, L. Garren and Q. Lu, *The Art Framework*, *J. Phys. Conf. Ser.* **396** (2012) 022020. 159
- 4770 [170] H. Bethe, *Zur theorie des durchgangs schneller korpuskularstrahlen durch materie*, *Annalen der Physik* **397** (1930) 325. 165
- 4772 [171] E. Fermi, *The ionization loss of energy in gases and in condensed materials*, *Physical Review* **57** (1940) 485. 165
- 4774 [172] 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. 165
- 4777 [173] W.W.M. Allison and J.H. Cobb, *Relativistic Charged Particle Identification by Energy Loss*, *Ann. Rev. Nucl. Part. Sci.* **30** (1980) 253. 165
- 4779 [174] W. Blum, L. Rolandi and W. Riegler, *Particle detection with drift chambers*, *Particle Acceleration and Detection* (2008), 10.1007/978-3-540-76684-1. 166
- 4781 [175] ALICE TPC collaboration, *Particle identification of the ALICE TPC via dE/dx*, *Nucl. Instrum. Meth. A* **706** (2013) 55. 166
- 4783 [176] L. Landau, *On the energy loss of fast particles by ionization*, *J. Phys. (USSR)* **8** (1944) 201. 169
- 4785 [177] W. Ulmer and E. Matsinos, *Theoretical methods for the calculation of Bragg curves and 3D distributions of proton beams*, *The European Physical Journal Special Topics* **190** (2010) 1. 170
- 4788 [178] E. Aprile, A.E. Bolotnikov, A.L. Bolozdynya and T. Doke, *Noble Gas Detectors*, Wiley (Oct., 2008), 10.1002/9783527610020. 172

BIBLIOGRAPHY

- 4790 [179] D.S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press. 210
- 4791 [180] ALICE collaboration, *Production of pions, kaons and protons in pp collisions at*
4792 $\sqrt{s} = 900 \text{ gev with alice at the lhc}$, 1101.4110. 214
- 4793 [181] U. Einhaus, *Charged hadron identification with de/dx and time-of-flight at future*
4794 *higgs factories*, 2110.15115. 214
- 4795 [182] R. Frühwirth, *Application of filter methods to the reconstruction of tracks and*
4796 *vertices in events of experimental high energy physics*, Ph.D. thesis, Technischen
4797 Universität Wien, 1988. 220
- 4798 [183] P. Astier, A. Cardini, R.D. Cousins, A. Letessier-Selvon, B.A. Popov and
4799 T. Vinogradova, *Kalman filter track fits and track breakpoint analysis*, *Nuclear*
4800 *Instruments and Methods in Physics Research Section A: Accelerators,*
4801 *Spectrometers, Detectors and Associated Equipment* **450** (2000) 138. 220
- 4802 [184] T2K UK collaboration, *The Electromagnetic Calorimeter for the T2K Near*
4803 *Detector ND280*, *JINST* **8** (2013) P10019 [1308.3445]. 228
- 4804 [185] J.E. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'*,
4805 Ph.D. thesis, Stanford University, 1982. 233
- 4806 [186] GENIE collaboration, *Neutrino-nucleon cross-section model tuning in GENIE v3*,
4807 *Phys. Rev. D* **104** (2021) 072009 [2104.09179]. 237
- 4808 [187] J. Nieves, J.E. Amaro and M. Valverde, *Inclusive quasi-elastic neutrino reactions*,
4809 *Phys. Rev. C* **70** (2004) 055503 [nucl-th/0408005]. 237
- 4810 [188] J. Nieves, I. Ruiz Simo and M.J. Vicente Vacas, *Inclusive Charged-Current*
4811 *Neutrino-Nucleus Reactions*, *Phys. Rev. C* **83** (2011) 045501 [1102.2777]. 237,
4812 274
- 4813 [189] C. Berger and L.M. Sehgal, *Lepton mass effects in single pion production by*
4814 *neutrinos*, *Phys. Rev. D* **76** (2007) 113004 [0709.4378]. 238, 274

BIBLIOGRAPHY

- 4815 [190] C. Berger and L.M. Sehgal, *PCAC and coherent pion production by low energy*
 4816 *neutrinos*, *Phys. Rev. D* **79** (2009) 053003 [[0812.2653](#)]. 238
- 4817 [191] A. Bodek and U.K. Yang, *Modeling deep inelastic cross-sections in the few GeV*
 4818 *region*, *Nucl. Phys. B Proc. Suppl.* **112** (2002) 70 [[hep-ex/0203009](#)]. 238, 275
- 4819 [192] T. Golan, L. Aliaga and M. Kordosky, *Minerva's flux prediction*, in *Proceedings of*
 4820 *the 10th International Workshop on Neutrino-Nucleus Interactions in Few-GeV*
 4821 *Region (NuInt15)*, Journal of the Physical Society of Japan, DOI. 271
- 4822 [193] A. Bashyal, H. Schellman and L. Fields, *Ppfx implementation in deep underground*
 4823 *neutrino experiment*, . 272
- 4824 [194] A. Bodek and J.L. Ritchie, *Fermi Motion Effects in Deep Inelastic Lepton*
 4825 *Scattering from Nuclear Targets*, *Phys. Rev. D* **23** (1981) 1070. 272
- 4826 [195] C. Wilkinson et al., *Testing charged current quasi-elastic and multinucleon*
 4827 *interaction models in the NEUT neutrino interaction generator with published*
 4828 *datasets from the MiniBooNE and MINERνA experiments*, *Phys. Rev. D* **93**
 4829 (2016) 072010 [[1601.05592](#)]. 272
- 4830 [196] T2K collaboration, *Search for CP Violation in Neutrino and Antineutrino*
 4831 *Oscillations by the T2K Experiment with 2.2×10^{21} Protons on Target*, *Phys. Rev.*
 4832 *Lett.* **121** (2018) 171802 [[1807.07891](#)]. 274
- 4833 [197] MINERνA collaboration, *Identification of nuclear effects in neutrino-carbon*
 4834 *interactions at low three-momentum transfer*, *Phys. Rev. Lett.* **116** (2016) 071802
 4835 [[1511.05944](#)]. 274
- 4836 [198] C. Colle, O. Hen, W. Cosyn, I. Korover, E. Piasetzky, J. Ryckebusch et al.,
 4837 *Extracting the mass dependence and quantum numbers of short-range correlated*
 4838 *pairs from $a(e, e' p)$ and $a(e, e' pp)$ scattering*, *Phys. Rev. C* **92** (2015) 024604. 274
- 4839 [199] D. Rein and L.M. Sehgal, *Neutrino Excitation of Baryon Resonances and Single*
 4840 *Pion Production*, *Annals Phys.* **133** (1981) 79. 274

BIBLIOGRAPHY

- 4841 [200] M. Sanchez, "NOvA Results and Prospects." June, 2018. 275
- 4842 [201] D.H. Stork, *First operation of the tpc facility at peplbl, ucla, uc riverside , johns*
4843 *hopkins , tokyo, yale (pep4 collaboration)*, . 276
- 4844 [202] H. Aihara, M. Alston-Garnjost, D.H. Badtke, J.A. Bakken, A. Barbaro-Galtieri,
4845 A.V. Barnes et al., *Spatial resolution of the pep-4 time projection chamber*, . 276
- 4846 [203] I. Lehraus, *Progress in particle identification by ionization sampling, Nucl.*
4847 *Instrum. Meth.* **217** (1983) 43. 277