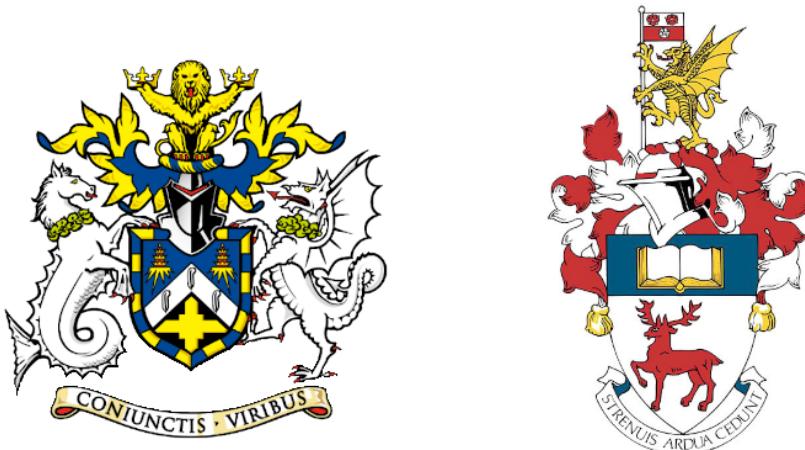


<sup>1</sup> ADVANCING NEUTRINO  
<sup>2</sup> DETECTION AND TRIGGERING IN  
<sup>3</sup> DUNE



<sup>4</sup>  
<sup>5</sup> Francisco Martínez López

<sup>6</sup> Submitted in partial fulfillment of the requirements  
<sup>7</sup> of the Degree of Doctor of Philosophy

<sup>8</sup> School of Physical and Chemical Sciences  
<sup>9</sup> Queen Mary University of London

<sup>10</sup> School of Physics and Astronomy  
<sup>11</sup> University of Southampton

<sup>12</sup> 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

35

36	Statement of originality . . . . .	3
37	Abstract . . . . .	5
38	Acknowledgements . . . . .	9
39	<b>List of Figures</b>	<b>15</b>
40	<b>List of Tables</b>	<b>33</b>
41	<b>List of Abbreviations</b>	<b>35</b>
42	<b>1 Introduction</b>	<b>37</b>
43	<b>2 Neutrino physics</b>	<b>39</b>
44	2.1 Neutrinos in the SM . . . . .	39
45	2.2 Trouble in the neutrino sector . . . . .	43
46	2.2.1 The solar neutrino problem . . . . .	43
47	2.2.2 The atmospheric neutrino problem . . . . .	45
48	2.3 Massive neutrinos . . . . .	47
49	2.4 Neutrino oscillation formalism . . . . .	49
50	2.4.1 Oscillations in vacuum . . . . .	51
51	2.4.2 Oscillations in matter . . . . .	54
52	2.4.3 Current status of neutrino oscillations . . . . .	56
53	2.5 Open questions in the neutrino sector . . . . .	57
54	2.6 Neutrino interactions . . . . .	58

## CONTENTS

55	<b>3 The Deep Underground Neutrino Experiment</b>	<b>63</b>
56	3.1 Overview . . . . .	63
57	3.2 Physics goals of DUNE . . . . .	65
58	3.3 LBNF beamline . . . . .	67
59	3.4 Near Detector . . . . .	68
60	3.4.1 ND-LAr . . . . .	71
61	3.4.2 TMS/ND-GAr . . . . .	72
62	3.4.3 PRISM . . . . .	73
63	3.4.4 SAND . . . . .	74
64	3.5 A More Capable Near Detector . . . . .	74
65	3.5.1 Requirements . . . . .	75
66	3.5.2 Reference design . . . . .	75
67	3.5.3 R&D efforts . . . . .	78
68	3.6 Far Detector . . . . .	81
69	3.6.1 Horizontal Drift . . . . .	82
70	3.6.2 Vertical Drift . . . . .	84
71	3.6.3 FD Data Acquisition System . . . . .	86
72	<b>4 Matched Filter approach to Trigger Primitives</b>	<b>89</b>
73	4.1 Motivation . . . . .	89
74	4.2 Signal-to-noise ratio definition . . . . .	91
75	4.3 Low-pass FIR filter design . . . . .	93
76	4.4 Matched filters . . . . .	96
77	4.5 Using simulated samples . . . . .	102
78	4.5.1 Angular dependence . . . . .	108
79	4.5.2 Distortion and peak asymmetry . . . . .	110
80	4.5.3 Hit sensitivity . . . . .	113
81	<b>5 DM searches with neutrinos from the Sun</b>	<b>123</b>
82	5.1 Gravitational capture of DM by the Sun . . . . .	123

## CONTENTS

83	5.2 Neutrino flux from DM annihilations . . . . .	130
84	5.3 Computing limits from solar neutrino fluxes . . . . .	131
85	5.4 Example: Kaluza-Klein Dark Matter . . . . .	135
86	5.5 High energy DM neutrino signals . . . . .	138
87	5.5.1 DIS events . . . . .	141
88	5.5.2 Single proton QEL events . . . . .	145
89	5.5.3 Results . . . . .	149
90	5.6 Example: Leptophilic Dark Matter . . . . .	151
91	5.7 Systematic uncertainties . . . . .	155
92	5.7.1 Systematic uncertainties in the solar WIMP signal . . . . .	156
93	5.7.2 Systematic uncertainties in the atmospheric background . . . . .	157
94	5.7.3 Common systematic uncertainties . . . . .	157
95	<b>6 Particle ID in ND-GAr</b>	<b>159</b>
96	6.1 GArSoft . . . . .	160
97	6.1.1 Event generation . . . . .	160
98	6.1.2 Detector simulation . . . . .	161
99	6.1.3 Reconstruction . . . . .	162
100	6.2 $dE/dx$ measurement in the TPC . . . . .	165
101	6.2.1 Energy calibration . . . . .	167
102	6.2.2 Truncated $dE/dx$ mean . . . . .	177
103	6.2.3 Mean $dE/dx$ parametrisation . . . . .	180
104	6.2.4 Particle identification . . . . .	184
105	6.3 Muon and pion separation in the ECal and MuID . . . . .	184
106	6.3.1 Track-ECal matching . . . . .	187
107	6.3.2 Classification strategy . . . . .	191
108	6.3.3 Feature selection and importance . . . . .	195
109	6.3.4 Hyperparameter optimisation . . . . .	206
110	6.3.5 Probability calibration . . . . .	210

## CONTENTS

111	6.3.6 Performance . . . . .	212
112	6.4 ECal time-of-flight . . . . .	212
113	6.4.1 Arrival time estimations . . . . .	214
114	6.4.2 Proton and pion separation . . . . .	216
115	6.5 Charged pion decay in flight . . . . .	216
116	6.5.1 Track breakpoints . . . . .	219
117	6.6 Neutral particle identification . . . . .	226
118	6.6.1 ECal clustering . . . . .	226
119	6.6.2 $\pi^0$ reconstruction . . . . .	229
120	<b>7 Event selection in ND-GAr</b>	<b>233</b>
121	7.1 Data sample . . . . .	233
122	7.2 $\nu_\mu$ CC selection . . . . .	233
123	7.2.1 Selection optimisation . . . . .	236
124	7.2.2 Selection performance . . . . .	242
125	7.2.3 Primary muon kinematics . . . . .	244
126	7.3 Charged pion identification . . . . .	245
127	7.3.1 $\nu_\mu$ CC $1\pi^\pm$ selection . . . . .	245
128	7.4 Neutral pion identification . . . . .	245
129	7.5 Detector systematic uncertainties . . . . .	245
130	<b>8 Conclusion and outlook</b>	<b>247</b>
131	<b>A An appendix</b>	<b>249</b>
132	<b>Bibliography</b>	<b>251</b>

# List of Figures

133

134	2.1 Solar neutrino fluxes for the solar model BS05(OP). . . . .	44
135	2.2 Zenith angle distributions for the selected $\nu_e$ and $\nu_\mu$ events in the SK	
136	detector. . . . .	46
137	2.3 $K^0 \rightleftharpoons \bar{K}^0$ mixing through $W^\pm$ exchange. . . . .	50
138	2.4 Feynman diagrams of the four most relevant CC interactions to long	
139	baseline neutrino oscillation experiments. . . . .	59
140	2.5 Total muon neutrino CC cross section per nucleon as a function of the	
141	neutrino energy. . . . .	60
142	2.6 Schematic representation of a $\nu_\mu$ CCQE interaction with a neutron inside	
143	a nucleus. . . . .	61
144	3.1 Schematic diagram of the DUNE experiment and the LBNF beamline. .	64
145	3.2 Schematic longitudinal section of the LBNF beamline at Fermilab. . .	67
146	3.3 Predicted neutrino fluxes at the FD in FHC mode and RHC mode. . .	68
147	3.4 Representation of the ND hall in Phase II, showing the different subcomponents.	69
148	3.5 Schematic representation of the external components of ND-LAr, including	
149	the cryostat and the PRISM movable system and detailed drawing of one	
150	ArgonCube module. . . . .	70
151	3.6 Schematic view of the TMS detector, highlighting its main parts. . . .	71
152	3.7 Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and	
153	magnet. . . . .	72
154	3.8 Predicted beam muon neutrino flux at the ND location for different off-axis	
155	positions. . . . .	73

## LIST OF FIGURES

156	3.9	Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers. . . . .	76
157	3.10	Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers. . . . .	77
158	3.11	Photographs of the TOAD pressure vessel at RHUL. . . . .	79
159	3.12	Electron microscope image and schematic diagram of a GEM electrode. . . . .	80
160	3.13	Schematic diagram showing the operating principle of a LArTPC with wire readout. . . . .	81
161	3.14	Proposed design for the FD-1 and FD-2 modules following the HD principle. . . . .	82
162	3.15	Schematic representation of an APA frames showing the U, V, X and G wires. . . . .	83
163	3.16	A PDS module containing 24 X-ARAPUCAs and the location of the modules on the APAs. . . . .	84
164	3.17	Proposed design for the FD-3 module following the VD principle. . . . .	85
165	3.18	Schematic representation of the electrode strip configuration for a top and bottom CRU. . . . .	86
166	3.19	Detailed diagram of the DUNE FD DAQ system. . . . .	87
167	4.1	<i>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.</i> . . . . .	90
168	4.2	<i>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 <math>\pm 3\sigma_{raw}</math>. The resulting noise waveform is also shown (red line). Top right panel: ADC distribution for channel 7840, where the green shaded region represents <math>\pm \sigma_{raw}</math>. Bottom right panel: noise ADC distribution for channel 7840, where the green shaded region represents <math>\pm \sigma_{noise}</math>.</i> . . . . .	91

## LIST OF FIGURES

184	4.3 <i>Left panel:</i> Zoomed filtered waveform corresponding to channel 7840 from the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i> (blue line). The filter used was the current implementation of the low-pass FIR filter in <i>dtp-firmware</i> . The green dashed lines mark the region $\pm 3\sigma_{\text{raw}}$ . The resulting noise waveform is also shown (red line). <i>Top right panel:</i> ADC distribution for channel 7840 after filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$ . <i>Bottom right panel:</i> noise ADC distribution for channel 7840 after filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$	93
192	4.4 Power spectrum in decibels for the current implementation of the low-pass FIR filter in <i>dtp-firmware</i> (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).	94
198	4.5 Relative change in the S/N for the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i> , using different values of the cutoff frequency $f_c$ and the transition width $\delta f$ . The optimal Chebyshev filters were applied using just the integer part of the coefficients given by the Parks-McClellan algorithm.	95
203	4.6 Distribution of the relative change of the S/N on the different wire planes from the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i> after the optimal Chebyshev filter was applied. The filter was computed with the Parks-McClellan algorithm using a cutoff of $f_c = 0.068 \text{ ticks}^{-1}$ and a transition width $\delta f = 0.010 \text{ ticks}^{-1}$ .	96

## LIST OF FIGURES

208	4.7 Left panel: Zoomed match filtered waveform corresponding to channel 7840 209 from the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i> 210 (blue line). The filter used was directly extracted from the data, being 211 the 32 values around the first peak in the original waveform. The green 212 dashed lines mark the region $\pm 3\sigma_{\text{raw}}$ . The resulting noise waveform is 213 also shown (red line). Top right panel: ADC distribution for channel 7840 214 after match filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$ . 215 Bottom right panel: noise ADC distribution for channel 7840 after match 216 filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$ . . . . .	97
217	4.8 Relative improvement in the S/N for the raw data capture <i>felix-2020-07-17-21:31:44</i> , 218 using the matched filter following the parametrisation in Eq. (4.17). The 219 black crosses in both panels denote the location of the maximum ratio value.	100
220	4.9 Left panel: Optimal matched filter coefficients for the U (blue line) and V 221 (red line) planes. The filters were computed with our parametrisation in 222 Eq. (4.17) for the parameter values $\delta = 0.035$ , $\sigma = 0.191$ and $\delta = 0.018$ , 223 $\sigma = 0.191$ respectively. Right panel: Distribution of the relative change of 224 the S/N on the two induction wire planes from the ProtoDUNE-SP raw 225 data capture <i>felix-2020-07-17-21:31:44</i> after their respective optimal 226 matched filters were applied. . . . .	103
227	4.10 Left panel: distributions of the particles track length in the liquid argon 228 for the generated $E_k = 100$ MeV monoenergetic samples, electrons (blue), 229 muons (red), protons (green) and neutral pions (purple). Right panel: 230 distribution of the length of the longest photon in the neutral pion sample 231 after the decay process $\pi^0 \rightarrow \gamma\gamma$ . . . . .	104

## LIST OF FIGURES

## LIST OF FIGURES

257	4.15 Selected consecutive waveforms corresponding to two monoenergetic $E_k =$	
258	100 MeV muon events, one is parallel to the APA and to the wires in	
259	the U plane (left panel) and the other is normal to the APA plane and	
260	perpendicular to the U plane wires (right panel). The solid lines represent	
261	the raw waveforms whereas the dashed lines correspond to the waveforms	
262	after the matched filter was applied. The waveforms on the left panel have	
263	been scaled by a factor of 0.15 to have similar amplitude to the ones on	
264	the right panel. . . . .	111
265	4.16 Left panel: peak asymmetry distribution for the case of the monoenergetic	
266	$E_k = 100$ MeV muon sample. Each value corresponds to a single bipolar	
267	signal peak from a channel in any event. The blue distribution represents	
268	the peaks on U plane channels, whereas the red corresponds to signal peaks	
269	in V wires. Right panel: relation between the mean peak asymmetry per	
270	event with the S/N for U channel waveforms from the $E_k = 100$ MeV	
271	muon sample. The top subplot shows the decimal logarithm of the mean	
272	S/N for the raw (red) and the matched filtered (blue) waveforms. The	
273	bottom subplot contains the mean S/N improvement ratio after the matched	
274	filter was applied. . . . .	112
275	4.17 Raw data display in the plane time (in firmware ticks) vs. offline channel	
276	number for an $E_k = 100$ MeV electron event. The produced true hits are	
277	superimposed (black boxes) as well as the hits comming from the standard	
278	hit finder chain (blue circles) and the hit finder using the matched filter	
279	(green triangles). . . . .	114

## LIST OF FIGURES

- 280    4.18 *Dependence of the precision (blue), sensitivity (red) and  $F_1$  (green) scores*  
281    *on the threshold values used in the hit finder, for the FIR (left panel)*  
282    *and matched filter (right panel) cases. The results were obtained after*  
283    *matching the hits to the true hits in the case of the isotropic muon sample*  
284    *with kinetic energy in the range 5 to 100 MeV, taking only into account*  
285    *the induction plane channels. The points represent the mean value while*  
286    *the error bars indicate one standard deviation around that mean value.* . . 116
- 287    4.19 *Dependence of the averaged hit sensitivity on the kinetic energy of the*  
288    *events for the matched filter (blue) and standard (red) hits, for the case of*  
289    *the muon (left panel) and electron (right panel) samples, separated between*  
290     *$U$  (top plots) and  $V$  (bottom plots) induction wire planes. The top subplots*  
291    *contain the hit sensitivities for the two hit finder alternatives, while the*  
292    *bottom subplots show the ratio between the two. The horizontal lines sit at*  
293    *the mean value and represent the size of the energy bins, while the vertical*  
294    *error bars indicate one standard deviation around that mean value.* . . . 118
- 295    4.20 *Distributions of the hit sensitivity in the  $U$  (top panels) and  $V$  (bottom*  
296    *panels) planes versus the hit sensitivity in the  $X$  plane, both for the*  
297    *standard hits (left panels) and the matched filter hits (right panels), in the*  
298    *case of the electron sample and a threshold of 30 ADC.* . . . . . 119
- 299    4.21 *Top panels: standard residual plots of the hit sensitivities between the  $X$*   
300    *and  $U$  planes. Bottom panels: quantile-quantile plots of the hit sensitivity*  
301    *standard residuals between the  $X$  and  $U$  planes. In all cases, the left*  
302    *panel corresponds to the standard hits while the right panel represents the*  
303    *matched filter case, all from the electron sample with a 30 ADC threshold.* 120

## LIST OF FIGURES

304	5.1 <i>Input solar parameters used in our capture rate computation as functions of the Sun's 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 [1]. . . . .</i>	126
305		
306		
307		
308		
309	5.2 <i>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. [2]. All the rates are shown for a choice of scattering cross section of <math>\sigma_i = 10^{-40} \text{ cm}^2</math>. . .</i>	128
310		
311		
312		
313		
314	5.3 <i>NuWro computed <math>\nu_\mu - {}^{40}\text{Ar}</math> charged-current scattering cross section as a function of the neutrino energy <math>E_\mu</math>. 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). . . . .</i>	131
315		
316		
317		
318		
319	5.4 <i>Expected atmospheric neutrino flux as a function of the neutrino energy <math>E_\nu</math> at Homestake at solar minimum, taken from Ref. [3]. The blue solid (dashed) line correspond to muon neutrinos (antineutrinos) and the red solid (dashed) line correspond to electron neutrinos (antineutrinos). . .</i>	133
320		
321		
322		
323	5.5 Feynman diagrams for $B^1B^1$ annihilation into SM fermions. . . . .	135
324		
325	5.6 Feynman diagrams for $B^1B^1$ annihilation into a Higgs boson pair. . . .	136
326		
327	5.7 <i>Computed spectra of muon neutrinos at the DUNE FD site from <math>B^1</math> annihilations in the Sun for three different values of <math>M_{\text{LKP}}</math>, plotted in relative energy units for legibility. . . . .</i>	137

## LIST OF FIGURES

<p>328    5.8    <i>Projected 90% confidence level upper limit for DUNE (400 kT yr) on the</i>      329    <i>spin-dependent <math>B^1</math>-proton scattering cross section as a function of <math>M_{\text{LKP}}</math></i>      330    <i>(green dots). I also show the previous limits from IceCube [4] (blue line)</i>      331    <i>and Antares [5] (red line) on the LKP cross section. The shaded area</i>      332    <i>represents the disfavoured region (at 95% confidence level) on the mass of</i>      333    <i>the LKP from LHC data [6]. . . . .</i></p>	138
<p>334    5.9    <i>Computed spectra of muon neutrinos at the DUNE FD site from <math>\tau^+\tau^-</math></i>      335    <i>(left panel) and <math>b\bar{b}</math> (right panel) annihilations in the Sun for the DM</i>      336    <i>masses <math>m_{\text{DM}} = 10</math> GeV (red line), 50 GeV (green line) and 100 GeV</i>      337    <i>(blue line), plotted in relative energy units. . . . .</i></p>	139
<p>338    5.10    <i>Distribution of the muon neutrino energies from the <math>\tau^+\tau^-</math> (left panel)</i>      339    <i>and <math>b\bar{b}</math> (right panel) annihilation channels, for <math>m_{\text{DM}} = 10</math> GeV, separated</i>      340    <i>by CC interaction type: QEL (blue), MEC (orange), RES (green) and</i>      341    <i>DIS (red). . . . .</i></p>	140
<p>342    5.11    <i>Distributions of <math>\theta_\mu</math> (left panel), <math>\theta_j</math> (central panel) and <math>\theta_{\text{plane}}</math> (right panel)</i>      343    <i>for the <math>b\bar{b}</math> sample with <math>m_{\text{DM}} = 10</math> GeV (blue) and the atmospheric</i>      344    <i>background (red). . . . .</i></p>	141
<p>345    5.12    <i>Left panel: signal efficiencies (blue lines) and background rejections (red</i>      346    <i>lines) for events passing the cuts <math>\theta &lt; \theta_{\text{cut}}</math> for the jet (solid lines) and</i>      347    <i>muon (dashed lines) angles. Right panel: signal efficiency (blue line) and</i>      348    <i>background rejection (red line) for events passing the cut <math>\theta_{\text{plane}} &lt; \theta_{\text{cut}}</math> for</i>      349    <i>the momentum conservation plane deviation. . . . .</i></p>	143
<p>350    5.13    <i>Signal efficiencies for the <math>\tau^+\tau^-</math> (blue line) and <math>b\bar{b}</math> (red line) DIS samples</i>      351    <i>as functions of the DM mass, <math>m_{\text{DM}}</math>, obtained by applying the optimal</i>      352    <i>angular cuts <math>\theta_\mu &lt; 27^\circ</math>, <math>4^\circ &lt; \theta_j &lt; 26^\circ</math> and <math>\theta_{\text{plane}} &lt; 3.5^\circ</math>. . . . .</i></p>	144
<p>353    5.14    <i>Distributions of <math>\cos \theta_\mu</math> (left panel), <math>\cos \theta_p</math> (central panel) and <math>\cos \theta_N</math></i>      354    <i>(right panel) for the <math>\tau^+\tau^-</math> QEL sample with <math>m_{\text{DM}} = 5</math> GeV (blue) and</i>      355    <i>the atmospheric background (red). . . . .</i></p>	145

## LIST OF FIGURES

356	5.15 Left panel: value of the loss function for the training sample (blue line) and accuracy for the validation sample (red line) versus the number of iterations for the MLP classifier training. Right panel: distributions of the predicted probabilities assigned by the MLP classifier to the test sample for the $\tau^+\tau^-$ QEL signal with $m_{\text{DM}} = 5$ GeV (blue) and the atmospheric background (red). . . . .	146
362	5.16 Signal efficiencies for the $\tau^+\tau^-$ (blue line) and $b\bar{b}$ (red line) single proton QEL samples as functions of the DM mass, $m_{\text{DM}}$ , 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%. . . . .	148
366	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 $m_{\text{DM}}$ , for the annihilation channels $\tau^+\tau^-$ (blue) and $b\bar{b}$ (red) separated by interaction type (up triangles denote DIS interactions whereas down triangles represent QEL interactions). I also show the previous limits from IceCube [7] (solid lines) and the projected sensitivities for Pingu [8] (dashed lines) and Hyper-Kamiokande [9] (dash-dotted lines), as well as the direct detection limits from PICASSO [10] (solid green line) and PICO-60 C <sub>3</sub> F <sub>8</sub> [11] (dashed green line). . . . .	149
375	5.18 Left panel: Projected 90% confidence level sensitivity of DUNE (400 kT yr) to the scale $\Lambda$ of an EFT containing only leptophilic DM axial-axial interactions (blue line). Right panel: . In both cases the corresponding limits from DarkSide-50 [12] (dotted green line) and XENON1T [13] (dashed red line) are also shown, together with the configurations for which the correct relic density is achieved (black line), all for the coupling values $c_A^e = 10^3$ and $c_A^\nu = 10^{-2}$ . . . . .	154
382	6.1 Schematic diagram showing the different modules involved in the ND-GAr production. . . . .	164

## LIST OF FIGURES

384	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
388	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
391	6.4	Distribution of the Geant4-simulated energy losses per unit length versus residual range for the stopping proton sample. . . . .	171
393	6.5	Fitted most probable $dQ/dx$ values for each $dE/dx$ bin, together with best fit to the logarithmic calibration function. . . . .	172
395	6.6	Fitted most probable $dQ/dx$ values for each $dE/dx$ bin for three different ADC bit limits. . . . .	174
397	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. . . . .	175
400	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. . . . .	177
404	6.9	Estimated values of the mean $dE/dx$ bias and resolution for the stopping proton sample at different values of the truncation factor. . . . .	178
406	6.10	Examples of the truncated mean $dE/dx$ LanGauss fits for various $\beta\gamma$ bins, from a simulated FHC neutrino sample. . . . .	180
408	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

## LIST OF FIGURES

411	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
414	6.13 Distribution of the 60% truncated mean $dE/dx$ versus reconstructed momentum for the FHC neutrino sample. . . . .	183
416	6.14 Estimated values of the mean $dE/dx$ bias and resolution obtained for the true protons in a FHC neutrino sample. . . . .	184
418	6.15 True momentum distribution for the primary muon in $\nu_\mu$ CC $N\pi^\pm$ interactions inside the fiducial volume of ND-GAr, compared to the post FSI charged pion spectrum. . . . .	185
421	6.16 Distributions of energy deposits in the ECal for a muon and a charged pion with similar momenta. . . . .	186
423	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 $\chi^2$ -based direction estimator (squares) and cheating the directions (circles), for different values of the $\chi^2$ cut. Right panel: comparison of the performance of the new algorithm when applying the cluster $t_0$ correction (squares) and when (circles). . . . .	188
429	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
433	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

## LIST OF FIGURES

<p>438    6.20 Predicted truncated mean <math>dE/dx</math> versus momentum, for electrons, muons,          439    charged pions and protons, obtained using the ALEPH parametrisation.          440    The vertical dashed lines represent the boundaries of the six regions used          441    for the muon and pion classification training. . . . .</p> <p>442    6.21 Example ECal feature distributions for muons and charged pions in the          443    five different momentum ranges considered. . . . .</p> <p>444    6.22 Example MuID feature distributions for muons and charged pions in the          445    three different momentum ranges considered. . . . .</p> <p>446    6.23 Left panel: cumulative explained variance for the first three principal          447    components (top panel) and contribution of the different features to the          448    principal axes in feature space (bottom panel). Right panel: Shapley          449    (blue) and Gini (red) feature importances for the different input features.          450    Both figures correspond to the samples in the momentum range <math>0.3 \leq</math>          451    <math>p &lt; 0.8 \text{ GeV}/c</math>. . . . .</p> <p>452    6.24 Left panel: cumulative explained variance for the first three principal          453    components (top panel) and contribution of the different features to the          454    principal axes in feature space (bottom panel). Right panel: Shapley          455    (blue) and Gini (red) feature importances for the different input features.          456    Both figures correspond to the samples in the momentum range <math>0.8 \leq</math>          457    <math>p &lt; 1.5 \text{ GeV}/c</math>. . . . .</p> <p>458    6.25 Evolution of the SHAP importance for the top six most important features          459    across all five momentum ranges. . . . .</p> <p>460    6.26 Permutation importances for the ten most important features in the          461    different momentum ranges (from left to right, top to bottom, in increasing          462    momentum order). The bars indicate the effect that permutations of          463    each feature have on the purity (blue) and the sensitivity (yellow), the          464    translucent regions representing one standard deviation around the central          465    value. . . . .</p>	<p>193</p> <p>197</p> <p>198</p> <p>200</p> <p>201</p> <p>203</p> <p>205</p>
--	--

## LIST OF FIGURES

466	6.27 Values of the precision and sensitivity obtained for 10000 BDT hyperparameter	
467	configurations, for the momentum regions I, III and V. . . . .	208
468	6.28 Reliability diagrams for the BDT classifier used in the momentum range	
469	$0.3 \leq p < 0.8 \text{ GeV}/c$ , both for the original (blue circles) and calibrated	
470	(yellow squares) responses. For reference, the response of a perfectly	
471	calibrated classifier is also shown (black dashed line). . . . .	210
472	6.29 Uncalibrated (left panel) and calibrated (right panel) predicted probabilities	
473	assigned by the BDT classifiers for true muons (blue) and charged pions	
474	(red) in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$ . . . . .	211
475	6.30 Schematic of the hit selection used for the ToF measurement. The grid	
476	represents the layers of the inner ECal, with coloured squares indicating	
477	the tiles with hits. Green squares indicate the selected hits. . . . .	213
478	6.31 Particle velocity versus momentum measured with different ECal arrival	
479	time estimations. From left to right: earliest hit time, average hit time,	
480	and fitted hit time. In all cases the time resolution is $\Delta\tau = 0.1 \text{ ns}$ . . . . .	215
481	6.32 Mass spectra for $p$ (blue) and $\pi^\pm$ (yellow) particles, using different ECal	
482	time resolution values (from top to bottom, in ascending order), and	
483	arrival time estimates. From left to right: earliest hit time, average hit	
484	time, and fitted hit time. The dashed lines indicate the true masses of	
485	the particles. . . . .	217
486	6.33 Efficiency (top panel) and purity (bottom panel) for the proton selection	
487	as a function of the momentum, for $\Delta\tau = 0.10 \text{ ns}$ . . . . .	218
488	6.34 Distributions of the velocities measured by ToF with the inner ECal, for	
489	different momentum bins, in a FHC neutrino interaction sample. The	
490	Gaussian fits are performed around the maxima for each particle species.	219

## LIST OF FIGURES

491	6.35 Left panel: number of non-decaying, decaying and decaying in the fiducial volume pions for a MC sample of 100000, $p = 500$ MeV/ $c$ isotropic positively charged pions inside the TPC. Right panel: event display for a positive pion decaying inside the fiducial volume, with a single reconstructed track for the pion and muon system. . . . .	220
496	6.36 Values of $\chi_k^2(FB)$ (top left panel), $F_k$ (top right panel), $D_k^{1/R}$ (bottom left panel) and $D_k^\phi$ (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. . .	221
500	6.37 Fractional residual distributions of the true and reconstructed decay position along the drift coordinate, using the position of the maximum of $\chi_k^2(FB)$ (left panel) and $F_k$ (right panel) as estimates of the decay position. Also shown are double Gaussian fits to these points (red lines). . .	222
504	6.38 Distributions of the extreme values of $\chi_k^2(FB)$ (top left panel), $F_k$ (top right panel), $D_k^{1/R}$ (bottom left panel) and $D_k^\phi$ (bottom right panel) for non-decaying reconstructed pion tracks (blue) and tracks which include the decay inside the fiducial volume (red). . . . .	224
508	6.39 Left panel: distributions of the predicted probabilities assigned by the BDT classifier to a test sample of decaying pion+muon tracks (blue) and non-decaying pion tracks (red). Left: signal efficiency versus background acceptance (ROC curve) obtained from the BDT for the test sample. . .	225
512	6.40 Left panel: dependence of the decay position finding resolution on the true value of the decay angle for the $\chi_k^2(FB)$ (red) and $F_k$ (blue) methods. Right panel: signal efficiency (blue line) and background rejection (red line) from the BDT classifier versus true decay angle. . . . .	225

## LIST OF FIGURES

516	6.41 Mean values of the $F_1$ -score marginal distributions for the different free parameters of the new clustering algorithm, with the error bars representing one standard deviation around the mean. The $F_1$ -score values were computed for the 6561 possible parameter configurations using 1000 $\nu_\mu$ CC interaction events. . . . .	228
517		
518		
519		
520		
521	6.42 Left panel: distributions of the number of ECal clusters per photon from $\pi^0$ decays for the standard (red) and new (blue) clustering algorithms. Right panel: reconstructed invariant mass distributions for photon pairs from single $\pi^0$ events using the standard (red) and new (blue) ECal clustering algorithms. . . . .	230
522		
523		
524		
525		
526	7.1 Schematic diagram of the HPgTPC including the fiducial volume. . . . .	235
527		
528		
529	7.2 True positive, false positive, and false negative true neutrino energy distributions for a $\nu_\mu$ CC selection. . . . .	236
530		
531	7.3 Efficiency and purity for the $\nu_\mu$ CC selection as a function of the muon score cut, FV half-length cut, and radial cut. . . . .	238
532		
533	7.4 Significance for the $\nu_\mu$ CC selection as a function of the muon score cut, FV half-length cut, and radial cut. . . . .	238
534		
535		
536		
537	7.5 Normalised 2D distributions of efficiency, purity and significance for the $\nu_\mu$ 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. . . . .	239
538		
539		
537	7.6 Cumulative efficiency and purity for the $\nu_\mu$ CC selection. . . . .	240
538		
539		
537	7.7 True neutrino energy spectra for the $\nu_\mu$ CC selection. . . . .	241
538		
539		
537	7.8 True neutrino energy spectra for the $\nu_\mu$ CC selection. . . . .	242
538		
539		

## LIST OF FIGURES

540	7.9 True neutrino energy spectra for the $\nu_\mu$ CC selection with (blue) and	
541	without (red) sign selection. The selected events are broken down by true	
542	positives (signal) and false positives (background). The true distribution	
543	is also shown (black data points). The bottom panel shows the ratios	
544	between the number of false positives and total selected events per bin. .	243
545	7.10 Distributions for the reconstructed versus truth generated primary muon	
546	momentum (top left panel), longitudinal momentum (top right panel),	
547	transverse momentum (bottom left panel), and beam angle (bottom right	
548	panel). The reconstructed values correspond to the selected primary muon	
549	candidate, whereas the truth values come from the true primary muon in	
550	the event. . . . .	244
551	7.11 Efficiency and purity of the $\nu_\mu$ CC selection as a function of the primary	
552	muon true momentum and beam angle. . . . .	245
553	7.12 Distributions of the true $\nu_\mu$ CC vertex positions for the full HPgTPC and	
554	the FV. . . . .	246



# List of Tables

555

556	2.1	Values of $T_3$ and $Y/2$ assigned to the first generation of fermions. . . . .	41
557	2.2	Neutral current couplings. . . . .	42
558	2.3	Summary of neutrino oscillation parameters determined in the Neutrino Global Fit of 2020 [14]. . . . .	57
560	3.1	Summary of the two-phased plan for DUNE. . . . .	65
561	3.2	Exposure and time required to achieve the different physics milestones of the two phases. . . . .	65
563	4.1	<i>Characteristic parameters of the two monoenergetic muon events selected, relative to the U plane: projected angles in the <math>xz'</math> and <math>y'z'</math> planes, S/N values for the raw and filtered waveforms, mean improvement of the S/N and peak asymmetry.</i> . . . . .	111
567	5.1	Systematic uncertainties for the solar WIMP signal events. . . . .	156
568	5.2	Systematic uncertainties for the solar WIMP atmospheric background events. . . . .	157
570	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
572	6.2	Momentum ranges and description of the PID approach assumed for the muon and pion classification task. . . . .	194
574	6.3	Optimal values of the hyperparameters used by the BDT, for each momentum range. . . . .	209

## LIST OF TABLES

576	6.4	Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges. . . . .	209
577	6.5	Summary of parameters and sampled values used in the optimisation of the clustering algorithm. . . . .	228
578	7.1	Event rates in ND-GAr. . . . .	234
579	7.2	Step-by-step $\nu_\mu$ CC selection cuts and cumulative passing rates. . . . .	240
580			
581			
582			

## List of Abbreviations

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



1

584

## Introduction

585



# Neutrino physics

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

591

– Terry Pratchett, *Sourcery*

592        Ever since they were postulated in 1930 by Wolfgang Pauli to explain the continuous  
593         $\beta$  decay spectrum [15] and later found by F. Reines and C. Cowan at the Savannah  
594        River reactor in 1953 [16], neutrinos have had a special place among all other elementary  
595        particles. They provide a unique way to probe a wide range of quite different physics,  
596        from nuclear physics to cosmology, from astrophysics to colliders. Moreover, there is  
597        compelling evidence to believe that the study of neutrinos may be key to unveil different  
598        aspects of physics beyond the SM, difficult to test elsewhere.

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

## 602        2.1 Neutrinos in the SM

603        The SM of fundamental interactions was initially proposed in 1967 by S. Glashow, S.  
604        Weinberg and A. Salam[17–19]. This theoretical framework describes the dynamics  
605        of leptons and quarks, by introducing a collection of mediating gauge vector bosons  
606        and one scalar particle, known as the Higgs boson. It assumes that the local  $SU(3) \times$

## CHAPTER 2. NEUTRINO PHYSICS

607 SU(2)<sub>L</sub> × U(1)<sub>Y</sub> gauge symmetry is an internal symmetry of the system, with SU(3)  
 608 describing quantum chromodynamics, and SU(2)<sub>L</sub> × U(1)<sub>Y</sub> being the gauge groups of  
 609 the electroweak sector. For a detailed overview of the SM of electroweak interactions,  
 610 see Ref. [20].

611 In the SM, neutrinos appear in three flavours, namely  $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$ . These are  
 612 associated with the corresponding charged leptons,  $e$ ,  $\mu$ , and  $\tau$ . Neutrinos exist only  
 613 as left-handed particles, grouped in doublets with the charged leptons, while the later  
 614 come in both chirality states:

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

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

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

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

623 Left and right-handed fermions transform differently under SU(2)<sub>L</sub> × U(1)<sub>Y</sub> rotations,  
 624 as the right-handed particles are singlets under SU(2)<sub>L</sub>. Applying a local transformation,  
 625 they change as:

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

626 where  $Y/2$  and  $T_a$  are the generators of SU(2)<sub>L</sub> and U(1)<sub>Y</sub>, 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.

	$e_L$	$\nu_e$	$e_R$	$u_L$	$d_L$	$u_R$	$d_R$
$T_3$	-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

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

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

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

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

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

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

643 where  $W_\mu^i$ ,  $i = 1, 2, 3$  and  $B_\mu$  are the gauge bosons for the  $SU(2)_L$  and  $U(1)_Y$  factors,  
 644 respectively, and  $g$  and  $g'$  are the corresponding gauge couplings. It can be shown that  
 645 these fields transform in the adjoint representation of the gauge group.

## CHAPTER 2. NEUTRINO PHYSICS

**Table 2.2:** Neutral current couplings.

	$u$	$d$	$\nu_e$	$e$
$2v_f$	$1 - \frac{8}{3}\sin^2\theta_W$	$-1 + \frac{4}{3}\sin^2\theta_W$	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  $A_\mu$  is the photon field, and  $Z_\mu$  and  $W_\mu^\pm$  are the neutral and the charged weak boson fields, respectively. The Weinberg angle,  $\theta_W$ , 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

661 with the currents defined as:

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

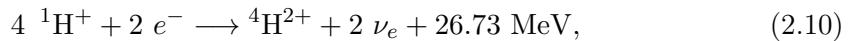
662 where  $f$  denotes any SM fermion,  $\ell$  and  $\nu_\ell$  a charged lepton and a neutrino of any flavour,  
 663 and  $u_f$  and  $d_f$  an up-like and a down-like quarks of any flavour. For the NC case, the  
 664 values of the  $v_f$  and  $a_f$  couplings are given in Tab. 2.2.

665 As seen in Eq. (5.8), in the electroweak theory neutrinos are coupled to the  $Z$  boson  
 666 in a universal way. Therefore, by measuring the so-called invisible decay width of the  $Z$   
 667 boson we have an estimate of the number of light (i.e. lighter than the  $Z$  boson) neutrino  
 668 flavours. This number was measured by LEP in a combined analysis of  $e^+ e^- \rightarrow \mu^+ \mu^-$   
 669 and  $e^+ e^- \rightarrow$  hadrons to be  $N_\nu = 2.9840 \pm 0.0082$  [21].

## 670 2.2 Trouble in the neutrino sector

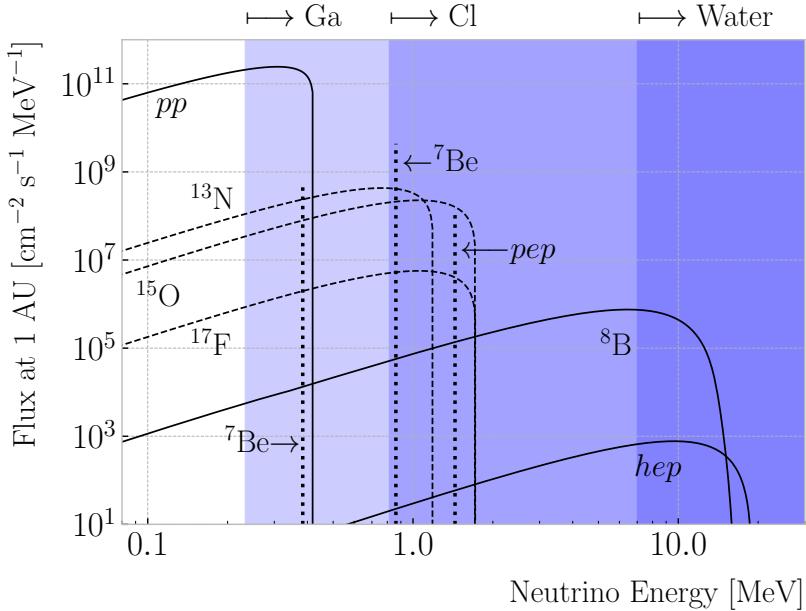
### 671 2.2.1 The solar neutrino problem

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



675 where part of the released energy is lost to the neutrinos. The electron neutrinos  
 676 produced are often labelled after the processes that generate them. Figure 2.1 shows the  
 677 solar neutrino flux as a function of the neutrino energy, broken down by the production  
 678 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. [1].

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

680 data taking with the goal of measuring the solar neutrino flux [22]. The experiment

681 used a tank containing  $380 \text{ m}^3$  of tetrachloroethene ( $\text{C}_2\text{Cl}_4$ ), a liquid commonly used

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

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



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

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

686 ground state transition.

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

688 J. Bahcall [23]. During its operation from 1968 to 2002, the experiment observed a solar

689  $\nu_e$  flux that was approximately a third of the total prediction [24].

690 In the early 1990s, the SAGE [25] and GALLEX [26] experiments started operations.

## 2.2. TROUBLE IN THE NEUTRINO SECTOR

691 The detection principle used for both experiments was similar to that of the Homestake  
 692 experiment, but using  $^{71}\text{Ga}$  instead of  $\text{C}_2\text{Cl}_4$ . With a detection threshold of 0.233 MeV,  
 693 the Gallium-based experiments were able to observe the  $pp$  neutrino flux. Both  
 694 experiments measured a solar electron neutrino flux that was a factor of two lower  
 695 than the predictions, demonstrating that this deficit was energy-dependent.

696 In the early 2000s, the SNO experiment put an end to the solar neutrino puzzle  
 697 [27, 28]. Thanks to its directionality capabilities, being a Cherenkov light detector, as  
 698 well as to its heavy water target, SNO measured the total solar neutrino flux through  
 699 the NC process:



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



702 they were able to establish that the  $\nu_\mu$  and  $\nu_\tau$  solar fluxes are in fact non-zero, revealing  
 703 that electron neutrinos were transitioning into different flavours.

### 704 2.2.2 The atmospheric neutrino problem

705 When cosmic-rays interact with the atoms in the upper atmosphere, a plethora of  
 706 hadrons, mainly  $\pi$  and  $K$  mesons, are produced. In particular, for the charged pions,  
 707 we have the following decay chain dominates:

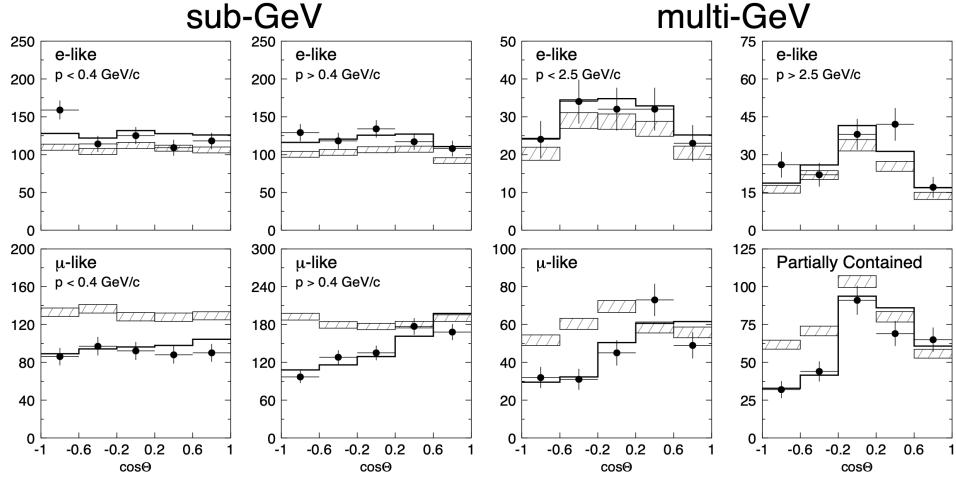


708 and similar for the antiparticles. For neutrino energies  $< 1$  GeV, the ratio:

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

709 of produced neutrinos and antineutrinos is, in good approximation, equal to two [29].

## CHAPTER 2. NEUTRINO PHYSICS



**Figure 2.2:** Zenith angle distributions for the selected  $\nu_e$  (top row) and  $\nu_\mu$  (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. [34].

During the 1980s, several proton decay experiments, like Kamiokande [30], IMB [31],

MACRO [32], and Soudan-2 [33], measured the flux of atmospheric neutrinos. This was an important part of their research programme, as the atmospheric neutrinos constitute their main background. All these experiments reported an atmospheric neutrino ratio lower than the predictions.

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

measured the atmospheric  $\nu_e$  and  $\nu_\mu$  spectra as a function of the zenith angle [34]. Upward-going particles have negative zenith angle,  $\cos \Theta < 0$ , indicating that they entered from the bottom of the detector. These upward-going neutrinos had to travel through the Earth in order to reach the detector, allowing SK to probe a broad range of baselines. Figure 2.2 shows the reported distributions (black dots), compared to the no oscillations prediction (hatched region). This measurement confirmed that muon neutrinos transition to other flavours, and that this phenomenon depends both on the energy and the path length of the neutrino.

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

oscillations, and therefore non-zero neutrino masses. This constitutes one of the groundbreaking discoveries of modern physics and has acted as driving force for beyond

## 2.3. MASSIVE NEUTRINOS

the Standard Model (BSM) physics. The minimal extension of the SM we can do to address these phenomena is introducing different masses for at least two of the neutrinos. This way, we are left with three neutrino mass eigenstates  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ , with masses  $m_1$ ,  $m_2$ , and  $m_3$  respectively, which in general will not coincide with the flavour eigenstates,  $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$ .

### 2.3 Massive neutrinos

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

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

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

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

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

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

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

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

752 with:

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

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

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

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

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

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

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

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

762 where the states  $\nu_{Mi}$ , commonly referred to as Majorana neutrinos, are defined as:

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

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

## 2.4. NEUTRINO OSCILLATION FORMALISM

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

768 If the eigenvalues of the Majorana mass matrix,  $M_N$ , are much larger than the  
 769 electroweak symmetry breaking scale, the diagonalisation of  $M_\nu$  leads to 3 light and  $m$   
 770 heavy neutrino states:

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

771 where the two mass matrices are given by:

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

772 with  $V_l$  and  $V_h$  two unitary matrices.

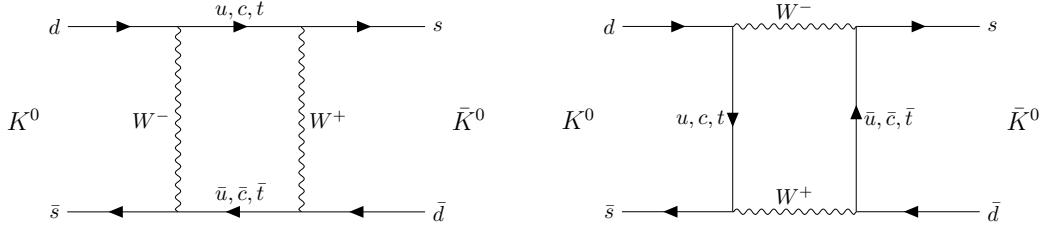
773 This scenario represents the so-called see-saw mechanism [35–39]. The name comes  
 774 from the fact that the masses of the heavy states are proportional to  $M_N$ , whereas for  
 775 the light states they are proportional to  $M_N^{-1}$ . While both the heavy and the light  
 776 neutrinos are Majorana particles, it can be shown that the heavy states are mainly  
 777 right-handed, whereas the light ones are mostly left-handed.

## 778 2.4 Neutrino oscillation formalism

779 Neutrino oscillations were first proposed in 1958 by B. Pontecorvo [40], inspired by the  
 780 neutral kaon oscillation phenomenon [41]. Neutral kaons,  $K^0$  and  $\bar{K}^0$ , have opposite  
 781 strangeness ( $\pm 1$ ) and are produced in strong processes. It was observed that, when  
 782 having a beam initially pure of neutral kaons of one type, these would transition into  
 783 their antiparticles while propagating. Because the weak interaction does not conserve  
 784 strangeness, neutral kaons can change their identity via the processes shown in Fig. 2.3.

785 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.

786 antineutrino states, as only one neutrino flavour was known at the time. After the  
 787 discovery of the muon neutrino, the mixing between flavours was also explored [42].

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

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

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

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

803 Considering the extended SM without any additional sterile neutrino states, the  
 804 resulting  $3 \times 3$  mixing matrix is unitary. This matrix, often called the Pontecorvo-Maki-

## 2.4. NEUTRINO OSCILLATION FORMALISM

805 Nakagawa-Sakata (PMNS) matrix [43, 44], relates the set of active neutrinos and the  
 806 three mass eigenstates as:

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

807 where the Greek index  $\alpha$  denotes the flavour  $\{e, \mu, \tau\}$  and the Latin index  $i$  the mass state  
 808  $\{1, 2, 3\}$ . This leptonic mixing matrix may be parametrized in terms of 6 parameters,  
 809 of which are mixing angles  $\theta_{12}$ ,  $\theta_{13}$  and  $\theta_{23}$ , one CP-violating phase  $\delta_{CP}$  and 2 Majorana  
 810 phases  $\alpha$  and  $\beta$ :

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

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

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

### 820 2.4.1 Oscillations in vacuum

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

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

823 in the plane wave approximation, as the mass eigenstates are also eigenstates of the free  
 824 Hamiltonian.

## CHAPTER 2. NEUTRINO PHYSICS

825 This way, the probability for the neutrino to transition from flavour  $\alpha$  to flavour  $\beta$   
 826 will be given by:

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

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

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

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

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

832 where  $\Delta m_{ij}^2$  is the difference of the squared masses of the  $j$ th and  $i$ th neutrino mass  
 833 eigenvalues. At this point, it is usual to write the phase responsible for the oscillations  
 834 as:

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

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

## 2.4. NEUTRINO OSCILLATION FORMALISM

838 CP symmetry violation:

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

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

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

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

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

842 where the sum includes all flavours, including  $\alpha$ . From these two constraints, one can  
 843 probe that:

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

844 and in particular:

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

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

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

851 The Jarlskog invariant can be used to compare the amount of CP-violation in the lepton  
 852 and the quark sectors, where  $J = 3.12^{+0.13}_{-0.12} \times 10^{-5}$  in the latter [45].

## CHAPTER 2. NEUTRINO PHYSICS

### 2.4.2 Oscillations in matter

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

The first proposed model to account for neutrino oscillations in matter was proposed by Mikhaev, Smirnov and Wolfenstein (MSW) [46]. It relies on the fact that, as the only charged lepton present in ordinary matter is the electron, electron neutrinos can undergo both charged and neutral-current interactions with matter whereas for muon and tau neutrinos just neutral currents are possible.

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

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

with a vacuum Hamiltonian given by:

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

where  $\Delta m^2$  is the mass splitting between the two neutrino states and  $\theta$  the only mixing angle. For simplicity, I omit the terms of the Hamiltonian that are proportional to the identity, as they do not affect the oscillation phenomenology.

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

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

## 2.4. NEUTRINO OSCILLATION FORMALISM

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

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

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

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

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

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

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

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

886 with:

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

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

## CHAPTER 2. NEUTRINO PHYSICS

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

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

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

### 894 2.4.3 Current status of neutrino oscillations

895 A wide range of neutrino experiments provide experimental input to the neutrino  
896 oscillation framework, both using natural or synthetic neutrino sources. The results  
897 from one of the neutrino global fit analyses, shown in Tab. 2.3 <sup>1</sup>, summarise well our  
898 current understanding of the different oscillation parameters.

899 **Solar neutrino experiments** detect neutrinos produced in thermonuclear reactions  
900 inside the Sun, mainly from the so-called *pp* chain and the CNO cycle. These neutrinos  
901 have a typical energy in the range from 0.1 to 20 MeV. These experiments (Homestake  
902 [47], GALLEX [26], SAGE [25], Borexino [48], Super-Kamiokande [49] and SNO [50])  
903 provide the best sensitivities to  $\theta_{12}$  and  $\Delta m_{21}^2$ .

904 **Atmospheric neutrino experiments** detect the neutrino flux produced when  
905 cosmic rays scatter with particles in Earth's atmosphere. These collisions generate particle  
906 showers that eventually produce electron and muon neutrinos (and antineutrinos). Their  
907 energies range from few MeV to about  $10^9$  GeV. Experiments, like Super-Kamiokande  
908 [51] and IceCube [52] use atmospheric neutrinos to measure oscillations and are specially  
909 sensitive to  $\theta_{23}$  and  $\Delta m_{32}^2$ .

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

---

<sup>1</sup>These 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 [14].

Parameter	Best fit $\pm 1\sigma$	$3\sigma$ range
$\Delta m_{21}^2$ [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
$\sin^2 \theta_{12}/10^{-1}$	$3.04 \pm 0.16$	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
$\delta_{CP}/\pi$ (NO)	$1.12^{+0.16}_{-0.12}$	0.76 – 2.00
$\delta_{CP}/\pi$ (IO)	$1.50^{+0.13}_{-0.14}$	1.11 – 1.87

long-baseline experiments like KamLAND [53] are sensitive to the solar mass splitting  
 $\Delta m_{21}^2$  whereas much shorter baseline experiment such as RENO [54] or DayaBay [55]  
measure  $\theta_{13}$  and  $\Delta m_{31}^2$ .

**Accelerator experiments** measure neutrino fluxes generated in particle accelerators.  
Usually mesons are produced in the accelerator to be focused into a beam, then some  
decay to muon neutrinos and the rest are absorbed by a target. Depending on the  
configuration one can obtain a beam made of mostly neutrinos or antineutrinos. The  
typical energies of these neutrinos are in the GeV range. Experiments such as NOvA  
[56], T2K [57], MINOS [58], OPERA [59] and K2K [60] (and in the future DUNE [61])  
are primarily sensitive to  $\theta_{13}$ ,  $\theta_{23}$  and  $\Delta m_{32}^2$ . Also, in the coming years DUNE [61] and  
Hyper-Kamiokande [62] will be sensitive to  $\delta_{CP}$ .

## 2.5 Open questions in the neutrino sector

A crucial question that remains open these days, and is of vital importance for oscillation  
phenomena, is whether the mass eigenvalue  $\nu_3$  is the heaviest (what we call normal  
ordering) or the lightest (referred to as inverted ordering) of the mass eigenstates. In

## CHAPTER 2. NEUTRINO PHYSICS

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

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

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

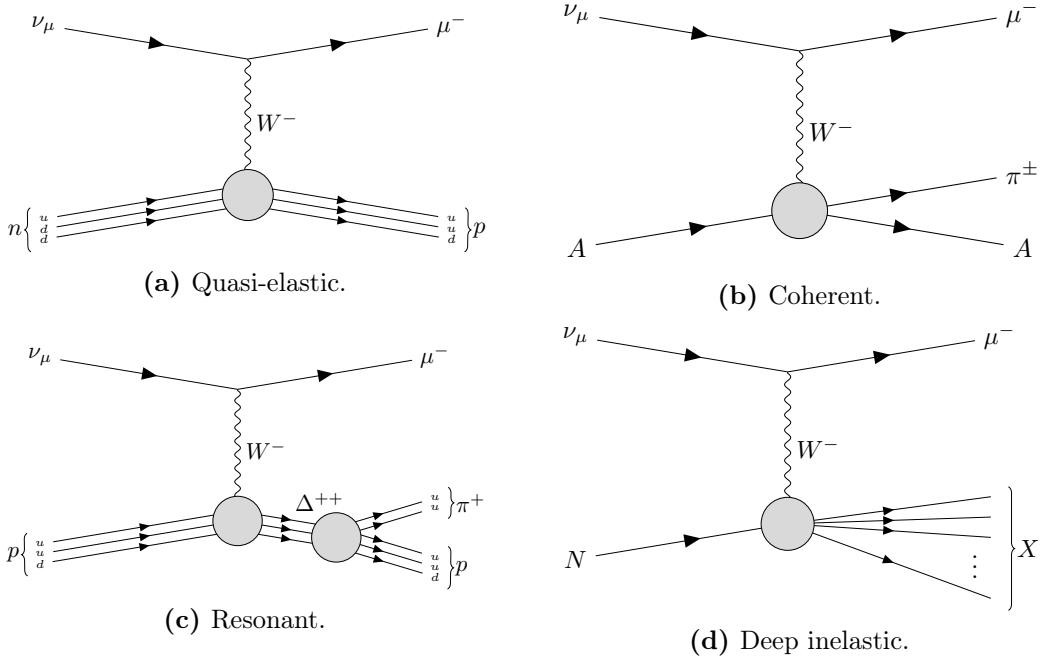
Notwithstanding, there are other mysteries that can not be unveiled just by conducting oscillation experiments, as certain quantities do not influence these phenomena. Among these there is the question of the absolute values of the neutrino masses. Depending on the value of the lightest of the neutrino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

Other open question concerns the nature itself of the neutrinos. If neutrinos are Dirac particles then their mass term can be generated through the usual Higgs mechanism by adding right-handed neutrino fields. However, if they are Majorana particles and therefore their own antiparticles, there is no need to add extra fields to have the mass term in the Lagrangian. Experiments like SuperNEMO [63], SNO+ [64] and NEXT [65], which search for neutrino-less double beta decay, will be able to determine whether neutrinos are Dirac or Majorana.

## 2.6 Neutrino interactions

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

## 2.6. NEUTRINO INTERACTIONS

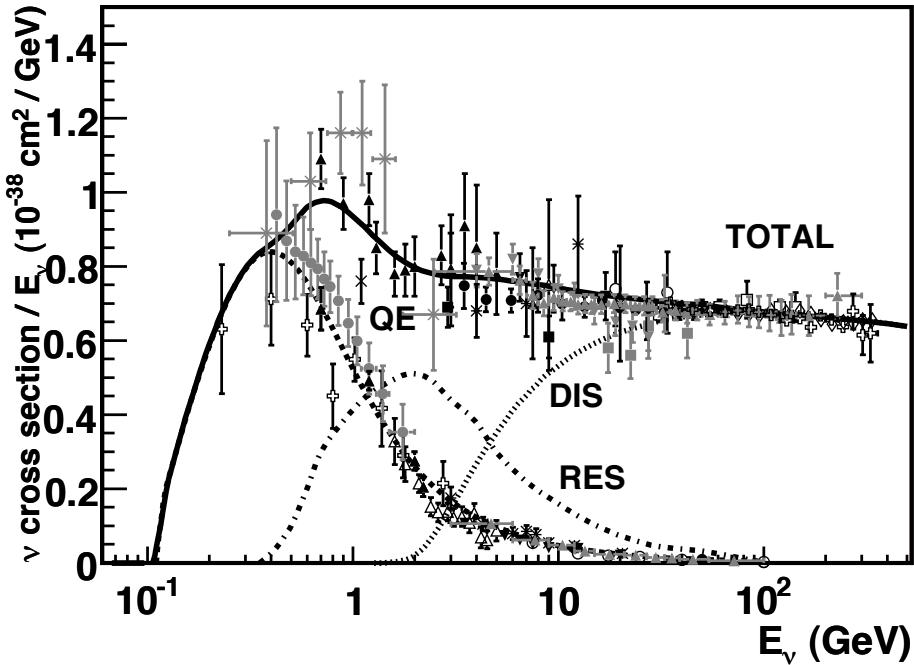


**Figure 2.4:** Feynman diagrams of the four most relevant CC interactions to long baseline neutrino oscillation experiments.

the energy of the incoming neutrino and the final state particles after the interaction.  
 Because in this kind of experiments neutrinos are obtained as secondary decay products of mesons, typically charged pions and kaons, their energies are not known a priori. Not only that, but the kinematics of the interacting nucleon are also unknown. Therefore, we rely on the neutrino interaction models to provide this relation between the observables in the detector and the true kinematics of the neutrino. Interaction modelling is expected to be the one of the leading sources of systematic uncertainties in the next generation of long baseline experiments [66–68].

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

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

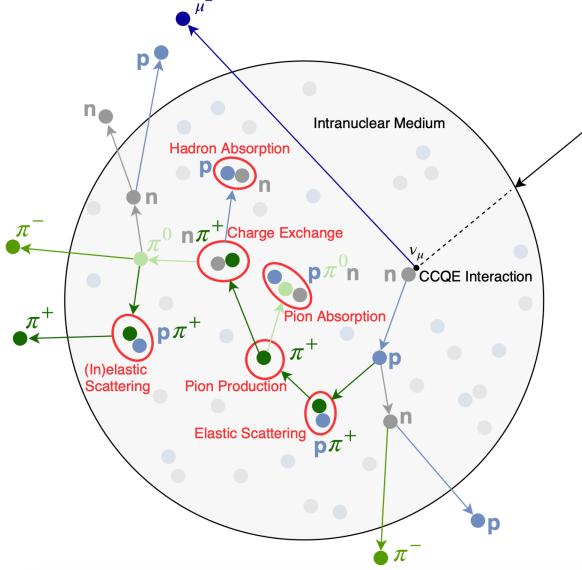


**Figure 2.5:** Total  $\nu_\mu$  CC cross section per nucleon as a function of the neutrino energy. The contributions from the different channels are shown. Figure taken from Ref. [69].

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

Figure 2.5 shows a compilation of measurements of the total  $\nu_\mu$  CC cross section (see Ref. [69] for the details of the different experimental results). Also shown are the contributions from the different interaction modes. The contribution of the CCCOH interaction is omitted, as it is negligible compared to the others. This shows how the

## 2.6. NEUTRINO INTERACTIONS



**Figure 2.6:** Schematic representation of a  $\nu_\mu$  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. [70].

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

Nuclear effects alter the neutrino cross section, as well as the multiplicities of the 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 [71], spectral functions [72] or the random phase approximation [73]. 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 [74]. Figure 2.6 illustrates the effects of FSI on the observable particle content in the detector after a  $\nu_\mu$  CCQE interaction.

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

## CHAPTER 2. NEUTRINO PHYSICS

cross sections. The list of such experiments in the recent years include MiniBooNE [75], MINERvA [76], MicroBooNE [77] and SBND [78]. Additionally, thanks to their near detectors, long baseline experiments can perform cross section measurements. Some recent examples are NOvA [79] or T2K [80]. 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 [81, 82].

# The Deep Underground Neutrino Experiment

1010 The Deep Underground Neutrino Experiment (DUNE) is a next generation long-baseline  
1011 neutrino experiment [83]. It will aim to address several questions in neutrino physics,  
1012 study neutrinos from astrophysical sources and search for beyond the standard model  
1013 physics.

1014 This chapter reviews the main goals of the DUNE experiment, the design of the far  
1015 detector modules and their data acquisition (DAQ) system, and the role that the near  
1016 detector plays in the physics program of DUNE.

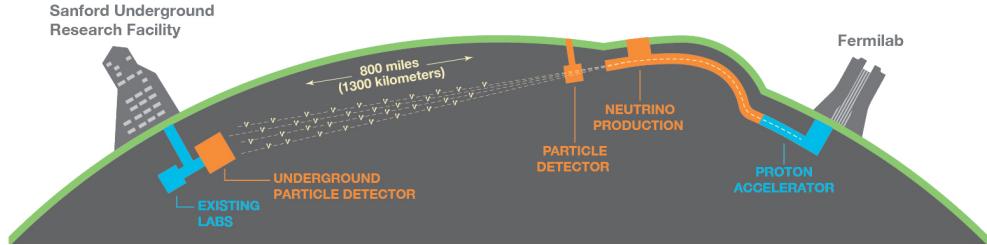
## 1017 3.1 Overview

1018 The main physics goals of DUNE are:

- 1019 • measure the neutrino mass hierarchy, the amount of CP violation in the leptonic  
1020 sector and the  $\theta_{23}$  octant,
- 1021 • detect rare low energy neutrino events, like neutrinos from supernova bursts, and
- 1022 • search for proton decay and other beyond the standard model phenomena.

1023 The design of DUNE has been tailored with these goals in mind. It will consist  
1024 of two neutrino detectors. A near detector (ND) complex will be placed at Fermilab,  
1025 574 m downstream of the neutrino production point, whereas a larger far detector

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT



**Figure 3.1:** Schematic diagram of the DUNE experiment and the LBNF beamline [83].

1026 (FD) will be built in the Sandford Underground Research Facility (SURF), South  
 1027 Dakota, approximately 1300 km away. Figure 3.1 shows a simplified view of the various  
 1028 components of DUNE (not to scale).

1029 The beam neutrinos will be provided by the Long-Baseline Neutrino Facility (LBNF)  
 1030 beamline, the multi-megawatt wide-band neutrino beam planned for Fermilab. It will  
 1031 produce neutrinos travelling in the direction of SURF, with the capability to switch  
 1032 between neutrino and antineutrino mode.

1033 Before arriving to the FD, the neutrino beam meets the ND complex, which serves  
 1034 as the experiment's control. The design of the DUNE ND is mainly driven by the  
 1035 needs of the oscillation physics program, as its main role is to measure the unoscillated  
 1036 neutrino energy spectra. From these we can predict the unoscillated spectra at the FD,  
 1037 which can be compared to the spectra measured at the FD to extract the oscillation  
 1038 parameters. Additionally, the ND has a physics programme of its own, including cross  
 1039 section measurements and BSM physics searches.

1040 The technology chosen for the FD modules of DUNE is the liquid Argon time  
 1041 projection chamber (LArTPC). Its four modules will record neutrino interactions from  
 1042 the accelerator-produced beam arriving at predictable times. As it also aims at recording  
 1043 rare and low energy events, like supernovae and solar neutrinos, the FD requires trigger  
 1044 schemes which can deal with both kinds of physics, and also maximum uptime.

1045 DUNE is planned to be built using a staged approach consisting on two phases,  
 1046 which are summarised in Tab. 3.1. Phase I consists of a FD with 50% of the total

### 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

**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	$5\sigma$ MO ( $\delta_{CP} = -\pi/2$ )	16	1-2
	$5\sigma$ MO (100% of the $\delta_{CP}$ values)	66	3-5
	$3\sigma$ CPV ( $\delta_{CP} = -\pi/2$ )	100	4-6
Phase II	$5\sigma$ CPV ( $\delta_{CP} = -\pi/2$ )	334	7-8
	$\delta_{CP}$ resolution of 10 degrees ( $\delta_{CP} = 0$ )	400	8-9
	$5\sigma$ CPV (50% of the $\delta_{CP}$ values)	646	11
	$3\sigma$ CPV (75% of the $\delta_{CP}$ values)	936	14
	$\sin^2(2\theta_{13})$ resolution of 0.004	1079	16

1047 fiducial mass, a reduced version of the ND complex and a 1.2 MW proton beam. It will  
 1048 be sufficient to achieve some early physics goals, like the determination of the neutrino  
 1049 mass ordering. For its Phase II, DUNE will feature the full four FD modules, a more  
 1050 capable ND and a 2.4 MW proton beam. The physics milestones for the two phases are  
 1051 given in Tab. 3.2, in a staging scenario which assumes that Phase II is completed after  
 1052 6 years of operation.

1053 A summary of the DUNE science program can be found in the DUNE FD Technical  
 1054 Design Report (TDR) Volume I [83]. For a detailed discussion on the two-phased  
 1055 approach the reader is referred to the DUNE Snowmass 2021 report [84].

## 1056 3.2 Physics goals of DUNE

1057 As noted in the literature (see for instance Ref. [14] for a review), the parameter space of  
 1058 the neutrino oscillation phenomena within the three-flavour picture is quite constrained

### CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

1059 by current experimental data. However, there are still crucial open questions, like the  
1060 mass ordering, the value of  $\delta_{CP}$  or the  $\theta_{23}$  octant. One of the main goals of DUNE is to  
1061 determine precisely the values of these parameters [85].

1062 To address these questions DUNE can look to the subdominant oscillation channel  
1063  $\nu_\mu \rightarrow \nu_e$  ( $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ) and study the energy dependence of the  $\nu_e$  ( $\bar{\nu}_e$ ) appearance probability.  
1064 When we focus on the antineutrino channel  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  there is a change in the sign of  $\delta_{CP}$ ,  
1065 thus introducing CP-violation. Moreover, due to the fact that there are no positrons in  
1066 the composition of Earth, there is a sign difference for the matter effect contribution  
1067 when looking to the antineutrino channel. This asymmetry is proportional to the baseline  
1068 length  $L$  and is sensitive to the sign of  $\Delta m_{31}^2$ , and thus to the neutrino mass ordering.

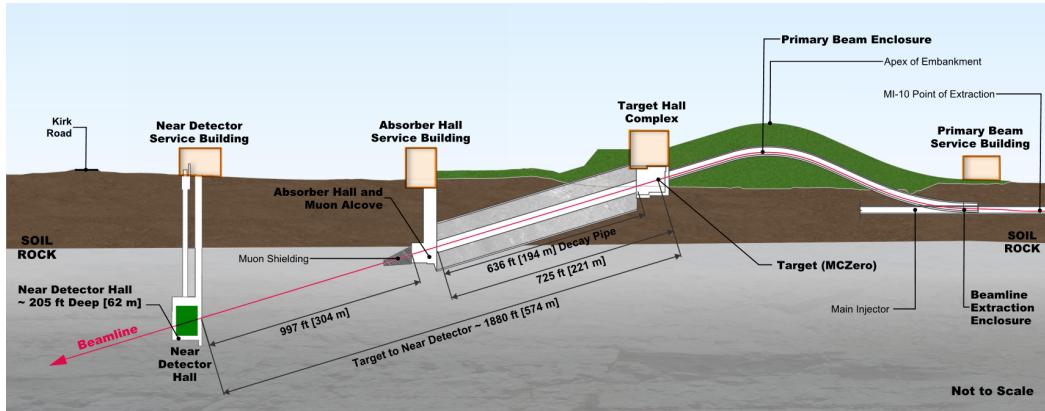
1069 Another of the main physics goals of DUNE is the search for baryon-number violating  
1070 processes. Specifically, it will try to answer the question of whether protons are stable  
1071 or not. There is no symmetry argument that forbids protons from decaying, but its  
1072 apparent stability seems to suggest that baryon number is conserved [86]. However,  
1073 proton decay is a usual feature of grand-unified theories, where electromagnetic, weak  
1074 and strong interactions are unified above a certain energy scale [87].

1075 As the energy deposition scale for this kind of searches is nearly the same as the one  
1076 for long-baseline neutrino oscillations, DUNE will be able to look for them. It has several  
1077 advantages over other experiments, such as excellent imaging and particle identification,  
1078 which can be translated to lower backgrounds.

1079 The last of the main objectives of DUNE is the detection of neutrinos originated in  
1080 supernovae explosions, what is called a supernova neutrino burst (SNB). These neutrinos  
1081 carry with them information about the core-collapse process, from the progenitor to the  
1082 explosion and the remnant; but also may have information about new exotic physics. So  
1083 far, the only neutrino events ever recorded from such a process were a few dozens of  $\bar{\nu}_e$   
1084 events from the 1987A supernova located in the Magellanic Cloud, 50 kpc away from  
1085 Earth [88, 89].

1086 DUNE aims to collect SNB events. Although these are quite rare, as the expected  
1087 supernovae explosion events are about one every few decades for our galaxy and

### 3.3. LBNF BEAMLINE



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

1088 Andromeda, the long lifetime of the experiment (around a few decades as well) makes it  
 1089 reasonable to expect some. Nowadays the main sensitivity to SNB of most experiments  
 1090 is to the  $\bar{\nu}_e$  through inverse beta decay. One of the advantages of DUNE is its expected  
 1091 sensitivity to  $\nu_e$ , since the dominant channel will be  $\nu_e$  CC scattering.

1092 Moreover, due to the stringent requirements that the main physics goals set for  
 1093 DUNE, it will allow also to perform searches for all kind of BSM physics. Among  
 1094 others, DUNE will be able to look for: active-sterile neutrino mixing, non-unitarity of  
 1095 the PMNS matrix, non-standard interactions, Lorentz and CPT violations, neutrino  
 1096 trident production, light-mass DM, boosted DM and heavy neutral leptons. The reader  
 1097 is referred to the DUNE FD TDR Volume II [85] for a full discussion of the physics  
 1098 scope of DUNE.

### 1099 3.3 LBNF beamline

1100 The LBNF project is responsible for producing the neutrino beam for the DUNE detectors.  
 1101 A detailed discussion of the LBNF program can be found in the DUNE/LBNF CDR  
 1102 Volume III [90].

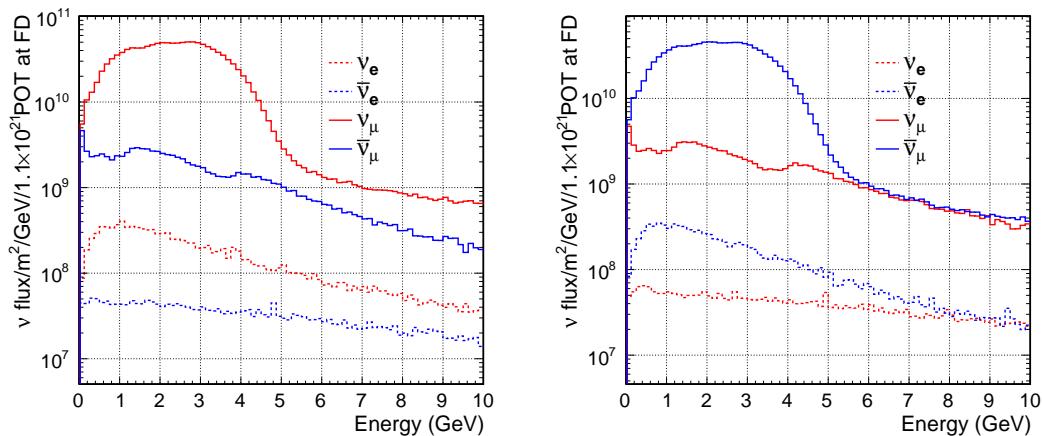
1103 A schematic diagram of the longitudinal section of the LBNF beamline is shown in  
 1104 Fig. 3.2. First, a beam of 60 – 120 GeV protons is extracted from the Fermilab Main  
 1105 Injector. This beam is aimed towards the target area, where it collides with a cylindrical

### CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

graphite target to produce pions and kaons.

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

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

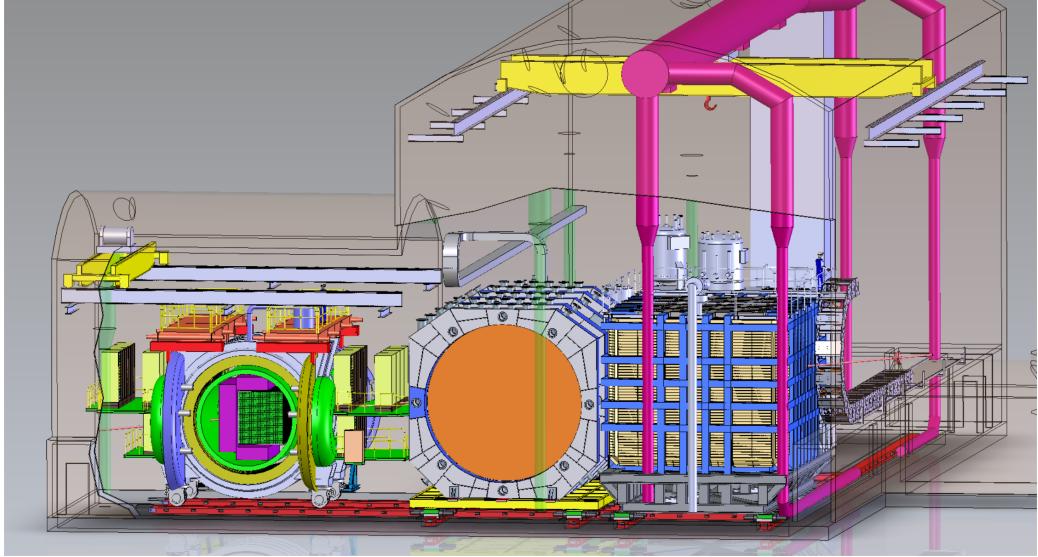


**Figure 3.3:** Predicted neutrino fluxes at the FD in FHC mode (left panel) and RHC mode (right panel). Figures taken from Ref. [85].

### 3.4 Near Detector

To estimate the oscillation parameters we measure the neutrino energy spectra at the FD. This reconstructed energy arises from a convolution of the neutrino flux, cross section, detector response and the oscillation probability. Using theoretical and empirical models

### 3.4. NEAR DETECTOR



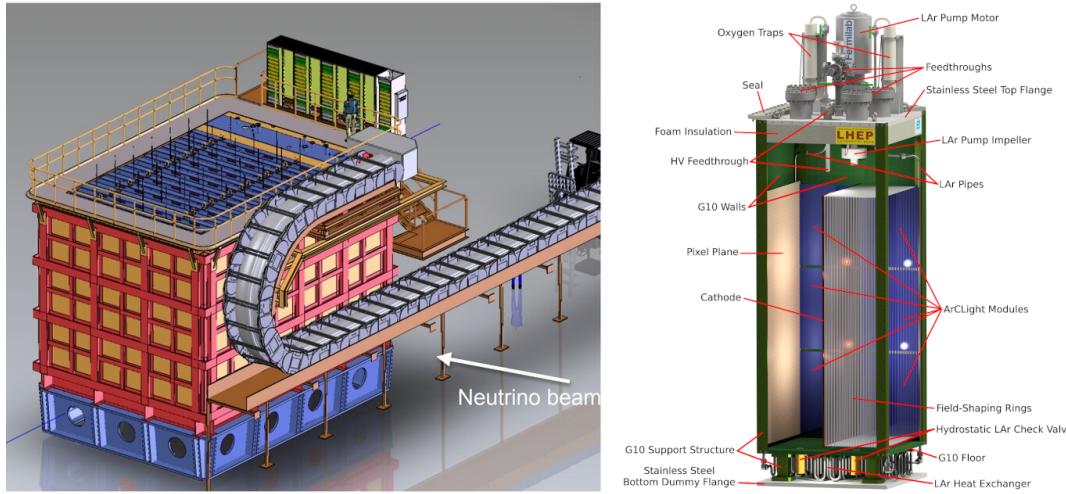
**Figure 3.4:** Representation of the ND hall in Phase II, showing the different subcomponents. From right to left, in the direction of the beam, we have ND-LAr, ND-GAr and SAND. Figure taken from Ref. [91].

to account for the other effects, one can extract the oscillation probability using the measurement. However, these models have associated a number of uncertainties that are then propagated to the oscillation parameters.

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

Nevertheless, having a highly capable ND, DUNE can minimise the systematic uncertainties affecting the observed neutrino energy. The ND data can be used to tune the model parameters by comparison with the prediction. Then, one uses the

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT



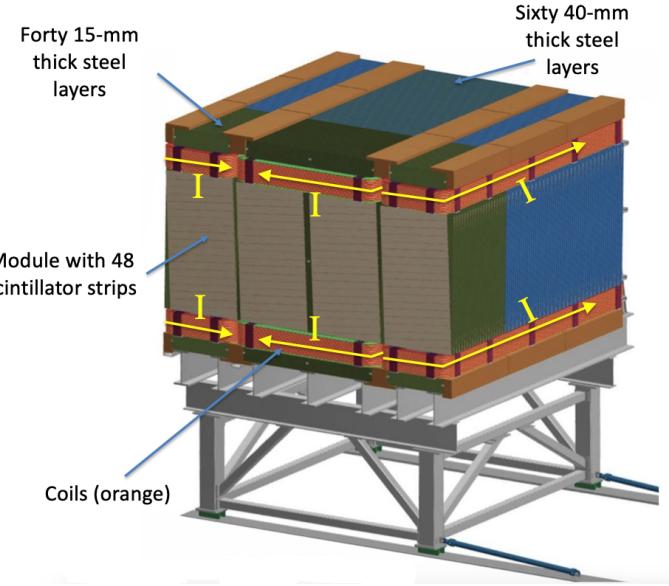
**Figure 3.5:** Schematic representation of the external components of ND-LAr, including the cryostat and the PRISM movable system (left) and detailed drawing of one ArgonCube module (right). Figure adapted from Ref. [83].

1139 tuned model to predict the unoscillated FD spectra. Comparing the prediction with the  
 1140 measured spectra it is possible to extract the oscillation parameters.

1141 Additionally, the ND will have a physics program of its own. In particular, it will  
 1142 measure neutrino cross sections that will then be used to constrain the model used in  
 1143 the long-baseline oscillation analysis. It will also be used to search for BSM phenomena  
 1144 such as heavy neutral leptons, dark photons, millicharged particles, etc.

1145 The DUNE ND can be divided in three main components, a LArTPC known as ND-  
 1146 LAr, a magnetised muon spectrometer, which will be the Temporary Muon Spectrometer  
 1147 (TMS) in Phase I and ND-GAr in Phase II, and the System for on-Axis Neutrino  
 1148 Detection (SAND). The layout of the Phase II DUNE ND can be seen in Fig. 3.4. The  
 1149 first two components of the ND will be able to move off-axis, in what is called the  
 1150 Precision Reaction-Independent Spectrum Measurement (PRISM) concept. More details  
 1151 on the purpose and design of the ND can be found in the DUNE ND Conceptual Design  
 1152 Report (CDR) [91].

### 3.4. NEAR DETECTOR



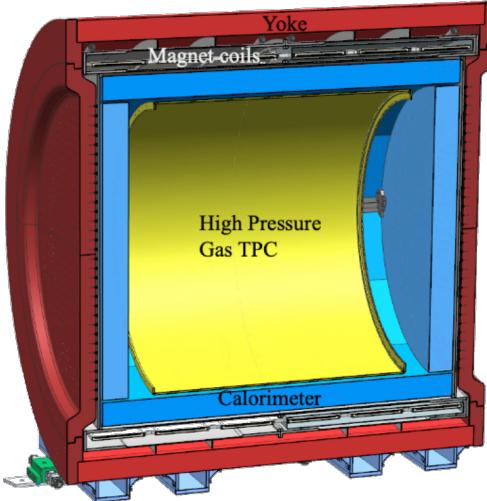
**Figure 3.6:** Schematic view of the TMS detector, highlighting its main parts. Figure adapted from Ref. [83].

#### 1153 3.4.1 ND-LAr

1154 ND-LAr is a LArTPC, as the ND needs a LAr component to reduce cross section and  
1155 detector systematic uncertainties in the oscillation analysis. However, its design differs  
1156 significantly from those proposed for the FD modules. Because of the high event rates  
1157 at the ND, approximately 55 neutrino interaction events per  $10 \mu\text{s}$  spill, ND-LAr will be  
1158 built in a modular way. Each of the modules, based on the ArgonCube technology, is a  
1159 fully instrumented, optically isolated TPC with a pixelated readout. The pixelisation  
1160 allows for a fully 3D reconstruction and the optical isolation reduces the problems due  
1161 to overlapping interactions. Figure 3.5 shows a representation of the external parts of  
1162 ND-LAr (left) and a detailed diagram of an ArgonCube module (right).

1163 With a fiducial mass of 67 t and dimensions  $7 \text{ m} (\text{w}) \times 3 \text{ m} (\text{h}) \times 5 \text{ m} (\text{l})$ , ND-LAr  
1164 will be able to provide high statistics and contain the hadronic systems from the beam  
1165 neutrino interactions, but muons with a momentum higher than 0.7 GeV will exit the  
1166 detector.

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT



**Figure 3.7:** Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet. Figure adapted from Ref. [92].

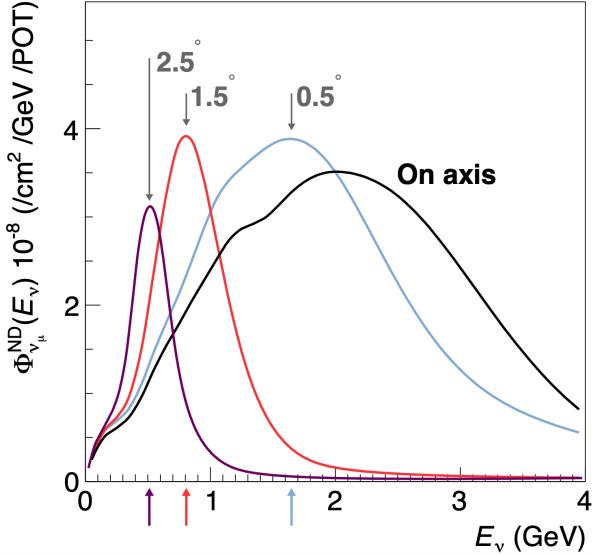
### <sup>1167</sup> 3.4.2 TMS/ND-GAr

<sup>1168</sup> To accurately estimate the neutrino energy, the momentum of the outgoing muons needs  
<sup>1169</sup> to be determined. That is the reason why a muon spectrometer is needed downstream  
<sup>1170</sup> of ND-LAr.

<sup>1171</sup> In Phase I that role will be fulfilled by TMS. It is a magnetised sampling calorimeter,  
<sup>1172</sup> with alternating steel and plastic scintillator layers. Figure 3.6 shows a schematic view  
<sup>1173</sup> of the TMS detector. The magnetic field allows a precise measurement of the sign of the  
<sup>1174</sup> muon, so one can distinguish between neutrino and antineutrino interactions.

<sup>1175</sup> After the Phase II upgrade, TMS will be replaced with a more capable near detector.  
<sup>1176</sup> The current technology considered is ND-GAr. This detector is a magnetised, high-  
<sup>1177</sup> pressure GAr TPC (often denoted as HPgTPC) surrounded by an electromagnetic  
<sup>1178</sup> calorimeter (ECal) and a muon tagger. A cross section of its geometry can be seen  
<sup>1179</sup> in Fig. 3.7. ND-GAr will be able to measure the momenta of the outgoing muons  
<sup>1180</sup> while also detect neutrino interactions inside the GAr volume. This allows ND-GAr  
<sup>1181</sup> to constrain the systematic uncertainties even further, as it will be able to accurately  
<sup>1182</sup> measure neutrino interactions at low energies thanks to the lower tracking thresholds of  
<sup>1183</sup> GAr.

### 3.4. NEAR DETECTOR



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

#### 1184 3.4.3 PRISM

1185 In general, the observed peak neutrino energy of a neutrino beam decreases as the  
 1186 observation angle with respect to the beam direction increases. This feature has been  
 1187 used in other long-baseline neutrino experiments, like T2K ( $2.5^\circ$  off-axis) and NOvA  
 1188 ( $0.8^\circ$  off-axis), to achieve narrower energy distributions. The DUNE PRISM concept  
 1189 exploits this effect using a movable ND. Within PRISM both ND-LAr and the muon  
 1190 spectrometer (TMS in Phase I and ND-GAr in Phase II) can be moved up to  $3.2^\circ$   
 1191 off-axis, equivalent to move the detectors 30.5 m laterally through the ND hall.

1192 This allows to record additional data samples with different energy compositions.  
 1193 Figure 3.8 compares the on-axis muon neutrino flux at the ND with the fluxes at different  
 1194 off-axis positions. As the off-axis position increases the neutrino flux becomes closer to  
 1195 a monoenergetic beam with a lower peak energy. These samples can be used to perform  
 1196 a data-driven determination of the relation between true and reconstructed neutrino  
 1197 energy, to reduce the dependence on the interaction model. The off-axis samples are  
 1198 linearly combined to produce a narrow Gaussian energy distribution centered on a target  
 1199 true energy. From the combination coefficients one can build a sample of reconstructed

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

1200 neutrino events that will determine the energy mapping.

1201 The PRISM samples will be used to form a flux at the ND location similar in shape  
1202 to the oscillated flux measured by the FD. This method can be used to extract the  
1203 oscillation parameters with minimal input from the neutrino interaction model.

### 1204 3.4.4 SAND

1205 The role of SAND is to monitor the beam stability by measuring the on-axis neutrino  
1206 energy spectra. As the PRISM program requires that ND-LAr and its downstream  
1207 muon spectrometer spend about half of the time in off-axis positions, it is not possible  
1208 to monitor the stability with the movable detectors. Moreover, for the success of PRISM  
1209 it is essential to have a stable beam configuration, or, at least, a quick assessment and  
1210 modeling of the distortions.

1211 The SAND detector is magnetised, and features an inner low density tracker, a LAr  
1212 target with optical readout and surrounding sampling calorimeter.

## 1213 3.5 A More Capable Near Detector

1214 In DUNE Phase II, a more capable near detector is needed to achieve the ultimate physics  
1215 goals of the experiments. The current leading proposal for this detector is ND-GAr.  
1216 As mentioned previously, it will fulfill the role of TMS, measuring the momentum and  
1217 sign of the charged particles exiting ND-LAr. Additionally, it will be able to measure  
1218 neutrino interactions inside the HPgTPC, achieving lower energy thresholds than those  
1219 of the ND and FD LArTPCs. By doing so ND-GAr will allow to constrain the relevant  
1220 systematic uncertainties for the LBL analysis even further. A detailed discussion on the  
1221 requirements, design, performance and physics of ND-GAr can be found in the DUNE  
1222 ND CDR [91] and the ND-GAr white paper [93].

### 3.5. A MORE CAPABLE NEAR DETECTOR

#### 3.5.1 Requirements

The primary requirement for ND-GAr is to measure the momentum and charge of muons from  $\nu_\mu$  and  $\bar{\nu}_\mu$  CC interactions in ND-LAr, in order to measure their energy spectrum. To achieve the sensitivity to the neutrino oscillation parameters described in the DUNE FD TDR Volume II [85], ND-GAr should be able to constrain the muon energy within a 1% uncertainty or better. The main constraint will come from the calibration of the magnetic field, which will be performed using neutral kaon decays in the HPgTPC.

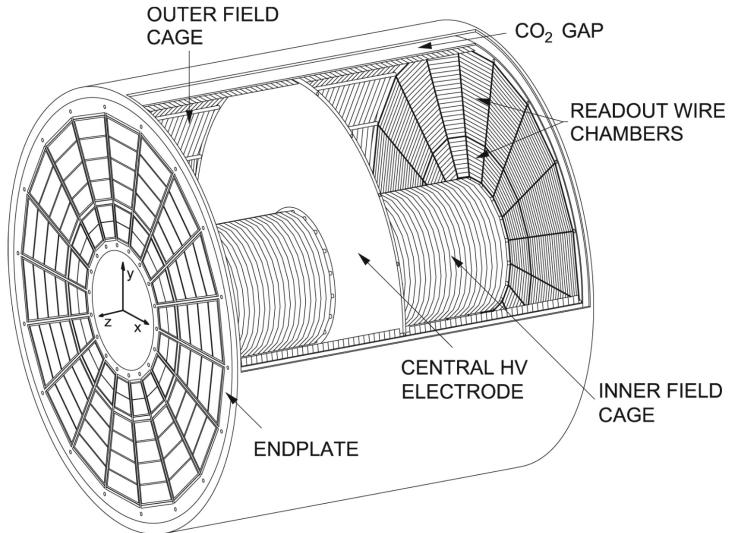
Another requirement for ND-GAr is the precise measurement of neutrino interactions on argon for the energies relevant to the neutrino oscillation program. The goal is to constrain the cross section systematic uncertainties in the regions of phase space that are not accessible to ND-LAr. This requires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

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

#### 3.5.2 Reference design

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

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT



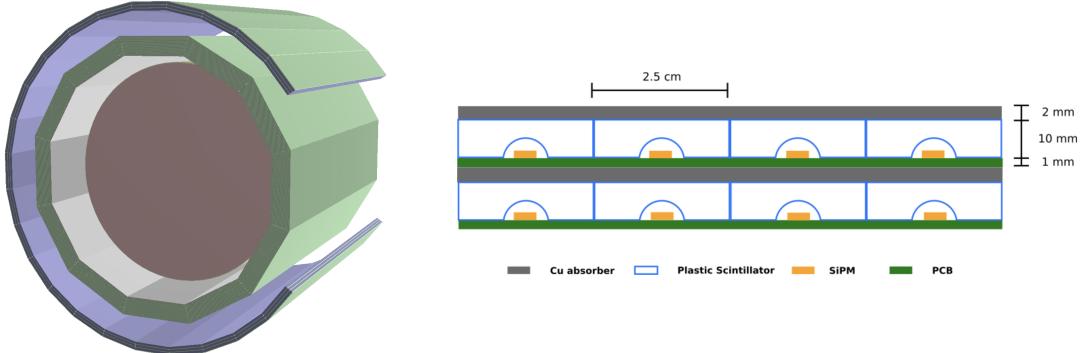
**Figure 3.9:** Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers. Figure taken from Ref. [91].

### 1249 HPgTPC

1250 The reference design for the ND-GAr HPgTPC follow closely that of the ALICE TPC.  
 1251 It is a cylinder with a central high-voltage cathode, generating the electric field for  
 1252 the two drift volumes, with a maximum drift distance of 2.5 m each. The anodes will  
 1253 be instrumented with charge readout chambers. The original design repurposed the  
 1254 multi-wire proportional readout chambers (MWPCs) of ALICE, however some of the  
 1255 current R&D efforts focus on a gas electron multiplier (GEM) [94] option instead. Figure  
 1256 3.9 shows a schematic diagram of the ALICE TPC design. The basic ND-GAr geometry  
 1257 will resemble this, except for the inner field cage.

1258 It will use a 90:10 molar fraction Ar: $\text{CH}_4$  mixture at 10 bar. With this baseline gas  
 1259 mixture light collection is not possible, as the quenching gas absorbs most of the VUV  
 1260 photons. Additional R&D efforts are underway, to understand if different mixtures allow  
 1261 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].

#### 1262    **ECal**

1263    The main role of the ND-GAr ECal is the calorimetric measurement of the electron  
 1264    energies and the reconstruction of photons, in particular those from neutral pion decays.  
 1265    Also, the ECal is able to provide a  $t_0$  timestamp for neutrino interactions, by associating  
 1266    its activity to the tracks in the HPgTPC. The ECal will also be able to perform  
 1267    neutron reconstruction using time of flight and reject external backgrounds, thanks to  
 1268    its sub-nanosecond time resolution.

1269       The ECal design features three independent subdetectors, two end caps at each side  
 1270    and a barrel surrounding the HPgTPC. Each of the detectors is divided in modules,  
 1271    which combine alternating layers of plastic scintillator and absorber material readout  
 1272    by SiPMs. The inner scintillator layers consist of  $2.5 \times 2.5 \text{ cm}^2$  high-granularity tiles,  
 1273    whereas the outer ones are made out of 4 cm wide cross-strips spanning the whole  
 1274    module length. The current barrel geometry consists of 8 tile layers and 34 strip layers,  
 1275    while the end caps feature 6 and 36 respectively. The thickness of the scintillator layers  
 1276    is 7 mm and 5 mm for the Pb absorber layers. The 12-sided geometry of the ECal barrel  
 1277    (left) and the layout of the tile layers (left)<sup>1</sup> can be seen in Fig. 3.10.

---

<sup>1</sup>The 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

### 1278 Magnet

1279 The ND-GAr magnet design, know as the Solenoid with Partial Yoke (SPY), consists of  
1280 two coupled solenoids with an iron return yoke. The idea behind the design is to have a  
1281 solenoid as thin as possible, as well as a return yoke mass distribution that minimises  
1282 the material budget between ND-LAr and ND-GAr. The magnet needs to provide a  
1283 0.5 T field in the direction perpendicular to the beam, parallel to the drift electric field.  
1284 It needs to host the pressure vessel and the surrounding ECal, which points to a inner  
1285 diameter of  $\sim 6.4$  m.

1286 The solenoid is a single layer coil, based on niobium titanium superconducting  
1287 Rutherford cable. The total length of the coil is 7.5 m. The bobbin will be split in four  
1288 segments grouped in pairs with two identical cryostats, connected in series. The iron  
1289 yoke features an aperture in the upstream side to allow the muons coming from ND-LAr.  
1290 Still, its material will be enough to reduce the magnetic field reaching SAND, and also  
1291 stop the charged pions produced inside the HPgTPC.

### 1292 Muon system

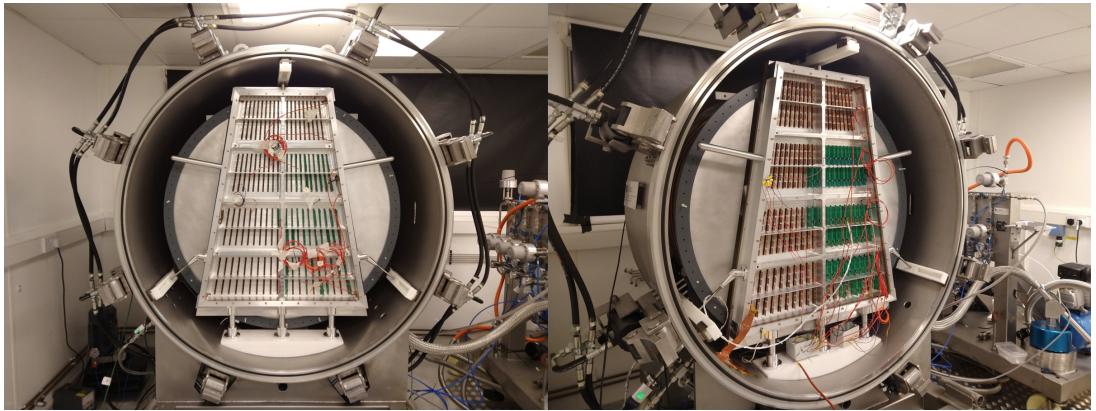
1293 The design of the ND-GAr muon system is still in a preliminary stage. Its role is to  
1294 distinguish between muons and pions punching through the ECal. This is especially  
1295 important for wrong-sign determination, to separate these from neutral current events.

1296 In its current form, the muon system consists of three layers of longitudinal sampling  
1297 structures. It alternates 10 cm Fe absorber slabs with 2 cm plastic scintillator strips.  
1298 The transverse granularity required is still under study.

### 1299 3.5.3 R&D efforts

1300 There are several ND-GAr-related prototypes, mostly focused on the TPC charge  
1301 readout and electronics. The priority is to test the full readout chain, in a high-pressure  
1302 environment, using a gas mixture with high argon fraction. A detailed summary of these  
1303 can be found in the DUNE Phase II white paper [92].

### 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. [95].

#### **1304 Multi-Wire Proportional Chambers**

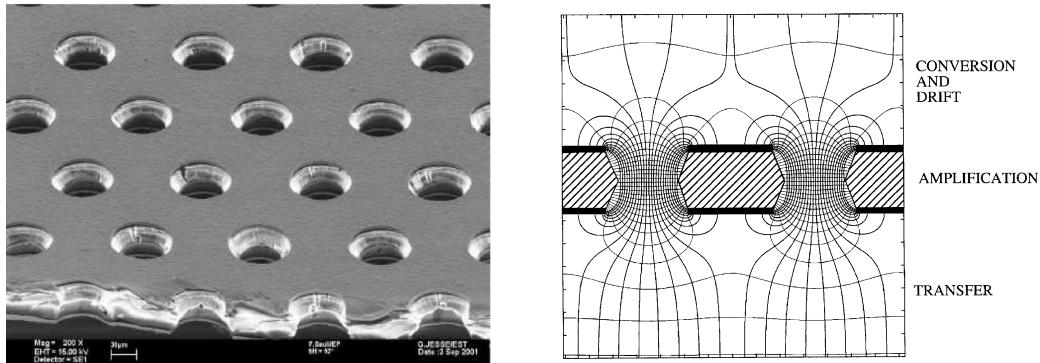
1305 As mentioned before, the original ND-GAr design repurposes the MWPCs of the ALICE  
1306 TPC, which became available after the recent upgrade [96]. These were operated using  
1307 a 90:10:5 Ne:CO<sub>2</sub>:N<sub>2</sub> gas mixture at 1 atm. Therefore, their performance needed to be  
1308 studied in an argon gas environment at high pressure.

1309 The Gas-argon Operation of ALICE TPC (GOAT) test stand tested the ALICE  
1310 readout chambers at high pressure. In particular, it used one of the previously operated  
1311 ALICE inner MWPCs, also known as IROCs, in a pressure vessel rated to 10 atm. It  
1312 measured the gas gain at various pressure points, voltages and gas mixtures.

1313 The Test stand of an Overpressure Argon Detector (TOAD) tested an ALICE outer  
1314 MWPC, also known as OROC, up to 5 atm. During its time at RHUL, it was used to  
1315 study the achievable gas gain of the OROC [95]. At the moment, it is being commissioned  
1316 at Fermilab for a full detector test of the readout electronics and the DAQ.

1317 Figure 3.11 shows the interior of the TOAD pressure vessel. The TPC is mounted  
1318 inside the vessel on three rails. The back of the OROC, supported by an aluminium  
1319 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. [97].

### 1320 Gas Electron Multiplier

1321 An alternative to the MWPC option is the use of GEMs. These are a type of micro-patter  
 1322 detector, where the ionisation electrons passing through the holes in the GEM layers  
 1323 are accelerated by a high intensity electric field. The acceleration causes the electrons  
 1324 to ionise the medium, resulting in an avalanche which increase the signal exponentially  
 1325 [94]. GEMs are used in numerous experiments that need a high spatial resolution, like  
 1326 ALICE [98] and CMS [99] after their upgrades.

1327 Figure 3.12 (left panel) shows an electron microscope picture of a  $50\ \mu\text{m}$  thick GEM  
 1328 electrode, with a pitch between neighbouring holes of  $140\ \mu\text{m}$  and a hole diameter of  
 1329  $70\ \mu\text{m}$ . A schematic representation of the cross section of a GEM layer is shown in Fig.  
 1330 3.12 (left panel).

1331 The Gaseous Argon T0 (GAT0<sup>2</sup>) prototype studies the use of thick GEMs made out  
 1332 of glass to achieve optical imaging of the primary ionisation. Using a 10 atm pressure  
 1333 vessel, the goal is to study different argon-based mixtures that allow for a precise  $t_0$   
 1334 determination.

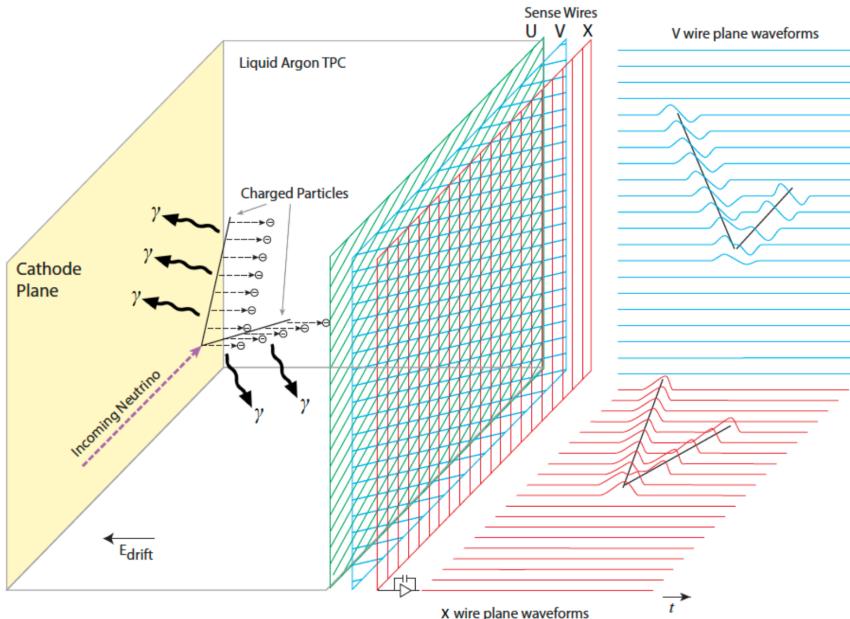
1335 The GEM Over-pressurized with Reference Gases (GORG<sup>3</sup>) test stand is currently  
 1336 testing a GEM-based charge readout, using a triple-GEM stack.

---

<sup>2</sup>Spanish for cat.

<sup>3</sup>Persian 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. [83].

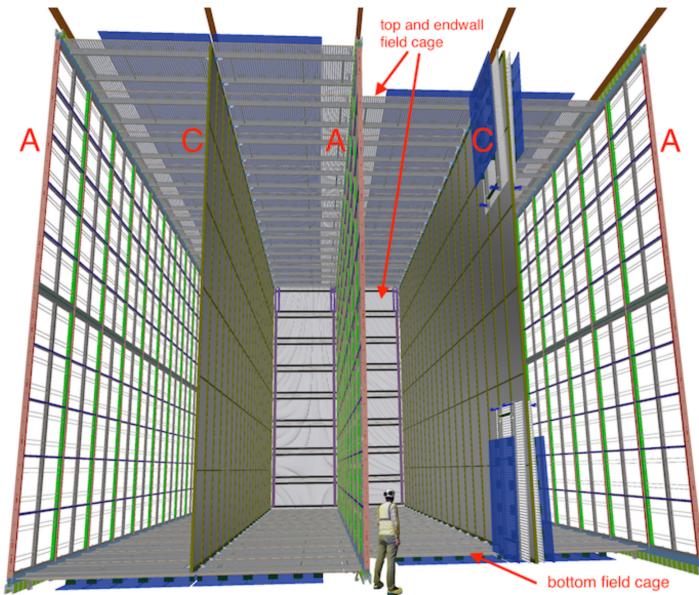
## 1337 3.6 Far Detector

1338 The DUNE FD complex will sit 1300 km away from the beam target and 1.5 km  
 1339 underground at SURF, South Dakota. Two caverns will host the four FD modules, two  
 1340 of them per cavern, each embedded in cryostats of dimensions 18.9 m (w)  $\times$  17.8 m (h)  $\times$   
 1341 65.8 m (l). A central, smaller cavern will host the cryogenic system.

1342 Three out of the four modules will be liquid argon (LAr) time projection chamber  
 1343 detectors, often refer to as LArTPCs, with a LAr fiducial mass of at least 10 kt each.  
 1344 The first and second FD modules, FD-1 and FD-2, will use a Horizontal Drift (HD)  
 1345 technology, whereas the third module, FD-3, will have a Vertical Drift (VD) direction.  
 1346 The technology for the fourth module is still to be decided,

1347 For each event, with energies ranging from a few MeV to several GeV, these detectors  
 1348 collect both the scintillation light and the ionisation electrons created when the charged  
 1349 particles produced in neutrino-nucleus interactions ionise the argon nuclei. In both HD  
 1350 and VD designs the characteristic 128 nm scintillation light of argon is collected by a  
 1351 photon detection system (PDS). This light will indicate the time at which electrons

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT



**Figure 3.14:** Proposed design for the FD-1 and FD-2 modules following the HD principle. Figure taken from Ref. [83].

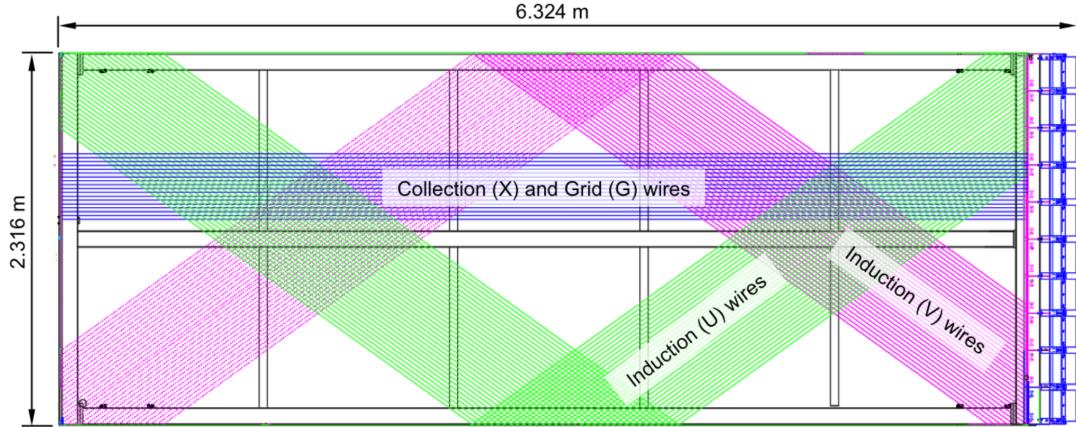
1352 start to drift, thus enabling reconstruction over the drift coordinate when compared  
 1353 to the time when the first ionisation electron arrives to the anode. Reconstruction of  
 1354 the topology in the transverse direction is achieved using the charge readout. Fig. 3.13  
 1355 illustrates the detection principle described, for the case of a HD detector with a wire  
 1356 readout.

### 1357 3.6.1 Horizontal Drift

1358 The HD design the ionisation electrons produced as charged particles traverse the LAr  
 1359 drift horizontally towards the anode planes, due to the effect of an electric field. These  
 1360 anode planes are made out of three layers of wire readout. This design, previously  
 1361 known as single-phase (SP), was tested by the ProtoDUNE-SP detector at CERN. The  
 1362 prototype collected data from a hadron beam and cosmic rays, providing high-quality  
 1363 data sets for calibration and performance studies.

1364 Each FD HD detector module is divided in four drift regions, with a maximum drift  
 1365 length of 3.5 m, by alternating anode and cathode walls. The surrounding field cage  
 1366 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. [83].

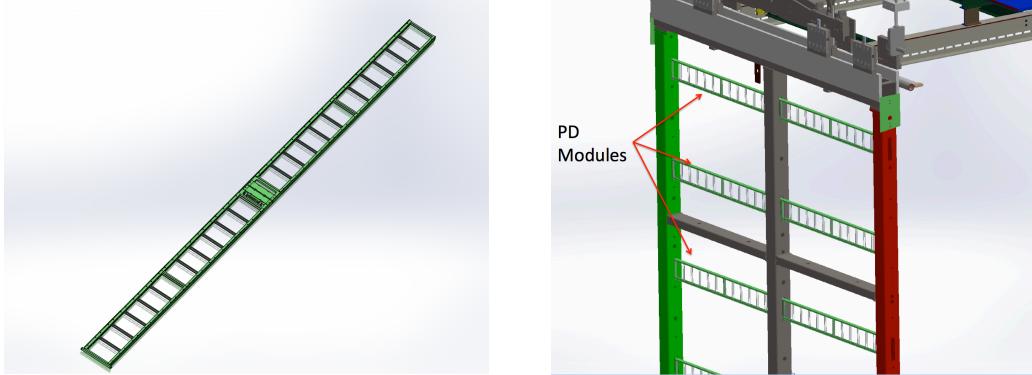
1367 The three anode walls, which constitute the charge readout of the detector, are built by  
 1368 stacking anode plane assemblies (APAs), 2 high times 25 wide. The design of the HD  
 1369 modules is shown in Fig. 3.14.

1370 Each APA is made of 2560 active wires arranged in three layers, plus an extra grid  
 1371 layer, wrapped around a metal frame. The two induction wire planes, U and V, sit at  
 1372  $\pm 35.7^\circ$  to the vertical on each side of the APA. The collection and shielding plane wires,  
 1373 X and G, run parallel to the vertical direction. The ionisation electrons drift past the  
 1374 induction planes, generating bipolar signals on those wires, and are collected by the  
 1375 collection plane, producing a monopolar positive signal. The spacing between the wires  
 1376 is  $\sim 5$  mm, and it defines the spatial resolution of the APA.

1377 The front-end readout electronics, or cold electronics as they are immerse in the LAr,  
 1378 are attached to the top of the up APAs and the bottom of the down APAs. Mounted on  
 1379 the front-end mother boards we have a series of ASICs that digitize the signals from the  
 1380 collection and induction planes. Each wire signal goes to a charge-sensitive amplifier,  
 1381 then there is a pulse-shaping circuit and this is followed by the analogue-to-digital  
 1382 converter. This part of the process happens inside the LAr to minimise the number of  
 1383 cables penetrating the cryostat. The digitised signals come out finally via a series of

## CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

- 1384 high-speed serial links to the warm interface boards (WIBs), from where the data is sent  
1385 to the back-end DAQ through optical fibers.



**Figure 3.16:** A PDS module containing 24 X-ARAPUCAs (left) and the location of the modules on the APAs (right). Figure taken from Ref. [83].

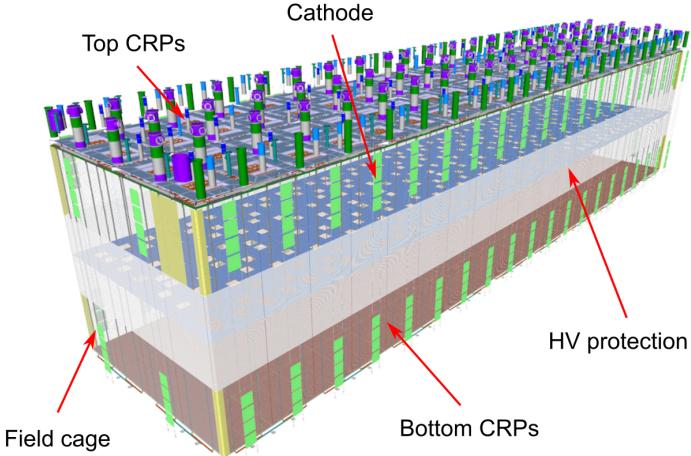
1386 The PDS uses modules of X-ARAPUCA devices, mounted on the APA frames  
1387 between the wire planes. Each X-ARAPUCA consists of layers of dichroic filter and  
1388 wavelength-shifter. They shift the VUV scintillation light into the visible spectrum,  
1389 sending then the visible photons to silicon photomultiplier (SiPM) devices. The PDS  
1390 modules are 209 cm × 12 cm × 2 cm bars, containing 24 X-ARAPUCAs. There are 10  
1391 of these PDS modules per APA. Fig. 3.16 shows a PDS module (left) and the placement  
1392 of the modules on the APAs (right).

### 1393 3.6.2 Vertical Drift

1394 In the VD case the ionisation electrons will drift vertically until they meet a printed  
1395 circuit board-based (PCB) readout plane. It is based on the original dual-phase (DP)  
1396 design deployed at CERN, known as ProtoDUNE-DP, used a vertical drift design with  
1397 an additional amplification of the ionization electrons using a gaseous argon (GAr) layer  
1398 above the liquid phase. The VD module incorporates the positive features of the DP  
1399 design without the complications of having the LAr-GAr interface.

1400 The current design of the FD VD module counts with two drift chambers with a  
1401 maximum drift distance of 6.5 cm. A cathode plane splits the detector volume along the  
1402 drift direction while the two anode planes are connected to the bottom and top walls

### 3.6. FAR DETECTOR



**Figure 3.17:** Proposed design for the FD-3 module following the VD principle. Figure adapted from Ref. [100].

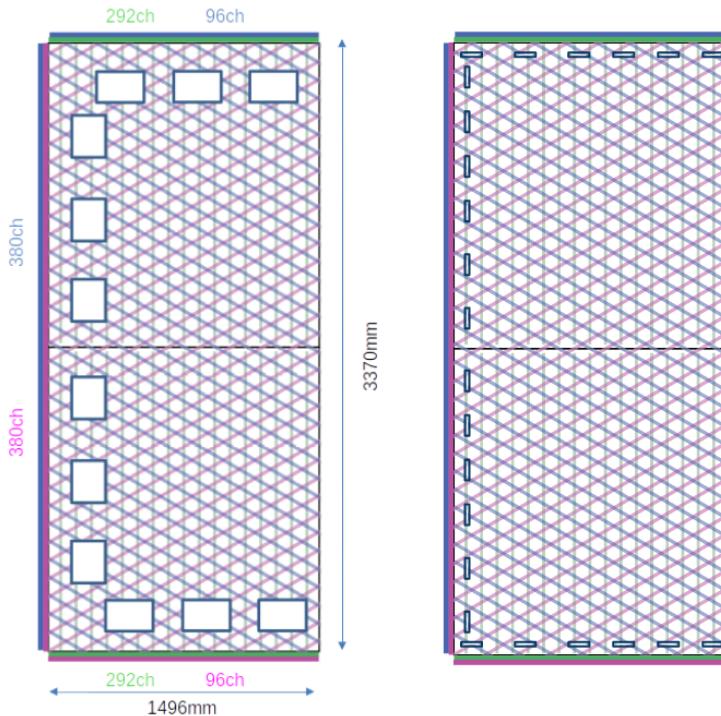
1403 of the detector. The layout of the VD module is shown in Fig. 3.17. Compared with  
 1404 the HD design, the VD option offers a slightly larger instrumented volume and a more  
 1405 cost-effective solution for the charge readout.

1406 As in the HD design, each drift volume features a 500 V/cm electric field and a  
 1407 field cage that ensures its uniformity. The anode planes are arrays of 3.4 m × 3 m  
 1408 charge-readout planes (CRPs). These are formed by a pair of charge-readout units  
 1409 (CRUs), which are built from two double-sided perforated PCBs, with their perforations  
 1410 aligned. The perforations allow the drift electrons to pass between the layers.

1411 The PCB face opposite to the cathode has a copper guard plane which acts as  
 1412 shielding, while its reverse face is etched with electrode strips forming the first induction  
 1413 plane. The outer PCB has electrode strips on both faces, the ones facing the inner PCB  
 1414 form the second induction plane while the outermost ones form the collection plane. Fig.  
 1415 3.18 shows the layout of the electrode strips for the top (left) and bottom (right) CRUs.  
 1416 The magenta and blue lines represent the first and second induction planes respectively,  
 1417 and the green lines correspond to the collection plane.

1418 The PDS in the VD module will use the same X-ARAPUCA technology developed  
 1419 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. [100].

1420 the cathode, in order to maximise the photon yield.

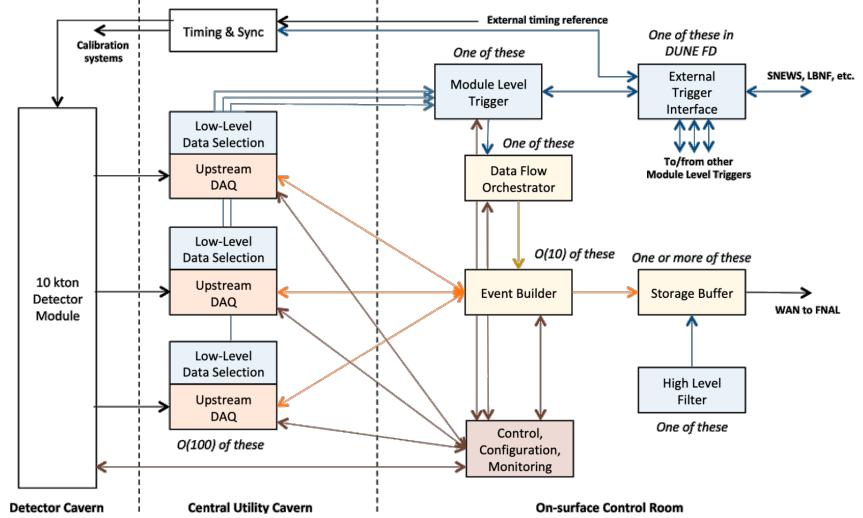
1421 **3.6.3 FD Data Acquisition System**

1422 The data acquisition (DAQ) system receives, processes and stores data from the detector  
 1423 modules. In the case of DUNE the DAQ architecture is designed to work for all FD  
 1424 modules interchangeably, except some aspects of the upstream part which may depend  
 1425 on the specific module technology.

1426 The enormous sample rate and the number of channels in TPC and PD readouts  
 1427 will produce a very large volume of data. These pose really strong requirements and  
 1428 challenges to the DUNE FD DAQ architecture. It will be required to read out data of  
 1429 the order of ten thousand or more channels at rates of a few MHz. To cope with the  
 1430 huge data volume, segmented readouts and compression algorithms are used to reduce  
 1431 the data rate to manageable levels.

1432 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. [101].

1433 first one is the upstream DAQ, which receives the raw data from the detector, buffers it  
 1434 and perform some low-level pre-processing. The minimally processed data is then fed  
 1435 into a hierarchical data selection system, which then performs a module level trigger  
 1436 decision. In case of a positive decision a trigger command is produced and executed by  
 1437 the data flow orchestrator, located in the back-end (BE) DAQ subsystem. Subsequently  
 1438 the DAQ BE retrieves the relevant data from the buffers located in the upstream DAQ,  
 1439 adds all the data into a cohesive record and saves it to permanent storage. Watching  
 1440 over all the other subsystems we also have the control, configuration and monitoring  
 1441 subsystem and the time and synchronization subsystem. Figure 3.19 shows a schematic  
 1442 diagram of the DAQ system, showing the different subsystems and their relations.

1443 A notorious challenge for the DUNE DAQ system comes from its broad physics  
 1444 goals. We must be prepared to process events spanning a wide range of time windows  
 1445 (from 5 ms in the case of beam and cosmic neutrinos and nucleon decay to 100 s in the  
 1446 case of SNBs) and therefore this requires a continuous readout of the detector modules.  
 1447 Moreover, because of the off-beam measurements we need to ensure the capabilities  
 1448 of online data processing and self-triggering. Having this into account, together with  
 1449 the technical constraints, the DUNE FD DAQ faces a series of challenges: it needs to

### CHAPTER 3. THE DEEP UNDERGROUND NEUTRINO EXPERIMENT

- 1450 be fault tolerant and redundant to reduce downtime, accommodate new components
- 1451 while it keeps serving the operational modules, have large upstream buffers to handle
- 1452 SNB physics, be able to support a wide range of readout windows and last reduce the
- 1453 throughput of data to permanent storage to be at most 30 PB/year.

# 4

1454

1455

## Matched Filter approach to Trigger

1456

## Primitives

### 1457 4.1 Motivation

1458 The filter implemented in the firmware of the upstream DUNE FD DAQ is a 32nd-order  
1459 low-pass finite impulse-response (FIR) filter. The output of such filter for a discrete  
1460 system can be written as:

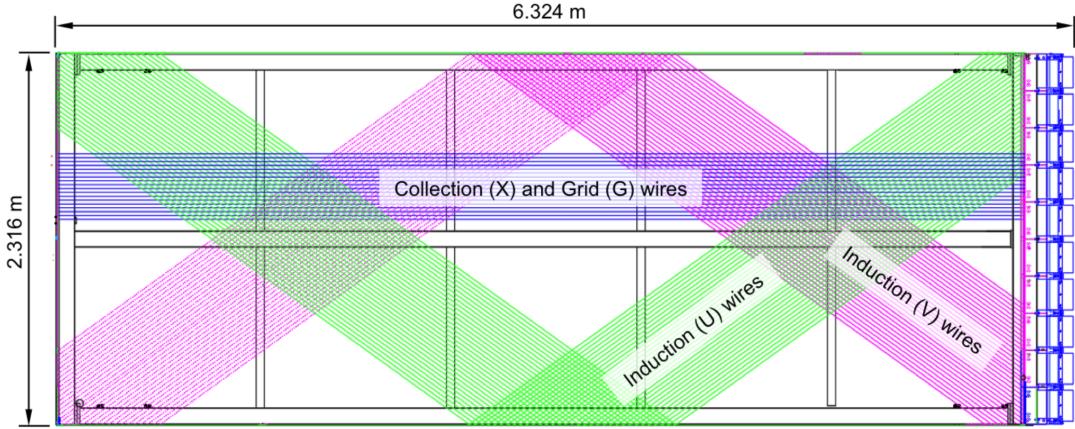
$$y[i] = \sum_{j=0}^N h[j]x[i-j], \quad (4.1)$$

1461 where  $N$  is the order of the filter,  $y$  is the output sequence,  $x$  is the input sequence and  $h$   
1462 is the set of coefficients of the filter. The current implementation within `dtp-firmware`  
1463 [102] uses a set of 16 non-zero integer coefficients.

1464 Filtering is a vital step in the hit finder chain. It helps to suppress the noise and  
1465 enhance the signal peaks with respect to the noiseless baseline. A good filtering strategy  
1466 allows us to use lower thresholds when forming the trigger primitives (TPs) and thus  
1467 increasing the sensitivity of our detector to low energy physics events. In such events,  
1468 the hits produced by the ionisation electrons tend to have lower amplitudes than those  
1469 of interest to the baseline physics programme of the DUNE experiment.

1470 This is particularly important for the induction planes. In general, signal peaks in  
1471 the induction wires have smaller amplitude than the ones in the induction plane. This,  
1472 together with the fact that the pulse shapes are bipolar, reduces our capacity to detect

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES



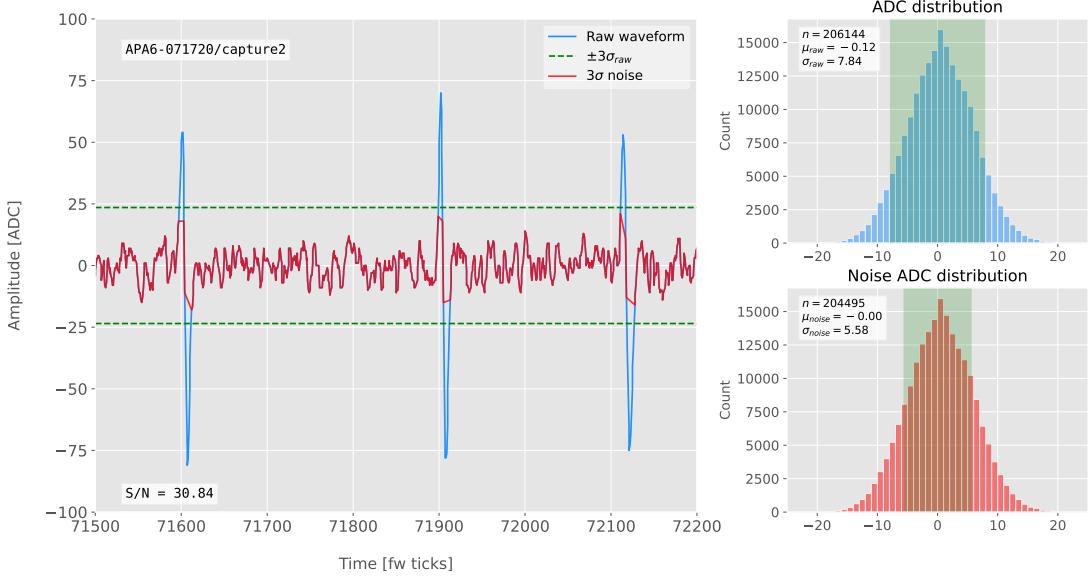
**Figure 4.1:** Schematic representation of an APA. The black lines represent the APA steel frame. The green and magenta lines correspond to the direction of the U and V induction wires respectively. The blue lines indicate the direction of the X collection wires and the wire shielding G.

1473 the hits on these channels. The inefficiency of detecting TPs in the induction planes  
 1474 (denoted as U and V planes) lead trigger algorithms to focus mainly on the TPs from  
 1475 the collection plane (so-called X plane). As a result, the possibility of making trigger  
 1476 decisions based on the coincidence of TPs across the three wire planes remains nowadays  
 1477 unexploited in DUNE. Fig. 4.1 shows a schematic view of an anode plane assembly  
 1478 (APA), with the different wire plane orientations highlighted.

1479 A possible improvement of the current hit finder chain could require optimising  
 1480 the existing or choosing a new filter implementation. A filter strategy which improves  
 1481 the induction signals may be able to enhance the detection efficiency of TPs from the  
 1482 induction planes and ideally make it comparable to that of the collection plane.

1483 The goal is to implement a better finite-impulse response filter design and to evaluate  
 1484 its performance relative to the current filter. To do so, we need to take into account the  
 1485 limitations of the firmware: the FIR filter shall have maximum 32 coefficients (so-called  
 1486 taps) whose values are 12-bit unsigned integers. Although it is technically possible to  
 1487 include non-integer coefficients, it would be a technical challenge as we have 40 FIR  
 1488 instances per APA, as there are 4 FIR per optical link and 10 optical links per APA.  
 1489 With these restrictions, the task is to provide a set of 32 coefficients which yield an

## 4.2. SIGNAL-TO-NOISE RATIO DEFINITION



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

1490 optimal filter performance for the induction wires.

## 1491 4.2 Signal-to-noise ratio definition

1492 I introduce the signal to noise ratio (S/N) as a measure of the FIR filter performance  
1493 and demonstrate how to extract its value for a set of ProtoDUNE-SP data. The S/N  
1494 metrics allow us to compare different filter implementations and serve as a basis for more  
1495 detailed studies presented later in this document. Specifically, I use the ADC capture  
1496 `felix-2020-07-17-21:31:44` (data capture taken for firmware validation purposes). I  
1497 defined S/N as the height of the signal peaks relative to the size of the noise peaks.  
1498 To quantify this quantity channel by channel one first need to estimate the standard  
1499 deviation of the ADC data for each channel,  $\sigma_{\text{ADC}}$ . Then, I define the corresponding  
1500 noise waveform to be the ADC values in the range  $\pm 3\sigma_{\text{ADC}}$ . From this new noise data  
1501 one can estimate again the mean and standard deviation,  $\mu_{\text{noise}}$  and  $\sigma_{\text{noise}}$ , so I can

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1502 write the S/N for any given channel as:

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

1503 where  $\max[ADC]$  is simply the maximum ADC value found in the corresponding channel.

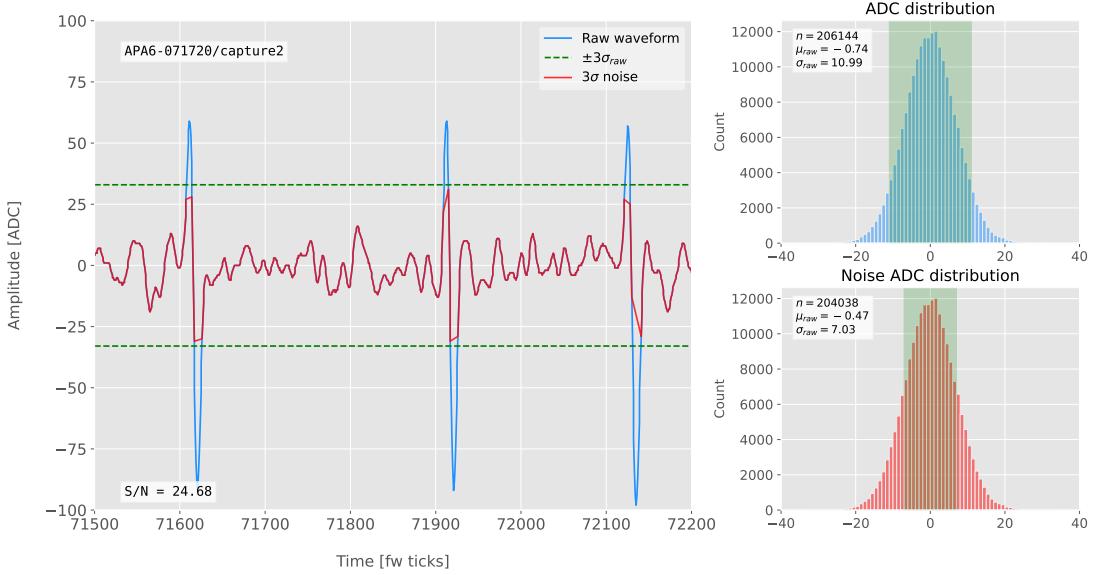
1504 One can apply this definition of the S/N with a waveform from one of the channels  
 1505 of the data capture<sup>1</sup>. Fig. 4.2 shows a zoomed region of the waveform corresponding to  
 1506 channel 7840 (blue line), where one can clearly see three signal peaks and continuous  
 1507 additive noise (we actually see 6 peaks, 3 positive and 3 negative, but, because by design  
 1508 for induction channels the expected signal pulse shapes are bipolar, I treat them as a  
 1509 collection of 3 individual signal peaks). I estimated the standard deviation of this raw  
 1510 waveform to be  $\sigma_{raw} = 7.84$  ADC, so I am able to define the noise waveform (red line)  
 1511 as the ADC values in the range  $\pm 23.52$  ADC. This way one obtains  $\mu_{noise} = 0$  and  
 1512  $\sigma_{noise} = 5.58$  ADC, which gives  $S/N = 30.84$ .

1513 We can repeat this calculation now for the corresponding filtered waveform (using the  
 1514 current firmware FIR filter). In Fig. 4.3 I plotted the same time window for the filtered  
 1515 waveform from channel 7840 (blue line). In this case, the standard deviation of the  
 1516 waveform is larger than before, giving  $\sigma_{raw} = 10.99$  ADC. The resulting noise waveform  
 1517 (red line) results from selection the ADC values in the range  $\pm 32.91$  ADC, giving now  
 1518  $\mu_{noise} = -0.47$  ADC and  $\sigma_{noise} = 7.03$  ADC. Finally, one obtains  $S/N = 24.68$ . Notice  
 1519 that the value of S/N decreases after the filtering. Clearly, one can see that the noise  
 1520 baseline has increased by a factor of 1.35 when we applied the FIR filter and at the same  
 1521 time the amplitude of the signal peaks has remained almost unchanged, leading to this  
 1522 poorer S/N value.

---

<sup>1</sup>All the original work was done within the `dtp-simulation` package [103], which offers a variety of tools to read raw data and emulate the TPG block (pedestal subtraction, filtering and hit finder). However, the results shown in this report were re-worked later using the C++ based `dtpemulator` package [104]. Its main purpose is the emulation of the TPG block and, in the same way as its predecessor, it has been cross-checked against the current firmware implementation.

### 4.3. LOW-PASS FIR FILTER DESIGN



**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  $\pm \sigma_{\text{noise}}$

## 1523 4.3 Low-pass FIR filter design

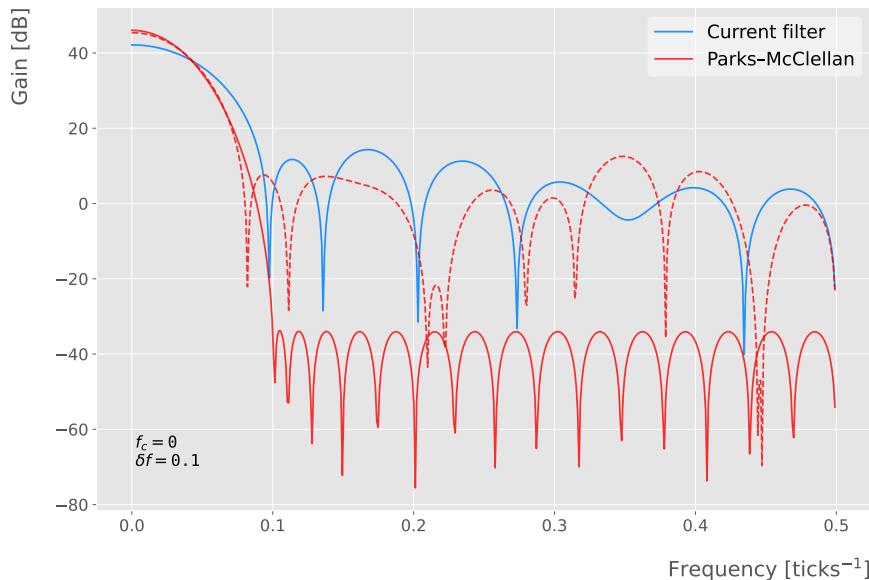
1524 In general, when one uses a method to optimize the frequency response of a digital filter,  
 1525 such as the Parks-McClellan algorithm, one finds a set of  $N$  real coefficients that give  
 1526 the best response for the specified pass-band and order of the filter [105].

1527 In our case, as the sampling frequency is defined as  $1 \text{ ticks}^{-1}$ , the Nyquist frequency  
 1528 will simply be  $1/2 \text{ ticks}^{-1}$ . The current implementation of the filter seems to have as  
 1529 pass-band the range  $[0, 0.1] \text{ ticks}^{-1}$ . This can be seen in Fig. 4.4, where I show the  
 1530 power spectrum, in decibels, of such filter implementation (blue solid line). For instance,  
 1531 the Park-McClellan algorithm finds the optimal Chebyshev FIR filter taking as input  
 1532 the boundaries of the target pass-band and stop-band, which can be written in the form:

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

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

where  $f_c$  is the cut-off frequency,  $\delta f$  is the transition width and  $f_N$  is the aforementioned Nyquist frequency. A similar behaviour to the one in the current filter 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 included 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.

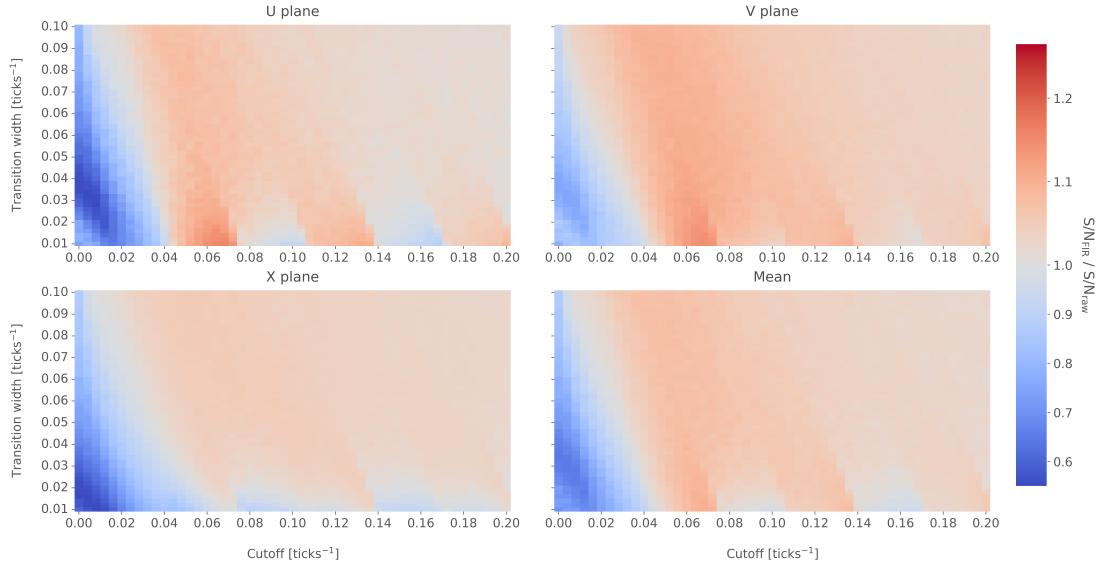


**Figure 4.4:** Power spectrum in decibels for the current implementation of the low-pass FIR filter in `dtp-firmware` (blue line), compared to the response of an optimal filter obtained using the Parks-McClellan algorithm for the same pass-band (red line). Also for comparison I include the spectrum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

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

Fig. 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`,

### 4.3. LOW-PASS FIR FILTER DESIGN

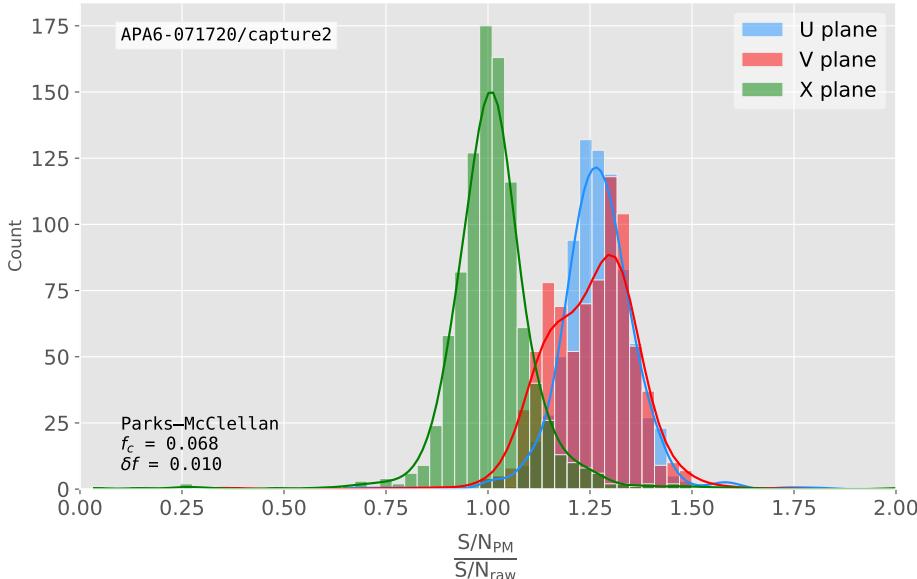


**Figure 4.5:** Relative change in the S/N for the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44`, using different values of the cutoff frequency  $f_c$  and the transition width  $\delta f$ . The optimal Chebyshev filters were applied using just the integer part of the coefficients given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

when using filters designed with the Parks-McClellan algorithm for the specified values of the cut-off frequency  $f_c$  and the transition width  $\delta f$ , 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).

Using the configuration which gives the best mean performance for the three planes (see bottom right panel of Fig. 4.5), i.e.  $f_c = 0.068 \text{ ticks}^{-1}$  and  $\delta f = 0.010 \text{ ticks}^{-1}$ , we can see how such filter affects the different channels. Fig. 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 induction wire planes, obtaining a mean change of 1.25 and 1.30 for them respectively. However, in the case of the collection plane X the mean of this distribution is roughly 1, meaning that a good fraction of channels in that plane get a slightly worse S/N after the

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES



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

filter is applied. In any case, this is not a big issue as the S/N for collection channels is usually much higher than the one for induction channels.

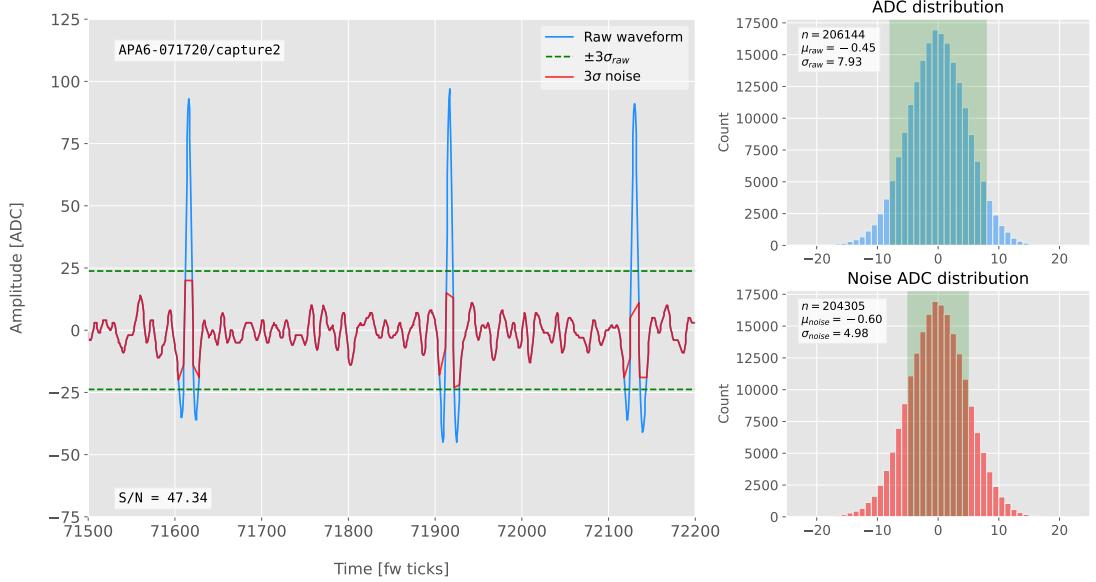
The results I obtained optimising the low pass filter with the Parks-McClellan method are promising. Nonetheless, the improvement found is rather marginal so I wondered if there could be an alternative approach to the filtering problem which yields better outputs. At this point, I found a possible alternative 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 signal-to-noise ratio (S/N) in the presence of additive noise, obtained by convolving a conjugated time-reversed known template with an unknown signal to detect the presence of the template in the signal [106].

Given a known signal sequence  $s(t)$  and another (a priori unknown) noise sequence

#### 4.4. MATCHED FILTERS



**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  $\pm \sigma_{\text{noise}}$

1575  $n(t)$ , the input signal can be written as:

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

1576 Now, considering a linear time-invariant filter, whose impulse-response function I  
1577 will refer to as  $h(t)$ , one can write the output signal as:

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

1578 where  $y_s(t)$  and  $y_n(t)$  are simply the outputs of the filter due to the signal and the noise  
1579 components respectively.

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1580 The goal of the matched filter is to detect the presence of the signal  $s(t)$  in the input  
 1581 sample  $x(t)$  at a certain time  $t_0$ , which effectively means we need to maximise the S/N.  
 1582 This way, what one wants is to have a filter which gives a much bigger output when the  
 1583 known signal is present than when it is not. Putting it in other words, the instantaneous  
 1584 power of the signal output  $y_s(t)$  should be much larger than the average power of the  
 1585 noise output  $y_n(t)$  at some time  $t_0$ .

1586 For the case of the filtered signal, one can easily re-write it as an inverse Fourier  
 1587 transform:

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

1588 where  $H(\omega)$  and  $S(\omega)$  are the Fourier transforms of the impulse-response function (i.e.  
 1589 the transfer function of the filter) and of the input signal, respectively.

1590 Now focusing on the noise, we can use the Wiener-Khinchin theorem [107] to write  
 1591 the mean power of the noise after filtering as:

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

1592 where  $S_n(\omega)$  is the power spectral density of the noise.

1593 Having these, one can write the instantaneous S/N at time  $t_0$  as:

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

1594 Once we have this expression, we need to find the upper limit of it to determine what  
 1595 would be the optimal choice for the transfer function. One can use the Cauchy-Schwarz  
 1596 inequality, which in the present case takes the form:

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

#### 4.4. MATCHED FILTERS

1597 for any two analytical functions  $f(x)$  and  $g(x)$ . One can prove that making the choice:

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

1598 leads to the following upper bound for the S/N:

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

1599 From Eqs. (4.8), (4.9) and (4.10) one can also derive the form of the transfer function  
1600 such that the upper bound is exactly reached [108]:

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

1601 From this last expression we can clearly see the way the matched filter acts. As the  
1602 transfer function is proportional to the Fourier transform of the signal it will try to only  
1603 pick the frequencies present in the signal [109].

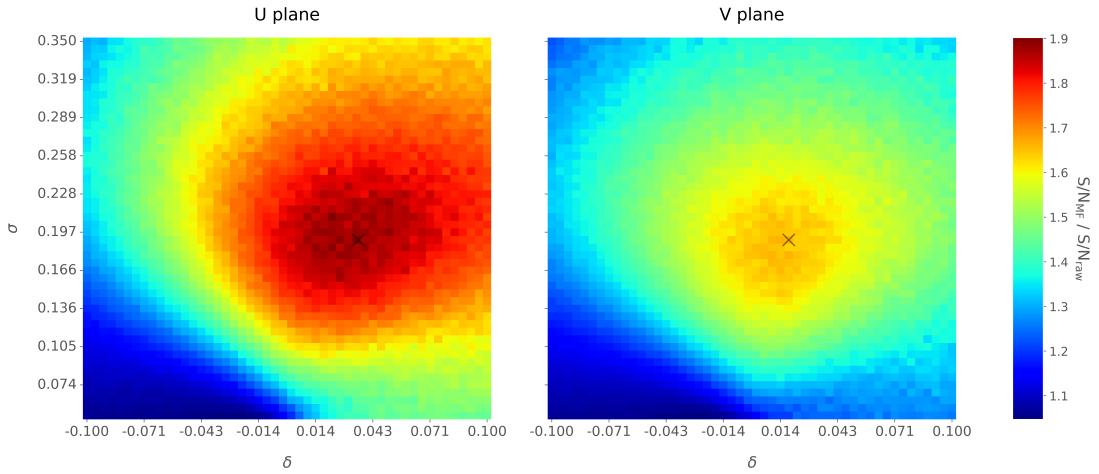
1604 The matched filter transfer function can be greatly simplified if the input noise is  
1605 Gaussian. In that case, the power spectral density of the noise is a constant, so it can be  
1606 re-absorbed in the overall normalisation of the transfer function. Moreover, considering  
1607 that the input signal is a real function, one can simply set  $S^*(\omega) = S(-\omega)$ , which gives:

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

1608 For a discrete signal, one can think of the input and impulse-response sequences as  
1609 vectors of  $\mathbb{R}^N$ . Then, the matched filter tries to maximise the inner product of the signal  
1610 and the filter while minimising the output due to the noise by choosing a filter vector  
1611 orthogonal to the later. In the case of additive noise, that leads to the impulse-response  
1612 vector:

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

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES



**Figure 4.8:** Relative improvement in the S/N for the raw data capture *felix-2020-07-17-21:31:44*, using the matched filter following the parametrisation in Eq. (4.17). The black crosses in both panels denote the location of the maximum ratio value.

1613 where  $s$  is a reversed signal template sequence of length  $N$  equal to the order of the filter  
 1614 and  $R_n$  is the covariance matrix associated with the noise sequence  $n$ . For the Gaussian  
 1615 noise case, the covariance matrix is simply the unit matrix, so the above expression  
 1616 simplifies again to:

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

1617 For this first stage of the study, I use a definition of the S/N per channel given by:

$$\text{S/N} = \frac{\max [ADC] - \mu_{noise}}{\sigma_{noise}}, \quad (4.16)$$

1618 where the subscript *noise* refers to a subset of the data obtained by only taking into  
 1619 account waveform values within a  $\pm 3\sigma$  range around the mean of the data and  $\max [ADC]$   
 1620 is the maximum of the original waveform. This definition is further discussed in App.  
 1621 4.2, where I also show examples of its application to raw data and to a waveform filtered  
 1622 with the current low-pass FIR filter.

1623 To test whether this choice of filter is appropriate one needs to choose a signal  
 1624 template. As an example of how a matched filter would affect our signal, I simply took  
 1625 the filter coefficients to be the 32 ADC values around a signal peak present in the data.

#### 4.4. MATCHED FILTERS

1626 In Fig. 4.7 (left panel) I plotted a zoomed region for channel 7840 in the raw data capture  
1627 `felix-2020-07-17-21:31:44`, after applying the matched filter described before (blue  
1628 line). When compared to the raw and FIR filtered case (see App. 4.2), after applying  
1629 the match filter the standard deviation of the noise waveform (red line) decreases and at  
1630 the same time the signal peaks are enhanced. This leads to an improvement of the S/N  
1631 by a factor of 1.92 when compared to the raw waveform.

1632 In order to obtain the matched filter that is more suitable for our data, I explored  
1633 different configurations of signal templates. In order to perform this exploration, I  
1634 parametrised the signal using the bipolar function:

$$f(x) = -A(x + \delta) e^{-x^2/\sigma^2}, \quad (4.17)$$

1635 where the parameter  $\delta$  controls the asymmetry between the positive and negative peaks  
1636 and  $\sigma$  controls their width. The amplitude parameter  $A$  is set such that it keeps the  
1637 height of the biggest peak to be less than 200 ADC in absolute value.

1638 As this parametrisation is only adequate for bipolar signals I will focus exclusively  
1639 on the induction channels. Also, the optimal configurations I found for the U and V  
1640 plane will be kept separate, i.e. I will have two sets of coefficients that will be applied to  
1641 either the U and V planes of wires. I do so as I found this was the choice giving the  
1642 best performance. Even so, as I will discuss, the differences are not very pronounced. In  
1643 case it is not technically possible to separate channels in the firmware according to the  
1644 wire plane they come from and use different sets of filter coefficients for them, we can  
1645 just find a common unique set of coefficients. In such case, I do not expect our results  
1646 to change dramatically.

1647 In Fig. 4.8 I present the results of our parameter scan, for channels in the induction  
1648 planes U (left panel) and V (right panel). For each configuration of  $\sigma$  and  $\delta$  the resulting  
1649 matched filter was applied to all channels in the corresponding plane within the data  
1650 capture `felix-2020-07-17-21:31:44`, the S/N improvement was computed with respect  
1651 to the raw waveforms and then the S/N mean value was kept as a score for such filter.

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

1652 One can see that the improvement obtained for the U plane is in general higher than the  
1653 one for the V plane. In any case, I got substantially higher ratios than the ones obtained  
1654 for the low-pass FIR filters. For the optimal configurations I attained improvements up  
1655 to a factor of 1.85 for the U plane and 1.65 for the V plane.

1656 The sets of optimal matched filter coefficients were obtained for the parameters  
1657  $\delta = 0.035$ ,  $\sigma = 0.191$  for the U plane and  $\delta = 0.018$ ,  $\sigma = 0.191$  for the V plane. I  
1658 show these two sets of coefficients in Fig. 4.9 (left panel). Also in Fig. 4.9 (right  
1659 panel) I plot the distribution of the S/N improvement after the optimal match filters  
1660 for the U and V were applied to the corresponding channels in the raw data capture  
1661 `felix-2020-07-17-21:31:44`. As mentioned before, the mean improvement achieved  
1662 for the U plane channels is slightly bigger than the one for the V channels. Note, however,  
1663 that the spread of the distribution for the V plane is also smaller than the one for the U  
1664 plane.

1665 I also performed a similar scan for the case of a low-pass FIR filter using the Parks-  
1666 McClellan algorithm. In that case, the parameters to check were the cutoff frequency  
1667 and the transition width of the filter. A summary of the results is given in App. 4.3.

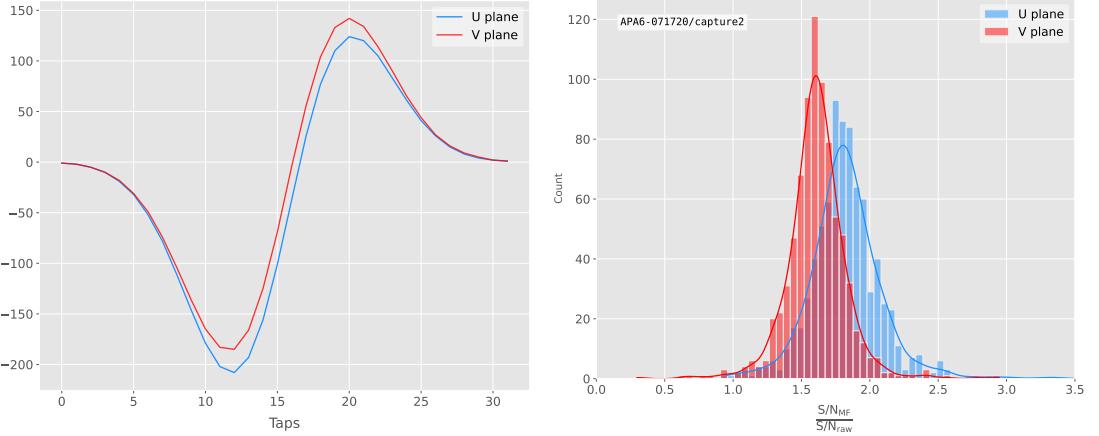
1668 Overall, one can see that the improvements on the S/N are much more significant in  
1669 the case of the matched filter than it is for the low-pass FIR filters. The analysis of this  
1670 and other raw data captures from ProtoDUNE-SP suggest that matched filters increase  
1671 the S/N of induction channels by a factor of 1.5 more than the optimal low-pass FIR  
1672 filters.

1673 Although these results are by themselves great points in favour of the matched  
1674 filter, more studies are needed to completely assess the robustness of this approach. I  
1675 proceeded then to test the matched filter with simulated data samples.

## 1676 4.5 Using simulated samples

1677 In order to further test the matched filter, the next step was to generate and process  
1678 data samples using *LArSoft* [110]. In this way, one can control the particle content of

## 4.5. USING SIMULATED SAMPLES



**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. (4.17) 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.

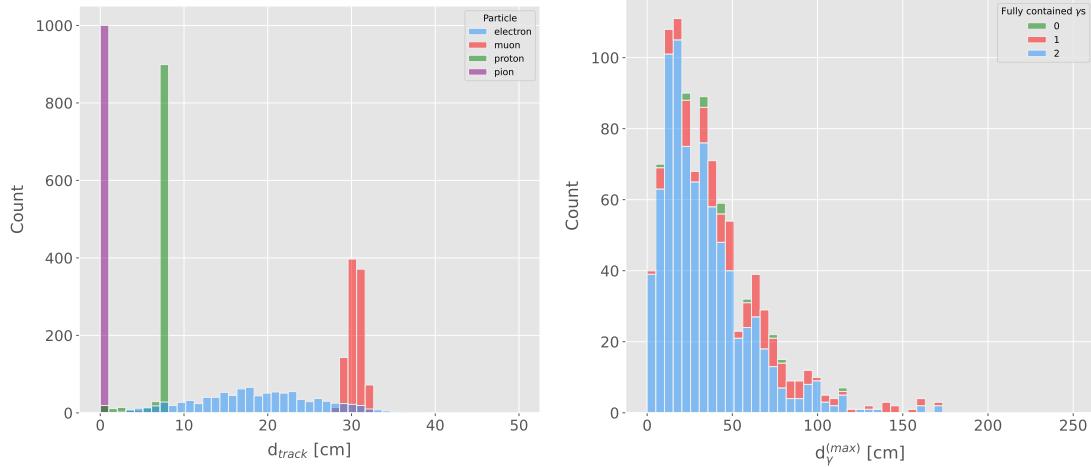
1679 the samples, the orientation of the tracks and their energy, and therefore see how the  
1680 matched filter behaves in various situations.

1681 To begin with, I prepared different monoenergetic and isotropic samples containing  
1682 a single particle per event. Each sample contains a different particle species, namely  
1683 electrons, muons, protons and neutral pions all with a kinetic energy of  $E_k = 100$  MeV.  
1684 I chose these because of the fairly different topologies they generate in the liquid argon,  
1685 ranging from shower-like to track-like. The procedure I followed to generate the samples  
1686 and process them is discussed in detail in App. ??.

1687 These were generated with the single particle gun and the Geant4 stage of the  
1688 *LArSoft* simulation [110] was performed with the standard configuration for the DUNE  
1689 FD 10kt module.

1690 For simplicity, I restricted the particles to start drifting in a single TPC volume  
1691 (in this case TPC 0), so I can focus exclusively on the signals coming from one APA.  
1692 The chosen kinetic energy for all the particles in my first trial is  $E_k = 100$  MeV, so a  
1693 necessary check is to see if all our tracks will be typically contained in one TPC volume.  
1694 Fig. 4.10 (left panel) shows the distributions of the track lengths in the liquid argon

## 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  $\pi^0 \rightarrow \gamma\gamma$ .

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  $\sim 30$  cm. The case of neutral pions can be misleading, as they decay promptly the track length associated with the true Monte Carlo 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 both). One can see that the vast majority of events has both photons contained and that just a negligible number of them has none of them contained in the TPC volume. In any case, for the sake of caution, I will only keep the pion events with both photons contained.

Once I have prepared a sample at the Geant4 level, I need to process it through the detector simulation. In order to make adequate estimations of the noise levels and run the filtering and hit finder as I did with the ProtoDUNE data, one needs to turn off the default zero-suppression of the waveforms produced by the simulation. At this first stage I am only concerned with the waveforms with the noise added, so I keep the noise

## 4.5. USING SIMULATED SAMPLES

addition option as true in the configuration. However, for studies related to the hit finder performance one will also need to store the noiseless waveforms in order to retrieve the truth information of the hits. I will discuss this approach next.

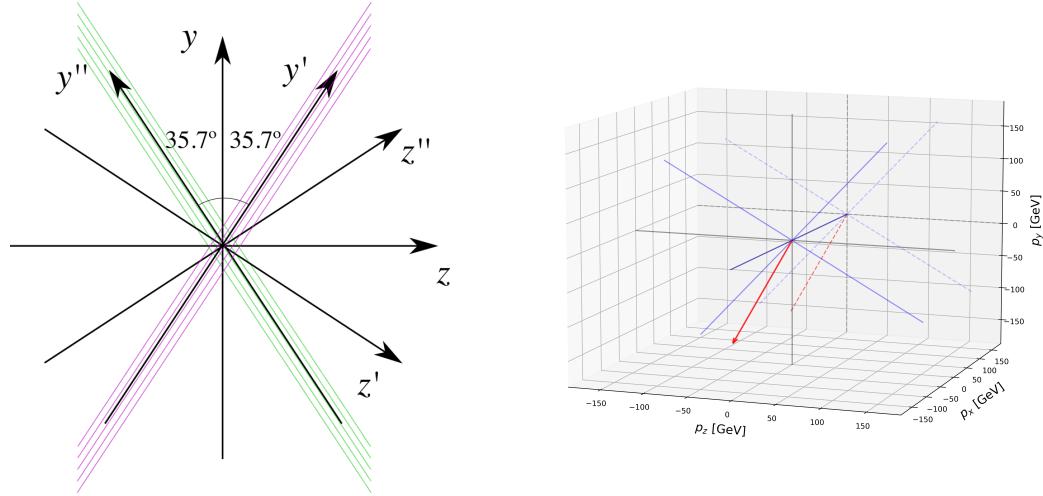
After the detector simulation stage, one needs to extract the no zero-suppressed noisy waveforms, along with their offline channel numbers, and store them in a certain format to be analysed later. To reduce the amount of data that will go for processing, I used the information from the Geant4 step of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

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

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

Additionally, for the analysis of the samples it was necessary to use two different reference frames, to study separately the signals coming from the U and V induction wire planes. As I am focussing on a single APA, the U and V wires have a different orientation in the  $yz$  plane. In the case of U wires, these are tilted  $35.7^\circ$  clockwise from the vertical ( $y$  direction), whereas the V wires are at the same angle but in the counter clockwise direction. Because of this, the best option is to deal with two new coordinate systems 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

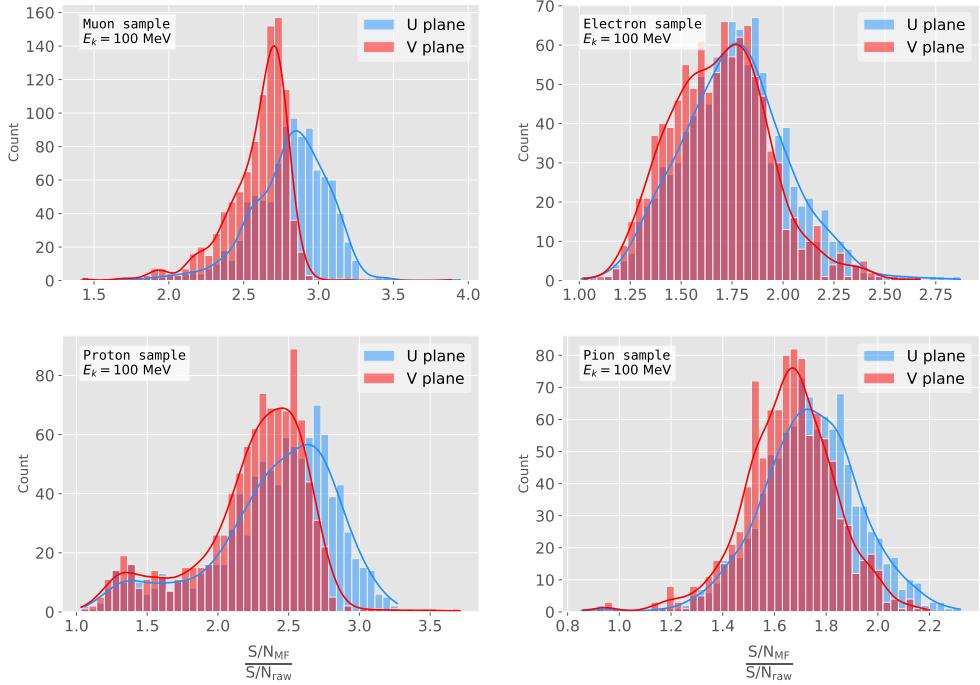


**Figure 4.11:** Left panel: schematic representation of the two new rotated reference frames used in this analysis (denoted as prime and double prime), viewed from the  $yz$  plane. The magenta stack of lines represent the wires in the  $U$  plane, whereas the green lines correspond to the wires in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

1741 U and V induction wires. Fig. 4.11 (left panel) shows a schematic representation of  
 1742 the original reference frame together with the two rotated ones (denoted by primed and  
 1743 double primed). This way, one can easily understand how parallel was a track to the  
 1744 wires in the two induction planes. Fig. 4.11 (right panel) shows a 3D representation of  
 1745 the momentum of a track (red arrow) in the original reference frame (black lines), along  
 1746 with the new reference frame for  $U$  wires (blue lines). I added the projection in the  $yz$   
 1747 plane of this three, to show the usefulness of the new reference frame to tell whether a  
 1748 track is parallel or normal to the wires in the induction plane.

1749 Fig. 4.12 shows the distribution of the average S/N improvement per event when one  
 1750 applies the optimal matched filters. I produced separate distributions for the channels  
 1751 in the  $U$  (red) and  $V$  (blue) induction wire planes. Notice that the S/N distributions  
 1752 for the track-like particles, i.e. muons (top left panel) and protons (bottom left panel),  
 1753 have significantly larger mean values than the distributions of the shower like particles,  
 1754 i.e. electrons (top right panel) and neutral pions (bottom right panel). An important

## 4.5. USING SIMULATED SAMPLES



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

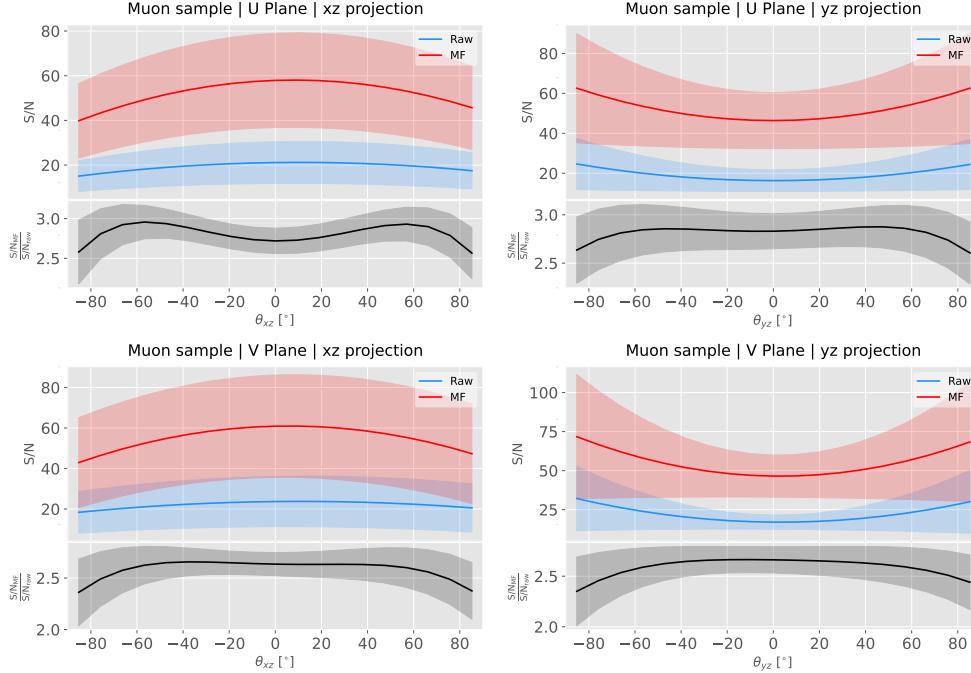
difference between these results and the ones seen before for ProtoDUNE data is that, overall, the improvements that I get for simulated data are bigger. This could be due either to the default noise model used in the *LArSoft* simulation or to the simulated hits having higher energy than the ones in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

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

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

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

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

1765 However, for the ratio of the raw and filtered S/N (what I called the S/N improvement)  
 1766 per event I am not just taking the ratio of the previous two quantities but computing  
 1767 the average of the individual ratios per channel in the event:

$$\left( \frac{S/N_{fir}}{S/N_{raw}} \right)_{event} = \frac{\sum_{i=0}^{N_{chan}} \left( \frac{S/N_{fir}}{S/N_{raw}} \right)_i}{N_{chan}}, \quad (4.19)$$

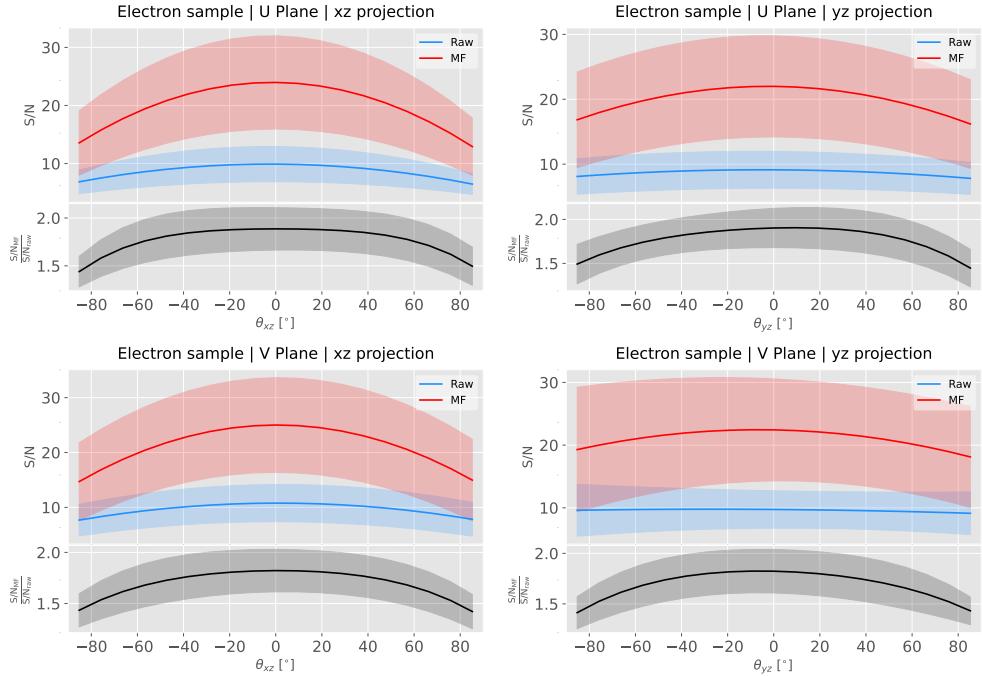
1768 and so:

$$\left( \frac{S/N_{fir}}{S/N_{raw}} \right)_{event} \neq \frac{(S/N_{fir})_{event}}{(S/N_{raw})_{event}}. \quad (4.20)$$

### 1769 4.5.1 Angular dependence

1770 Having these monoenergetic samples, one can also study the angular dependence of the  
 1771 performance of the matched filter. This is an important point, as it is a well established

## 4.5. USING SIMULATED SAMPLES



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

fact that for certain configurations (an extreme case configuration being signals normal to the wire plane and perpendicular to the induction wires at the same time) the S/N is much lower than average as the corresponding waveforms are severely distorted. In this sense, I am interested to see how the matched filter behaves for these cases and how the S/N improvement on those compare to the average.

Fig. 4.13 shows the angular dependence of the S/N for the monoenergetic  $E_k = 100$  MeV isotropic muons, for the different induction wire 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  $xy''z''$ ). 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

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

each plot, the top subplot represents the mean values of the S/N for the raw (blue) and matched filtered (red) signals, and the bottom subplot the averaged S/N improvement (black). The solid lines represent the mean value obtained for the corresponding angular value, whereas the semitransparent bands represent one standard deviation around the mean at each point.

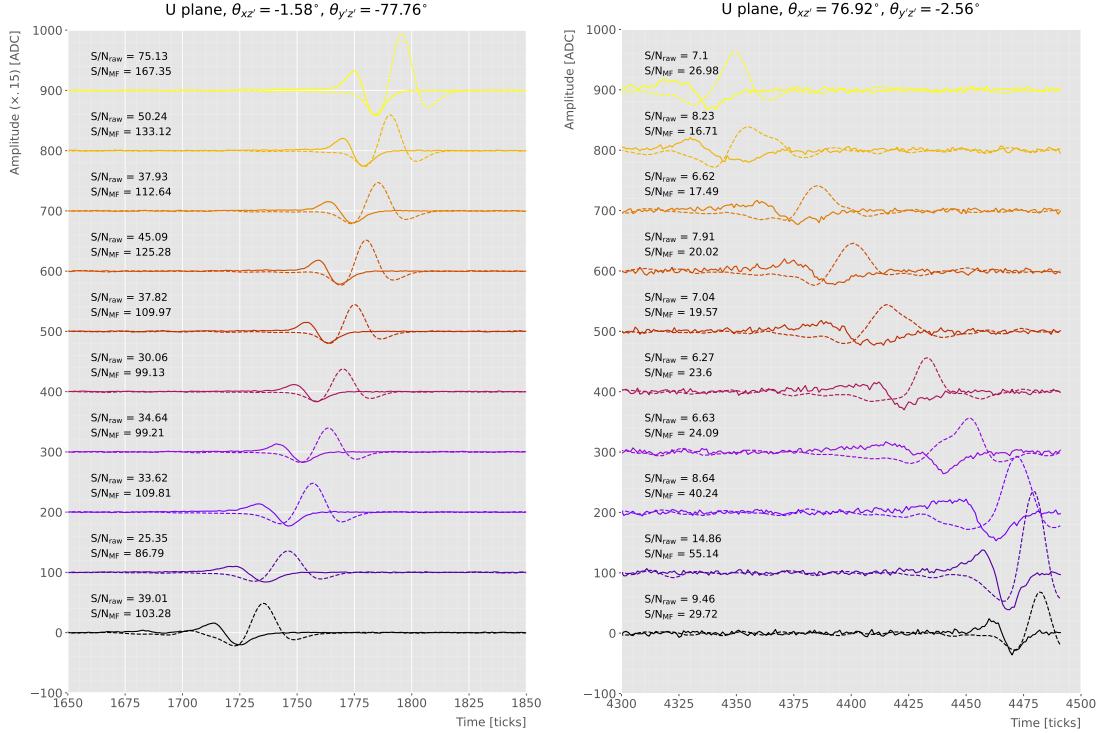
As expected, the S/N is in general higher when tracks are parallel to the APA (i.e.  $\theta_{xz} \sim 0$ ) and lower when it is normal to the plane ( $\theta_{xz} \sim \pm 90^\circ$ ). In the same way, tracks parallel to the wires ( $\theta_{yz} \sim \pm 90^\circ$ ) tend to have higher S/N than those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

Fig. 4.14 shows the corresponding angular dependence information for the  $E_k = 100$  MeV electrons sample. Notice that, in this case, the S/N behaviour discussed above does not hold. A possible explanation can be that, because most hits in these events are produced by the secondary particles generated in the EM shower, the signal peaks whose S/N ratios were computed do not correspond to the directional information of the primary electron.

### 4.5.2 Distortion and peak asymmetry

As a little case of study, I selected 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 (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 bigger S/N than the other, it is the one with the smallest S/N the one that got the biggest averaged S/N improvement. In Table 4.1 I included all the relevant parameters of these two  $E_k = 100$  MeV muon events I am considering, namely, the angles with respect to the  $xy'z'$  reference frame, the values of the S/N, the S/N improvement and also the so-called peak asymmetry  $\Delta_{peak}$  that I will discuss next.

## 4.5. USING SIMULATED SAMPLES



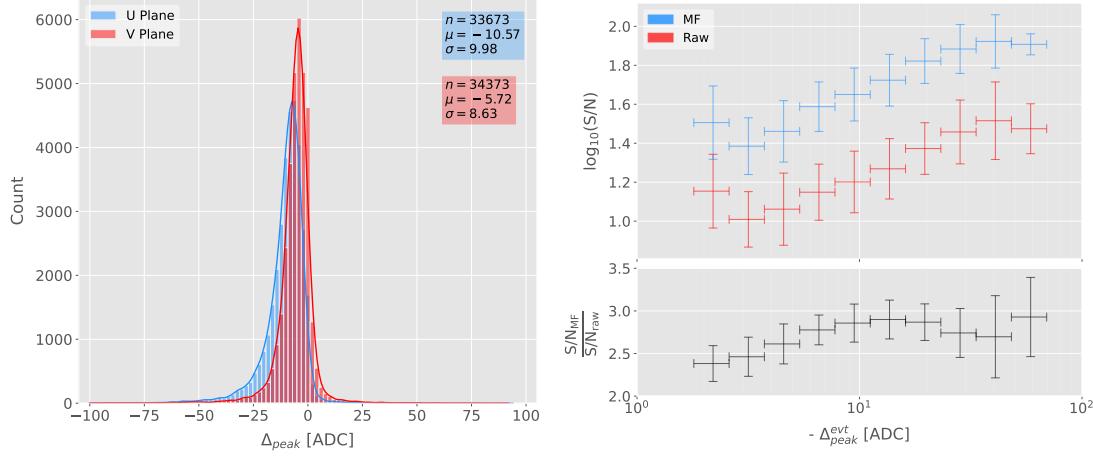
**Figure 4.15:** Selected consecutive waveforms corresponding to two monoenergetic  $E_k = 100$  MeV muon events, one is parallel to the APA and to the wires in the U plane (left panel) and the other is normal to the APA plane and perpendicular to the U plane wires (right panel). The solid lines represent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

**Table 4.1:** Characteristic parameters of the two monoenergetic muon events selected, relative to the U plane: projected angles in the  $xz'$  and  $yz'$  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 <sub>raw</sub>	S/N <sub>MF</sub>	$\frac{S/N_{MF}}{S/N_{raw}}$	$\Delta_{peak}$ (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

1813 One can try to understand better what is going on with these two events by looking  
 1814 at the raw and filtered data from some of their active channels. Fig. 4.15 shows a  
 1815 selection of consecutive raw and filtered U plane waveforms from the event with high S/N  
 1816 (left panel) and the one with low S/N (right panel). Notice that to show both collections  
 1817 of waveforms at a similar scale I had to apply a factor of 0.15 to the waveforms with  
 1818 high S/N. Additionally, next to each waveform I included the values of the raw and

## 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.

1819 matched filtered S/N for the corresponding channel. The first thing to notice in this plot  
 1820 is that the amplitude of the signal peaks from the normal track have a much smaller  
 1821 amplitude, and also appear quite distorted when compared to the others. On the other  
 1822 hand, although the matched filtered S/N is still smaller, the relative improvement is  
 1823 bigger than in the parallel case.

1824 A way I found to quantify the difference between the shapes within these two events  
 1825 is their different peak asymmetry. One can define the peak asymmetry as the (signed)  
 1826 difference between the positive and the negative peaks of the bipolar shape, i.e.:

$$\Delta_{peak} \equiv h_+ - h_-, \quad (4.21)$$

1827 where both heights  $h_+$  and  $h_-$  are positive defined. Fig. 4.16 (left panel) shows the  
 1828 distribution of this peak asymmetry for all the waveforms corresponding to channels  
 1829 in the U (blue) and V (red) planes for the monoenergetic muon sample. One can  
 1830 see that these distributions are clearly shifted to negative values (with mean values

## 4.5. USING SIMULATED SAMPLES

1831  $\mu_{\Delta}^U = -10.57$  ADC and  $\mu_{\Delta}^V = -5.72$  ADC respectively). It is interesting to notice  
1832 that the peak asymmetry value of the sample with high S/N sits at the left tail of the  
1833 distribution whereas the corresponding value of the sample with low S/N lies around  
1834 the mean.

1835 Now, one can try to correlate the peak asymmetry with the S/N and the S/N change  
1836 per event. Fig. 4.16 (right panel) shows the result of comparing (minus) the mean  
1837 peak asymmetry per event to the averaged raw (red) and matched filtered (blue) S/N  
1838 per event (top subplot). The horizontal lines sit at the mean value obtained in the fit  
1839 and represent the width of the  $-\Delta_{peak}$  bins used, while the vertical lines indicate one  
1840 standard deviation around that mean value. Notice that, when taking decimal logarithm  
1841 on both, there is an approximate linear relation between these quantities, except for  
1842 peak asymmetry values bigger than  $-5$  ADC where the S/N remains constant.

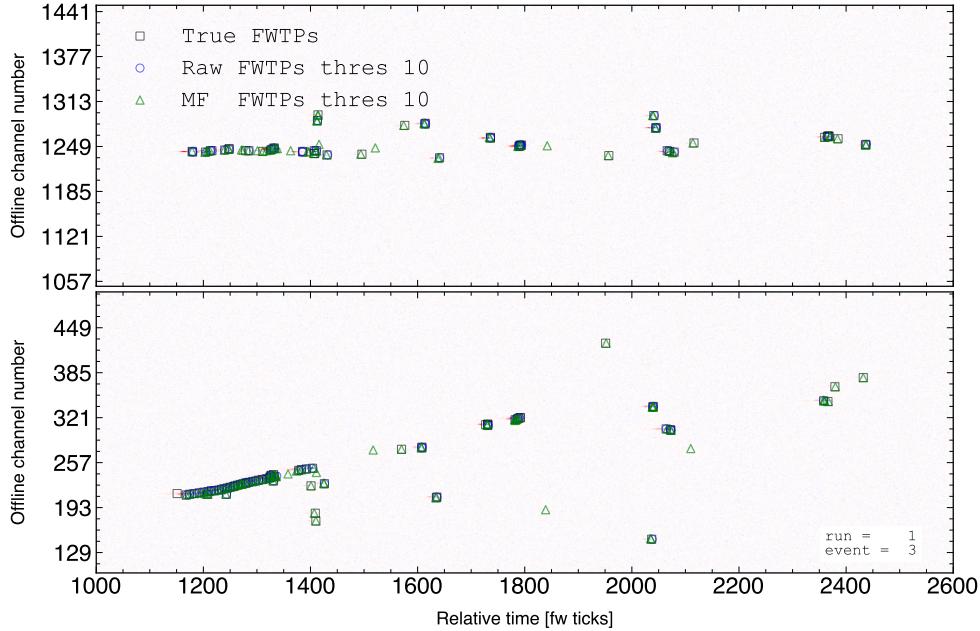
1843 Also, in the bottom subplot of Fig. 4.16 (right panel) I show the relation between  
1844 the peak asymmetry and the mean S/N improvement. In this case, one see that there is  
1845 a maximum at  $\Delta_{peak} \sim -10$  ADC. As mentioned previously, this is also the value of the  
1846 mean of the peak asymmetry distribution. In fact, it is expected that our filter favours  
1847 the signal peaks with the most common values of the peak asymmetry, as this was one  
1848 of the features I target in our filter coefficient optimisation through the parameter  $\delta$ .

1849 These results suggest that events with poorer values of the mean S/N, usually  
1850 associated to non-favourable track orientations, tend to have smaller values of the mean  
1851 peak asymmetry (in absolute value). Nonetheless, because our matched filters have  
1852 been optimised to account for these asymmetries, the improvement on the S/N for these  
1853 events is sizeable if not better than the one for events which already had a high S/N.

### 1854 4.5.3 Hit sensitivity

1855 One of the advantages of the matched filter, directly related to increasing the S/N, is  
1856 the capability of picking hits that before fell below the threshold. For instance, Fig. 4.17  
1857 shows the raw ADC data from an example event (electron,  $E_k = 100$  MeV) with the  
1858 produced true hits superimposed (black boxes), together with the hits produced by the

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES



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

standard hit finder chain (blue circles), i.e. using the current FIR filter, and the hits obtained using the matched filters (green triangles). Both the standard and the matched filter hit finders were run with a threshold of 10 ADC. Notice that the standard hits match well the true ones at the initial part of the event (where we have a track-like object), but they miss most of the hits produced by the EM shower at later times. On the other hand, the hits produced with the matched filter have a better agreement with the true hits even for the more diffuse shower activity.

Notwithstanding that now I get more hits with this combination of matched filter and low threshold as a results of the enhancement of the signal peaks relative to the noise level, it is also true that I pick some spurious hits not related to any real activity if one lowers the thresholds too much. Therefore, some optimisation of the threshold is needed. Basically one will need to make a trade-off between precision and sensitivity.

Having this in mind, I tried to compare the produced hits one gets from the standard hit finder and the ones resulting from applying the matched filter with the true hits.

## 4.5. USING SIMULATED SAMPLES

1873 By running the hit finders on our samples with different values of the threshold one  
1874 can understand, for instance, how low one can set the threshold without getting mostly  
1875 spurious hits and then evaluate the gains obtained from this.

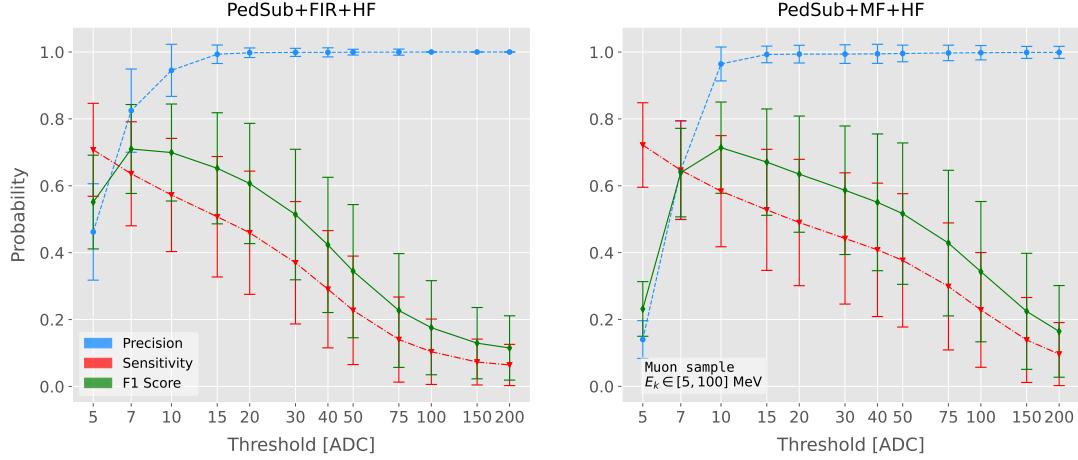
1876 Because now I am also interested in seeing how the hit sensitivity changes with the  
1877 energy, I prepared new isotropic samples with the same types of particles as before  
1878 (muons, electrons, protons and neutral pions) but with a flat kinetic energy distribution  
1879 ranging from 5 to 100 MeV.

1880 In order to estimate the hit sensitivity, given a certain sample, one needs to recover  
1881 the set of true hits to be able to compare these with the ones produced. To do so,  
1882 a modification in the procedure I was using to extract the raw waveforms is needed.  
1883 For this kind of study I run the detector simulation in two steps, first I produce the  
1884 waveforms without noise and extract them in the same format I used for the raw data,  
1885 then the noise is added and the noisy waveforms are then written to a file as well.

1886 To have a better comparison between the true hits and the ones produced from  
1887 the raw waveforms after applying the two filters, I applied also the FIR filter and the  
1888 matched filters to the noiseless waveforms and then I run the hit finder with a minimal  
1889 threshold (in this case I used 1 ADC) on these noiseless filtered waveforms. In this way  
1890 I generated two sets of true hits, I will refer to them as standard true hits (with the  
1891 current/default FIR filter) and matched filter true hits respectively. This allows a more  
1892 precise matching between the different groups of hits produced, as it will account for  
1893 any delays and distortions introduced by the FIR and the matched filters.

1894 In the case of the raw waveforms (with noise), I run the hit finder on them, with  
1895 different values of the threshold, after applying either the FIR or the matched filters. I  
1896 will name them simply standard hits and matched filter hits respectively. Then, I match  
1897 the generated hits to the true hits (the standard hits with the standard true hits and  
1898 the matched filter hits with the matched filter true hits). The matching is performed by  
1899 comparing the channel number and the timestamp of the hits. To count as a match,  
1900 I require that all hits with the same channel number and timestamp have overlapping  
1901 hit windows, i.e. the time windows between their hit end and hit start times need to

## 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.

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 hit with closer hit peak time value.

The generation of the samples, the procedure to produce the standard hits (with the default FIR filter) and matched filter hits and the matching of these with the true hits is described in detail in App. ??.

To quantify the performance of the two hit finder approaches, I use a classical method from statistical classification known as confusion matrix [111]. This is basically a way of sorting the outputs of a binary classifier, considering the true values of the classification and the predicted values. It divides the outputs in four categories: true positive (TP, both true and predicted values are 1), false negative (FN, true value is 1 but predicted is 0), false positive (FP, true value is 0 but predicted is 1) and true negative (TN, both true and predicted values are 0)).

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

## 4.5. USING SIMULATED SAMPLES

1916 metrics, namely the precision or positive predictive value:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4.22)$$

1917 the sensitivity or true positive rate:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (4.23)$$

1918 and the  $F_1$  score [112]:

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (4.24)$$

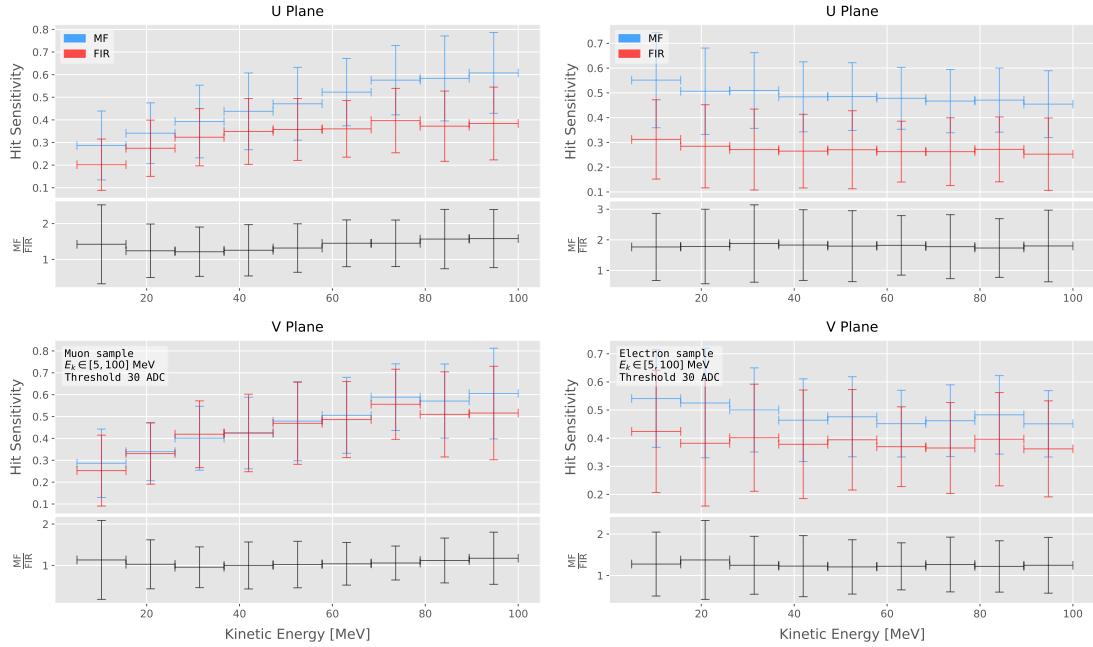
1919 which is the harmonic mean of the precision and the sensitivity.

1920 In our specific case I am not going to make use of the true negative value, as its  
1921 definition in this context can be ambiguous because one does not have clear instances in  
1922 the classification process. This way, I will only count the number of true positives as the  
1923 total amount of hits I can match between true and raw populations, the number of false  
1924 negatives will be the number of missing true hits and the false positive the number of  
1925 hits which do not match any true hit.

1926 In Fig. 4.18 I show the precision (blue), sensitivity (red) and  $F_1$  (green) scores I  
1927 obtained for different values of the threshold used in the hit finder for the case of the  
1928 muon sample. Because the matched filters are only applied to induction channels, I only  
1929 consider here hits coming from the U and V planes. The panel on the left corresponds  
1930 to the scores I got when I ran the hit finder on the FIR filtered waveforms, whereas the  
1931 right panel contains the scores for the matched filter case. The points are centered at  
1932 the threshold value used and represent the mean value obtained for each score using all  
1933 the generated events, while the error bars indicate one standard deviation around the  
1934 mean value.

1935 One can see that the precision for the matched filter case is lower when the thresholds  
1936 are very low, as the noise baseline is slightly amplified, but then rises to high values  
1937 quicker than for the FIR case. The other difference one can spot is that the sensitivity

## CHAPTER 4. MATCHED FILTER APPROACH TO TRIGGER PRIMITIVES

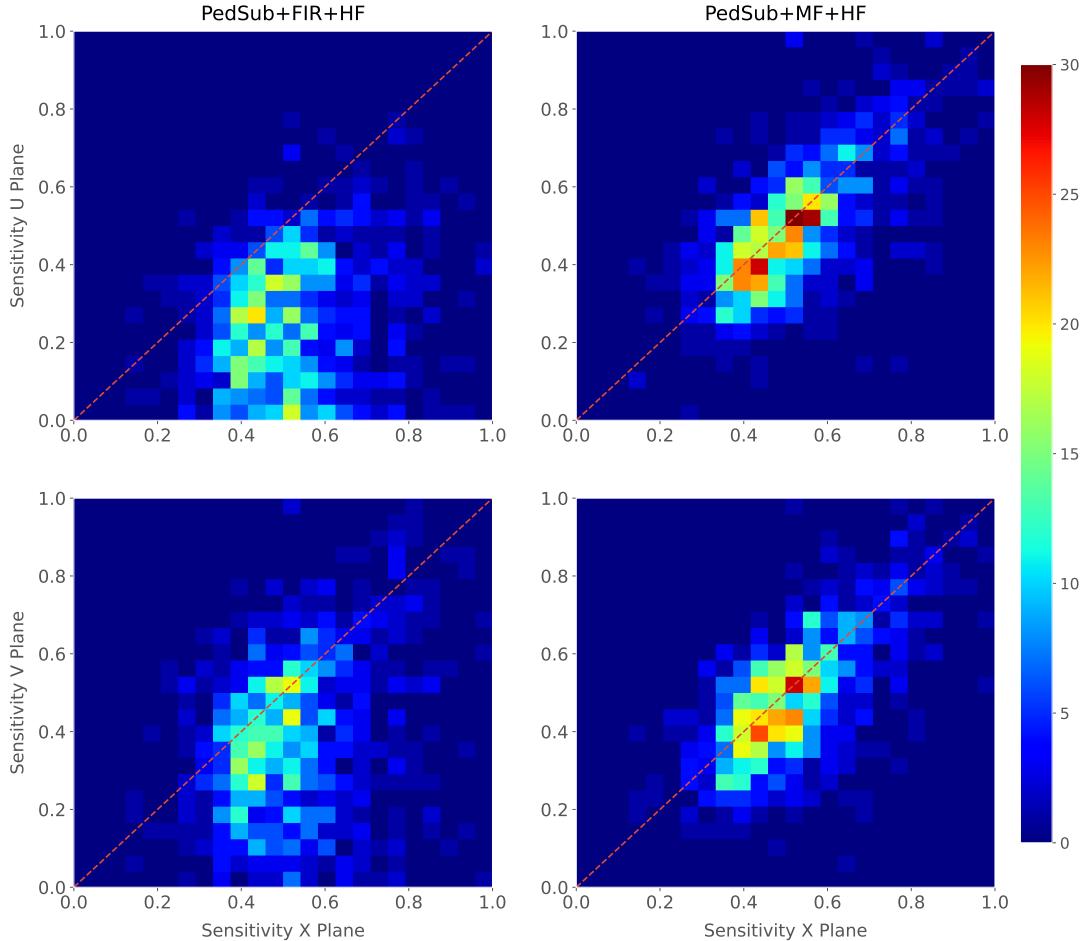


**Figure 4.19:** Dependence of the averaged hit sensitivity on the kinetic energy of the events for the matched filter (blue) and standard (red) hits, for the case of the muon (left panel) and electron (right panel) samples, separated between U (top plots) and V (bottom plots) induction wire planes.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

in the FIR case starts dropping faster at around the same threshold values where the precision stabilizes 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 averaged 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.

## 4.5. USING SIMULATED SAMPLES

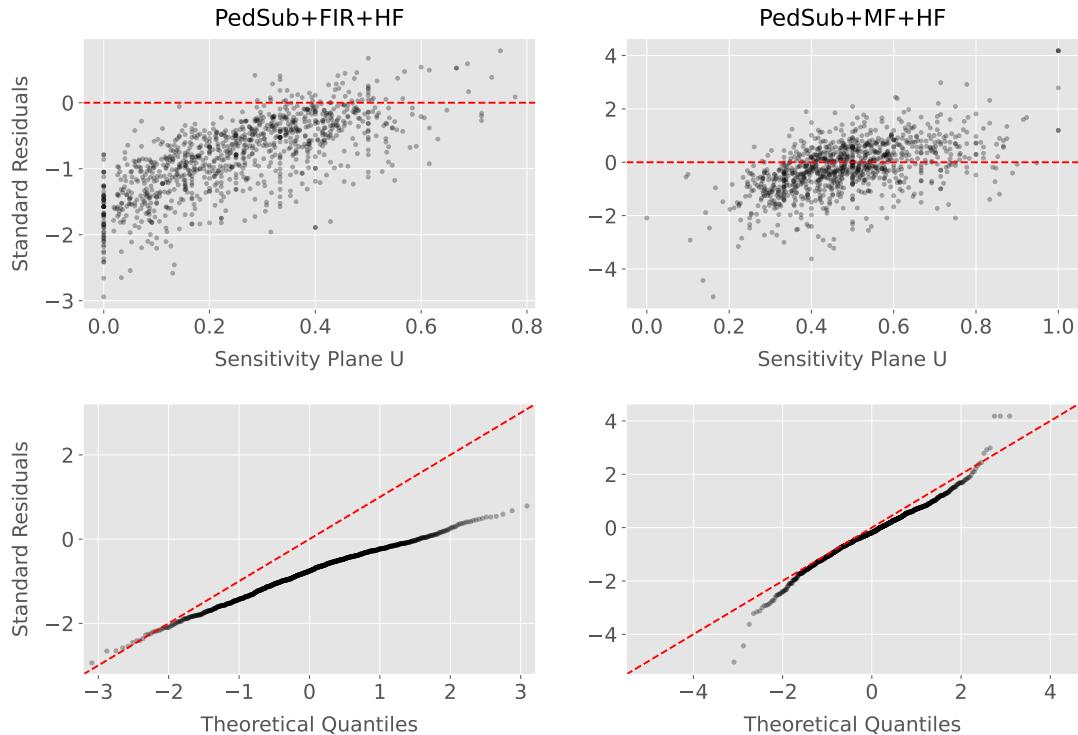


**Figure 4.20:** Distributions of the hit sensitivity in the  $U$  (top panels) and  $V$  (bottom panels) planes versus the hit sensitivity in the  $X$  plane, both for the standard hits (left panels) and the matched filter hits (right panels), in the case of the electron sample and a threshold of 30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

1950 The horizontal lines are placed at the mean value obtained in the fit and represent the  
 1951 width of the  $E_k$  bins used, while the vertical error bars indicate one standard deviation  
 1952 around that mean value. In both cases the threshold used was 30 ADC, as I required  
 1953 the precision to be higher than 0.99 for both matched filter and standard cases.

1954 One can see that, in general, the improvements are better for the  $U$  than for the  $V$   
 1955 plane. While for the  $U$  channels I achieved a mean improvement of 50% and 80% for  
 1956 muons and electrons respectively, the improvement in the  $V$  plane is stalled at 10% and  
 1957 25%. Nevertheless, if I look at the sensitivities for the matched filter hits in both planes  
 1958 one can see these have similar mean values for each energy bin, while on the contrary

## 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.

1959 for the standard hits the sensitivity remains relatively high for the  $V$  plane. This way, it  
 1960 looks there was a less significant gain because the hit sensitivity was already high.

1961 Another interesting observation is the different behaviors for muons and electrons.  
 1962 While hit sensitivity for muons grows significantly with energy, in the case of electrons  
 1963 this slightly decreases the higher the kinetic energy of the event is. In any case, when it  
 1964 comes to the improvement on the sensitivities, this remains almost constant in all cases.

1965 Furthermore, we can look at how the concurrence of hits between the different wire  
 1966 planes has changed. For any given event, I expect to have a similar number of hits in the  
 1967 three planes. As the ionisation electrons need to cross the  $U$  and  $V$  planes prior to reach  
 1968 the collection plane  $X$  they will induce current in those wire planes. A way to check the  
 1969 concurrence of hits across planes is looking at the relation between the hit sensitivities  
 1970 for each individual event. One cannot expect the sensitivities to be exactly equal across

## 4.5. USING SIMULATED SAMPLES

1971 planes, but ideally they should be normally distributed around the diagonal.

1972 Fig. 4.20 shows the hit sensitivity in the U (top panels) and V (bottom panels)  
1973 planes versus the hit sensitivity in the X plane, for the case of the standard hits (left  
1974 panels) and the matched filter hits (right panels). All plots were generated for the  
1975 electron sample and a threshold of 30 ADC. From these one can see a clear trend,  
1976 when I use the standard hit finder chain the sensitivities in the induction planes are  
1977 systematically lower than the hit sensitivity in the X plane, i.e. most of the points sit  
1978 below the diagonal (red dashed line). In contrast, when the matched filters are applied,  
1979 the majority of the events are distributed around the diagonal. This points out that the  
1980 concurrence of hits across planes has improved.

1981 To exemplify the improvement I obtained, one can consider the residuals of the hit  
1982 sensitivities for the X and U planes. Assuming the diagonal hypothesis, i.e. given a  
1983 dataset of the form  $(x, y)$  for any  $x$  I take the predicted  $y$  value to be equal to the value  
1984 of  $x$ , I can compute the standard residuals for the hit sensitivities in U given the ones for  
1985 X. In Fig. 4.21 (top panels) I show these standard residuals against the corresponding  
1986 values of the hit sensitivity in the U plane, for our electron sample with kinetic energy  
1987 between 5 and 100 MeV. If I compare the scatter points in the case of the standard  
1988 hits (left panel) and the matched filter hits (right panel), I see that the residuals of the  
1989 standard hit finder case follow a certain pattern and their mean deviates from 0.

1990 To see clearly if the residuals are normally distributed, in Fig. 4.21 (bottom panels)  
1991 I plot the corresponding quantile-quantile plot for both the standard (left panel) and  
1992 matched filter (right panel) standard residuals. One can clearly see that the points for  
1993 the standard case follow a strongly non-linear pattern, suggesting that the residuals  
1994 do not follow a normal distribution. In contrast, for the matched filter hits the points  
1995 conform to a roughly linear path, implying that in this case the normality condition is  
1996 fulfilled.

1997 All these results hint at the fact that the concurrence of hits across the wire planes  
1998 can be strengthened by applying the matched filters.



## 2000 DM searches with neutrinos from the Sun

2001 The idea of detecting neutrino signals coming from the Sun’s core to probe DM is not  
2002 new. The main focus of these searches has usually been high-energy neutrinos originated  
2003 from DM annihilations into heavy particles [113–116], although recent studies have  
2004 proposed to look at the low-energy neutrino flux arising from the decay of light mesons  
2005 at rest in the Sun [117–120] previously thought undetectable.

2006 In this Chapter I try to demonstrate the capability of DUNE to constrain different  
2007 DM scenarios. I used the neutrino fluxes arising from DM annihilations in the core  
2008 of the Sun to compute the projected limits that DUNE would be able to set on the  
2009 annihilation rates in the Sun and the DM scattering cross sections.

### 2010 5.1 Gravitational capture of DM by the Sun

2011 The Sun and the centre of the Earth are possible sources of DM annihilations, specially  
2012 interesting because of their proximity. Their gravitational attraction ensured the capture  
2013 of DM from the local halo through repeated scatterings of DM particles crossing them.  
2014 Only neutrinos produced from DM annihilations can escape the dense interior of these  
2015 objects. Therefore, neutrino telescopes are the most useful experimental layouts to  
2016 pursue DM searches from their cores.

2017 The neutrino flux from DM annihilations inside the Sun depends on the DM capture  
2018 rate, which is proportional to the DM scattering cross section, and the annihilation rate,  
2019 which is proportional to the velocity-averaged DM annihilation cross-section. The total

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2020 number of DM particles inside the Sun follows the Boltzmann equation [117]:

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

2021 where  $C_{\odot}$  and  $A_{\odot}$  are the total Sun DM capture and annihilation rates respectively.  
 2022 In this expression I neglected the evaporation term, proportional to  $N_{DM}$ , which only  
 2023 contribute for  $m_{DM} \lesssim 4$  GeV [121]. As the current threshold of neutrino telescopes  
 2024 is a few GeV, this region falls below the probed range but can be important in future  
 2025 low-energy projects.

2026 This equation has an equilibrium solution:

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

2027 which represents the amount of DM inside the Sun if the capture and annihilation have  
 2028 reached equilibrium. As the Sun is approximately 4.6 Gyr old, it is usually assumed that  
 2029 equilibrium has been achieved. Therefore, the anomalous neutrino flux from the Sun  
 2030 would only depend on the DM scattering cross section, enabling us to set limits on this  
 2031 quantity. If one does not assume equilibrium, some assumptions on the DM annihilation  
 2032 cross section are necessary to extract predictions from neutrino signals.

2033 Here, I am going to consider three possible scenarios for the DM interactions: DM  
 2034 scattering off electrons, spin-dependent (SD) and spin-independent interactions off nuclei.  
 2035 For the case of these last two, the cross sections will be given in terms of the SD and  
 2036 SI elastic scattering DM cross section off protons (assuming that DM interactions off  
 2037 protons and neutrons are identical),  $\sigma_p^{\text{SD}}$  and  $\sigma_p^{\text{SI}}$ , as [2, 117]:

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

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

2038 where  $\tilde{\mu}_{A_i}$  is the reduced mass of the DM-nucleus  $i$  system,  $\tilde{\mu}_p$  is the reduced mass of

## 5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

2039 the DM-proton system,  $A_i$  and  $J_i$  the mass number and total angular momentum of  
 2040 nucleus  $i$  and  $\langle S_{p,i} \rangle$  and  $\langle S_{n,i} \rangle$  the expectation value of the spins of protons and neutrons  
 2041 averaged over all nucleons, respectively (see Ref. [122] for a review on spin expectation  
 2042 values).

2043 Since the Sun is mainly composed of Hydrogen, the capture of DM from the halo  
 2044 is expected to occur mainly through spin-dependent scattering. However, since the  
 2045 spin-independent cross section is proportional to the square of the atomic mass, heavy  
 2046 elements can contribute to the capture rate (even though they constitute less than 2%  
 2047 of the mass of the Sun). Heavy elements can also contribute to the spin-dependent cross  
 2048 section if the DM has also momentum-dependent interactions.

2049 DM particles can get captured by the Sun if after repeated scatterings off solar  
 2050 targets their final velocity is lower than the escape velocity of the Sun. In the limit of  
 2051 weak cross sections, this capture rate can be approximately written as [2]:

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

2052 where the summation extends over all possible nuclear targets. In this expression,  $R_{\odot}$   
 2053 is the radius of the Sun,  $\rho_{\chi}$  is the local DM density,  $m_{\chi}$  the mass of the DM particle,  
 2054  $f_{v_{\odot}}(u_{\chi})$  the DM velocity distribution seen from the Sun's reference frame,  $R_i^-(\omega \rightarrow v)$   
 2055 is the differential rate at which a DM particle with velocity  $v$  scatters a solar target of  
 2056 mass  $m_i$  to end up with a velocity  $\omega$  and  $|F_i(q)|$  is the nuclear form factor of target  $i$ .

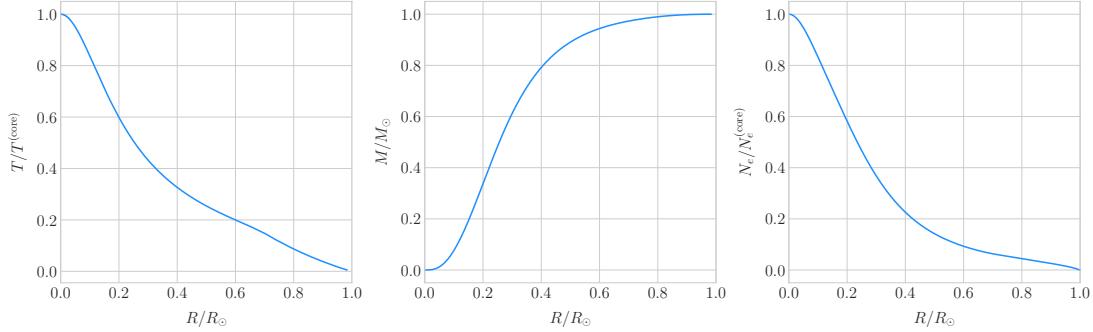
2057 The differential scattering rate takes a rather simple form when considering velocity-  
 2058 independent and isotropic cross sections. In that case, this quantity is given by [2, 123]:

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

2059 where  $\mu_i$  is the ratio between the DM mass and the mass of target  $i$ ,  $\mu_{i,\pm}$  is defined as:

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

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



**Figure 5.1:** Input solar parameters used in our capture rate computation as functions of the Sun's 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 [1].

2060  $n_i(r)$  is the density profile of target  $i$  in the solar medium,  $u_i(r)$  is the most probable  
2061 velocity of target  $i$  given by:

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

2062 where  $T_\odot(r)$  is the temperature of the Sun, the quantities  $\alpha_\pm$  and  $\beta_\pm$  are defined as:

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

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

2063 and the function  $\chi(a, b)$  is a Gaussian integral of the form:

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

2064 Finally, if one assumes the DM halo velocity distribution in the galactic rest frame  
2065 to be a Maxwell-Boltzmann distribution, one can write the halo velocity distribution for  
2066 an observer moving at the speed of the Sun with respect to the DM rest frame as:

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

## 5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

2067 where:

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

2068 is the DM velocity squared,  $v_\odot$  the relative velocity of the Sun from the DM rest frame  
 2069 and  $v_d \simeq \sqrt{3/2}v_\odot$  the velocity dispersion.

2070 For the case of strong scattering cross section, Eq. (5.5) ceases to be valid, as it  
 2071 escalates indefinitely with the cross section. In that limit, the capture rate saturates to  
 2072 the case where the probability of interaction is equal to one, which can be written as:

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

2073 where  $v_d = \sqrt{8/3\pi}v_d$  is the mean velocity in the DM rest frame and the factor  $\xi(v_\odot, v_d)$   
 2074 accounts for the suppression due to the motion of the Sun:

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

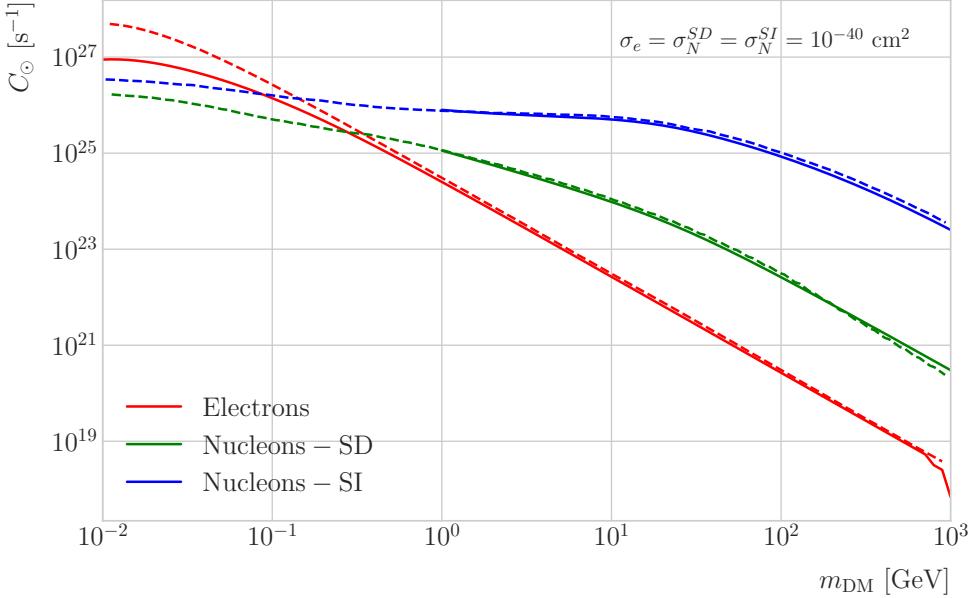
2075 Having these into account, one can write the total capture rate as a combination of  
 2076 both contributions, allowing a smooth transition between the two, as:

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

2077 I computed the capture rate from Eq. (5.16) in the case of interactions with  
 2078 electrons. To do so, I used the standard solar model BS2005-OP [1]. Fig. 5.1 shows the  
 2079 three parameters from the solar model that are needed for the computation, the solar  
 2080 temperature (left panel), mass (central panel) and electron density (right panel) profiles.

2081 For the case of the interactions off nuclei, the computations are more convoluted  
 2082 as one needs to add up the contributions of the different most abundant nuclei in  
 2083 the Sun. Also, in contrast to the electron scenario where the form factor is trivially  
 2084  $|F_e(q)|^2 = 1$ , for any nucleus  $i$  one would need to consider some appropriate nuclear  
 2085 density distribution (either a Gaussian approximation, a Woods-Saxon distribution, etc)

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



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

which would complicate the calculations even further.

That is the reason why, at this stage of our study, I decided to take an alternative approach to the computation of the DM-nucleus capture rates. I used the **DarkSUSY** software, that allows us to compute these quantities performing a full numerical integration over the momentum transfer of the form factors. The default standard solar model used by **DarkSUSY** is BP2000<sup>1</sup> [124].

In Fig. 5.2 I show the results I obtained for the capture rates, for the case of interactions off electrons (red solid line), SD (green solid line) and SI (blue solid line) interactions of nucleons. In all cases I used a value of the scattering cross sections of  $\sigma_i = 10^{-40} \text{ cm}^2$ . Note here one of the limitations of the **DarkSUSY** approach, one can not extend the computation below  $m_{\text{DM}} = 1 \text{ GeV}$ . Nevertheless, this is not something to worry about in this case, as I will discuss next. As a comparison, I added also the

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

## 5.1. GRAVITATIONAL CAPTURE OF DM BY THE SUN

values computed in Ref. [2] (same color scheme, dashed lines). One can see there is good agreement between these and the **DarkSUSY** computation of the SD and SI interactions for  $m_{\text{DM}} \geq 1$  GeV. In this regime their computations also matches quite well our result for the electron capture rate. However, these start to differ significantly below  $m_{\text{DM}} = 1$  GeV, being their estimate up to a factor of 5 bigger than ours for low masses.

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

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

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

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

and  $\kappa$  is defined as:

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

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

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

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

In this way, one can define the evaporation mass as the mass for which the number of DM particles in equilibrium approaches Eq. (5.20) at 10% level:

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

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

It was reported in Ref. [2] that, in the case of both SD and SI DM interactions

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

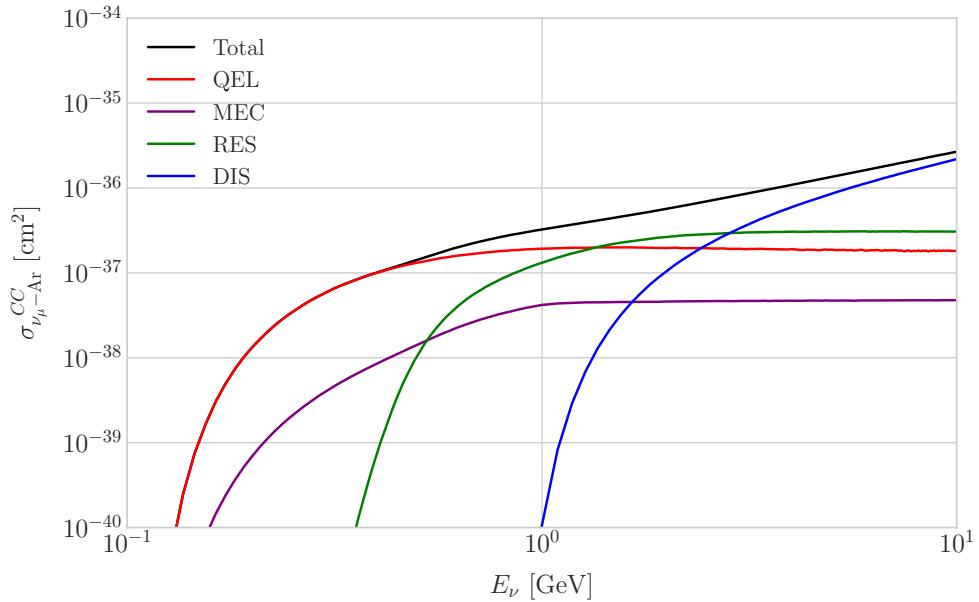
off nuclei, this value ranges from 2 to 4 GeV depending on the specific scattering cross section value, compatible with the usual assumptions in the literature. What is interesting is the case of the electron capture. It was found that, when one applies a cutoff in the velocity distribution of the DM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

## 5.2 Neutrino flux from DM annihilations

When WIMPs annihilate inside the Sun a flux of high-energy neutrinos is expected from heavy quarks, gauge bosons and  $\tau^+\tau^-$  final states, which decay before losing energy in the dense solar medium, as they will produce a continuum spectra up to  $E_\nu \sim m_\chi$  (in the case of direct annihilation to neutrinos one would have a line at  $E_\nu = m_\chi$ ) [118]. This kind of signal has been extensively studied in the literature, allowing to put strong limits on the SD WIMP-proton cross section for large  $m_\chi$ . However, the number of high-energy neutrinos per WIMP annihilation is small and the spectrum depends on the unknown final state. Moreover, background rejection is easier for large  $m_\chi$  but neutrinos with  $E_\nu \gtrsim 100$  GeV are significantly attenuated by interactions in the Sun.

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 [117]. 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  $E_\nu = 236$  MeV  $\nu_\mu$  while in the case of pions one would have a  $E_\nu = 29.8$  MeV  $\nu_\mu$ . In practice only  $K^+$  and  $\pi^+$  contribute to these signals, as  $K^-$  and  $\pi^-$  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

### 5.3. COMPUTING LIMITS FROM SOLAR NEUTRINO FLUXES



**Figure 5.3:** *NuWro* computed  $\nu_\mu - {}^{40}\text{Ar}$  charged-current scattering cross section as a function of the neutrino energy  $E_\mu$ . 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).

2144 pion decays, leptonic decays of other hadrons and heavy leptons or even directly from  
 2145 WIMP annihilations, which can decay at rest and contribute to the previous low-energy  
 2146 neutrino flux with a well known spectrum below 52.8 MeV.

2147 These monoenergetic MeV neutrinos were previously considered undetectable but,  
 2148 due to the large yield, the known spectra and the modern advances in the detector  
 2149 technology, these low-energy neutrino flux can be a good probe of the SD WIMP-proton  
 2150 cross-section in standard solar WIMP capture scenario, as it is sensitive to low WIMP  
 2151 masses and insensitive to the particular final state. A good place to look for these signals  
 2152 are next-generation neutrino experiments such as DUNE.

### 2153 5.3 Computing limits from solar neutrino fluxes

2154 In order to use the neutrino fluxes from DM annihilations in the Sun, the first thing I  
 2155 need to do is to determine the expected number of atmospheric background events, for

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2156 a given exposure, after directionality selection has been applied. I can write this number

2157 as:

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

2158 where  $\eta_B$  is the background efficiency,  $E_{min}$  and  $E_{max}$  the minimum and maximum

2159 energies to integrate over,  $d^2\Phi_{atm}^\mu/dE_\nu d\Omega$  the differential flux of atmospheric muon

2160 neutrinos,  $A_{eff}^{(\mu)}$  is the effective area of DUNE to muon neutrinos and  $T$  is the exposure

2161 time. The effective area can be expressed as the product of the neutrino-nucleus scattering

2162 cross section and the number of nuclei in the fiducial volume of the detector. This way

2163 for DUNE I can write:

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

2164 where  $\sigma_{\nu-\text{Ar}}^{(\mu)}$  is the  $\nu_\mu - {}^{40}\text{Ar}$  charged-current scattering cross section. In Fig. 5.3 I

2165 show the computed value of this cross section as a function of the neutrino energy  $E_\nu$ ,

2166 in the range of interest both for the atmospheric background and signal events. It was

2167 computed using the NuWro Monte Carlo neutrino event generator [125], including the

2168 charged-current contributions of the quasi-elastic scattering (red line), resonant pion

2169 exchange (green line), deep inelastic scattering (blue line) and meson exchange current

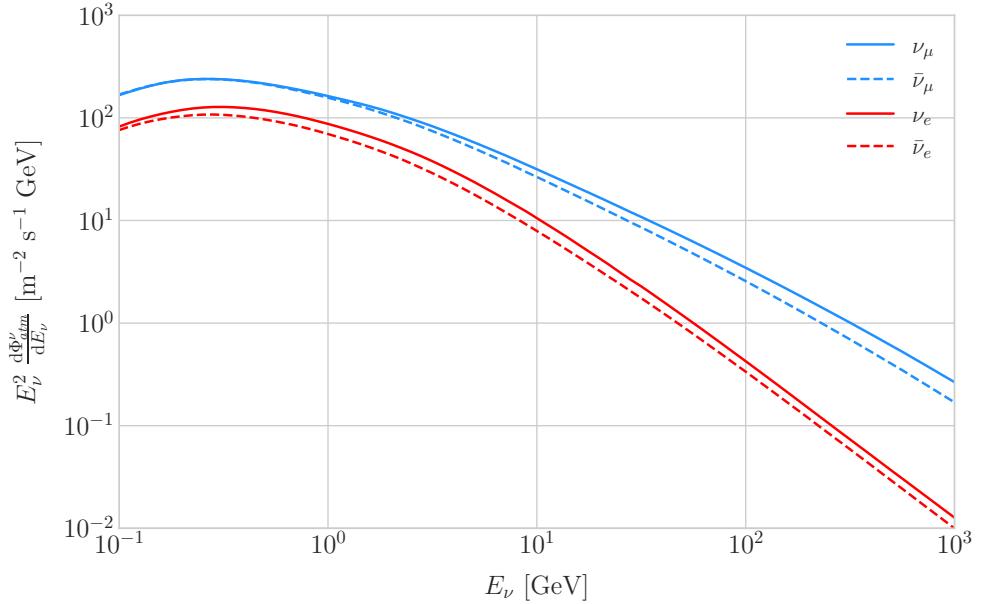
2170 (purple line).

2171 The background rejection will depend on the resolution of the detector and the  
 2172 selection one applies on the events. A geometry argument can be used to estimate  
 2173 the maximum background rejection one can achieve in this case, considering one can  
 2174 efficiently discriminate all events coming from a direction different from that of the  
 2175 Sun. In that case, the optimal background efficiency will simply be the relative angular  
 2176 coverage of the Sun. Taking the angular diameter of the Sun as seen from the Earth to  
 2177 be  $0.5^\circ$ , I have:

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

2178 This value will give a very optimistic estimate of the number of background events.

### 5.3. COMPUTING LIMITS FROM SOLAR NEUTRINO FLUXES



**Figure 5.4:** Expected atmospheric neutrino flux as a function of the neutrino energy  $E_\nu$  at Homestake at solar minimum, taken from Ref. [3]. The blue solid (dashed) line correspond to muon neutrinos (antineutrinos) and the red solid (dashed) line correspond to electron neutrinos (antineutrinos).

2179 However, it can be regarded as an lower limit, as it represents the best case scenario.

2180 In Fig. 5.4 I show the fluxes of atmospheric neutrinos at the Homestake mine during  
 2181 solar minimum, taken from Ref. [3]. The values are averaged over the two angular  
 2182 directions. In blue I have the flux of muon neutrinos while in red I indicate the flux  
 2183 of electron neutrinos. Additionally, the dashed lines correspond to both antineutrino  
 2184 species.

2185 Using these values for the muon neutrino and the corresponding total CC cross  
 2186 section, one can compute the number of expected background events by integrating over  
 2187 the given energy range (as in this case the angular integral is trivial). As for the energy  
 2188 range to integrate over, I choose the range for DUNE specified in [85],  $E_{min} = 10^{-1}$  GeV  
 2189 and  $E_{max} = 10$  GeV. Taking all these into account, I found the number of background  
 2190 events to be:

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

2191 In order to estimate the sensitivity of DUNE to this kind signal, one can consider a

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2192 hypothetical data set where the number of observed neutrinos is taken to be the expected  
 2193 number of background events rounded to the nearest integer,  $N_{obs} = \text{round}(N_B)$  [126].  
 2194 Now, if I assume that the number of signal and background events seen by DUNE are  
 2195 given by Poisson distributions with means equal to the expected number of signal and  
 2196 background events,  $N_S$  and  $N_B$ , one can denote by  $N_S^{90}$  to the number of expected  
 2197 signal events such that the probability of having an experimental run with a number of  
 2198 events greater than  $N_{obs}$  is 90%. This number can be obtained as the numerical solution  
 2199 to the equation:

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

2200 where  $\Gamma(x, y)$  is the upper incomplete gamma function.

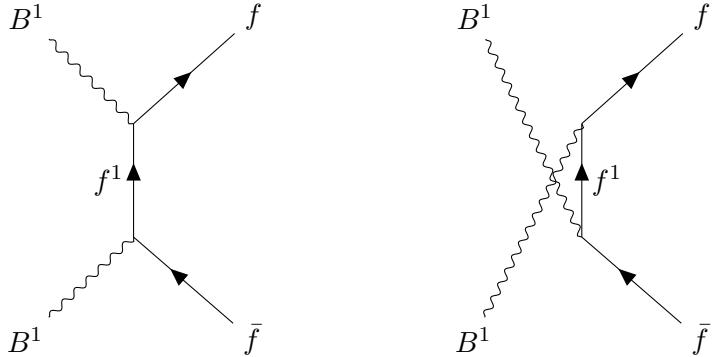
2201 The number of signal events is related to the neutrino flux from DM annihilations in  
 2202 a similar way as the background events to the atmospheric neutrino flux. In this case I  
 2203 have:

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

2204 where  $\eta_S$  is the signal efficiency,  $\Gamma_A^{eq}$  is the total annihilation rate of DM particles at  
 2205 equilibrium,  $\Gamma_A^{eq} = A_\odot (N_{DM}^{eq})^2$ ,  $z_{min}$  and  $z_{max}$  the minimum and maximum relative  
 2206 energies to integrate over (in such a way that  $z_{min,max} \leq E_{min,max}/m_{DM}$  for each  $m_{DM}$ )  
 2207 and  $dN_\nu/dAdN_A dz$  the muon neutrino flux per DM annihilation in the Sun.

2208 Knowing  $N_S^{90}$  one can use the relation in Eq. (5.27) to obtain  $\Gamma_A^{eq,90}$  for different  
 2209 values of the DM mass. From there I can directly translate those values into the  
 2210 upper limits for DUNE on the DM scattering cross sections, for a given exposure. The  
 2211 relation between the annihilation rate and the DM-nucleon cross section comes from the  
 2212 equilibrium condition through the solar DM capture rate. The details of the evolution  
 2213 of the number of DM particles inside the Sun and the computation of the capture rates  
 2214 are discussed in App. 5.1.

## 5.4. EXAMPLE: KALUZA-KLEIN DARK MATTER



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

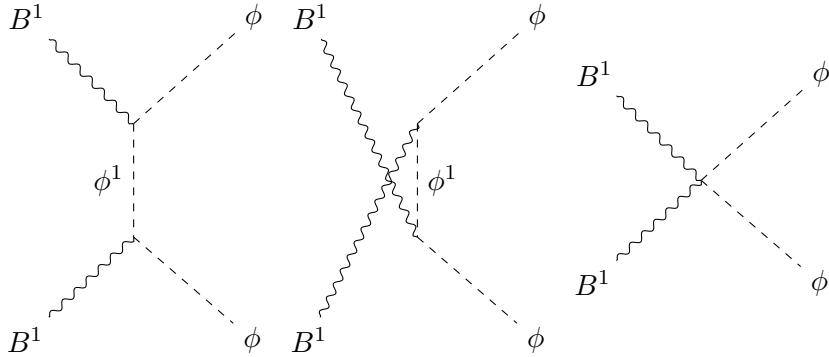
### **2215 5.4 Example: Kaluza-Klein Dark Matter**

2216 Even though there are plenty of BSM theories which provide viable dark matter  
 2217 candidates, Kaluza-Klein type of models [127, 128] within the universal extra dimensions  
 2218 (UED) paradigm naturally predict the existence of a massive, stable particle that can  
 2219 play the role of the dark matter. In the UED scenario all the SM fields can propagate  
 2220 in one or more compact extra dimensions [129], as opposed to the idea of brane worlds  
 2221 [130, 131], where just gravity can propagate in the bulk while SM particles live at fixed  
 2222 points.

2223 Furthermore, in UED there is no violation of the translational invariance along the  
 2224 extra dimensions, thus leading to degenerate KK modes masses and also the conservation  
 2225 of the KK number in the effective four dimensional theory. At loop level, radiative  
 2226 corrections and boundary terms shift the masses of the KK modes and break KK  
 2227 number conservation into a KK parity. As a result, this theory only contains interactions  
 2228 between an even number of odd KK modes and therefore the lightest among the first KK  
 2229 excitations will be stable. This particle is usually denoted as the lightest Kaluza-Klein  
 2230 particle (LKP) and its mass is proportional to  $1/R$ , being  $R$  the size of the extra  
 2231 dimension.

2232 A viable DM candidate needs to be electrically neutral and non-baryonic, therefore  
 2233 good candidates among the first Kaluza-Klein excitations would be the KK neutral

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



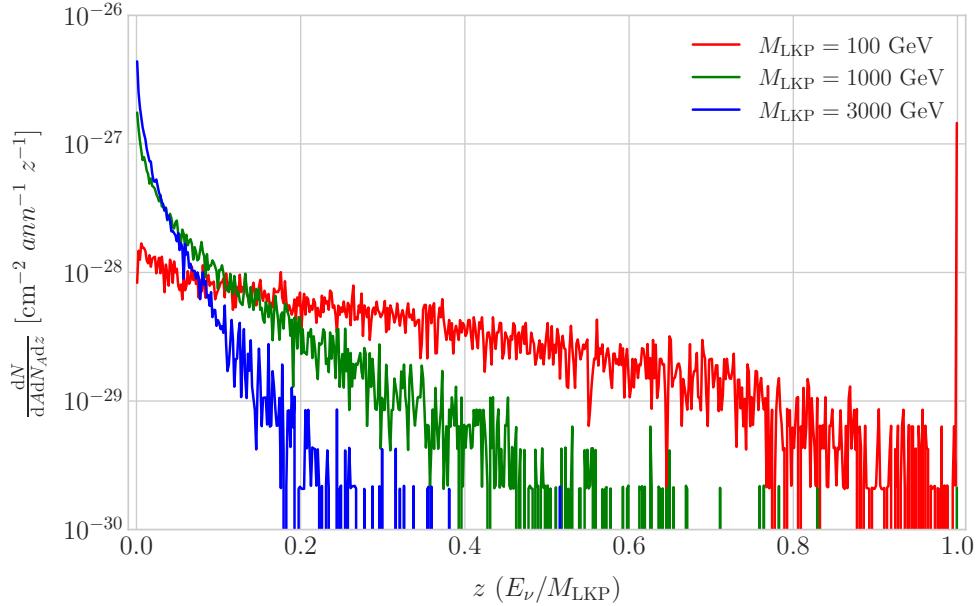
**Figure 5.6:** Feynman diagrams for  $B^1B^1$  annihilation into a Higgs boson pair.

gauge bosons and the KK neutrinos [132]. Another possible candidate is the first KK excitation of the graviton, which receives negligible radiate contributions and therefore has a mass almost equal to  $1/R$ , but it has been shown that the lightest eigenstate from the mixing of the gauge mass states  $(B^1, W_3^1)$  would be lighter, as  $B^1$  and  $W_3^1$  receive negative radiate corrections [133]. It is also understood that, when these corrections become sizeable, the eigenstates become approximately pure  $B^1$  and  $W_3^1$  states as the Weinberg mixing angle grows small with the KK number [133]. In that case, the LKP can be well-approximated as being entirely  $B^1$ .

I need to compute the neutrino flux produced by the annihilations of the LKP in the core of the Sun, taking into account their propagation in the solar medium, as well as neutrino oscillations. To this end I used `WimpSim` [134, 135] to generate one million annihilation events in the Sun over a time span of four years and propagate them to the DUNE FD location ( $44^\circ 20' \text{ N}, 103^\circ 45' \text{ W}$ ), for different values of  $M_{\text{LKP}}$ . In Fig. 5.7 I show the obtained muon neutrino spectra arriving to the detector from LKP annihilations in the Sun, per unit area and per annihilation, plotted in relative energy units for different values of the mass. As one could expect the spectra get steeper the higher is the mass, due to the absorption of high-energy neutrinos in the solar medium. Also, one can see the peak at  $z = 1$  due to the direct annihilation into neutrinos  $\chi\chi \rightarrow \nu\bar{\nu}$ .

Now, one can estimate the sensitivity of DUNE to this particular model by using

#### 5.4. EXAMPLE: KALUZA-KLEIN DARK MATTER



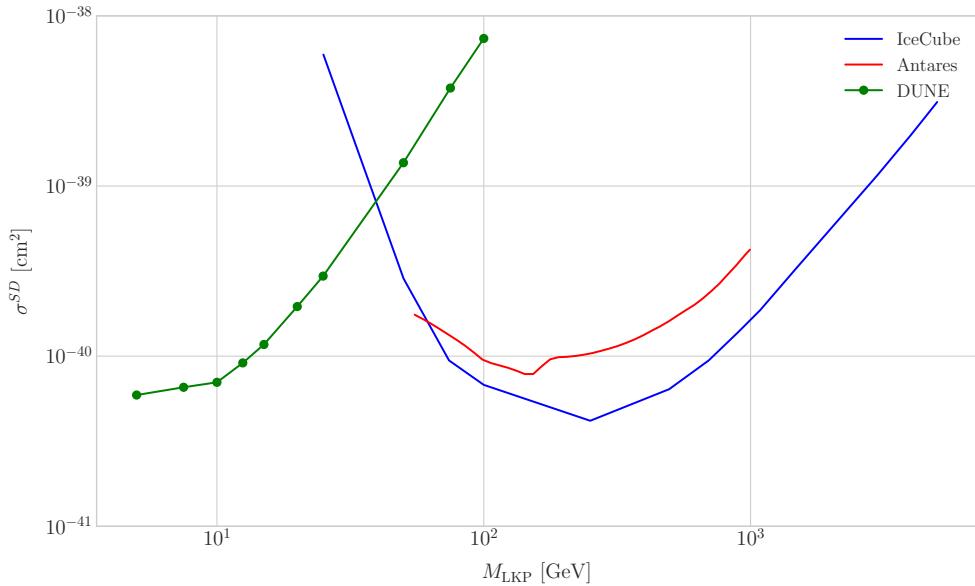
**Figure 5.7:** Computed spectra of muon neutrinos at the DUNE FD site from  $B^1$  annihilations in the Sun for three different values of  $M_{\text{LKP}}$ , plotted in relative energy units for legibility.

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

In Fig. 5.8 I show the projected sensitive for DUNE on the spin-dependent  $B^1$ -proton  
 scattering cross section versus the mass of the DM particle, for a exposure of 400 kT yr  
 (green dots). I also include the previous results from IceCube [4] (blue line) and Antares  
 [5] (red line). The shaded area represents the disfavoured region from combined searches  
 for UED by ATLAS and CMS [6].

From the experimental point of view, this estimation lacked a detailed simulation of  
 the detector response and thus this must be consider as a mere optimistic sensitivity

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



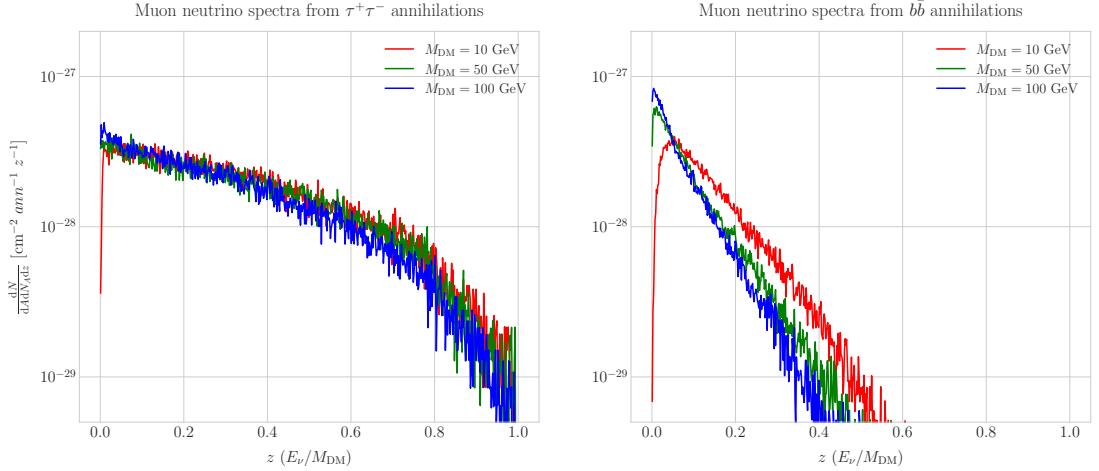
**Figure 5.8:** Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent  $B^1$ -proton scattering cross section as a function of  $M_{LKP}$  (green dots). I also show the previous limits from IceCube [4] (blue line) and Antares [5] (red line) on the LKP cross section. The shaded area represents the disfavoured region (at 95% confidence level) on the mass of the LKP from LHC data [6].

computation. However, it shows the potential of DUNE to constrain this kind of exotic scenarios, showing the region where it will be in a position to compete with other neutrino telescopes. A more detailed analysis is needed if I am to make a realistic estimation. Even though the region of the parameter space where DUNE would be sensitive to this particular model is quite constrained by collider searches [6] and other rare decay measurements [136, 137], it still constitutes an alternative indirect probe.

## 5.5 High energy DM neutrino signals

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

## 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),  $50 \text{ GeV}$  (green line) and  $100 \text{ GeV}$  (blue line), plotted in relative energy units.

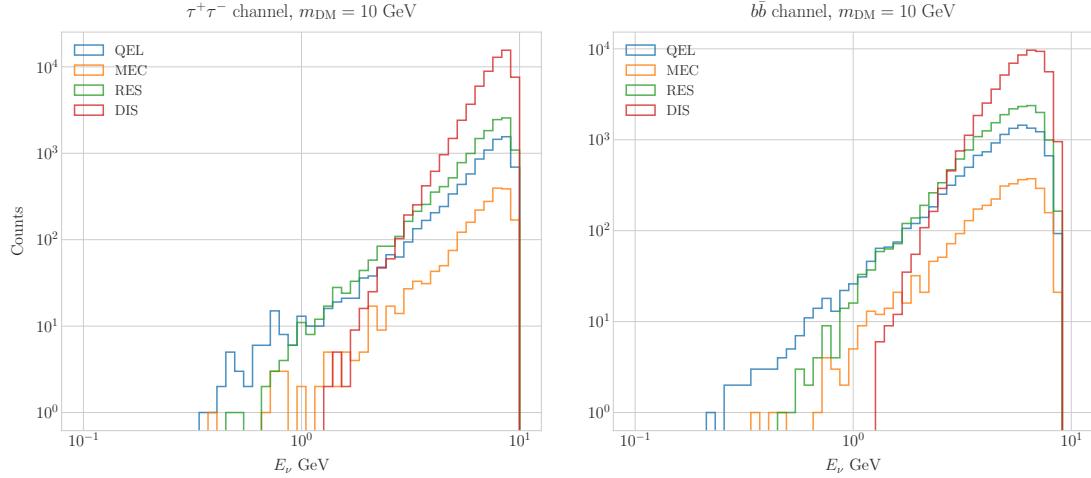
2282 fluxes. Other channels, like  $W^+W^-$  and  $ZZ$ , are open for more massive WIMPs, but  
 2283 those will produce usually a higher energy neutrino flux that will be out of reach for  
 2284 DUNE (usually the maximum neutrino energy is taken to be  $E_{\max} = 10 \text{ GeV}$ ).

2285 In Fig. 5.9 I show the `WimpSim` [134, 135] generated muon neutrino spectra at the  
 2286 DUNE FD location ( $44^\circ 20' \text{ N}, 103^\circ 45' \text{ W}$ ) from  $\tau^+\tau^-$  (left panel) and  $b\bar{b}$  (right panel)  
 2287 annihilations in the core of the Sun, for different DM masses. Here, one can clearly see  
 2288 the meaning of the previous distinction between hard and soft channels. For the same  
 2289 DM mass value, the muon neutrino spectrum from the  $\tau^+\tau^-$  channel is more flat and  
 2290 reaches higher energies than the one from the  $b\bar{b}$  channel, which drops faster.

2291 In this case, I prepared two sets of files, one for  $\tau^+\tau^-$  and the other for  $b\bar{b}$ , for DM  
 2292 masses in the range from 5 to 100 GeV (actually for  $b\bar{b}$  the first mass point I took is  
 2293  $7.5 \text{ GeV}$ , as a WIMP with  $m_{\text{DM}} = 5 \text{ GeV}$  can not kinematically self annihilate into  $b\bar{b}$ ).  
 2294 Then, I prepared the `WimpSim` output fluxes in a specific way to use them as inputs to  
 2295 `NuWro`, which simulates the neutrino interaction with the argon.

2296 Because `WimpSim` outputs an event list together with the fluxes, I can use the former  
 2297 to generate the events. The direction of these is given in terms of the azimuth and  
 2298 altitude angles viewed from the specified location, so first I need to convert these into the

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



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

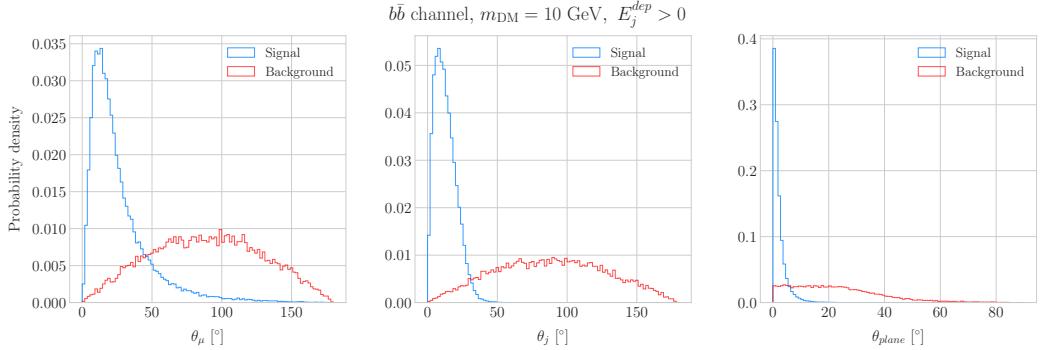
2299 DUNE FD coordinates. Once I have done it, each event can be processed with `NuWro`.  
 2300 To increase the number of samples and optimise the computation time, I generate 100  
 2301 interactions (i.e. `NuWro` events) for each `WimpSim` event<sup>2</sup>. I restrict the event generation  
 2302 to charged current interactions, but I allow all the different contributions to the CC  
 2303 cross section, i.e. quasielastic scattering (QEL), meson exchange current process (MEC),  
 2304 resonant pion production (RES) and deep inelastic scattering (DIS). I just take into  
 2305 account the CC contribution because I am only interested in final states with charged  
 2306 leptons, as we have better chances of reconstructing the kinematics of CC events.

2307 For the atmospheric fluxes I follow a similar procedure, only that this time I do not  
 2308 have a set of events but the fluxes binned in azimuth and altitude angles. This way, I  
 2309 transform these to DUNE coordinates and process the fluxes for each bin separated with  
 2310 `NuWro`.

2311 At this point, I have two sets of events with different energies and final states.  
 2312 In Fig. 5.10 one can see the distribution of the muon neutrino energies for the case  
 2313  $m_{\text{DM}} = 10$  GeV, both for the  $\tau^+\tau^-$  (left panel) and  $b\bar{b}$  (right panel) channels, separated

<sup>2</sup>This 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 5.11:** Distributions of  $\theta_\mu$  (left panel),  $\theta_j$  (central panel) and  $\theta_{\text{plane}}$  (right panel) for the  $b\bar{b}$  sample with  $m_{\text{DM}} = 10$  GeV (blue) and the atmospheric background (red).

by interaction. One can clearly see that there are different energy regimes where the primary interaction type is different. This leads to a plurality of event topologies, therefore making it difficult to implement a general approach to the selection of events in detriment of the background. As a way to proceed, I decided to split our samples, based on the different interaction modes and contents of the final state, into a CC DIS sample and a single proton CC QEL sample.

### 5.5.1 DIS events

To begin with, I consider the high energy part of the spectrum. In this region DIS events dominate, i.e. interactions of the form  $\nu_\mu + q_d(\bar{q}_u) \rightarrow \mu^- + q_u(\bar{q}_d)$ . Therefore, our final states will contain a muon and a hadronic jet from the fragmentation of the outgoing quark. As all these events have  $E_\nu \gtrsim 1$  GeV the momentum transfer to the remnant nucleus is negligible, for this reason the neutrino energy can be effectively reconstructed just taking into account the momenta of the muon and the jet. This technique was successfully used in Ref. [138] to select monoenergetic DM solar neutrino events from  $\nu\bar{\nu}$  annihilation channels.

Using momentum conservation one sees that the plane generated by the momenta of the muon and the jet needs to also contain the momentum of the neutrino. As we are interested in neutrinos coming from the Sun, the momentum of the neutrino can be regarded as known beforehand. This will allow us to define the angle of the outgoing

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2333 muon and jet with respect to the incoming neutrino. Moreover, one can also use that  
 2334 information to reject poorly reconstructed jets, checking for deviations of these from the  
 2335 momentum conservation plane.

2336 To account for the limited angular resolution of the detector, I smeared the momenta  
 2337 of the muons and hadrons. In a liquid argon TPC muons are expected to be tracked with  
 2338 high precision, therefore I take the associated angular resolution to be  $1^\circ$ . In the case of  
 2339 jets, it is expected that for the hadrons dominating the cascade a detector like DUNE  
 2340 has an angular resolution between  $1^\circ$  to  $5^\circ$  [85], so I take the latter, more conservative,  
 2341 estimate.

2342 As a first selection step, I will just take into account particles with kinetic energies  
 2343 above the detection threshold of DUNE. For muons and photons the specified threshold  
 2344 energy is 30 MeV, for charged pions 100 MeV and for other hadrons 50 MeV [85]. This  
 2345 way, if the outgoing muon in a certain event has an energy lower than the required  
 2346 threshold I will drop such event. For the case of hadrons and photons, I will only require  
 2347 to have at least one particle above the energy threshold, so then one can compute the  
 2348 jet momentum using the (smeared) momenta of the  $N$  particles above threshold as:

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

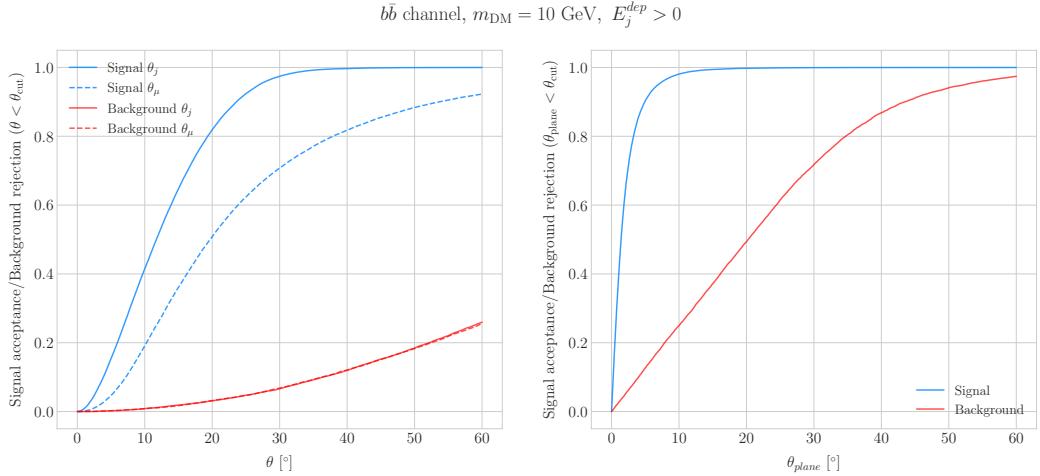
2349 Additionally, I will also define an estimation of the deposited hadronic energy as:

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

2350 This quantity is useful to select events with enough hadronic visible energy in the  
 2351 detector. For events where most of the hadronic energy is scattered across plenty of  
 2352 hadrons with individual energies below the detection threshold, this estimation will  
 2353 give  $E_j^{dep} \leq 0$ . In these cases it could be expected that the jet momentum is poorly  
 2354 reconstructed, and therefore I require events to pass the cut  $E_j^{dep} > 0$ .

2355 For the events I can compute the angles for the muon and jet with respect to the

## 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.

2356 incoming neutrino as:

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

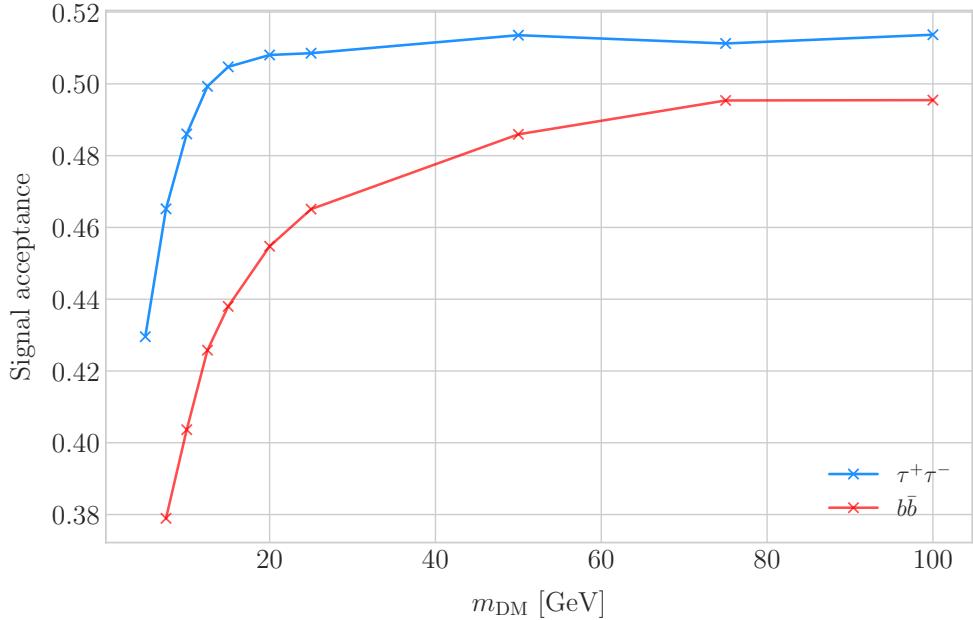
$$\cos \theta_j = \hat{p}_\nu \cdot \hat{p}_j, \quad (5.31)$$

2357 and the deviation from the momentum conservation plane as:

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

2358 In Fig. 5.11 I show some distributions of these quantities for the case of the  $b\bar{b}$  sample  
 2359 with  $m_{DM} = 10$  GeV (blue histograms) and for the atmospheric backgrounds (red).  
 2360 In order to select the atmospheric events I followed the same criteria as for the signal  
 2361 events. However, because in the signal case I used the true direction of the neutrino  
 2362 as input, as it should be that of the Sun at that time and therefore known, in the  
 2363 atmospheric case I used a set of solar positions as our ansatz for the neutrino direction.  
 2364 From the distributions, one can see that the muon and the jet for the signal events are  
 2365 predominantly forward and also that the deviations from the momentum conservation

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



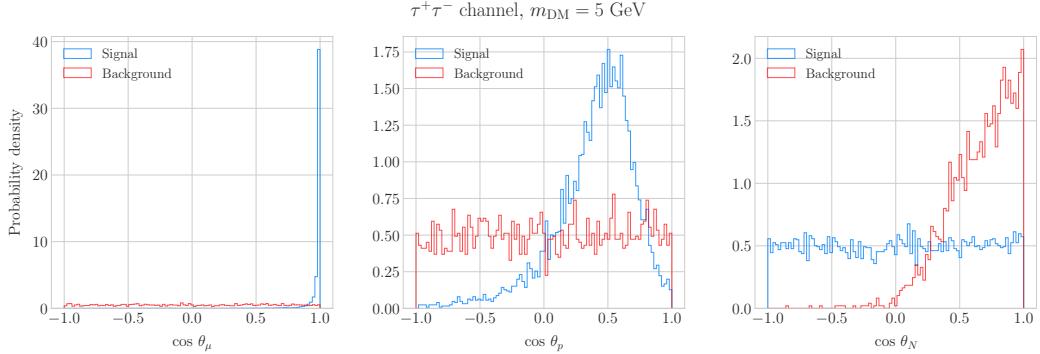
**Figure 5.13:** Signal efficiencies for the  $\tau^+\tau^-$  (blue line) and  $b\bar{b}$  (red line) DIS samples as functions of the DM mass,  $m_{\text{DM}}$ , obtained by applying the optimal angular cuts  $\theta_\mu < 27^\circ$ ,  $4^\circ < \theta_j < 26^\circ$  and  $\theta_{\text{plane}} < 3.5^\circ$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

2366 plane are peaked at zero, as one should expect.

2367 Now, I can start applying cuts to maximise our signal selection efficiency while at  
 2368 the same time I try to minimise the amount of atmospheric background events passing  
 2369 the selection. To this end, I will need to find some lower and upper cuts for  $\theta_j$  and  
 2370  $\theta_\mu$  and an upper bound for  $\theta_{\text{plane}}$ . In Fig. 5.12 I show how upper bound cuts in the  
 2371 different angular variables affect the signal efficiency (blue lines) and the background  
 2372 rejection (red lines). Notice that the signal efficiency behaves in a quite different way  
 2373 when I apply cuts in the jet and the muon angles. On the contrary, the cuts on both  
 2374 variables have a similar effect on the background rejection.

2375 In order to obtain the optimal set of cuts, I perform a multidimensional scan. I  
 2376 do this separately for the  $\tau^+\tau^-$  and the  $b\bar{b}$  samples. For each case, I scan the possible  
 2377 cuts for each mass point and then I take the mean value of the signal efficiency for  
 2378 each configuration, to get the mean efficiency for each set of cuts. I do a similar scan  
 2379 for the atmospheric sample independently. Then, I take the sets of cuts such that  
 2380 the background rejection achieved is greater than 99.8% and search for the one which

## 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  $\tau^+\tau^-$  QEL sample with  $m_{\text{DM}} = 5 \text{ GeV}$  (blue) and the atmospheric background (red).

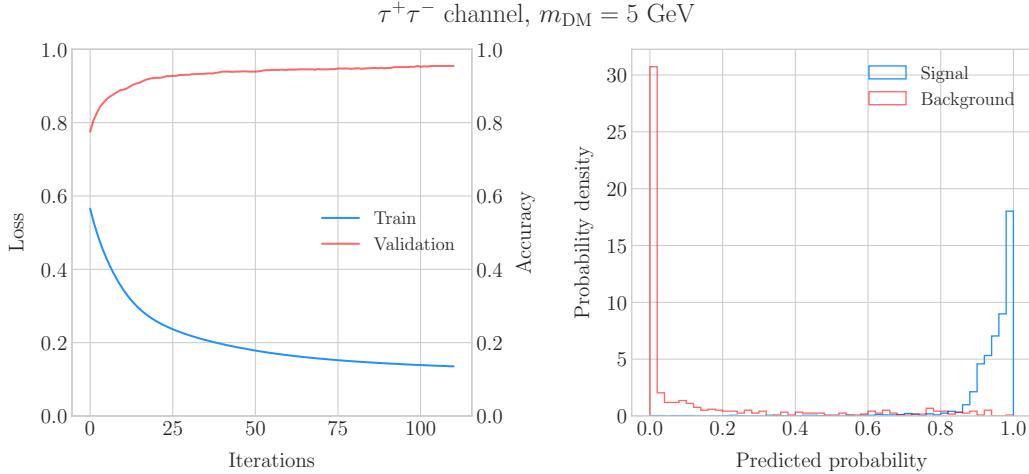
2381 maximises the  $\tau^+\tau^-$  and  $b\bar{b}$  sample mean efficiencies. I found that with the cuts  $\theta_\mu < 27^\circ$ ,  
 2382  $4^\circ < \theta_j < 26^\circ$  and  $\theta_{\text{plane}} < 3.5^\circ$  I get a background rejection of 99.80% while achieving  
 2383 a 49.40% and 44.92% mean signal efficiencies for the  $\tau^+\tau^-$  and  $b\bar{b}$  signals respectively.

2384 In Fig. 5.13 I show the signal efficiencies as a function of the DM mass for the  $\tau^+\tau^-$   
 2385 (blue line) and the  $b\bar{b}$  (red line) DIS events, after applying the cuts discussed above, as  
 2386 well as the energy threshold and hadronic visible energy selections. One can see that  
 2387 the efficiency grows with the mass, as annihilations of more massive DM particles will  
 2388 produce a neutrino spectrum centered at higher energies, where DIS events dominate.  
 2389 Notice also that the efficiency is higher for the  $\tau^+\tau^-$  case at every mass point, as in  
 2390 general this channel produces neutrinos at higher energies than the corresponding  $b\bar{b}$   
 2391 channel.

### 2392 5.5.2 Single proton QEL events

2393 Now, one can try to explore the low energy tail of the neutrino energy distributions. This  
 2394 regime is dominated by the QEL interactions, i.e. events of the type  $\nu_\mu + n \rightarrow \mu^- + p$ .  
 2395 In this case, as the typical energies are  $E_\nu \lesssim 1 \text{ GeV}$ , the momentum transfer to the  
 2396 remnant nucleus is sizeable. Therefore, I can not make the approximation I did before  
 2397 and assume that the momentum of the muon and the proton will give an adequate  
 2398 estimation of the reconstructed neutrino energy.

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



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

2399 In any case, as before, I can take the direction of the incoming neutrino as known.  
 2400 That way, one can estimate the energy of the neutrino as:

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

2401 and using momentum conservation I can write the momentum of the remnant nucleus  
 2402 as:

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

2403 As in the previous case, I need to drop the events where the muon or the proton fall  
 2404 below the kinetic energy detection threshold [85]. Also, I again apply a smearing to the  
 2405 momenta of the particles, a 1% for muons and 5% for protons.

2406 Having done that, one can compute the following angular variables for our selected  
 2407 events:

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

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

## 5.5. HIGH ENERGY DM NEUTRINO SIGNALS

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

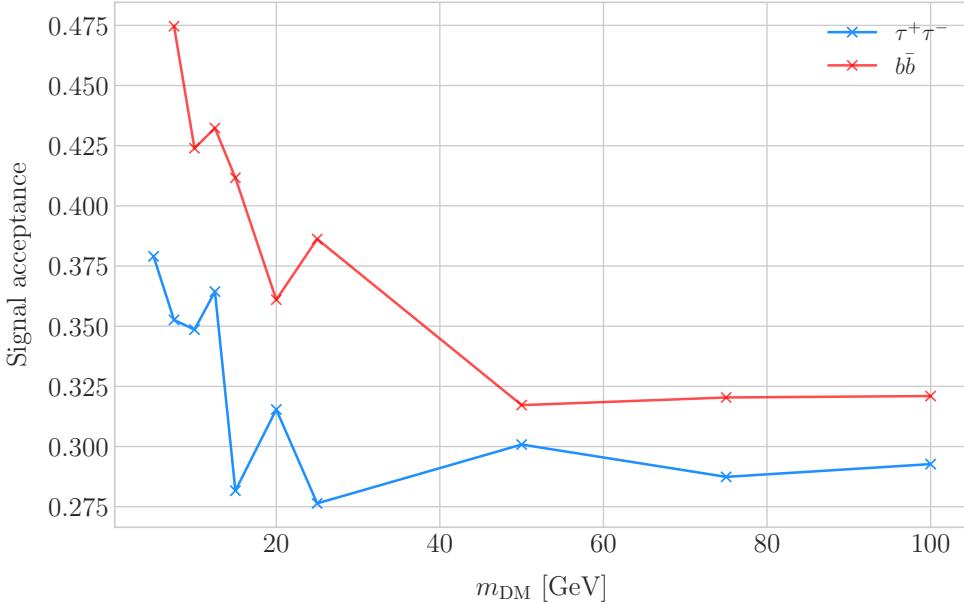
2408     Fig. 5.14 shows the distributions of these angular variables for the  $\tau^+\tau^-$  QEL  
 2409    sample with  $m_{\text{DM}} = 5$  GeV (blue) and the atmospheric background (red). Again, for  
 2410    the atmospheric events I used a random solar position as the ansatz for the incoming  
 2411    neutrino direction. Notice that now, opposed to the DIS case where the signal had very  
 2412    sharp distributions for the variables considered, the shapes of the angular distributions  
 2413    for signal and background are not that much different.

2414     This effectively means that the usual approach of applying simple angular cuts would  
 2415    not work as well as in the previous situation. Therefore, as a possible solution, I tried to  
 2416    use a multilayer perceptron (MLP) classifier to separate between signal and background  
 2417    events. Thus, the power of the hypothesis test will serve as an estimate of the signal  
 2418    efficiency, and in the same way one can take the size of the test to be our background  
 2419    rejection.

2420     For each DM mass value and channel, as well as for the background sample, I divide  
 2421    our events into training, validation and test samples. The input variables for the classifier  
 2422    were the reconstructed neutrino energy from Eq. (5.33) and the angular variables defined  
 2423    in Eqs. (5.35 - 5.37). I used the MLP classifier implemented in `scikit-learn` [139], with  
 2424    a total of five hidden layers, the rectified linear unit activation function and adaptive  
 2425    learning rate. In order to account for fluctuations due to artifacts in the training process I  
 2426    repeated the training a thousand times for each sample, redefining each time the training,  
 2427    validation and test subsets, so one can take as our signal efficiency and background  
 2428    rejection the mean values of the powers and sizes of the tests.

2429     The results of one of these training processes for the  $\tau^+\tau^-$  QEL signal with  $m_{\text{DM}} =$   
 2430    5 GeV is shown in Fig. 5.15. On the left panel I show the loss function values (blue) and  
 2431    accuracy (red) at each iteration for the training and the validation samples respectively.  
 2432    The training stops either when the maximum number of iterations is reached (1000 in  
 2433    this case) or when the accuracy for the validation sample reaches a certain tolerance  
 2434    (I chose  $10^{-4}$  as our tolerance). On the right panel I have the distributions for the

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

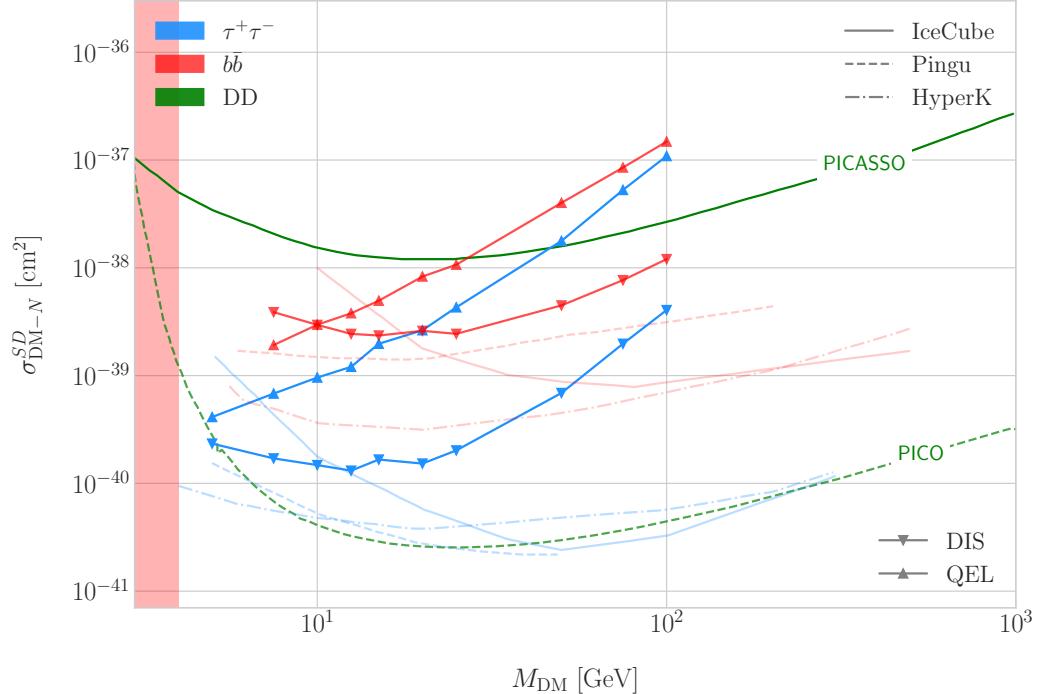


**Figure 5.16:** Signal efficiencies for the  $\tau^+\tau^-$  (blue line) and  $b\bar{b}$  (red line) single proton QEL samples as functions of the DM mass,  $m_{\text{DM}}$ , obtained by requiring a minimum predicted probability from the MLP classifier of 0.97 in order to achieve a background rejection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

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

## 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  $m_{\text{DM}}$ , for the annihilation channels  $\tau^+\tau^-$  (blue) and  $b\bar{b}$  (red) separated by interaction type (up triangles denote DIS interactions whereas down triangles represent QEL interactions). I also show the previous limits from IceCube [7] (solid lines) and the projected sensitivities for Pingu [8] (dashed lines) and Hyper-Kamiokande [9] (dash-dotted lines), as well as the direct detection limits from PICASSO [10] (solid green line) and PICO-60 C<sub>3</sub>F<sub>8</sub> [11] (dashed green line).

### 2449 5.5.3 Results

2450 In order to estimate the DM-nucleon cross section sensitivities in the present case I need  
 2451 again to compute the expected number of background events. As I am now separating  
 2452 events by interaction type Eq. (5.25) does not hold anymore, as in that case I integrated  
 2453 over the total neutrino-argon cross section. In this instance, the expected background  
 2454 events for DIS events is approximately given by:

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

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2455 whereas for QEL events we have:

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

2456 Now, using these together with Eqs. (5.26) and (5.27) one can obtain the 90% C.L.  
2457 upper limit on the total annihilation rate at equilibrium for both kind of events. Then,  
2458 applying the computed DM-nucleons capture rates I can translate these into limits on  
2459 the DM-nucleon cross section by means of Eqs. (5.2), (5.5) and (5.6).

2460 Fig. 5.17 shows the obtained limits on the SD DM-nucleon cross section for DUNE,  
2461 using the DIS (up triangles) and QEL (down triangles) events both for the  $\tau^+\tau^-$  (blue)  
2462 and the  $b\bar{b}$  (red) samples, for an exposure of 400 kT yr. I also include the corresponding  
2463 current limits from IceCube [7] (solid lines), as well as the projected sensitivities of  
2464 Pingu [8] (dashed lines) and Hyper-Kamiokande [9] (dash-dotted lines). For comparison,  
2465 I also show the reported direct detection limits from PICASSO [10] (solid green line)  
2466 and PICO-60  $C_3F_8$  [11] (dashed green line).

2467 Notice that, for most of the mass range, the limits one can set by using the DIS  
2468 events are stronger than those of the QEL interactions, except for the low mass part  
2469 of both the  $\tau^+\tau^-$  and the  $b\bar{b}$  curves where the QEL events dominate. In general, the  
2470 expected sensitivity of DUNE for DM masses  $\lesssim 25$  GeV surpasses the stronger current  
2471 indirect limits. However, experiments like Hyper-Kamiokande are foreseen to have an  
2472 overall better sensitivity in this kind of searches, as they have a bigger active volume  
2473 and accept a broader energy range.

2474 A pending question is what happens when we add the RES and MEC charged-current  
2475 interaction contributions. In that case it would probably be more convenient to split  
2476 the samples by final state interaction topologies. Also, another necessary improvement  
2477 would be adding a full detector simulation and reconstructions. This will also require  
2478 considering the effect of poorly reconstructed events or final states containing neutral  
2479 particles such that they mimic the desired topology at the reconstruction level.

## 5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

### **2480 5.6 Example: Leptophilic Dark Matter**

**2481** In general, the capture rate of DM particles by the Sun via interactions with electrons is  
**2482** several orders of magnitude smaller than the capture via DM-nucleus scattering. Thus,  
**2483** it would be sub-leading even when nucleon capture is loop suppressed. As I showed in  
**2484** Fig. 5.2, the capture rate via scattering off electrons only surpasses the capture rates  
**2485** via DM-nucleons interactions for DM masses  $\lesssim 100 - 500$  MeV.

**2486** However, if one considers a model where DM-nucleon interactions are forbidden even  
**2487** at loop level, then electron interactions will be the sole contributor to DM capture in  
**2488** the Sun. One can describe such scenario where the DM particles couple to leptons but  
**2489** not to the quark sector using effective operators.

**2490** In general, assuming that the DM particle is a Dirac fermion, the dimension six  
**2491** operators describing the interaction between two DM particles and two leptons can be  
**2492** written as:

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

**2493** where  $G = 1/\Lambda^2$  is the effective coupling strength,  $\Lambda$  the cut-off of the effective field  
**2494** theory and  $\ell$  denotes any lepton. In principle, one should consider all the possible  
**2495** Lorentz structures  $\Gamma_f^i$  in order to have a complete set of effective operators.

**2496** However, some combinations will induce interactions with nucleons at loop level.  
**2497** As we are specifically interested in interactions which forbid any communication with  
**2498** the quark sector, I will not consider those [140]. In addition, some of the effective  
**2499** operators give rise to velocity-suppressed scattering cross sections between DM particles  
**2500** and leptons. I will also neglect those, as the suppression goes with the square of the DM  
**2501** halo velocity which in units of the speed of light is  $\sim 10^{-6}$ .

**2502** This way, the only Lorentz tensor structure that do not induce interactions with  
**2503** quarks at loop level and gives a contribution to the scattering cross section that is not  
**2504** velocity suppress is the axial-axial interaction. The effective Lagrangian is then given

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

2505 by:

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

2506 where  $c_A^\chi$  and  $c_A^\ell$  are the couplings for the different species. As the DM coupling appears  
 2507 as a common factor for any lepton choice, I will redefine the corresponding coupling  $c_A^\ell$   
 2508 to absorb  $c_A^\chi$ . Also, for simplicity, I will assume that the couplings between the DM  
 2509 particles and the leptons are flavour independent, i.e. I have just two couplings,  $c_A^e$  for  
 2510 charged leptons and  $c_A^v$  for neutrinos.

2511 In the case of a scalar DM particle, the lowest order effective interaction with  
 2512 leptons happens through a dimension five operator, generating scalar and pseudoscalar  
 2513 interactions. However, the former induces interactions with quarks at two loop level  
 2514 whereas the latter gives a velocity suppressed scattering cross section.

2515 From the effective Lagrangian in Eq. (5.41) it can be shown that the axial-axial  
 2516 contribution to the scattering cross section for the fermionic DM and a charged lepton  
 2517 is given by:

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

2518 If the DM interacts exclusively with fermions, then the only annihilation channels  
 2519 that will give us a measurable neutrino flux coming out of the Sun are  $\tau^+ \tau^-$  and  $\nu \bar{\nu}$ . The  
 2520 former channel, already explored previously in the more mainstream scenario of the DM  
 2521 capture via scattering off nucleons, is open only for  $m_{DM} > m_\tau \simeq 1776.86 \pm 0.12$  MeV  
 2522 [141], a mass region where the solar DM capture by electrons is at least one order of  
 2523 magnitude smaller than the capture via interactions with nucleons. On the contrary, the  
 2524 latter allows us to explore a region where the capture rate via scattering off electrons  
 2525 dominates over the rest.

2526 One downside of focusing in such low mass range is that it falls below the usual  
 2527 limit of  $m_{evap} \sim 4$  GeV usually explored in the literature. The pretext to explore this  
 2528 region is the result discussed previously reported in Ref. [2], where DM evaporation in  
 2529 the Sun for the case of capture via electron scattering could be negligible for masses  
 2530 as low as  $m_{evap} \sim 200$  MeV. This result is quite sensitive to the high velocity tail of

## 5.6. EXAMPLE: LEPTOPHILIC DARK MATTER

2531 the DM velocity distribution in equilibrium inside the Sun, and therefore full numerical  
 2532 simulations would be needed to asses the impact of this effect. However, this falls out of  
 2533 the scope of our work.

2534 In this case, as I have an specific realisation of the interaction between the DM  
 2535 and leptons, one can estimate the relic density of our DM for different values of the  
 2536 couplings and the effective field theory scale  $\Lambda$ . The first step to do so is compute the  
 2537 self-annihilation cross section. Because I consider cold relics, at the freeze-out time our  
 2538 DM particles were non-relativistic and so one can expand the annihilation cross section  
 2539 in terms of the relative velocity  $v$  between two annihilating DM particles as [142]:

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

2540 where the sum includes all the possible lepton final states with mass  $m_{\ell}$ .

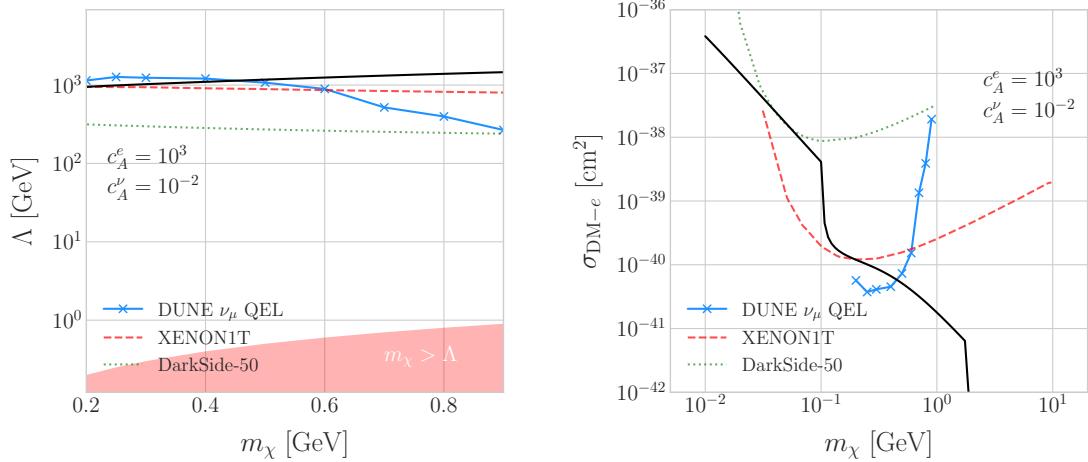
2541 Solving the Boltzmann equation for the evolution of the DM density gives as a  
 2542 solution a relic density of:

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

2543 where  $x_F = m_{\chi}/T_F$  being  $T_F$  the freeze-out temperature,  $g_*$  the number of relativistic  
 2544 degrees of freedom at freeze-out and  $a$  and  $b$  the terms in the annihilation cross section  
 2545 expansion  $\sigma_{ann}|v| \approx a + bv^2 + \mathcal{O}(v^4)$ . Using the current best fit for the relic DM density  
 2546  $\Omega_{\chi} h^2 = 0.1198 \pm 0.0012$  [143] one can use these relations to compute the required  
 2547 effective theory scale  $\Lambda$  at which the correct density is achieved for any combinations of  
 2548  $m_{\chi}$  and  $c_A^{\ell}$ .

2549 As discussed before, in the low DM mass region QEL interactions dominate. Moreover,  
 2550 if I focus on direct annihilation to neutrinos, the energy of the muon neutrino flux is  
 2551 known as it must be equal to the mass of the DM particle,  $E_{\nu} = m_{\chi}$ . That way, now  
 2552 I do not need to use Eq. (5.33) in order to estimate the momentum transfer to the

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN



**Figure 5.18:** Left panel: Projected 90% confidence level sensitivity of DUNE (400 kT yr) to the scale  $\Lambda$  of an EFT containing only leptophilic DM axial-axial interactions (blue line). Right panel: . In both cases the corresponding limits from DarkSide-50 [12] (dotted green line) and XENON1T [13] (dashed red line) are also shown, together with the configurations for which the correct relic density is achieved (black line), all for the coupling values  $c_A^e = 10^3$  and  $c_A^\nu = 10^{-2}$ .

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 features to train the classifier with, because the characteristic monoenergetic flux for each  $m_\chi$  value will help to distinguish between signal and background events.

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

## 5.7. SYSTEMATIC UNCERTAINTIES

Now, estimating the number of background events using Eq. (5.39) one can go on and apply Eqs. (5.26) and (5.27) together with Eq. (5.42) to derive the sensitivity of DUNE to this kind of model. Fig. 5.18 (left panel) shows the potential reach of DUNE to constrain the EFT scale  $\Lambda$  this model containing only leptophilic DM axial-axial interactions (blue line), for a choice of couplings  $c_A^e = 10^3$  and  $c_A^\nu = 10^{-2}$ . I also included the current limits on the DM-electron scattering cross section from DarkSide-50 [12] (dotted green line) and XENON1T [13] (dashed red line), reworked with Eq. (5.42) to show their implications for the EFT scale. The values of  $\Lambda$  for which the correct DM relic density value is achieved for each mass are also shown (black line). This tells us that, for that specific choice of couplings, DUNE would be sensitive to DM configurations allowed by the relic density constraint up to a mass of  $m_\chi \sim 400$  MeV.

In Fig. 5.18 (right panel) I show the same upper limits but for the DM-electron scattering cross section. From this view one can see that DUNE would be able to offer complementary information to the low energy DM-electron interaction searches performed by direct detection experiments, in a slightly higher mass range.

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

## 5.7 Systematic uncertainties

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

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

**Table 5.1:** Systematic uncertainties for the solar WIMP signal events. Table adapted from Ref. [144].

Systematic	Value
Form factor	Does not apply to SD [145]
Solar model	3% [145]
Local DM density	Not relevant for relative interpretations [145, 146]
Dynamics of solar system	Negligible [147]
Velocity distributions	20% at 20 GeV [145, 146]
Oscillation parameters	8% for $\tau^+\tau^-$ , 5% for $b\bar{b}$ [148]
Neutrino interactions in the Sun	10%
Matter effects in the Earth	10%

### 5.7.1 Systematic uncertainties in the solar WIMP signal

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

- **Uncertainties on the annihilation rate.** These include the astrophysical effects 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 [144, 146, 148].

## 5.7. SYSTEMATIC UNCERTAINTIES

**Table 5.2:** Systematic uncertainties for the solar WIMP atmospheric background events. Table adapted from Ref. [51].

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/ $\pi$ ratio	5% $E_\nu \leq 100$ GeV

**2610 5.7.2 Systematic uncertainties in the atmospheric background**

**2611** For the atmospheric background events, one needs to take into account the systematic  
**2612** uncertainties affecting the atmospheric  $\nu_\mu$  flux. These have been extensively studied  
**2613** in the context of atmospheric neutrino oscillation measurements. Among these, the  
**2614** energy-dependent flux normalisation uncertainty is the in the low energy regime. Other  
**2615** important contributions to the uncertainty come from the ratios between the muon to  
**2616** electron neutrino and the muon to anti-muon neutrino components of the flux. Additional  
**2617** uncertainty is introduced by the errors in the pion and kaon production rates calculated  
**2618** for the hadronic interactions of cosmic rays in the atmosphere [149].

**2619** Table 5.2 shows a summary of the leading contributions to the uncertainty on the  
**2620** atmospheric muon neutrino flux, in the energy range relevant for this analysis.

**2621 5.7.3 Common systematic uncertainties**

**2622** Finally, there are sources of uncertainty common to both signal and backgrounds. These  
**2623** have two different origins:

- 2624** • **Uncertainties on the neutrino cross section.** These are introduced by the  
**2625** modelling of the neutrino-nucleus interactions. In the context of the solar WIMP  
**2626** analysis, these have been estimated to be 10% for DM masses around 10 GeV  
**2627** [148].

## CHAPTER 5. DM SEARCHES WITH NEUTRINOS FROM THE SUN

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

2634

2635

## Particle ID in ND-GAr

2636 ND-GAr is a magnetised, high-pressure gaseous argon TPC (HPgTPC), surrounded by  
2637 an electromagnetic calorimeter (ECal) and a muon detector (commonly refer to as  $\mu$ ID).  
2638 A detailed discussion on the requirements, design, performance and physics of ND-GAr  
2639 can be found in the DUNE ND CDR [91] and the ND-GAr whitepaper (cite).

2640 In DUNE Phase II ND-GAr will fulfill the role of TMS, measuring the momentum  
2641 and sign of the charged particles exiting ND-LAr. Additionally, it will be able to measure  
2642 neutrino interactions inside the HPgTPC, achieving lower energy thresholds than those  
2643 of the ND and FD LArTPCs. By doing so ND-GAr will allow to constrain the relevant  
2644 systematic uncertainties for the LBL analysis even further.

2645 The goal of the present chapter is to review the requirements that the physics program  
2646 of DUNE impose on ND-GAr, present the current status of its design and describe the  
2647 GArSoft package, its simulation and reconstruction software.

2648 As decided during the DUNE Phase II workshop in June 2023 [reference], we want  
2649 to build ND-GAr physics case by showing:

- 2650 • That ND-GAr can constrain systematic uncertainties that ND-LAr might miss.
- 2651 • The impact on the neutrino oscillation results if such systematic uncertainties are  
2652 missed.
- 2653 • That ND-GAr is necessary to reach DUNE's main physics goals.

2654 This way, the design of ND-GAr will be physics driven.

## CHAPTER 6. PARTICLE ID IN ND-GAr

2655 In order to study the effects of final state interactions (FSI) in CC interactions,  
2656 ND-GAr should be able to measure the spectrum of protons and charged pions at low  
2657 energies. ND-GAr also needs to be able to measure the pion multiplicity, specially for  
2658 energies above 100 MeV as at these energies the pions shower in the LAr, to inform the  
2659 pion mass correction in the ND and FD LArTPCs.

2660 In order to correctly identify electrons, muons, pions, kaons and protons ND-GAr  
2661 can use a combination of:  $dE/dx$  measurements in the HPgTPC,  $E_{ECAL}/p$  using the  
2662 ECAL total energy and the momentum obtained from magnetic spectroscopy in the  
2663 HPgTPC and penetration information through the ECAL and muon tagger.

### 2664 6.1 GArSoft

2665 GArSoft is a software package developed for the simulation and reconstruction of events  
2666 in ND-GAr. It is inspired by the LArSoft toolkit used for the simulation of LArTPC  
2667 experiments, like the DUNE FD modules. It is based on `art`, the framework for event  
2668 processing in particle physics experiments [150]. Other of its main dependencies are  
2669 `ROOT`, `NuTools`, `GENIE` and `Geant4`. It allows the user to run all the steps of a generation-  
2670 simulation-reconstruction workflow using FHiCL configuration files.

#### 2671 6.1.1 Event generation

2672 The standard generator FHiCLs in GArSoft run the event generation and particle  
2673 propagation simulation (i.e. Geant4) in the same job by default. However, it is possible  
2674 to split them up if needed. The current version of GArSoft provides five different event  
2675 generators, each of them producing `simb::MCTruth` products defined in `NuTools`. The  
2676 available modules are:

- 2677 • **SingleGen**: particle gun generator. It produces the specified particles with a given  
2678 distribution of momenta, initial positions and angles.
- 2679 • **TextGen**: text file generator. The input file must follow the `hepevt` format<sup>1</sup>, the

---

<sup>1</sup>In brief, each event contains at least two lines. The first line contains two entries, the event number

## 6.1. GARSOFT

2680 module simply copies this to `simb::MCTruth` data products.

2681 • **GENIEGen**: GENIE neutrino event generator. The module runs the neutrino-nucleus  
2682 interaction generator using the options specified in the driver FHiCL file (flux file,  
2683 flavour composition, number of interactions per event,  $t_0$  distribution, ...). Current  
2684 default version is v3\_04\_00.

2685 • **RadioGen**: radiological generator. It produces a set list of particles to model  
2686 radiological decays. Not tested.

2687 • **CRYGen**: cosmic ray generator. The module runs the CRY event generator with a  
2688 configuration specified in the FHiCL file (latitude and altitude of detector, energy  
2689 threshold, ...). Not tested.

2690 The module **GArG4** searches for all the generated `simb::MCTruth` data products, using  
2691 them as inputs to the Geant4 simulation with the specified detector geometry. A constant  
2692 0.5 T magnetic field along the drift coordinate is assumed. The main outputs of this step  
2693 are `simb::MCParticle` objects for the generated Geant4 particles, `gar::EnergyDeposit`  
2694 data products for the energy deposits in the HPgTPC and `gar::CaloDeposit` data  
2695 products for the energy deposits in the ECal and muon system.

### 2696 6.1.2 Detector simulation

2697 The standard detector simulation step in GArSoft is all run with a single FHiCL, but  
2698 the different modules can be run independently as well. First the `IonizationReadout`  
2699 module simulates the charge readout of the HPgTPC, and later the `SiPMReadout` module  
2700 runs twice, once for the ECal and then for the muon system, with different configurations.

2701 The `IonizationAndScintillation` module collects all the `gar::EnergyDeposit`  
2702 data products, to compute the equivalent number of ionization electrons for each energy

---

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.

## CHAPTER 6. PARTICLE ID IN ND-GAr

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

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

### 6.1.3 Reconstruction

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

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

## 6.1. GARSOFT

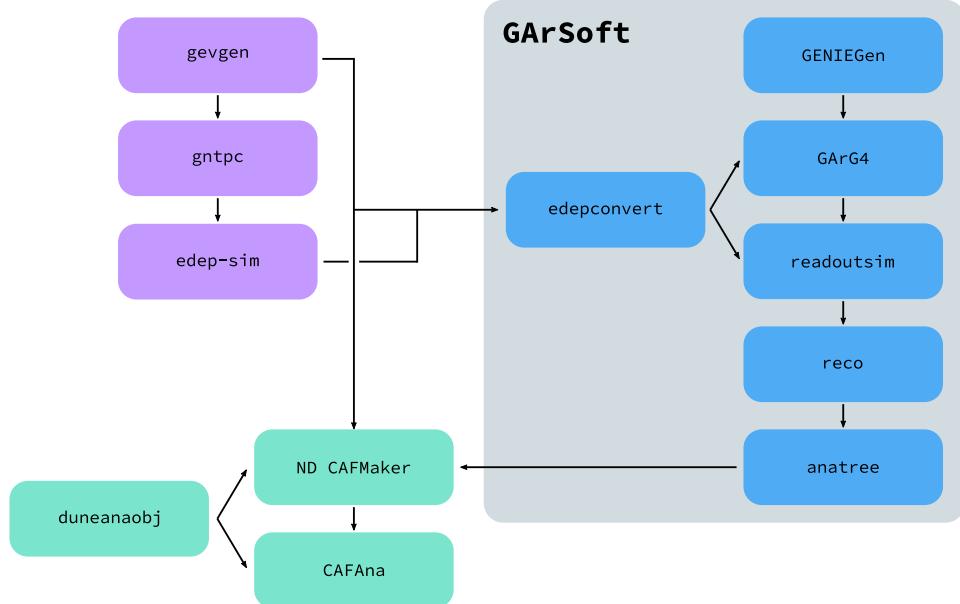
2731 hits.

2732 The following step prior to the track fitting is pattern recognition. The module  
2733 called `tpcvechitfinder2` uses the `gar::rec::TPCClusters` data products to find track  
2734 segments, typically called vector hits. They are identified by performing linear 2D fits  
2735 to the positions of the clusters in a 10 cm radius, one fit for each coordinate pair. A  
2736 3D fit defines the line segment of the vector hit, using as independent variable the one  
2737 whose sum of (absolute value) slopes in the 2D fits is the smallest. The clusters are  
2738 merged to a given vector hit if they are less than 2 cm away from the line segment. The  
2739 outputs are `gar::rec::VecHit` data products, as well as associations to the clusters. The  
2740 `tpcpatrec2` module takes the `gar::rec::VecHit` objects to form the track candidates.  
2741 The vector hits are merged together if their direction matches, their centers are within  
2742 60 cm and their direction vectors point roughly to their respective centers. Once  
2743 the clusters of vector hits are formed they are used to make a first estimation of the  
2744 track parameters, simply taking three clusters along the track. The module produces  
2745 `gar::rec::Track` data products and associations between these tracks and the clusters  
2746 and vector hits.

2747 The track is fitted by means of a Kalman filter in the `tpctrackfit2` module, using  
2748 the position along the drift direction as the independent variable. Two different fits are  
2749 performed per track, a forward and a backwards fit, each starting from one of the track  
2750 ends. The Kalman filter state vector ( $y, z, R, \phi, \tan\lambda$ ) is estimated at each point along  
2751 the track using a Bayesian update. The track parameters reported in the forward and  
2752 backwards fits are the ones computed at the opposite end where the fit started. The  
2753 main outputs of the track fit are the `gar::rec::Track` objects. Additionally, the module  
2754 stores the fitted 3D positions along the track in the `gar::rec::TrackTrajectory` data  
2755 products and the total charge and step sizes for each point also get stored in the form of  
2756 `gar::rec::TrackIonization` objects.

2757 After the tracking step, the `vertexfinder1` module looks at the reconstructed  
2758 `gar::rec::Track` products, creating vertex candidates with the track ends that are  
2759 within 12 cm of each other. The vertices are then fitted using linear extrapolations from

## CHAPTER 6. PARTICLE ID IN ND-GAr



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

2760 the different track ends associated. The results are `gar::rec::Vertex` data products,  
 2761 and associations to the tracks and corresponding track ends.

2762 For the ECal and muon tagger, the `SiPMHitFinder` module runs twice with different  
 2763 configurations, adapted to the particular capabilities of both. The module simply takes  
 2764 the `gar::raw::CaloRawDigit` products, applies a calibration factor to convert the ADC  
 2765 counts to MeV and for the strip layer hits it calculates the position along the strip using  
 2766 the times recorded of both SiPMs. This module produces `gar::rec::CaloHit` data  
 2767 products. Next, these objects are used as inputs to the `CaloClustering` module. It  
 2768 merges the hits based on a simple nearest neighbours (NN) algorithm. For the resulting  
 2769 clusters it also computes the total energy and position of the centroid. The results are  
 2770 stored as `gar::rec::Cluster` data products, with associations to the hits.

2771 The last step in the reconstruction is associating the reconstructed tracks in the  
 2772 HPgTPC to the clusters formed in the ECal and muon system. The `TPCECALAssociation`  
 2773 module checks first the position of the track end points, considering only the points  
 2774 that are at least 215 cm away from the cathode or have a radial distance to the center

## 6.2. $dE/dx$ MEASUREMENT IN THE TPC

2775 greater than 230 cm. The candidates are propagated up to the radial position, in the  
2776 case of clusters in the barrel, or the drift coordinate position, for the end cap cluster, of  
2777 the different clusters in the collection using the track parameters computed at the end  
2778 point. The end point is associated to the cluster if certain proximity criteria are met.  
2779 This module creates associations between the tracks, the end points and the clusters.  
2780 The criteria for the associations are slightly different for the ECal and the muon tagger.

### 2781 6.2 $dE/dx$ measurement in the TPC

2782 Among the parameters extracted from the track fitting, ionisation is particularly useful  
2783 for particle identification, as it is a function of the particle velocity. Although for the  
2784 case of relativistic particles this dependence is not very strong, measuring the track on  
2785 a large number of points may allow us to estimate the amount of ionisation accuratel.  
2786 This, paired with a measurement of the momentum, may allow us to identify the particle  
2787 type.

2788 The first calculation of the energy loss per unit length of relativistic particles using a  
2789 quantum-mechanical treatment is due to Bethe [151]. Using this approach, the mean  
2790 ionisation rate of a charged particle traveling through a material medium is (using  
2791 natural units  $G = \hbar = c = 1$ ):

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

2792 where  $N$  is the number density of electrons in the medium,  $e$  the elementary charge,  $m_e$   
2793 is the electron mass,  $z$  the charge of the particle in units of  $e$ ,  $\beta$  is the velocity of the  
2794 particle,  $\gamma = (1 - \beta^2)^{-1}$  and  $I$  denotes the effective ionisation potential averaged over  
2795 all electrons. This relation is known as the Bethe-Bloch formula.

2796 From Eq. (6.1) one can see that the ionisation loss does not depend explicitly on  
2797 the mass of the charged particle, that for non-relativistic velocities it falls as  $\beta^{-2}$ , then  
2798 goes through a minimum and increases as the logarithm of  $\gamma$ . This behaviour at high  
2799 velocities is commonly known as the relativistic rise. The physical origin of this effect

## CHAPTER 6. PARTICLE ID IN ND-GAr

2800 is partly due to the fact that the transverse electromagnetic field of the particle is  
2801 proportional to  $\gamma$ , therefore as it increases so does the cross section.

2802 It was later understood that the relativistic rise could not grow indefinitely with  $\gamma$ .  
2803 A way to add this feature in the Bethe-Bloch formula is by introducing the so-called  
2804 density effect term. It accounts for the polarisation effect of the atoms in the medium,  
2805 which effectively shield the electromagnetic field of the charged particle halting any  
2806 further increase of the energy loss [152]. Denoting the correction as  $\delta(\beta)$ , one can rewrite  
2807 Eq. (6.1) as:

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

2808 In general, the form of  $\delta(\beta)$  depends on the medium and its state of aggregation,  
2809 involving the usage of tabulated parameters and implicit relations [153].

2810 Another standard method to compute the amount of ionisation a charged particle  
2811 produces is the so-called photo-absorption ionisation (PAI) model proposed by Allison  
2812 and Cobb [154]. Within their approach, the mean ionisation is evaluated using a  
2813 semiclassical calculation in which one characterises the continuum material medium by  
2814 means of a complex dielectric constant  $\varepsilon(k, \omega)$ . However, in order to model the dielectric  
2815 constant they rely on the quantum-mechanical picture of photon absorption and collision.  
2816 Therefore, in the PAI model the computation of the ionisation loss involves a numerical  
2817 integration of the measured photo-absorption cross-section for the relevant material.

2818 In a particle physics experiment, the typical way of determining the energy loss  
2819 per unit length as a function of the particle velocity is studying identified particles  
2820 over a range of momenta. Once we have established this relation we can use it for  
2821 other, unknown particles. In this sense, it makes sense to have a regular mathematical  
2822 expression for this relation that one can use.

2823 It happens that neither the Bethe-Bloch theory nor the PAI model from Allison and  
2824 Cobb offer a close mathematical form for the ionisation curve. This is the reason why a  
2825 full parametrisation of the ionisation curves can be useful. A parametrisation originally  
2826 proposed for the ALEPH TPC [155] and later used by the ALICE TPC [156] group that

## 6.2. dE/dx MEASUREMENT IN THE TPC

2827 manages to capture the features of the ionisation energy loss is:

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

2828 where  $P_i$  are five free parameters. Hereafter, we will refer to Eq. (6.3) as the ALEPH  
2829 dE/dx parametrisation.

### 2830 6.2.1 Energy calibration

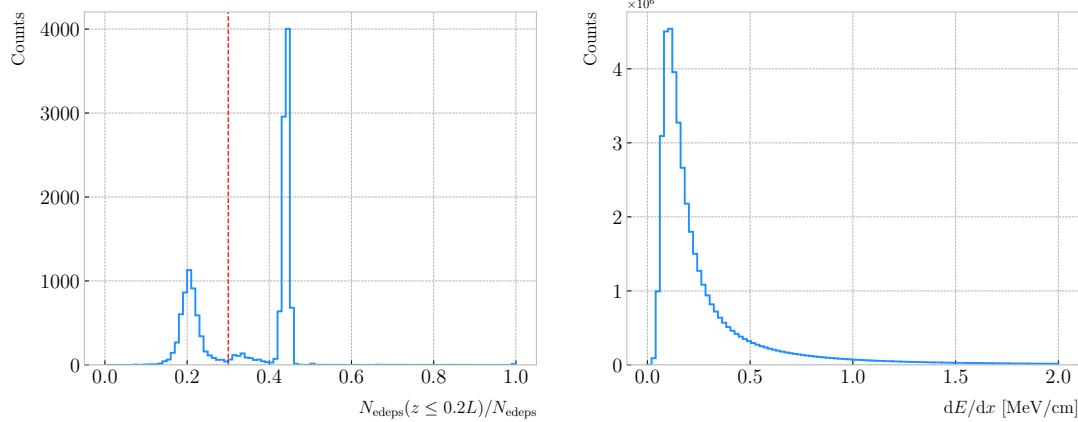
2831 In order to obtain the amount of energy loss by a charged particle due to ionisation  
2832 in our TPC we need to determine the conversion between the charge deposited in our  
2833 readout planes and the actual energy depositions. This procedure is known as energy  
2834 calibration.

2835 In a general, the first step of the calibration involves a non-uniformity correction,  
2836 to make sure that the detector response is uniform throughout the TPC. These are  
2837 typically divided into three categories, non-uniformities in the transverse  $YZ$  plane,  
2838 non-uniformities along the drift direction  $X$  and variations of the detector response  
2839 over time (would not apply to us as the detector is not built yet). These would correct  
2840 for effects such as electron diffusion and attenuation, space charge effects or channel  
2841 misconfiguration. However, because at the moment I am only interested in making sure  
2842 we recover a sensible result from our simulation, I will not apply uniformity corrections  
2843 to our charge deposits.

2844 Other effects, like electron-ion recombination or ADC saturation, lead to a non-linear  
2845 relation between the observed charge and the deposited energy in the detector, with the  
2846 observed readout charge saturating at high ionisation energies. In this case, because we  
2847 are dealing with gaseous argon and therefore recombination is not as important as in  
2848 liquid, we do not simulate recombination effects in the TPC. Even so, the simulation of  
2849 the electronic response will still introduce charge saturation, and one needs to correct  
2850 for it in order to obtain the exact amount of energy loss due to ionisation.

2851 By default, the track fitting algorithm in GArSoft provides a `TrackIonization`

## CHAPTER 6. PARTICLE ID IN ND-GAr



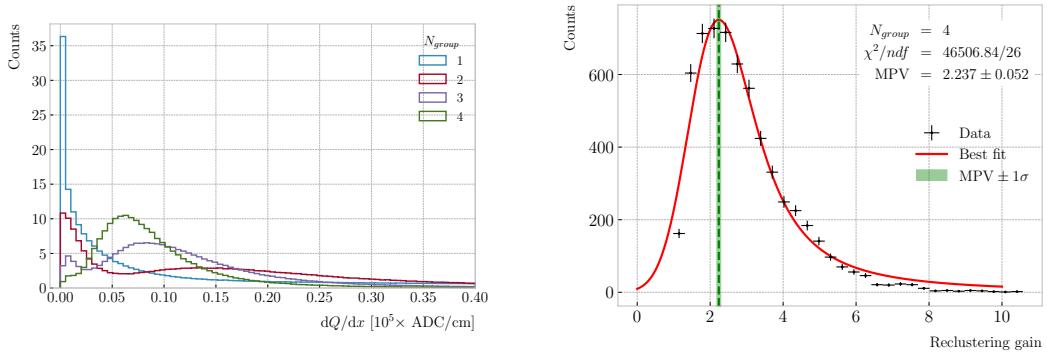
**Figure 6.2:** Left panel: distribution of the fraction of Geant4-level energy deposits per track with residual range less than 20% of the total track length, for the isotropic proton sample. Right panel: distribution of the ionisation per unit length of the energy deposits in the proton sample after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

object associated to each reconstructed track. It contains two collections of charge deposits, one for each fitting direction, consisting on pairs of charge values ( $dQ$ , in ADC) and step sizes ( $dx$ , in cm).

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

For studying the energy loss of the protons I select the reconstructed tracks that range out (i.e. slow down to rest) inside the TPC. A characteristic feature of the energy loss profile of any stopping ionising particle is the so-called Bragg peak, a pronounced peak that occurs immediately before the particle comes to rest. From Eq. (6.1) we can see that this behaviour is expected, as the energy loss for non-relativistic particles is inversely proportional to  $\beta^2$ . In data, a way of identifying the Bragg peak, and thus select the stopping particles, is checking the number of energy deposits towards the

## 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.

2869 end of the track. In this case, I count the fraction of the Geant4 simulated energy  
 2870 deposits with a residual range value (the distance from a given energy deposit to the  
 2871 last deposit in the track trajectory) less than a 20% of the corresponding track length<sup>2</sup>.  
 2872 The distribution of this fraction of energy deposits for our proton sample is shown in  
 2873 Fig. 6.2 (left panel). We can clearly see two well separated peaks in this distribution,  
 2874 one centered at 0.2 and another, narrower, one centered at a higher value. The first  
 2875 one corresponds to non-stopping protons, as in that case the number of energy deposits  
 2876 towards the end of the track is uniformly distributed due to the absence of the Bragg  
 2877 peak. In that way, I apply a cut in this distribution, requiring that at least 30% of the  
 2878 simulated energy deposits sit in the last 20% of the tracks, to ensure that the Bragg  
 2879 peak is present.

2880 Figure 6.2 (right panel) shows the distribution of the energy loss per unit length for  
 2881 the Geant4 simulated energy deposits of the selected stopping protons. We can see that  
 2882 it follows the expected shape of a Landau distribution, which describes the fluctuations of  
 2883 the ionisation energy losses [157]. This distribution has a characteristic asymmetric PDF,  
 2884 with a long right tail that translates into a high probability for high-energy ionisation  
 2885 losses. The origin of these fluctuations is mainly the possibility of transferring a high

<sup>2</sup>As 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.

## CHAPTER 6. PARTICLE ID IN ND-GAr

2886 enough energy to an electron, so it becomes a ionising particle itself.

2887 Now, from the point of view of the reconstruction, the objects that we have available  
 2888 to extract the ionisation information for the different reconstructed tracks are the  
 2889 collections of  $dQ$  and  $dx$  pairs, as stated before. The  $dQ$  values come from adding up  
 2890 the amplitude of all the reconstructed hits in a cluster, which is the input object to the  
 2891 Kalman fit.

2892 Figure 6.3 (left panel) shows the distribution of the ionisation charge deposits  
 2893 per unit length for the track in the stopping proton sample (blue line). As one can  
 2894 notice, this distribution does not resemble the expected shape of the Landau PDF. This  
 2895 distribution peaks sharply at 0 and has a heavy tailed behaviour. Notice, however, how  
 2896 the distribution changes its shape as we group together  $N_{group}$  consecutive charge deposit  
 2897 pairs (red, purple and green lines). The distribution in the  $N_{group} = 4$  case already has  
 2898 a shape which resembles that of the Geant4-level ionisation per unit length, so I will  
 2899 proceed using this amount of reclustering for the reconstruction-level depositions.

2900 An extra factor I need to account for, when reclustering is applied, is how the overall  
 2901  $dQ/dx$  per track changes. To do so, we can look at the ratio between the median  $dQ/dx$   
 2902 after and before the reclustering. Figure 6.3 (right panel) shows the median enhancement  
 2903 in  $dQ/dx$  per track for the stopping proton sample in the case  $N_{group} = 4$ . Fitting a  
 2904 Landau distribution convolved with a Gaussian<sup>3</sup>, I estimate the most probable value of  
 2905 this ratio to be  $G_{group} = 2.24 \pm 0.05$ .

2906 At this point, I am left with determining the conversion between the charge deposits  
 2907 per unit length  $dQ/dx$  and the energy deposits per unit length  $dE/dx$ . To this end, we  
 2908 need a way of comparing the two. I can use the residual range  $z$  to get a prediction of  
 2909 the most probable  $dE/dx$  by using the following empirical parametrisation [158]:

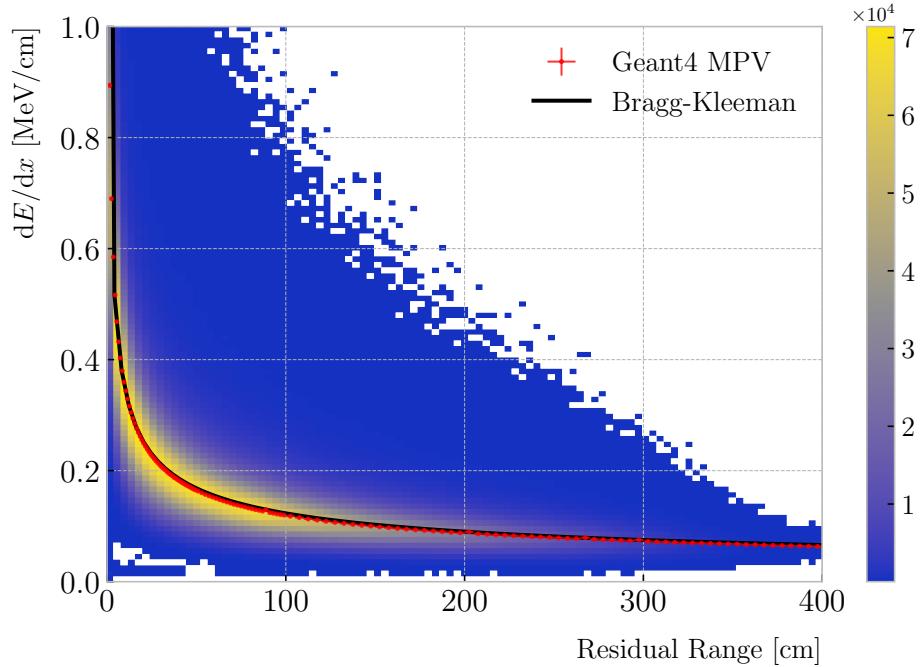
$$\frac{dE}{dx}(z) = \frac{z^{\frac{1}{p}-1}}{p\Lambda^{\frac{1}{p}}}, \quad (6.4)$$

2910 which is quoted in the literature as the Bragg-Kleeman formula. In order to obtain the

---

<sup>3</sup>In 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

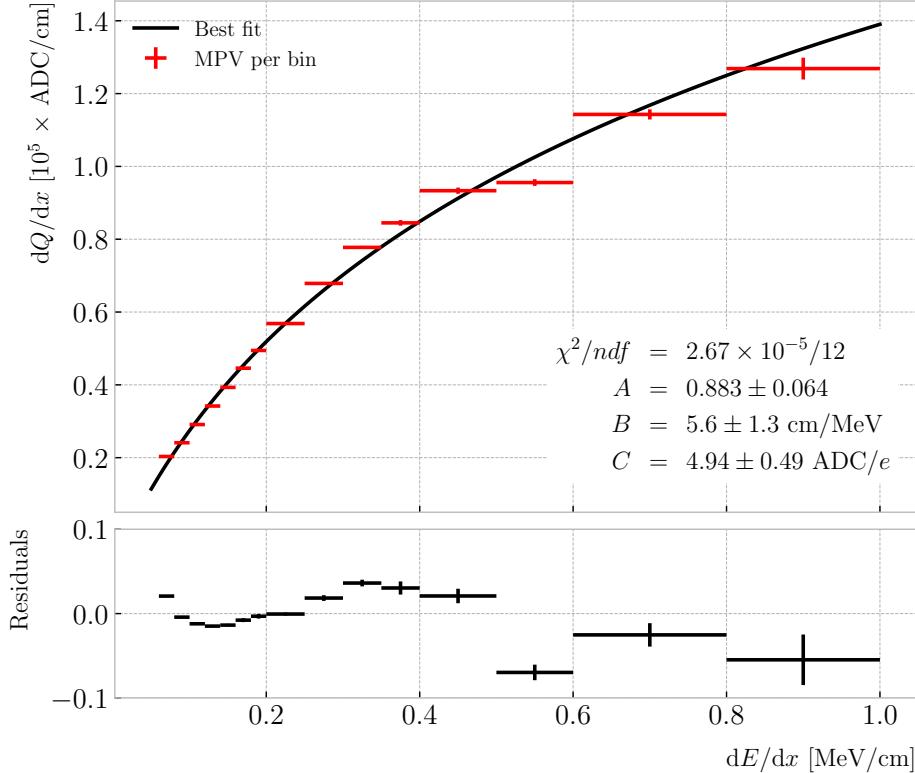


**Figure 6.4:** Distribution of the Geant4-simulated energy losses per unit length versus residual range for the stopping proton sample. The overlaid points represent the fitted most probable value of the  $dE/dx$  distribution in each residual range bin, whereas the curve is their best fit to the Bragg-Kleeman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

2911  $p$  and  $\Lambda$  parameters I perform a fit using the energy losses and the residual ranges given  
 2912 by the Geant4 stage of our proton sample.

2913 Within our simulation, the residual range is sampled with a maximum size of  
 2914 5 mm. Therefore, to perform the fit to the Bragg-Kleeman formula, we can use a  
 2915 fine-grained residual range binning. For each of the residual range bins I extract the  
 2916  $dE/dx$  distribution and fit it to a LanGauss distribution, to obtain the value of the  
 2917 most probable  $dE/dx$  in the bin together with a statistical uncertainty. I then fit Eq.  
 2918 (6.4) to these most probable values and the centres of the residual range bins. This  
 2919 procedure is depicted in Fig. 6.4, where I show the distribution of the energy loss per  
 2920 unit length versus the residual range, together with the most probable  $dE/dx$  values  
 2921 and their uncertainty in each bin (red points) and the curve with the best fit of the  
 2922 Bragg-Kleeman relation to those values (black line). The best fit is obtained for the

## CHAPTER 6. PARTICLE ID IN ND-GAr



**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).

2923 parameter values  $p = 1.8192 \pm 0.0005$  and  $\Lambda = 0.3497 \pm 0.0008 \text{ cm/MeV}^{p4}$ .

2924 Having an analytical expression that relates the residual range to  $dE/dx$ , I can take  
2925 our reconstruction-level residual ranges from the stopping proton sample and compute  
2926 the most probable energy loss associated.

2927 In order to parametrise the charge saturation, we can use the following logarithmic  
2928 function inspired by the modified box model for recombination:

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

2929 where  $A$  and  $B$  are the calibration parameters we need to determine,  $W_{ion}$  is the average  
2930 energy to produce an electron-ion pair,  $G_{group}$  is the gain from the reclustering discussed

---

<sup>4</sup>These strange units for  $\Lambda$  come from dimensional analysis, just to keep the Bragg-Kleeman formula (6.4) consistent.

## 6.2. $dE/dx$ MEASUREMENT IN THE TPC

**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.

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

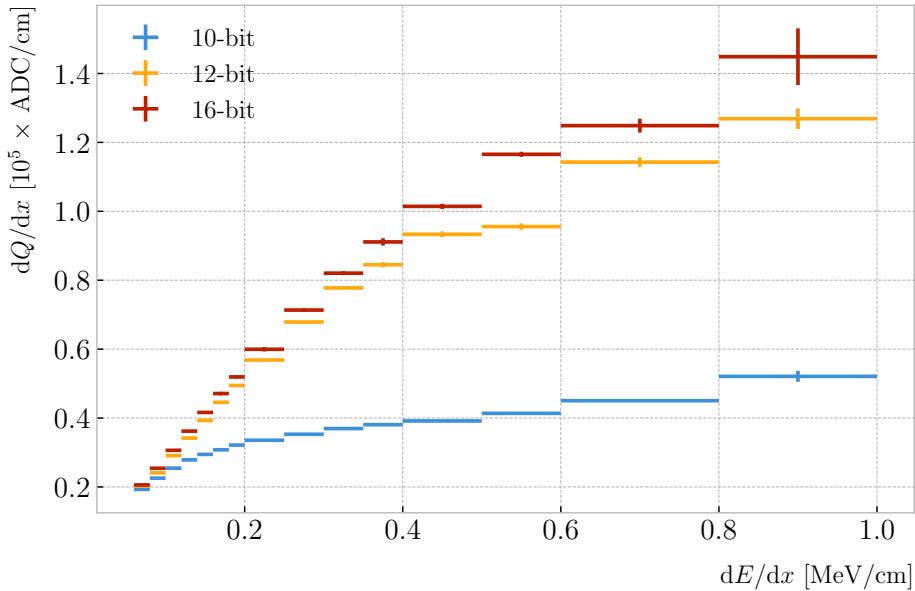
above and  $C$  is the calibration constant to convert number of electrons to ADC counts, commonly refer to as gain (also to be obtained in the fit). In this case, I use a value for the electron-ion production energy of  $W_{ion} = 26.4$  eV [159]. This value, used in our simulation as well, was measured for gaseous argon in normal conditions, and therefore should be checked in the future to describe correctly the high-pressure argon-CH<sub>4</sub> mixture of ND-GAr.

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

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

One interesting thing to check is what induces this non-linear relation between charge and energy. The only effects that modify the amount of electrons reaching the readout planes in the simulation are the transverse diffusion and the finite electron lifetime. Once the electrons reach the readout chambers, the pad response functions are applied,

## CHAPTER 6. PARTICLE ID IN ND-GAr



**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).

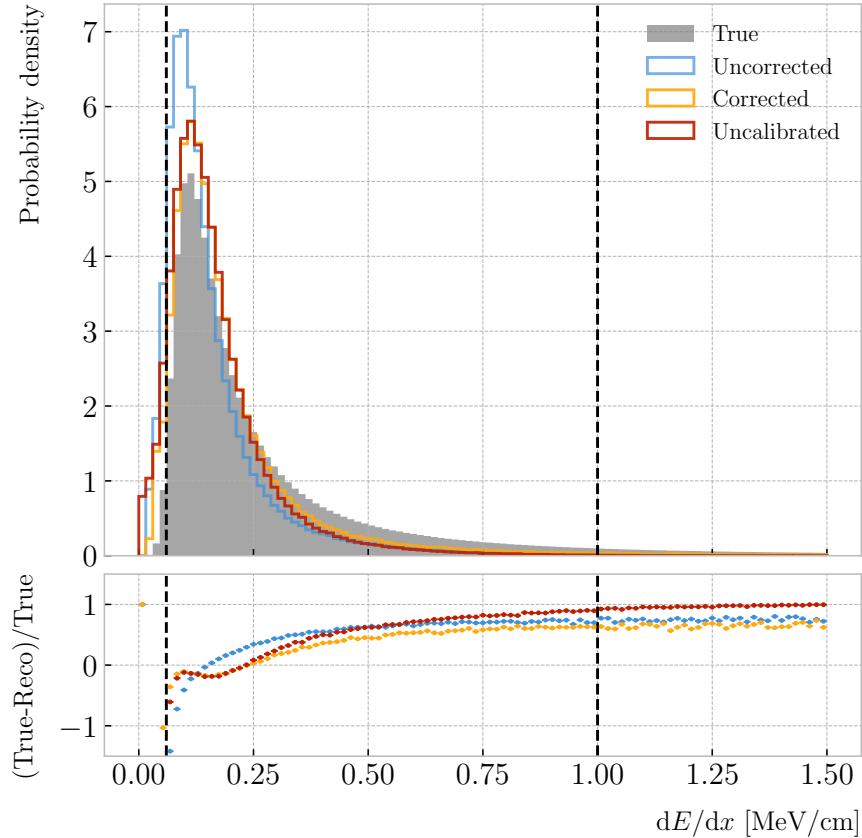
2953 together with an electrons-to-ADC conversion and the ADC saturation limit.

2954 By default, GArSot applies a 12-bit ADC limit, which can be changed in the  
 2955 simulation configuration. However, it can only be increased up to 16-bit, as we represent  
 2956 the ADC collection as a `std::vector<short>`. This way, I tried to change the saturation  
 2957 parameter to see how it affects the relation between reconstructed charge and energy.  
 2958 Figure 6.6 shows a comparison between the most probable  $dQ/dx$  for 10, 12 and 16-  
 2959 bit ADC limits. As expected, the lower the limit is the sooner the charge saturates.  
 2960 For higher ADC limits the relation between energy and charge remains linear up to  
 2961 higher  $dE/dx$  values, but even for the 16-bit limit the saturation is noticeable for values  
 2962  $\gtrsim 0.5$  MeV/cm.

2963 Table 6.1 shows the results of fitting the samples with 10 and 16-bits ADC limits to  
 2964 the calibration function from Eq. (6.5), using the weights based on their relative error  
 2965 as described previously. One interesting feature to notice is how different the best fit  
 2966 points look for the 10-bit ADC saturation when compared to the other two, which are  
 2967 consistent with each other.

2968 At this point we can compare the  $dE/dx$  distribution one gets from Geant4, i.e. the

## 6.2. $dE/dx$ MEASUREMENT IN THE TPC



**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.

2969 true energy loss distribution, and the distribution I found by applying the calibration  
 2970 function to our collection of reconstructed  $dQ/dx$  values. Figure 6.7 (top panel) shows  
 2971 the true (solid grey) and reconstructed (blue, labeled as uncorrected) distributions  
 2972 together. The dashed vertical lines indicate the region of validity of the calibration fit, i.e.  
 2973 the left and right edges of the first and last  $dE/dx$  bin respectively. Notice that these  
 2974 histograms are area-normalised, as the total number of true energy deposits is much  
 2975 higher than the number of reconstructed charge deposits. This is due to a combination  
 2976 of effects, like the finite spatial resolution of the detector, the hit clustering used in the  
 2977 track fitting and the reclustering we have applied here.

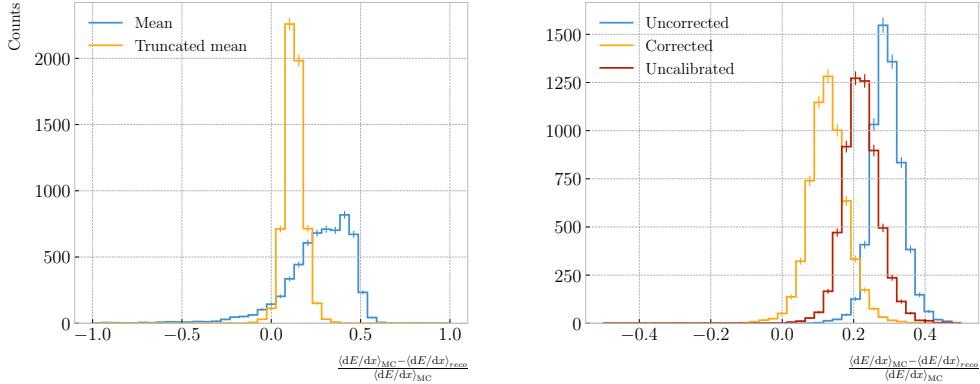
## CHAPTER 6. PARTICLE ID IN ND-GAr

2978        The two distributions are significantly different. That can be seen clearly when  
2979        looking at the fractional residuals, shown in Fig. 6.7 (bottom panel). In particular,  
2980        the position of the peak is off, which could bias the mean energy loss predictions. It  
2981        seems like the difference between these may be due to an overall scaling factor. One  
2982        possibility is to scale the most probable value of the reconstructed distribution to  
2983        the most probable value predicted by Geant4. I do this by fitting both distributions  
2984        using a LanGauss function, obtaining  $dE/dx_{MPV, true} = 0.1145 \pm 0.0005$  MeV/cm and  
2985         $dE/dx_{MPV, reco} = 0.0928 \pm 0.0005$  MeV/cm for the true and reconstructed most probable  
2986        values respectively. These can be translated into an scaling factor  $S = 0.579 \pm 0.006$ .

2987        The result of applying the scaling correction can be seen in Fig. 6.7 (top panel).  
2988        The corrected  $dE/dx$  distribution (yellow, labeled as corrected) peaks around the same  
2989        value the true distribution does, as expected. Moreover, the high energy region is also  
2990        slightly better described. For low ionisations, below the lower limit of the calibration  
2991        fit, the differences between true and reconstructed are still significant. This low energy  
2992        excess may be migration of some events from the peak region. The overall effect of the  
2993        correction can be seen in the fractional residual plot in Fig. 6.7 (bottom panel).

2994        One can also check what happens if instead of applying the logarithmic calibration we  
2995        simply scale the  $dQ/dx$  distribution (post reclustering) to have the same most probable  
2996        value as the true  $dE/dx$  distribution. In this case, following an analogous procedure to the  
2997        one described earlier, I found the scaling factor  $S_{uncalibrated} = 0.414 \pm 0.002$  MeV/ADC<sup>5</sup>.  
2998        The resulting distribution (red, labeled as uncalibrated) is also shown in in Fig. 6.7 (top  
2999        panel). The behaviour of the new distribution is similar to the corrected case at low  
3000        energy losses, around the peak of the true distribution, but it is worse at describing the  
3001        high energy tail. This is expected, it is in the high ionisation regime where saturation  
3002        effects apply and therefore calibration is needed.

## 6.2. $dE/dx$ MEASUREMENT IN THE TPC



**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.

### 3003 6.2.2 Truncated $dE/dx$ mean

3004 Once we have a collection of  $dE/dx$  values for each reconstructed track, we can compute  
 3005 the corresponding most probable ionisation loss per unit length of the particle. This  
 3006 is the value predicted by the Bethe-Bloch or the PAI models, and together with a  
 3007 measurement of the momentum it allows for particle identification.

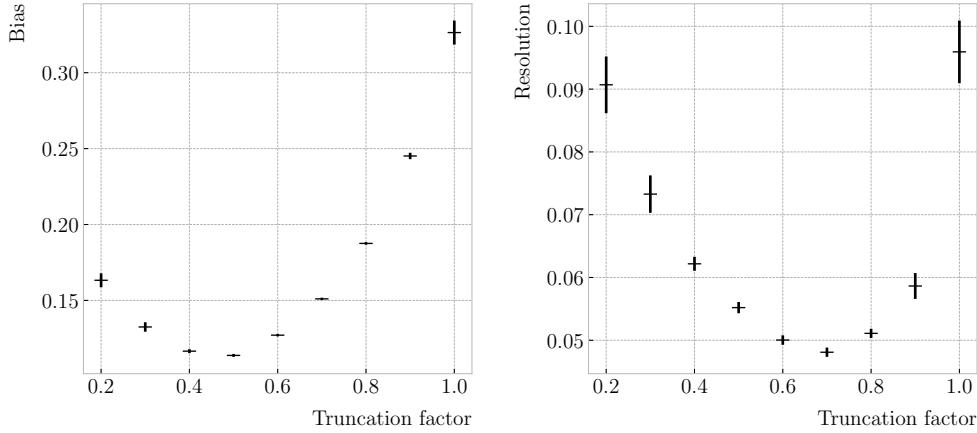
3008 However, estimating the most probable  $dE/dx$  value for each reconstructed track  
 3009 is not a trivial task. As mentioned before, the  $dE/dx$  distributions follow Landau-like  
 3010 distributions. Therefore, one should perform e.g. a LanGauss fit to correctly estimate  
 3011 the most probable values. Automating this kind of fits is often problematic, as they  
 3012 usually incur in convergence problems. Moreover, the reconstructed  $dE/dx$  distributions  
 3013 we obtain tend to have relatively small statistics, which may also produce poor fits. In  
 3014 practice, doing these unsupervised fits may degrade our performance, and a more robust  
 3015 method is preferred.

3016 A possibility could be taking the mean of the reconstructed  $dE/dx$  distribution for  
 3017 each particle. The problem with this approach is that the high energy Landau tail,  
 3018 combined with our limited statistics, can induce large fluctuations in the computation

---

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

## CHAPTER 6. PARTICLE ID IN ND-GAr



**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.

of the mean. Imagine you have two protons with the same kinetic energy, but due to reconstruction problems in one case you did not get as many charge deposits reconstructed in its high ionisation loss region. If you do not remove the tails the computed  $dE/dx$  means will be significantly different.

In order to avoid those fluctuations, one can compute the mean of a truncated  $dE/dx$  distribution instead. By keeping only a given fraction of the lowest energy deposits we obtain an estimate of the mean energy loss that is more resilient to reconstruction inefficiencies and statistical effects. Figure 6.8 (left panel) shows a comparison between the  $\langle dE/dx \rangle$  computed by taking the mean of the full distribution (blue line) and the 60% lowest energy clusters (yellow line), for the stopping proton sample. The fractional residuals are computed for each proton, taking the corresponding means using their collections of true and reconstructed energy deposits. One can see that using the simple mean translates into a high bias and uncertainty in the  $\langle dE/dx \rangle$  estimation, whereas applying the truncation reduces both significantly.

Additionally, I performed a comparison between the 60% truncated mean  $dE/dx$  obtained using the different calibration methods discussed earlier, namely the uncorrected (blue), corrected (yellow) and uncalibrated (red) distributions. The results are shown in Fig. 6.8 (right panel). While the widths of these distributions are similar, the bias

## 6.2. $dE/dx$ MEASUREMENT IN THE TPC

3037 obtained for the corrected sample, i.e. calibration function and correction factor applied,  
 3038 is a factor of  $\sim 2$  lower than in the uncalibrated case and almost three times smaller  
 3039 than for the uncorrected sample.

3040 The next step is to optimise the level of truncation we are going to apply to our  
 3041 data. To do so, I used different truncation factors, i.e. the percentage of energy-ordered  
 3042 reconstructed energy deposits we keep to compute the mean, on the corrected  $dE/dx$   
 3043 sample of the stopping protons. Then, following the same procedure of computing the  
 3044 fractional residuals as before, I fitted the resulting histograms using a double Gaussian  
 3045 function. This is simply the sum of two Gaussian functions of the type:

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

3046 I do not add the classical normalisation factor of the Gaussian,  $1/\sqrt{2\pi}\sigma$ , therefore  
 3047 the amplitude  $A$  simply represents the maximum of the function. One of the two  
 3048 Gaussian functions describes the core part of the distribution, while the other captures  
 3049 the behaviour of the tails.

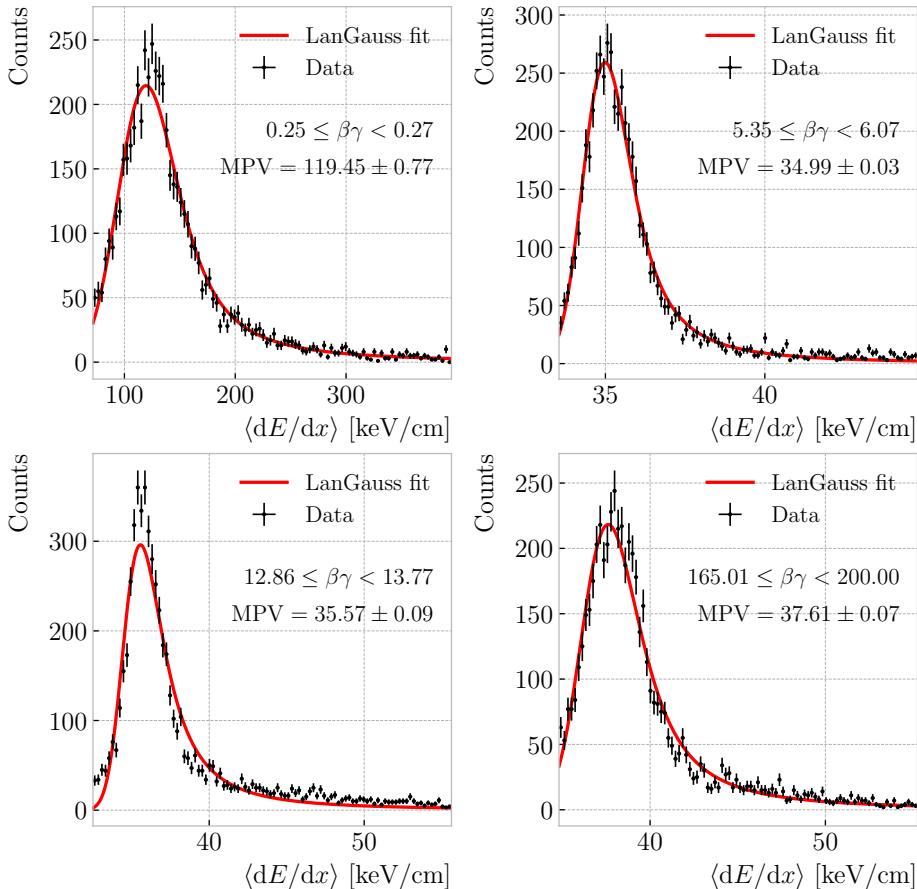
3050 For each truncation factor, I look at the bias and the resolution I obtain. I define  
 3051 these as the weighted means of the corresponding parameters in the fits:

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

3052 where  $A_{core}$  and  $A_{tail}$  are the amplitudes of the core and tail distributions respectively  
 3053 and  $x$  is either the mean  $\mu$  or the width  $\sigma$  of said distributions.

3054 Figure 6.9 shows the bias (left panel) and the resolution (right panel) I obtained  
 3055 for the stopping proton sample, using different values of the truncation. From these, it  
 3056 can be seen that a truncation factor of 50% minimises the bias in the estimation, while  
 3057 70% gives the best resolution. That way, I settled on the intermediate value of 60%  
 3058 truncation, which yields a  $\langle dE/dx \rangle$  resolution of  $5.00 \pm 0.08$  % for stopping protons.

## CHAPTER 6. PARTICLE ID IN ND-GAr



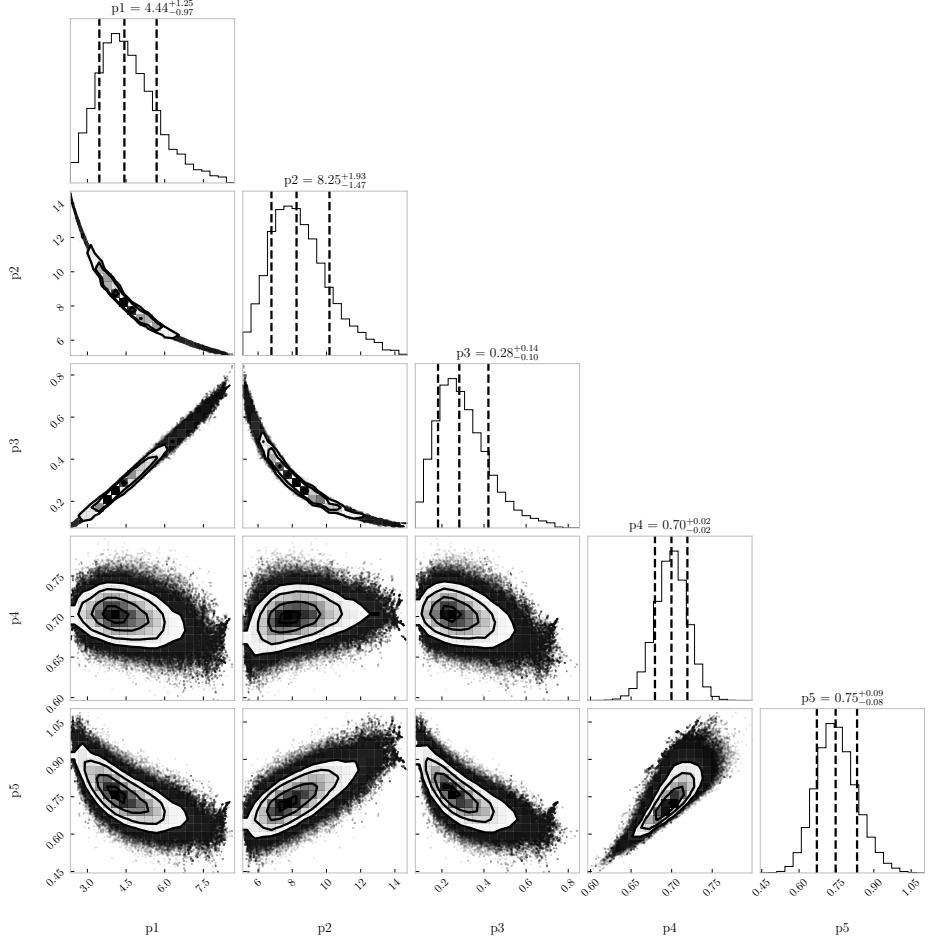
**Figure 6.10:** Examples of the truncated mean  $dE/dx$  LanGauss fits for various  $\beta\gamma$  bins, from a simulated FHC neutrino sample.

### 3059 6.2.3 Mean $dE/dx$ parametrisation

3060 Now that we have a way to estimate the mean energy loss of a particle in the HPgTPC,  
 3061 we can determine the value of the free parameters in the ALEPH formula, Eq. (6.3).  
 3062 For this, I used a sample of  $10^5$  reconstructed FHC neutrino events inside ND-GAr. In  
 3063 this case I cannot use the stopping proton sample, as we need to cover the full kinematic  
 3064 range of interest for the neutrino interactions in our detector.

3065 The original data does not contain an estimation of the velocity of the tracks, instead  
 3066 the tracks have a value for the reconstructed momentum and the associated PDG code  
 3067 of the Geant4-level particle that created the track. Therefore, one can select some of the  
 3068 particles in the data, in this case I selected electrons, muons, pions and protons, and

## 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.

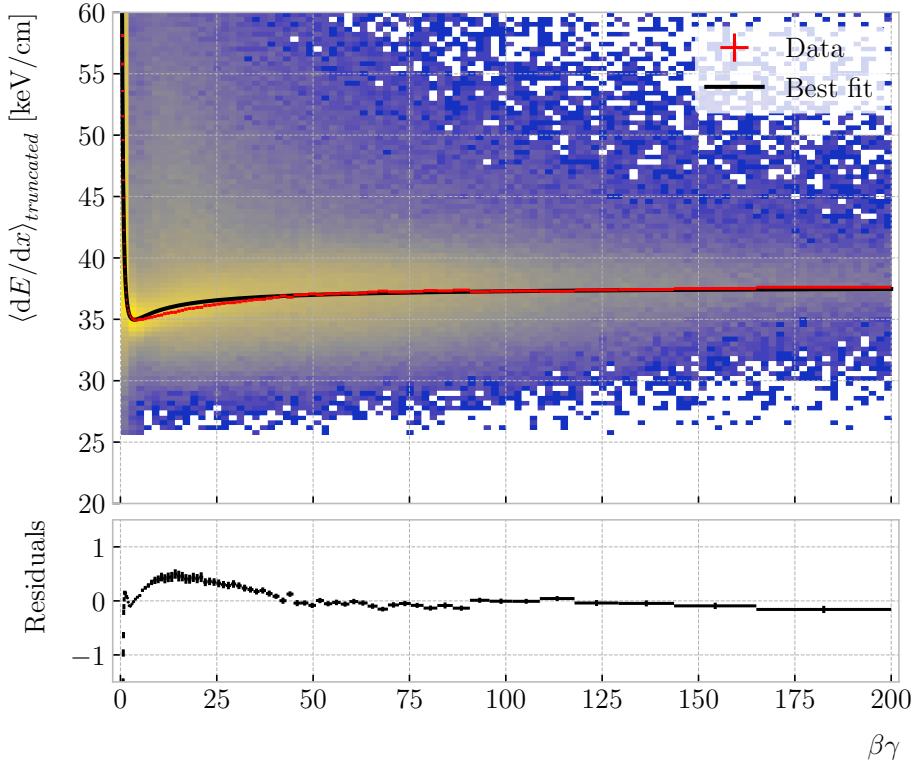
3069 compute  $\beta$  and  $\gamma$  using the reconstructed momentum and their mass. In terms of  $\beta\gamma$   
 3070 the mean  $dE/dx$  does not depend on the particle species, so one can consider all the  
 3071 dataset as a whole. For this fit, I will express  $\beta$  in terms of the  $\beta\gamma$  product as:

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

3072 which can be easily proven from the definition of  $\gamma$ .

3073 Next, I bin the data in  $\beta\gamma$ . I chose a fine binning so as to capture the different  
 3074 features of the ionisation curve. Instead of fixing the bin width, I select them so each one

## CHAPTER 6. PARTICLE ID IN ND-GAr

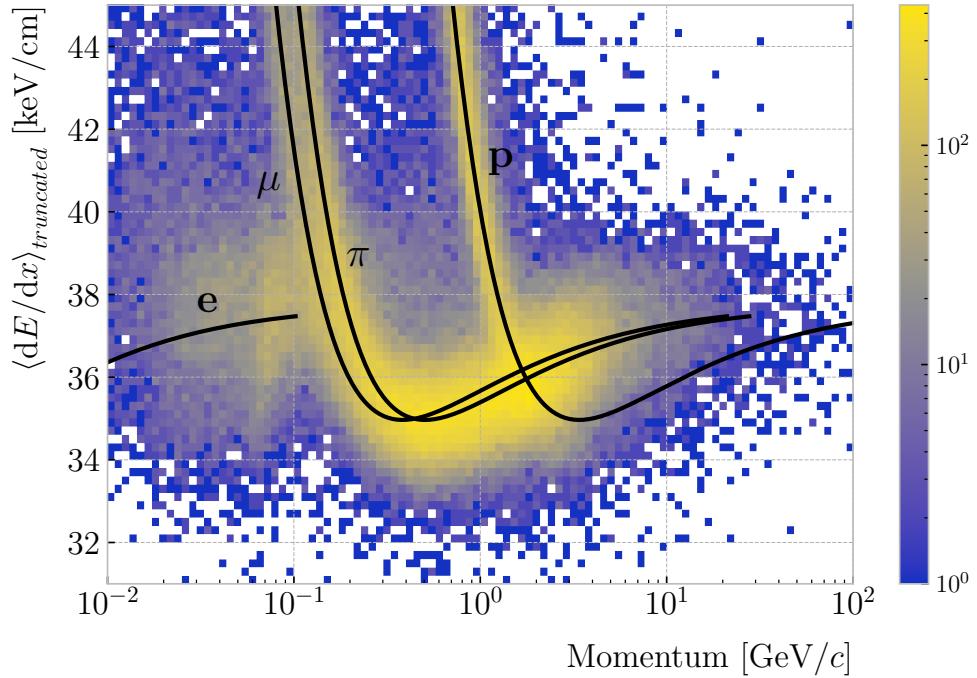


**Figure 6.12:** Truncated mean  $dE/dx$  obtained for the FHC neutrino sample as a function of the  $\beta\gamma$  product (upper panel). Also shown are the fitted most probable values for each  $\beta\gamma$  bin (red points) and the best fit obtained using the ALEPH parametrisation (black line). The residuals resulting from the fit are shown in the lower panel.

3075 has approximately the same statistics. Then, for each  $\beta\gamma$  slice, I compute the median  
 3076 and the interquartile range (IQR) of the  $\langle dE/dx \rangle$  distribution. Using these, I make a  
 3077 histogram in the range [median – IQR, median + 5 IQR], which I fit to a LanGauss  
 3078 function in order to extract the MPV. Using this range accounts for the asymmetric  
 3079 nature of the distributions, while also helps avoiding a second, lower maximum present  
 3080 at low  $\beta\gamma$ , probably a result of reconstruction failures.

3081 A few examples of these fits are shown in Fig. 6.10. The chosen values of  $\beta\gamma$  sit in  
 3082 very distinct points along the  $\langle dE/dx \rangle$  curve, going from the high ionisation region at  
 3083 low velocities (top left panel), to the minimum point (top right panel), the beginning of  
 3084 the relativistic rise (bottom left panel), and the plateau produced by the density effect  
 3085 (bottom right panel).

## 6.2. $dE/dx$ MEASUREMENT IN THE TPC

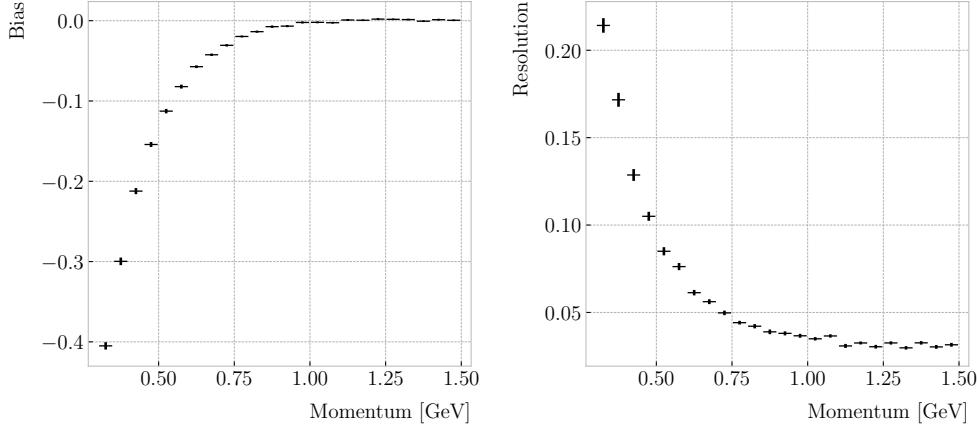


**Figure 6.13:** Distribution of the 60% truncated mean  $dE/dx$  versus reconstructed momentum for the FHC neutrino sample. The black lines indicate the predictions of the ALEPH parametrisation for electrons, muons, charged pions and protons.

3086 I used the resulting most probable  $\langle dE/dx \rangle$  values and the centres of the  $\beta\gamma$  bins as  
 3087 the points to fit to the ALEPH formula. For this particular fit I used the least-squares  
 3088 method to get a first estimation of the ALEPH parameters. Applying some uniform  
 3089 priors, I then used these values as the starting point of a 100000 steps MCMC. Figure 6.11  
 3090 shows the posterior probability distributions I obtain for each parameter. The reported  
 3091 best fit points are based on the 16th, 50th, and 84th percentiles in the marginalised  
 3092 distributions.

3093 The resulting fit (black line), compared to the data points (red points) and the  
 3094 underlying distribution is shown in Fig. 6.12 (top panel). The overall fit is good, with a  
 3095 reduced chi-squared of  $\chi^2/ndf = 1.02$ . However, there are some regions where the fit  
 3096 does not describe the data correctly, like the very low  $\beta\gamma$  regime, where the fit severely  
 3097 underestimates for energy losses  $\gtrsim 50$  keV/cm, and the start of the relativistic raise,  
 3098 where we have a slight overestimation. This is a result of those points having a larger  
 3099 uncertainty when compared to the ones around the dip or the plateau areas. These

## CHAPTER 6. PARTICLE ID IN ND-GAr



**Figure 6.14:** Estimated values of the mean  $dE/dx$  bias (left panel) and resolution (right panel) obtained for the true protons in a FHC neutrino sample.

3100 differences can be better seen in the residual plot, Fig. 6.12 (bottom panel).

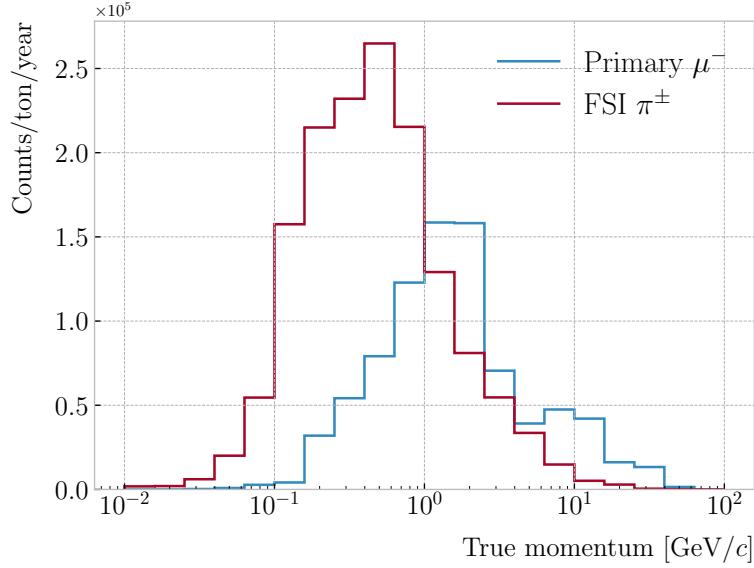
### 3101 6.2.4 Particle identification

## 3102 6.3 Muon and pion separation in the ECal and MuID

3103 As it could be seen from Fig. 6.13, it is not possible to separate muons and charged pions  
3104 in the HPgTPC using  $dE/dx$  for momenta  $\gtrsim 300 \text{ MeV}/c$ . In ND-GAr, approximately  
3105 70% of the interactions in FHC mode will be  $\nu_\mu$  CC (compared to the 47% of  $\bar{\nu}_\mu$  CC  
3106 interactions when operating in RHC mode), while 24% are neutral currents. Out of  
3107 these, around 53% and 47% of them will produce at least one charged pion in the final  
3108 state, respectively. Figure 6.15 shows a comparison between the spectra of the primary  
3109 muons and the charged pions for  $\nu_\mu$  CC interactions in ND-GAr producing one or more  
3110 charged pions. From this, one can see that (i) the majority of muons and charged pions  
3111 are not going to be distinguishable with a  $\langle dE/dx \rangle$  measurement, and that (ii) particle  
3112 identification is necessary both to classify correctly the  $\nu_\mu$  CC events and identify the  
3113 primary muon within them.

3114 ND-GAr features two other subdetectors which can provide additional information  
3115 for this task, namely the ECal and MuID. The current ECal design, described in (ref  
3116 section), consists of 42 layers, made of 5 mm of Pb, 7 mm of plastic scintillator and a

### 6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID



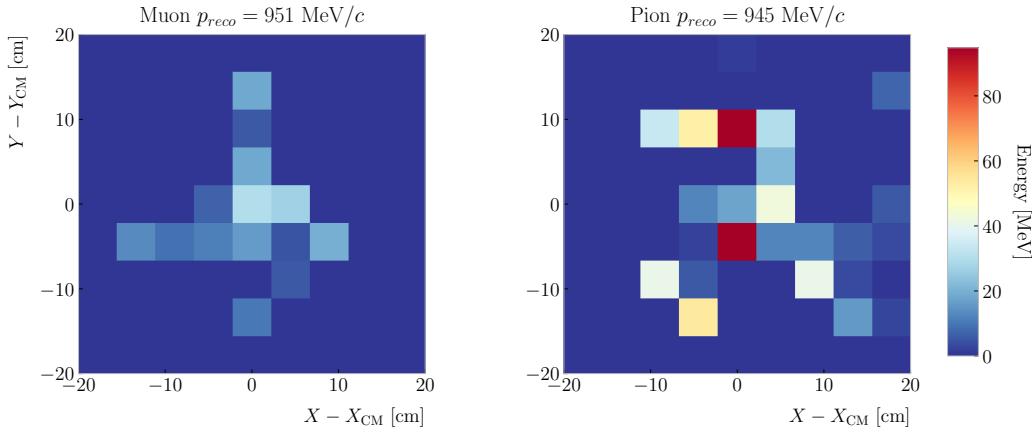
**Figure 6.15:** True momentum distribution for the primary muon in  $\nu_\mu$  CC  $N\pi^\pm$  interactions inside the fiducial volume of ND-GAr (blue line), compared to the post FSI charged pion spectrum (red line).

3117 1 mm PCB board. The total thickness of this calorimeter is 1.66 nuclear interaction  
 3118 lengths or 1.39 pion interaction lengths. The Muid design is in a more conceptual  
 3119 stage, however it is envisioned to feature layers with 10 cm of Fe and 2 cm of plastic  
 3120 scintillator<sup>6</sup>. With its three layers, it will have a thickness of 1.87 or 1.53 nuclear or pion  
 3121 interaction lengths, respectively.

3122 Because pion showers are dominated by inelastic nuclear interactions, the signatures  
 3123 of these particles in the calorimeter will look significantly different from those of muons.  
 3124 Although our ECal is not thick enough to fully contain the hadronic showers of the  
 3125 charged pions at their typical energies in FHC neutrino interactions, they can still be  
 3126 used to understand whether the original particle was more hadron-like or MIP-like. In  
 3127 Fig. 6.16 I show two examples of energy distributions created by a muon (left panel)  
 3128 and a charged pion (right panel) of similar momenta interacting in the ECal. These  
 3129 figures represent the transverse development of the interactions. For each of them, I  
 3130 computed the principal component and centre of mass of the interaction, projecting

<sup>6</sup>It 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 ID IN ND-GAr



**Figure 6.16:** Distributions of energy deposits in the ECal for a muon (left panel) and a charged pion (right panel) with similar momenta. The energy is projected onto the plane perpendicular to the principal component of the hits, and the positions are relative to the center of the interaction.

the position of the hits onto the plane perpendicular to that direction, and taking the distances relative to the centre. It can be seen that the muon follows an almost MIP-like behaviour, being the central bin in the histogram the one with the highest deposited energy. On the other hand, the pion not only deposits more energy overall, but also this energy is more spread-out among the different hits. It is this kind of information that would allow us to tell apart muons from pions.

This way, I identify three main action points that need to be addressed if one wants to use these detectors to distinguish between muons and charged pions. These are:

1. the way we make the associations between tracks in the HPgTPC to the activities (what in GArSoft we call clusters) in the ECal and the MuID,

2. what variables or features one can extract from the calorimeters that encapsulate the information we are interested about,

3. and how to carry out the classification problem.

### 6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

#### 3144 6.3.1 Track-ECal matching

3145 One of the main players in the muon and pion separation is the way we associate clusters  
3146 in the ECal to reconstructed tracks in the TPC. Missing some associations or making  
3147 wrong ones can bias the ECal quantities that we can use for classifying particles. The  
3148 current algorithm in GArSoft provides precise associations, i.e. most of the associations  
3149 that it produces are correct, but it appears to miss an important number of associations  
3150 (at least when using the default configuration).

3151 The current TPC track-ECal cluster association algorithm is divided in four parts.  
3152 It first checks whether the track end point fulfils certain conditions to be extrapolated.  
3153 There are two cut values in this step, one for the drift direction and other radial.

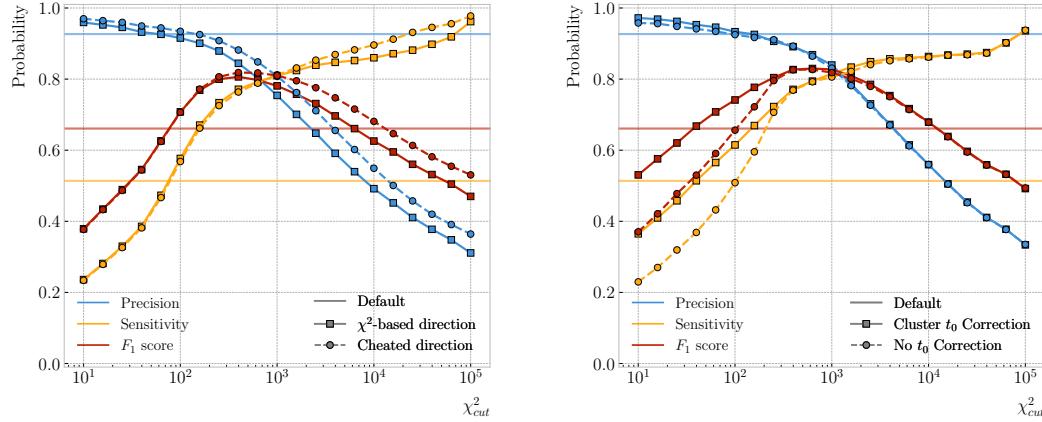
3154 If the point can be extrapolated, the code computes the coordinates of the centre  
3155 of curvature using the Kalman fit estimates at the track end ( $y$ ,  $z$ ,  $1/R$ ,  $\phi$ ,  $\tan\lambda$ ). It  
3156 then compares the distance between this and the cluster in the  $(z, y)$  plane with  $R$ . This  
3157 introduces another cut in the perpendicular direction.

3158 The next step is different for clusters in the barrel or in one of the end caps. If it  
3159 is a barrel cluster the algorithm extrapolates the track up to the radial distance of the  
3160 cluster. There are three possible outcomes, the extrapolated helix can cut the cylinder  
3161 of radius  $r_{clus}$  two, one or zero times. I get the cut point that is closer to the cluster and  
3162 check that it is either in the barrel or the end caps. Computing the difference between  
3163 the  $x$  coordinates of the cluster and the extrapolated point, the module checks that this  
3164 is not greater than a certain cut. If the cluster is in an end cap, I propagate the track  
3165 up to the  $x$  position of the cluster. Then, the algorithm computes the angle in the  $(z, y)$   
3166 plane between the centre of curvature and the cluster,  $\alpha$ , and the centre of curvature  
3167 and the propagated point,  $\alpha'$ . A cut is applied to the quantity  $(\alpha - \alpha')R$ .

3168 If the cluster contains more than a certain number  $N$  of hits, I apply an extra cut to  
3169 the dot product of the direction of the track at the propagated  $x$  value and the cluster  
3170 direction.

3171 The code makes sure to only associate one end of the track (if any) to a cluster.

## CHAPTER 6. PARTICLE ID 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  $\chi^2$ -based direction estimator (squares) and cheating the directions (circles), for different values of the  $\chi^2$  cut. Right panel: comparison of the performance of the new algorithm when applying the cluster  $t_0$  correction (squares) and when (circles).

3172 However, it can associate more than one track to the same cluster. This makes sense,  
3173 as different particles can contribute to the same cluster in the ECal, but it makes it  
3174 difficult to quantify the relative contributions of the tracks to a certain cluster.

3175 As a way of comparing the performance of this algorithm, a new, simpler association  
3176 module was written. The goal was to have a simple and robust algorithm, which depends  
3177 on as few parameters as possible and that can produce a one-to-one matching between  
3178 tracks and ECal clusters.

3179 For each reconstructed track, the new algorithms applies the same procedure to the  
3180 forward and the backward fits irrespective of their end point positions. It first gets the  
3181 Kalman fit parameters at the corresponding end point together with the  $X$  position,  $x_0$ ,  
3182 ( $y_0$ ,  $z_0$ ,  $1/R$ ,  $\phi_0$ ,  $\tan\lambda$ ).

3183 For each ECal cluster, I compute the radial distance to the centre of the TPC and  
3184 find the  $\phi$  value in the range  $[\phi_0, \phi_0 + \text{sign}(R)\phi_{max}]$  that makes the propagated helix  
3185 intersect with the circle defined with such radius. The  $(x, y, z)$  position of the helix for  
3186 the  $\phi$  value found (if any) is then computed. In case there are two intersections, I keep  
3187 the one that minimises the distance between  $(y, z)$  and  $(y_c, z_c)$ .

3188 I then calculate  $\chi^2$  value based on the Euclidean distance between the propagated

### 6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

3189 point and the cluster:

$$\chi^2/ndf = \frac{\sum_{n=0}^2 (x^{(n)} - x_c^{(n)})^2}{3}. \quad (6.9)$$

3190 If there was no intersection I store a  $-1$  instead. In the end, for each reconstructed track  
3191 in the event one ends up with two collections of  $\chi^2$  values, one for each ECal cluster  
3192 and fit directions.

3193 The current code only supports having ECal clusters associated to one end of each  
3194 track. We have two options to decide what track end to keep. The first one tries to  
3195 cheat the selection, looking at the distance between the two track ends and the true  
3196 start position of the associated MC particle. The second one keeps the track end with  
3197 more  $\chi^2$  entries below the cut.

3198 This feature of only considering one track end limits the algorithm, making it not  
3199 suitable for reconstructing events with particles originating outside the TPC. However,  
3200 as for the moment the main concern of the group is the study of neutrino interactions  
3201 off the gaseous argon, this is an acceptable assumption.

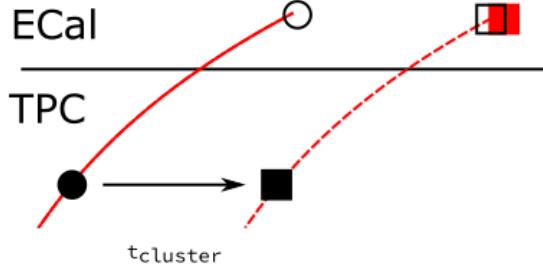
3202 In order to associate a cluster to a track, I take all clusters with a  $\chi^2$  value in the  
3203 range  $[0, \chi_{cut}^2]$ . If a cluster has been assigned to more than one track we leave it with  
3204 the one with the lowest  $\chi^2$ .

3205 This default behaviour of the algorithm can be modified to associate more than one  
3206 track to each cluster. Not only that, but the  $\chi^2$  values can be used to assign relative  
3207 weights to the different contributions.

3208 To evaluate the performance of the association method, I use a binary classification  
3209 approach. In this case, I check the leading MC Track IDs associated to the reconstructed  
3210 tracks and ECal clusters. I count an association as true positive (TP) if both Track  
3211 IDs coincide. An association is considered false positive (FP) when the Track IDs are  
3212 different. If a cluster has not been associated to any track but it shares the Track ID  
3213 with a reconstructed track it is counted as a false negative (FN).

3214 For the testing, I used a sample of 10000 FHC neutrino events inside the HPgTPC.  
3215 Figure 6.17 (left panel) shows the precision (blue line), sensitivity (yellow line), and  $F_1$

## CHAPTER 6. PARTICLE ID 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,  $t_{cluster}$ .

score (red line) I obtained for different values of  $\chi^2_{cut}$ . For comparison, the same metrics computed for the default algorithm with the current configuration are also shown (dashed lines). In the case of the new algorithm, I used both the  $\chi^2$ -based method to estimate the track direction described earlier (square markers) and the cheated direction from the Geant-level information (circle markers). For either of these we achieve similar values of the precision compared to the old code, while having a considerably higher sensitivity. It can be seen that cheating the direction of the tracks only makes a difference at high  $\chi^2_{cut}$ , past the optimal value of the cut around the  $F_1$  score maximum. Therefore, I set the  $\chi^2$  method as the default.

One of the possible weak points of this approach is that it relies on the position along the drift direction to make the decisions. Within the current ND-GAr design implemented in GArSoft, the timing information is provided by the ECal. That effectively means that prior to make the track-ECal associations the reconstructed  $x$  positions of the track trajectories differ from the simulated ones by an amount:

$$x_{reco}^{(n)} - x_{sim}^{(n)} = v_{drift} t_0, \quad (6.10)$$

where  $v_{drift}$  is the mean drift velocity in our medium and the initial time is in the range  $t_0 \in [0, t_{spill}]$  where  $t_{spill}$  is the spill length. For a  $10 \mu\text{s}$  spill this translates into a maximum 30 cm uncertainty on the drift direction position.

The current default in GArSoft sets  $t_0 = 0$ , but the functionality to randomly sample

### 6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

3234 this within the spill time is in place. Therefore, we need to understand what is the impact  
3235 of a non-zero  $t_0$  on the associations algorithm and foresee possible ways of minimising a  
3236 loss in performance.

3237 Figure 6.18 represents a possible option to tackle the association problem when  
3238 having events with a non-zero initial time  $t_0$ . The black and white circles represent the  
3239 original points, whereas the squares indicate the corrected positions. The end points of  
3240 the track and the propagated points up to the cluster radius are indicated using filled  
3241 and unfilled markers respectively. The red square represents the position of the cluster.

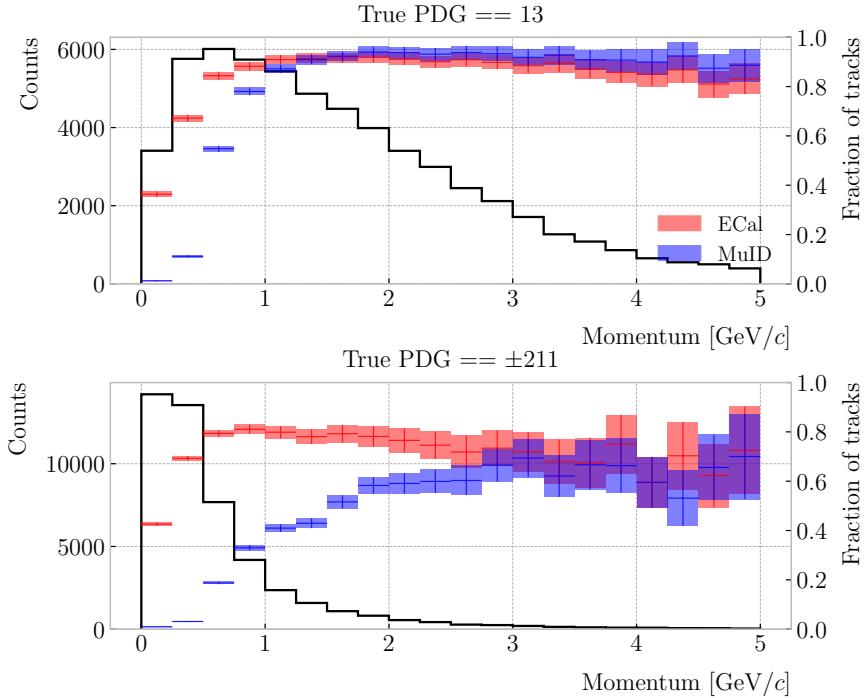
3242 Here I try to correct for the drift coordinate position using the time associated to the  
3243 cluster. Assuming that the drift time is much larger than the propagation time,  $t_{cluster}$   
3244 could be used as a good estimation of the  $t_0$ . An alternative can be using the earliest  
3245 time associated to a hit in said cluster. Doing this for each cluster before computing  
3246 the  $\chi^2$  value could be used as an alternative to knowing the specific value of the  $t_0$ , as  
3247 when the association is correct this will provide the right correction but its impact is  
3248 small enough to not change the position significantly in the case the cluster does not  
3249 correspond to a given track.

3250 I tested the effect of this correction again using a sample of 10000 FHC neutrino  
3251 events. Figure 6.17 (right panel) shows the precision (blue line), sensitivity (yellow line),  
3252 and  $F_1$  score (red line) for the case the cluster  $t_0$  correction is applied (square markers)  
3253 and for the no correction case (circle markers), as a function of  $\chi^2_{cut}$ . In this case, the  
3254 differences are particularly notorious at low values of the cut. It makes sense, as the  $t_0$   
3255 effect becomes subdominant when the distance we consider grows large. Overall, the  
3256 correction increases the sensitivity while keeping the precision almost unchanged. As a  
3257 result, I apply the  $t_0$  correction to the generated samples as the default.

#### 3258 6.3.2 Classification strategy

3259 The problem of the muon and charged pion separation has to be viewed in the broader  
3260 context of the particle identification in our detector. Focusing on the beam neutrino  
3261 interactions, it is clear that we are going to have muons and pions spanning a broad

## CHAPTER 6. PARTICLE ID IN ND-GAr

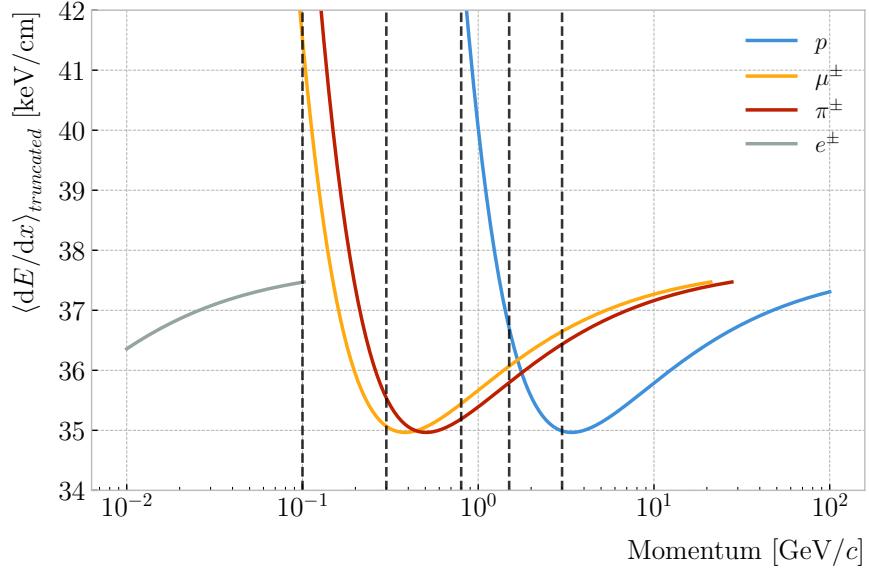


**Figure 6.19:** Momentum distribution for the reconstructed muons (top panel) and charged pions (bottom panel) in a FHC neutrino sample, together with the fraction of them reaching the ECal (red) and MuID (blue). Each entry corresponds to a reconstructed track, backtracked to a true muon or pion which has not produced any other reconstructed track.

3262 momentum range. Not only that, but we will also have other particles with similar  
 3263 characteristics that will make the classification even more challenging. Therefore, we are  
 3264 presented with a task that will depend heavily on the kinematic range we are looking at  
 3265 each time, as both the available information and the possible impurities of other particle  
 3266 species vary.

3267 For instance, distinguishing muons from pions could be difficult at low momenta, as  
 3268 a great number of them do not reach the ECal. Therefore, we could think of tailoring a  
 3269 version of the classification for that particular case, which could be complemented with  
 3270 a  $dE/dx$  measurement. Likewise, for momenta  $\gtrsim 1$  GeV muons and pions reach the  
 3271 calorimeters efficiently, but so do protons. Because of this, one can try to train another  
 3272 classifier for this energy range, and rely on other methods to remove as many of the  
 3273 protons as possible.

### 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.

3274     Figure 6.19 shows the momentum distribution of the reconstructed muons (top) and  
 3275     pions (bottom) in a FHC sample. It also contains the fraction of particles reaching the  
 3276     ECal (red) and MuID (blue), for the different momentum bins. In Fig. 6.20 I show the  
 3277     mean  $dE/dx$  of different particles as a function of the momentum, computed using the  
 3278     ALEPH parametrisation with the best fit parameters found in Subsec. 6.2.2.

3279     Using these two figures as references, I decided to approach the classification by  
 3280     dividing the problem into six different momentum regions. A summary of these can be  
 3281     found in Tab. 6.2. The basic idea is to exploit all the information that is available in  
 3282     each region and . For the problem at hand, I prepared separated samples of isotropic  
 3283     single muons and pions, with momenta uniformly distributed along the corresponding  
 3284     momentum range. Each sample contains 50000 events of the corresponding particle  
 3285     species. I did not generate samples for the first region, as it is assumed that the separation  
 3286     can be achieved using  $dE/dx$  only. For the last region, I generated particles up to a  
 3287     momentum of 10  $\text{GeV}/c$ , as that is well above the typical energies of muons and pions  
 3288     from FHC neutrino interactions in ND-GAr.

## CHAPTER 6. PARTICLE ID 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
$< 0.1 \text{ GeV}/c$	All tracks can be separated with $dE/dx$
$[0.1, 0.3) \text{ GeV}/c$	Use ECal for reaching muons and pions, $dE/dx$ for the rest
$[0.3, 0.8) \text{ GeV}/c$	Use ECal for muons and pions, $dE/dx$ for protons
$[0.8, 1.5) \text{ GeV}/c$	Use ECal and MuID for muons and pions, $dE/dx$ for protons
$[1.5, 3.0) \text{ GeV}/c$	Use ECal and MuID for muons and pions, ToF for protons
$\geq 3.0 \text{ GeV}/c$	Use ECal and MuID for muons and pions, $dE/dx$ and ToF for protons

3289     Additionally, I prepared another sample of 100000 FHC neutrino events. For each  
 3290    interaction, I select the reconstructed particles which were backtracked to true muons or  
 3291    charged pions. I use this dataset to perform validation checks, to see how the models  
 3292    trained with the single particle data generalise to a more realistic scenario.

3293     To tackle this classification problem, I make use of Boosted Decision Trees (BDT). A  
 3294    decision tree uses a flowchart-like structure to make decisions based on some input data.  
 3295    It starts from a root node, which represents the complete dataset, and then it splits  
 3296    this based on the variable or feature which gives the best separation between classes,  
 3297    creating two new nodes. The process repeats for each node until it reaches a certain  
 3298    limit, like a maximum number of splits or some tolerance criteria. The last set of nodes  
 3299    are often called leave nodes, and represent the final prediction of the classifier.

3300     Boosting refers to a family of methods to combine the predictions from multiple  
 3301    classifiers, following a sequential approach where each new model learns from the errors  
 3302    of the previous one. The process starts with a simple decision tree, which is used to  
 3303    make predictions on the training data. Then, the data points misclassified by the first  
 3304    model are assigned higher weights, and another decision tree is trained on the data with  
 3305    adjusted weights. The predictions of the two trees are then combined, and the cycle  
 3306    repeats for a predefined number of iterations. Gradient boosting uses the direction of  
 3307    the steepest error descent to guide the learning process and improve the accuracy with  
 3308    each iteration.

### 6.3. MUON AND PION SEPARATION IN THE ECal AND MuID

#### 3309 6.3.3 Feature selection and importance

3310 Using the reconstructed tracks as a starting point, I compute a number of ECal and  
3311 MuID variables for each of them. As there can be more than one cluster associated to a  
3312 track, what I do is collect all associated clusters and compute these variables from the  
3313 complete collection of associated hits. For the MuID, because it only features three layers  
3314 and typically there will be less hits, I also allow single hits to be associated with tracks<sup>7</sup>.  
3315 I can roughly divide the variables in three types: energy-related, geometry-related and  
3316 statistical. In the following, I briefly describe the variables related exclusively to the  
3317 ECal:

##### 3318 • Energy-related ECal

- 3319 – ECal total energy (ClusterTotalEnergy): sum of the energy of all the ECal  
3320 hits.
- 3321 – Mean ECal hit energy (HitMeanEnergy): mean of the hit energy distribution.
- 3322 – Standard deviation ECal hit energy (HitStdEnergy): standard deviation of  
3323 the hit energy distribution.
- 3324 – Maximum ECal hit energy (HitMaxEnergy): maximum of the hit energy  
3325 distribution.

##### 3326 • Geometry-related ECal

- 3327 – Mean distance hit-to-cluster (DistHitClusterMean): mean of the distance  
3328 distribution between the hits and the corresponding cluster's main axis.
- 3329 – RMS distance hit-to-cluster (DistHitClusterRMS): root mean square of the  
3330 distance distribution between the hits and the corresponding cluster's main  
3331 axis.

---

<sup>7</sup>At 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 ID IN ND-GAr

- 3332        – Maximum distance hit-to-centre (DistHitCenterMax): maximum of the  
3333                  distance distribution between the hits and the centre of the TPC.
- 3334        – Time-of-Flight velocity (TOFVelocity): slope obtained when fitting a straight  
3335                  line to the hit time versus hit distance to the centre (i.e.  $d = v \times t$ ).

### 3336        • Energy and geometry ECal

- 3337        – Radius 90% energy (Radius90E): distance in the hit-to-cluster distribution  
3338                  for which 90% of the total energy is contained in the hits that are closer to  
3339                  the axis (i.e. radius that contains 90% of the energy).

### 3340        • Statistical ECal

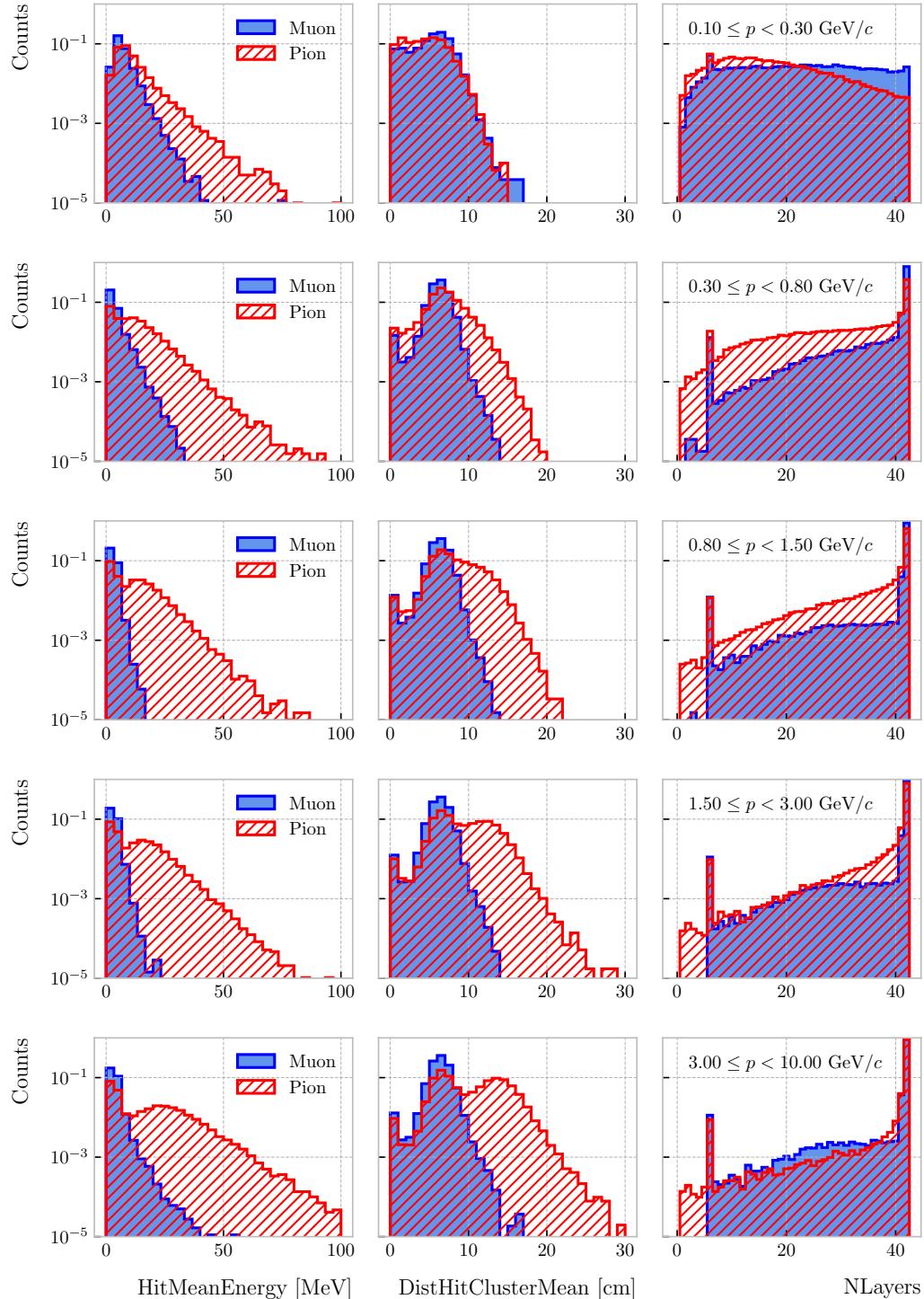
- 3341        – Number of hits (NHits): total number of hits associated to the track.
- 3342        – Number of layers with hits (NLayers): not really a count of all layers with  
3343                  hits but the difference between the last and the first layer with hits.

3344        Figure 6.21 shows the distributions of three different ECal variables, separating true  
3345        muons (blue) and charged pions (red), for the five momentum ranges considered. I chose  
3346        to show one feature from each category, namely the mean energy per hit (left column),  
3347        the mean distance between the hits and the centre of the cluster (middle column), and  
3348        the number of ECal layers with hits (right column). These give an idea of the separating  
3349        power of the different features, and how it changes considerably with the energy. In  
3350        the number of layers with hits distributions, the peak at 6 is due to the fact that the  
3351        first six ECal layers sit inside the pressure vessel<sup>8</sup>. Therefore, some of the particles get  
3352        stopped crossing it, never making it to the seventh layer.

3353        In the case of the MuID, because at low momenta a significant fraction of the particles  
3354        do not make it past the ECal, I only consider the information coming from this detector  
3355        for momenta  $\geq 0.8$  GeV/c, i.e. for the last three momentum regions. The variables I  
3356        extract from it are the following:

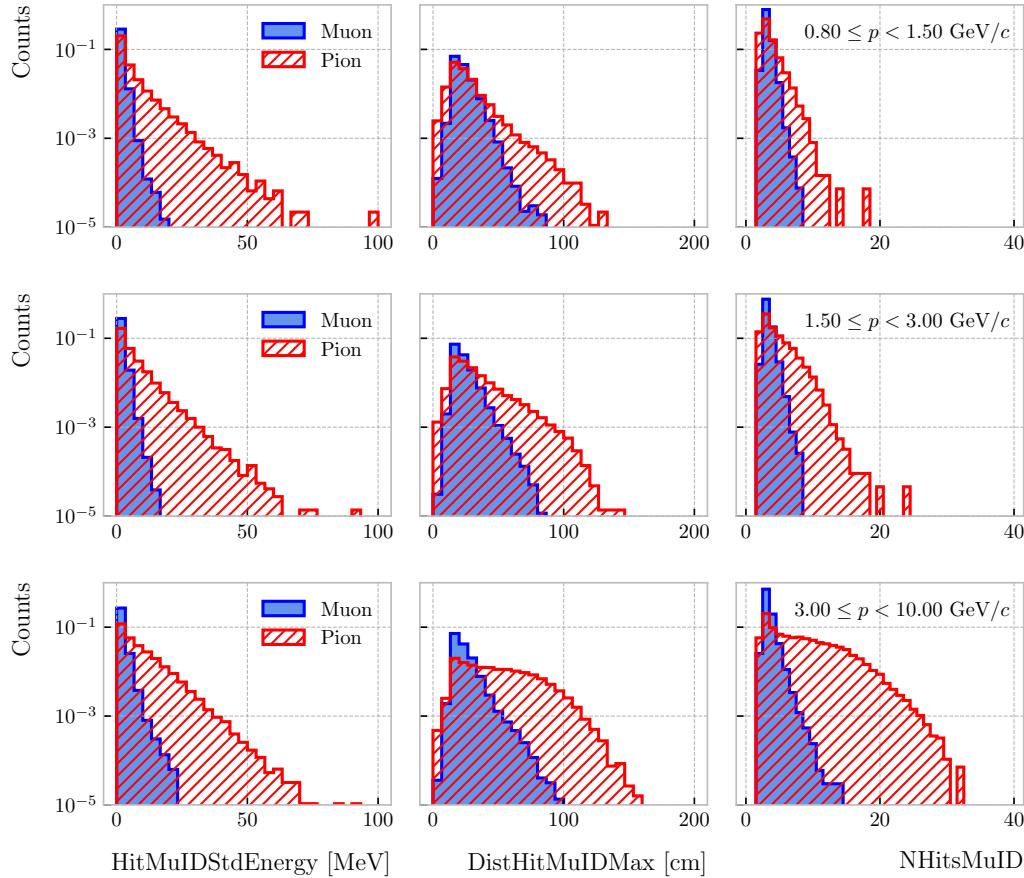
<sup>8</sup>Note to self: check this. I thought the ECal barrel had 8 layers of tiles, and that all of them were inside the pressure vessel.

### 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 ID 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.

3357

- Energy-related MuID

3358

- MuID total energy (ClusterMuIDTotalEnergy): sum of the energy of all the MuID hits.

3359

- Mean MuID hit energy (HitMuIDMeanEnergy): mean of the MuID hit energy distribution.

3360

- Standard deviation MuID hit energy (HitMuIDStdEnergy): standard deviation of the MuID hit energy distribution.

3361

- Maximum MuID hit energy (HitMuIDMaxEnergy): maximum of the MuID hit energy distribution.

3362

3363

3364

3365

### 6.3. MUON AND PION SEPARATION IN THE ECAL AND MUID

3366        • **Geometry-related MuID**

- 3367            – Maximum distance MuID hit-to-hit (DistHitMuIDMax): maximum distance  
3368              between pairs of MuID hits (not sure this is a good variable, distribution  
3369              looks nuts).
- 3370            – Maximum distance MuID hit-to-centre (DistHitCenterMuIDMax): maximum  
3371              of the distance distribution between the MuID hits and the centre of the  
3372              TPC.

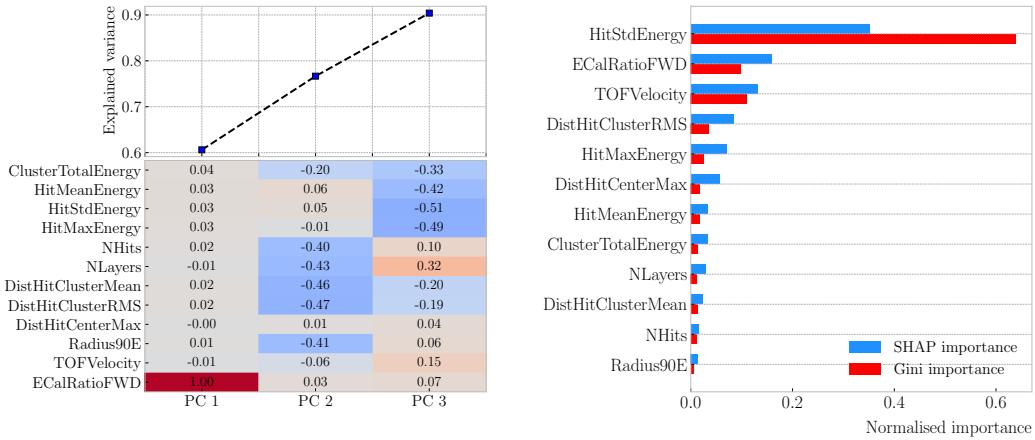
3373        • **Statistical MuID**

- 3374            – Number of hits (NHitsMuID): total number of MuID hits associated to the  
3375              track.
- 3376            – Number of layers with hits (NLayersMuID): not really a count of all layers  
3377              with MuID hits but the difference between the last and the first layer with  
3378              MuIDhits.

3379        Figure 6.22 shows the distributions of three different MuID variables, separating true  
3380        muons (blue) and charged pions (red), for the three momentum ranges which use the  
3381        muon tagger information. In this case I decided to standard deviation of the MuID hit  
3382        energy distribution (left column), the maximum distance between the MuID hit pairs  
3383        (middle column), and the number of MuID hits (right column). These variables are used  
3384        together with the ECal features at high momenta, providing additional disambiguation  
3385        power.

3386        Once our features have been defined, one can do some exploratory analysis to  
3387        understand how well the variables describe the target class, and avoid the black-box  
3388        approach by what features are most relevant for the learning process. This way, I  
3389        performed a feature analysis for each of the momentum ranges I divided this classification  
3390        problem into. It follows three steps: first a principal component analysis (PCA), followed  
3391        by a feature importance study using Gini and Shapley values, and finally a feature  
3392        permutation importance analysis.

## CHAPTER 6. PARTICLE ID 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  $0.3 \leq p < 0.8 \text{ GeV}/c$ .

The PCA is useful to understand the variance of the feature space. It is an unsupervised machine learning technique that allows the user to perform a dimensionality reduction. It uses a singular value decomposition of the input features to project them into a lower dimensional space. The idea is to find the matrix  $\mathbf{C}_m$ , whose columns are the first  $m$  orthonormal eigenvectors of the input covariance matrix. Consider the  $n \times p$  real matrix of input data  $\mathbf{X}$ , where  $n$  is the number of samples and  $p$  the number of features. If  $\mathbf{X}$  is centred, i.e. the means of its columns are equal to zero, we can write the covariance matrix of  $\mathbf{X}$  as  $\mathbf{C} = \mathbf{X}^\top \mathbf{X} / (n - 1)$ . This matrix can be diagonalised, yielding:

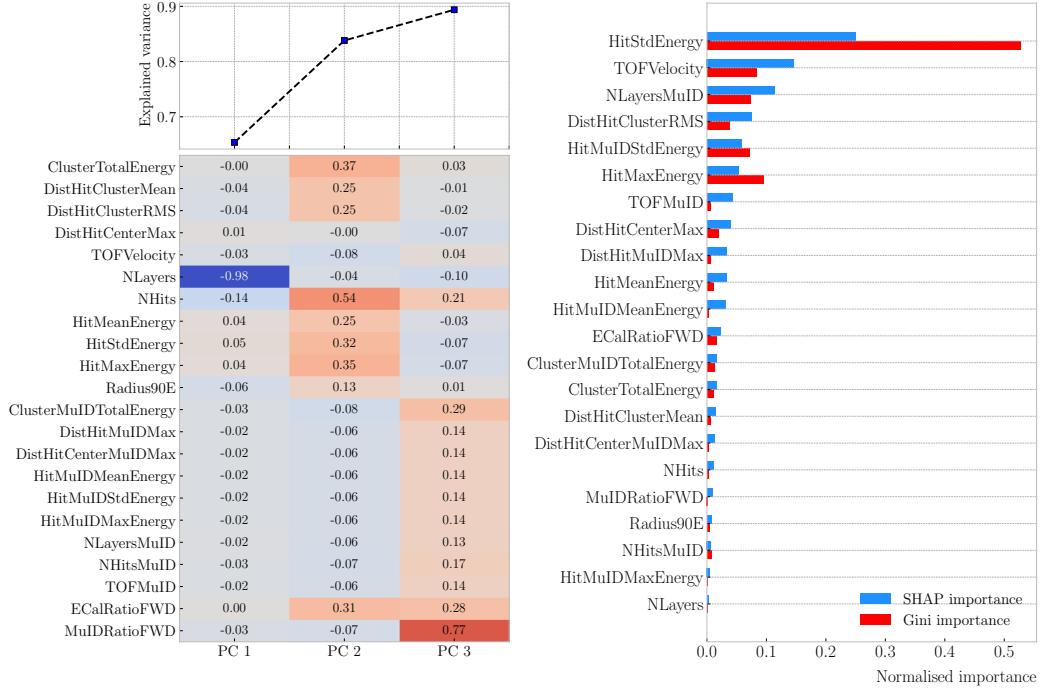
$$\mathbf{C} = \mathbf{V} \mathbf{L} \mathbf{V}^\top, \quad (6.11)$$

where  $\mathbf{V}$  is a matrix of eigenvectors and  $\mathbf{L}$  a diagonal matrix with eigenvalues  $\lambda_i$ . Then, performing SVD on  $\mathbf{X}$  gives us:

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{W}^\top, \quad (6.12)$$

where  $\mathbf{U}$  is a unitary matrix, whose columns are called left singular vectors,  $\mathbf{S}$  is a 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.

3405 as right singular vectors. This way, we can write:

$$\mathbf{C} = \mathbf{WSU}^\top \mathbf{USW}^\top / (n - 1) = \mathbf{W} \frac{\mathbf{S}^2}{n - 1} \mathbf{W}^\top. \quad (6.13)$$

3406 meaning that the right singular vectors are also the eigenvectors of the covariance matrix.

3407 The SVD can be computed numerically following an iterative approach.

3408 This way, taking an input data vector  $X \in \mathbb{R}^n$ , the resulting feature vector  $Y \in \mathbb{R}^m$

3409 is given by:

$$Y = \mathbf{C}_m^\top X. \quad (6.14)$$

3410 The new features capture most of the variance of the original sample, while being lower

3411 dimensional, as  $m < n$ .

3412 Before applying the PCA reduction one needs to centre and scale the input data.

## CHAPTER 6. PARTICLE ID IN ND-GAR

3413 Centring is necessary when using SVD to obtain the eigenvectors of the covariance  
3414 matrix, as only in that case we can do the identification with the right singular vectors  
3415 from the input data. Scaling is needed when variables are on different scales, as some  
3416 can then dominate the PCA procedure.

3417 I used the PCA module of `scikit-learn`, together with the `RobustScaler`, which  
3418 centres the data and scales it based on the interquartile range. In Fig. 6.23 (left panel)  
3419 and Fig. 6.24 (left panel) I show the results I obtained from the PCA for the momentum  
3420 ranges  $0.3 \leq p < 0.8 \text{ GeV}/c$  and  $0.8 \leq p < 1.5 \text{ GeV}/c$ , respectively. Notice that in  
3421 the second case the number of features increases considerably, as this is the first region  
3422 which uses the MuID variables. I found that, in all the cases, adding a fourth PC does  
3423 not add additional information. As it can be seen in the top panels of the figures, the  
3424 cumulative explained variance is already over 80% with three PCs.

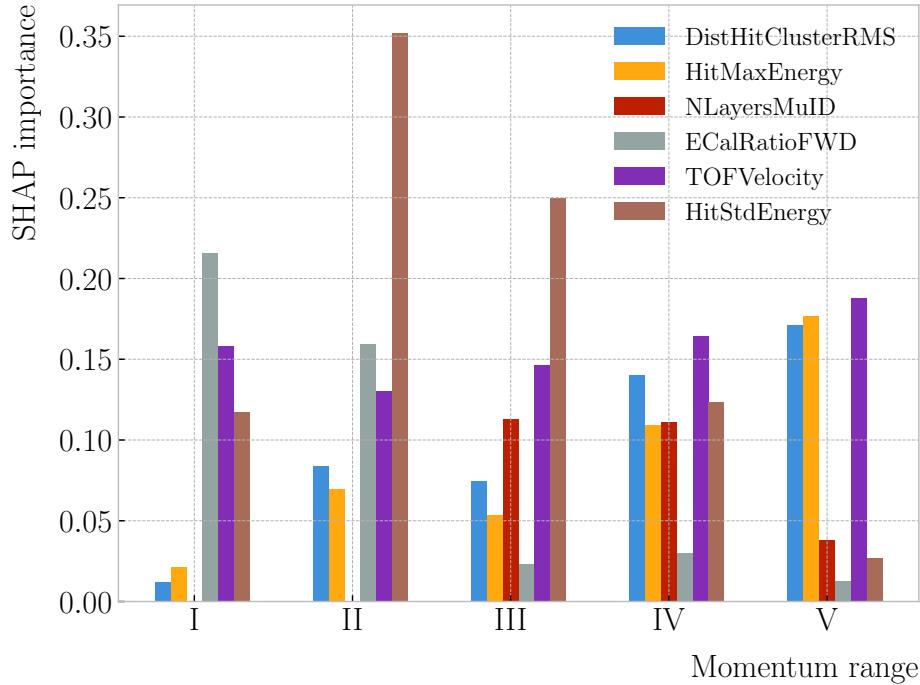
3425 The bottom panels show the contribution of the variables to the principal axes. For  
3426 the two first momentum regions, I observe a tendency of the energy-related and the  
3427 geometry-related ECal variables to be clustered together. For the other ranges, when  
3428 I include the MuID variables, there seems to be a division between ECal and MuID  
3429 variables. For these, it seems like the number of ECal layers with hits also plays an  
3430 important role.

3431 The next step in the analysis is to quantify the importance of the features based on  
3432 two additional metrics, namely the Gini and the Shapley values. The Gini importance,  
3433 often called mean decrease impurity, is based on how much a feature contributes to the  
3434 purity improvement at the splits in each decision tree. The purity is measured in terms  
3435 of the Gini impurity index, defined as:

$$I_G = 1 - \sum_i f_i, \quad (6.15)$$

3436 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.

3437 compute the weighted decrease in impurity as:

$$\Delta_G = \frac{N_t}{N} \left( I_G - \frac{N_t^R}{N_t} I_G^R - \frac{N_t^L}{N_t} I_G^L \right), \quad (6.16)$$

3438 where  $N$  represents the total number of samples,  $N_t$  the number of samples at the current  
 3439 node,  $N_t^R$  and  $N_t^L$  the number of samples in the right and left children respectively,  
 3440  $I_G$  is the Gini impurity at the current node, and  $I_G^R$  and  $I_G^L$  the Gini impurities of the  
 3441 resulting right and left children.

3442 For each decision tree, one will have a normalised vector with the accumulated  
 3443 decrease in Gini impurity for each feature. In the case of a BDT, the feature importances  
 3444 are simply the mean for all the estimators in the ensemble<sup>9</sup>.

3445 The concept of Shapley values originated in the context of game theory, and it  
 3446 measures the marginal contribution of a feature in enhancing the accuracy of a classifier.

---

<sup>9</sup>Note 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 ID IN ND-GAr

3447 Take  $F$  to be the set of all features in a problem, and  $S \subseteq F$  the subset of features. To  
 3448 compute the Shapley value of the  $i$ -th feature, one has to train a model with that feature  
 3449 present,  $f_{S \cup \{i\}}$ , and another model trained without it,  $f_S$ . This has to be repeated for  
 3450 all possible combinations of subsets  $S \subset F \setminus \{i\}$ , and evaluating the models predictions  
 3451 on the appropriate sets of data  $x_S$ . This way, the Shapley value results:

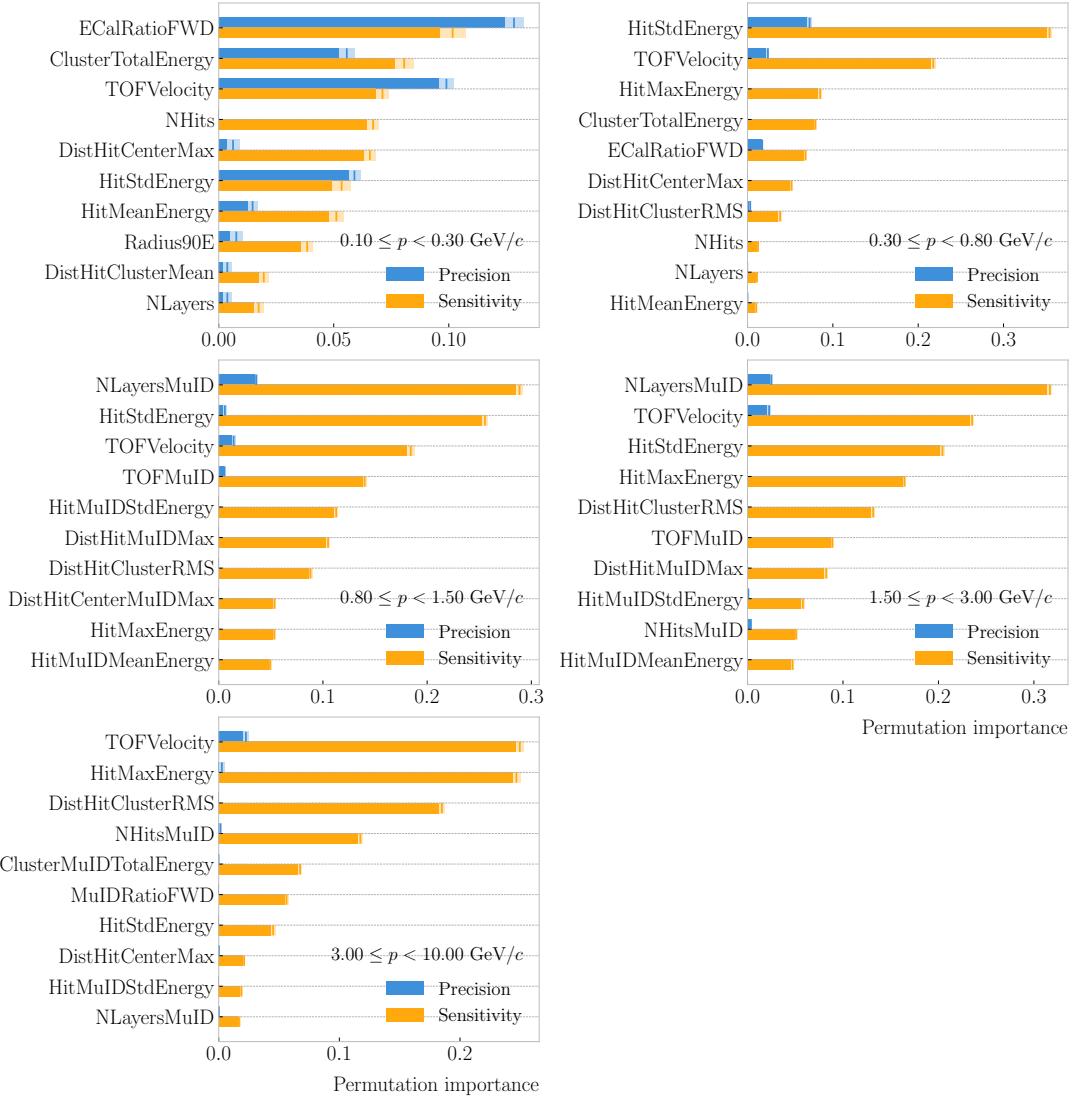
$$\varphi_i = \sum_{S \subset F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]. \quad (6.17)$$

3452 I trained the `GradientBoostingClassifier` from `scikit-learn` with the default  
 3453 configuration in order to evaluate both the Gini and Shapley importances. The Gini  
 3454 scores are automatically computed by `scikit-learn`, using the training data. For the  
 3455 Shapley importance, I used the implementation from the `SHAP` package, computing  
 3456 it using the test sample. The results can be seen in Fig. 6.23 (right panel) and  
 3457 Fig. 6.24 (right panel), again for the momentum ranges  $0.3 \leq p < 0.8 \text{ GeV}/c$  and  
 3458  $0.8 \leq p < 1.5 \text{ GeV}/c$ . The length of the bars denote either the SHAP (blue) or the Gini  
 3459 (red) importance of the feature. One interesting thing to notice is that, when looking at  
 3460 the Gini importance, there is always one feature that dominates over the rest. This is  
 3461 not the case for the SHAP importance, where importances tend to be more balanced.

3462 Across all momentum ranges, I observe that the most important features are. For  
 3463 the five momentum ranges considered, only six variables sit in the top five at least once.  
 3464 Figure 6.25 shows the evolution of the SHAP importance of these six features. It is  
 3465 interesting to see that the time-of-flight variable keeps its importance almost unchanged  
 3466 for all momenta. Also, it looks like the ECal energy ratio gets less relevant the higher the  
 3467 momentum is, but the RMS of the hit-to-cluster distance distribution and the maximum  
 3468 ECal hit energy become more important in the last momentum ranges.

3469 The last step in the feature selection analysis is the feature permutation. This  
 3470 technique measures the contribution of each feature to the performance of a model by  
 3471 randomly shuffling its values and checking how some scores degrade. For the present  
 3472 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 ID IN ND-GAR

3473 two are the most relevant metrics from a physics point of view. The `scikit-learn`  
3474 module provides the user with a method to perform the permutation scans.

3475 The results of these are shown in Fig. 6.26. For the different momentum ranges  
3476 I show the permutation importances for the ten most important features. For each  
3477 of the variables I report the effect the permutations have on the precision (blue) and  
3478 sensitivity (yellow) of the models. The bars indicate the importance value, with the  
3479 lighter part representing one standard deviation around the mean (hinted as an additional  
3480 vertical line). Something to notice is that, in the first momentum region, the feature  
3481 permutations have an effect on both the precision and the sensitivity. However, for the  
3482 rest the precision is almost unaffected, while the sensitivity changes are considerably  
3483 larger.

3484 It is also interesting to see that most of the variables identified as important here  
3485 are the same I found when looking at the Shapley values. The behaviour of these across  
3486 the momentum ranges is also similar, with the same patterns of some features being  
3487 important at low momenta and then dropping in importance for the high momentum  
3488 ranges.

3489 Wit this, I conclude the study of the features. I have prepared the training and  
3490 testing datasets and understood what features are likely to have the largest impact on  
3491 the performance of the classifiers.

### 3492 6.3.4 Hyperparameter optimisation

3493 Any BDT requires the user to specify a number of parameters that will dictate its  
3494 behaviour. They can be divided into two categories: (i) tree-specific parameters, which  
3495 affect each individual tree in the model, and (ii) boosting parameters, which control the  
3496 boosting operation in the model. The value of these so-called hyperparameters affect the  
3497 performance and predictive power of the models. Therefore, one needs to carefully select  
3498 their optimal values in order to extract as much information as possible from the data.

3499 From all the parameters used to define a tree in the `scikit-learn` implementation  
3500 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

3501 are mutually exclusive, but also because I noticed that others have little effect on the  
3502 problem at hand. Therefore, the parameters I investigate are the following:

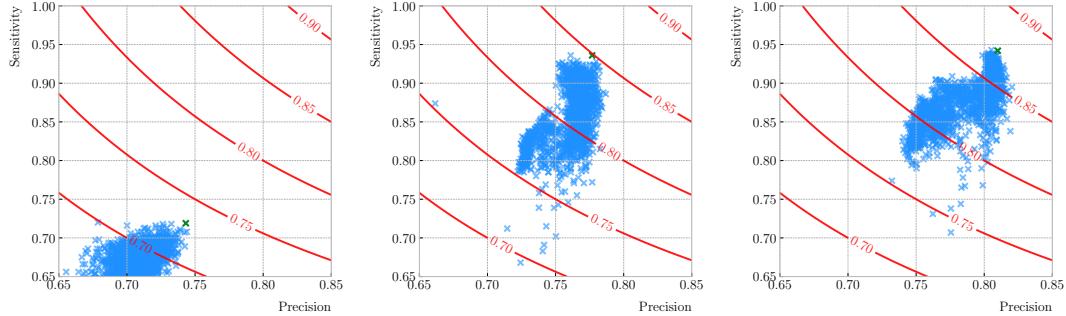
- 3503 • `min_samples_split`: defines the minimum number of samples required in a node  
3504 to be considered for splitting. Higher values prevent a model from learning relations  
3505 which might be highly specific to the particular sample, but may lead to under-fitting  
3506 if the value is too low.
- 3507 • `min_samples_leaf`: defines the minimum samples required in a leaf node. For  
3508 imbalanced problems it should take a low value, as there will not be many cases  
3509 where the minority class dominates.
- 3510 • `max_depth`: maximum depth of a tree. Useful to prevent over-fitting, as higher  
3511 depth will allow a model to learn relations specific to the training sample.

3512 In the case of the boosting parameters, the ones I look at are:

- 3513 • `learning_rate`: determines the impact of each tree on the final outcome. Low  
3514 values make the model robust to the specific characteristics of a tree, and thus  
3515 allow it to generalise well. However, that usually requires a large number of trees  
3516 to model the data properly.
- 3517 • `n_estimators`: number of sequential trees to be trained. In general, BDTs are  
3518 fairly robust at higher number of trees but it can still overfit at a point.
- 3519 • `subsample`: fraction of observations to be selected for each tree. Values slightly  
3520 less than 1 make the model robust by reducing the variance.

3521 In general, hyperparameters depend on each other. Thus, it is not possible to  
3522 optimise them independently. In the literature, we find two main strategies to explore  
3523 the hyperparameter space. We could use a grid search, in which one discretises a  
3524 portion of the space of hyperparameters and evaluates the model at each point. Another  
3525 approach is the randomised search, where a certain number of random configurations of  
3526 hyperparameters are explored.

## CHAPTER 6. PARTICLE ID 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.

3527 In this case, I used the random search to scan the hyperparameter space. Also,  
 3528 because it is not guaranteed that a set of hyperparameters can be efficiently applied  
 3529 across different datasets, I perform the optimisation for each of the momentum ranges  
 3530 considered. Table 6.3 shows the list of hyperparameters considered, and the range within  
 3531 which I let them vary. I decided to fix the number of estimators to 400 in all cases, as  
 3532 its value is correlated with that of the learning rate.

3533 I evaluate 10000 different hyperparameter configurations for each momentum range.  
 3534 For the hyperparameter tuning, I used subsamples containing 10% of the full datasets,  
 3535 keeping the original proportions between classes, in order to reduce the computational  
 3536 load. The performance of the models was assessed using a stratified 3-fold cross-validation  
 3537 with replacement. Cross-validation involves dividing the data in a number of subsets,  
 3538 training the model using some of them, and testing it with the rest. In our case, I  
 3539 divide the data in 3 equal-sized subsets, maintaining the class proportions of the original  
 3540 dataset. Then, for 3 consecutive iterations, I train the models using 2 of the subsets  
 3541 while I compute the precision and sensitivity scores with the other. This approach  
 3542 provides a more robust estimate of the performance on unseen data.

3543 Figure 6.27 shows the results in the precision versus sensitivity plane, for the  
 3544 momentum regions I, III and V (from left to right). The contours represent the curves  
 3545 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

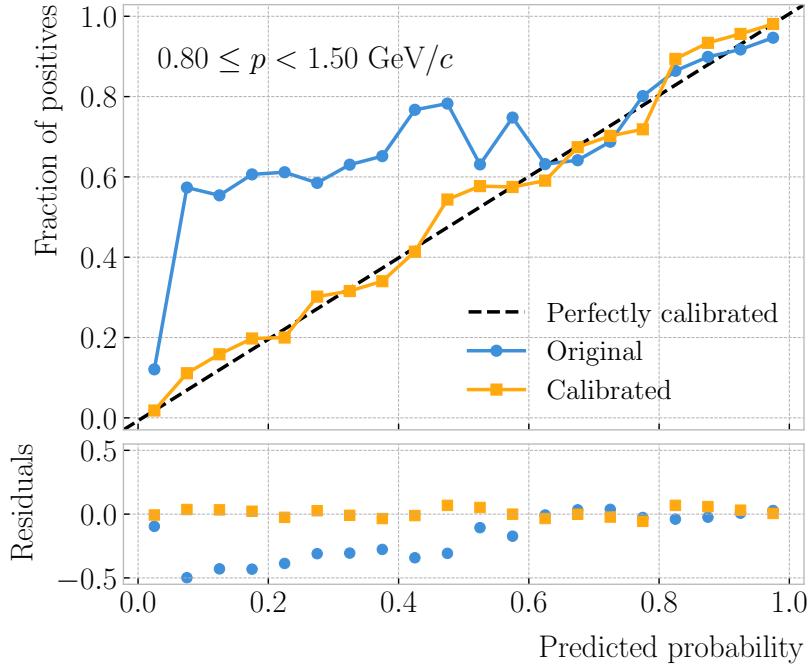
**Table 6.4:** Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges.

Metric	Value $\pm 1\sigma$				
	I	II	III	IV	V
Accuracy	$0.779 \pm 0.003$	$0.812 \pm 0.003$	$0.846 \pm 0.002$	$0.861 \pm 0.003$	$0.874 \pm 0.002$
Precision	$0.769 \pm 0.003$	$0.752 \pm 0.005$	$0.788 \pm 0.002$	$0.805 \pm 0.003$	$0.815 \pm 0.003$
Sensitivity	$0.745 \pm 0.009$	$0.921 \pm 0.006$	$0.965 \pm 0.002$	$0.967 \pm 0.002$	$0.976 \pm 0.001$
$F_1$ -score	$0.757 \pm 0.004$	$0.828 \pm 0.003$	$0.867 \pm 0.002$	$0.879 \pm 0.002$	$0.889 \pm 0.002$
ROC AUC	$0.868 \pm 0.003$	$0.865 \pm 0.003$	$0.899 \pm 0.002$	$0.902 \pm 0.002$	$0.911 \pm 0.001$

3546 to select the optimal configurations (indicated in the plots with a green cross), I chose  
 3547 the point with the highest  $F_1$ -score.

3548 The results for the different momentum ranges are summarised in Tab. 6.3. One  
 3549 can see some consistency in hyperparameter choices, with models generally preferring  
 3550 small values for the tree-specific parameters, small learning rate, and relatively large  
 3551 subsample sizes.

3552 Now that I have obtained the optimal values of the hyperparameters, I can train  
 3553 the different BDTs. In this case I use the complete datasets, keeping 20% of the data  
 3554 for testing. Table 6.4 shows the values of the different performance metrics obtained  
 3555 using the selected hyperparameters and 5-fold cross-validation. The last row indicates  
 3556 the value of the area under the receiver operating characteristic (ROC) curve. This  
 3557 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).

3558 included it here as it is a classic model metric used in the machine learning community.

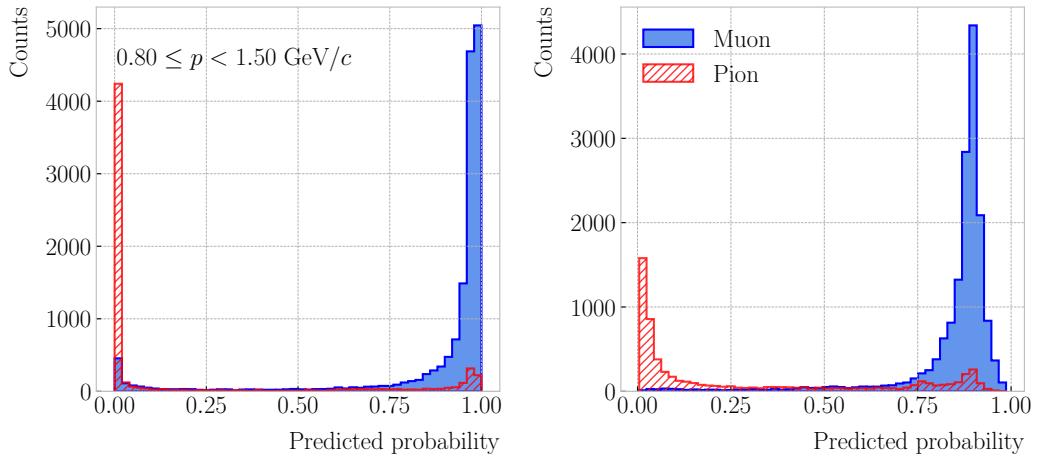
3559 Overall, there is a clear trend of models performing better at higher momentum.

### 3560 6.3.5 Probability calibration

3561 So far, the trained BDTs are able to provide predictions of the class labels. Ideally,  
 3562 one would like the output of a classifier to give a confidence level about the prediction.  
 3563 However, it is not straightforward to interpret the outputs of our BDTs in terms of  
 3564 probabilities.

3565 A way to visualise how well the predictions of a classifier are calibrated is using  
 3566 reliability diagrams [160]. They represent the probability of the positive label versus the  
 3567 probability predicted by the classifier. These can be obtained by binning the predicted  
 3568 probabilities, and then compute the conditional probability  $P(y_{true} = 1 | y_i \leq y_{pred} <$   
 3569  $y_{i+1})$  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  $0.3 \leq p < 0.8 \text{ GeV}/c$ .

3570 diagram of a perfectly calibrated classifier would be a diagonal line.

3571 In this case, I try to correct the raw response of the classifiers by applying a sigmoid  
3572 function:

$$\sigma(x; A, B) = \frac{1}{1 + e^{Ax+B}}, \quad (6.18)$$

3573 where the parameters  $A$  and  $B$  are real numbers determined using the method of least  
3574 squares.

3575 For each classifier, I perform a grid search to obtain the optimal values of  $A$  and  $B$ .  
3576 For any pair, I compute the predicted probabilities as  $y_{pred} = \sigma(y_{raw}; A, B)$ , where  $y_{raw}$   
3577 are the raw predictions of the classifier<sup>10</sup>. Then, I calculate the corresponding reliability  
3578 curve, and take the sum of the squared residuals between it and the response of the  
3579 perfectly calibrated classifier.

3580 Figure 6.28 shows the reliability diagrams for the original (blue) and calibrated  
3581 (yellow) probability predictions of the classifier for the III momentum range,  $0.3 \leq p <$   
3582  $0.8 \text{ GeV}/c$ . The original response of the classifier is given by  $y_{pred} = \sigma(y_{raw}; -2, 0)$ ,  
3583 which is the transformation applied by `scikit-learn` to produce the probability estimate.  
3584 Notice how the calibrated prediction matches the ideal response much better than the

---

<sup>10</sup>In `scikit-learn` these correspond to the outputs of the `decision_function` method.

## CHAPTER 6. PARTICLE ID IN ND-GAr

3585 original, across all the probability range.

3586 One can also compare the responses of the uncalibrated and calibrated classifiers  
3587 broken down by true particle type, as shown in Fig. 6.29. It can be seen that the  
3588 distributions for both muons (blue) and charged pions (red) smoothen after calibration,  
3589 but still the separating power of the classifier remains unchanged.

3590 At this point, having the trained classifiers and the probability calibration parameters,  
3591 I am able to assess the performance of the classification strategy in a physics-relevant  
3592 case.

### 3593 6.3.6 Performance

## 3594 6.4 ECal time-of-flight

3595 Looking at Fig. 6.20, it is clear that for momentum values in the range  $1.0 - 3.0 \text{ GeV}/c$   
3596 it is not possible to separate pions and protons using a  $\langle dE/dx \rangle$  measurement in the  
3597 HPgTPC. However, in the previous section I assumed that protons at those energies  
3598 could be identified by other means, and therefore were not an issue for the muon and  
3599 pion discrimination.

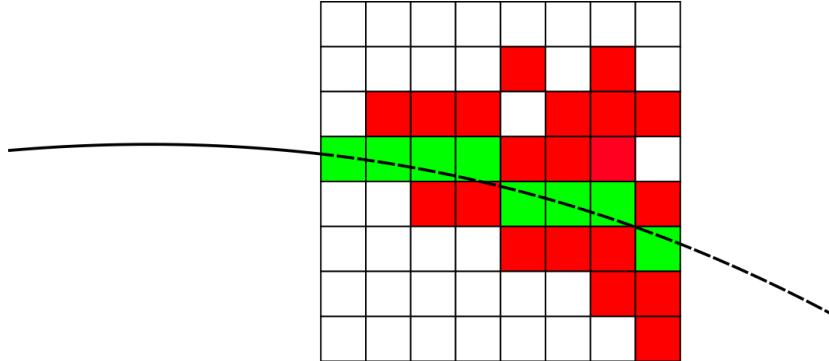
3600 Some detectors, like ALICE [161] or the ILD concept [162], complement the PID  
3601 capabilities of their gaseous trackers with time-of-flight measurements. The use of  
3602 fast timing silicon sensors, with hit time resolutions under 100 ps, would allow for the  
3603 identification of charged hadrons via a ToF measurement up to  $5.0 \text{ GeV}/c$ . In the case  
3604 of ND-GAr, one could think of using the inner layers of the ECal, the ones consisting of  
3605 high-granularity tiles, to obtain a ToF-based PID, with some inputs from the TPC.

3606 Measuring the momentum and the velocity of a charged particle allows for a  
3607 determination of the mass through the relativistic momentum formula:

$$m = \frac{p}{\beta} \sqrt{1 - \beta^2}. \quad (6.19)$$

3608 In our case, the momentum is measured in the TPC, using the curvature and the dip

## 6.4. ECAL TIME-OF-FLIGHT



**Figure 6.30:** Schematic of the hit selection used for the ToF measurement. The grid represents the layers of the inner ECAL, with coloured squares indicating the tiles with hits. Green squares indicate the selected hits.

3609 angle of the helix inside the magnetic field. The velocity of the particle can be written  
3610 as:

$$\beta = \frac{\ell_{track}}{c\tau}, \quad (6.20)$$

3611 where  $\ell_{track}$  is the length of the track, and  $\tau$  the arrival time to the ECAL.

3612 In GArSoft, the track length is computed at the Kalman filter stage. It is simply the  
3613 sum of the line segments along the track, either in the forward or backward fit. In this  
3614 case, because we are only interested in the particles that make it to the ECAL, I choose  
3615 the fit direction based on the results of the track-cluster associations.

3616 Additionally, because the last 30 cm of the TPC radius are uninstrumented<sup>11</sup>, I need  
3617 to correct for the length of the tracks. Using the track fit parameters to propagate the  
3618 helix to its entry point in the ECAL, one can write the total track length as:

$$\ell_{track} = \ell + \left| \frac{\phi_{EP} - \phi}{R^{-1}} \right| \sqrt{1 + \tan^2 \lambda}, \quad (6.21)$$

3619 where  $\phi_{EP}$  is the angle of rotation at the entry point to the calorimeter, and  $\ell$ ,  $\phi$ ,  $R$  and  
3620  $\lambda$  are the track length, angle of rotation, radius of curvature and dip angle at the last  
3621 point in the fit, respectively.

3622 To test the idea of performing a ToF measurement with the inner ECAL, I generated

---

<sup>11</sup>Note to self: check this number.

## CHAPTER 6. PARTICLE ID IN ND-GAr

3623 two data samples. Each consists of 10000 single particle events, either charged pions or  
3624 protons. Their momenta are uniformly distributed in the range  $0.5 - 5.0 \text{ GeV}/c$ , and  
3625 their directions are isotropic. I process each sample using different values of the time  
3626 resolution, from  $\Delta\tau = 0$ , the perfect time resolution case for comparison, to the current  
3627 nominal value of  $\Delta\tau = 0.7 \text{ ns}$ , and the worse scenario of  $\Delta\tau = 1.0 \text{ ns}$ .

### 3628 6.4.1 Arrival time estimations

3629 In the simulation, the limited time resolution of the ECal is taken into account by  
3630 applying a Gaussian smearing to the true hit times. Other effects, like the digitisation  
3631 of the signals, are not taken into account and fall beyond the scope of this study. After  
3632 the track-cluster, one ends up with a collection of ECal hits associated to each particle.  
3633 From these, the arrival time of the particle to the ECal can be extracted.

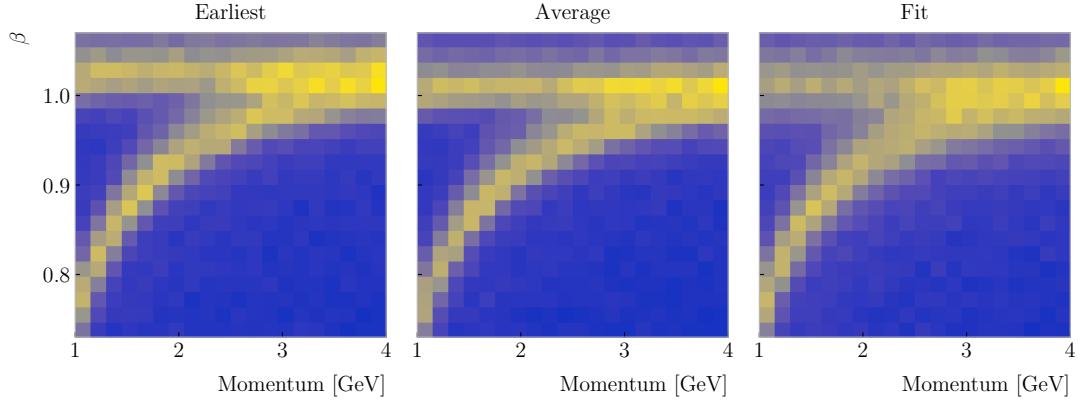
3634 The simplest possibilities are to either take the time of the earliest hit or the hit  
3635 closest to the entry point. Because these two coincide, in general, I focused only in  
3636 the earliest hit time. However, this needs to be corrected, to account for the distance  
3637 travelled from the entry point to the position of the hit:

$$\tau_{earliest} = \tau_{hit} - \frac{d_{EP-hit}}{c}, \quad (6.22)$$

3638 where  $\tau_{hit}$  is the time of the earliest hit, and  $d_{EP-hit}$  is the distance between that hit  
3639 and the entry point of the particle to the ECal. This is computed as the arc length  
3640 between the entry point and the point of the extrapolated helix up to the layer of the  
3641 hit. This way of correcting the time assumes  $c$  for the propagation of the particle, which  
3642 may lead to biased estimates.

3643 I also tried to estimate the arrival times using information from the rest of the hits.  
3644 In order to do this, as a simplifying assumption, I approximate the hadronic shower  
3645 considering only its MIP component. For each layer, I keep only the hit in the tile closest  
3646 to the point of the extrapolated track up to that layer. Figure 6.30 shows an example of  
3647 how this hit selection works. The dashed line represents the extrapolated track, while

## 6.4. ECAL TIME-OF-FLIGHT



**Figure 6.31:** Particle velocity versus momentum measured with different ECal arrival time estimations. From left to right: earliest hit time, average hit time, and fitted hit time. In all cases the time resolution is  $\Delta\tau = 0.1$  ns.

the coloured squares are the tiles containing hits. Green indicates the tiles closer to the track in each layer (in the sketch they correspond to the grid columns).

Now, I can use these collections of hits to estimate the arrival times. A possibility is to take the average of the times of the selected hits, denoted  $\tau_{average}$ . For that to work, one needs to correct these times, in a similar way as in Eq. (6.22), before taking the average. However, as before, this correction assumes that the particle travels at the speed of light inside the ECal. Another option is to perform a linear fit to the hit times and the distances to the entry point. In that case, the arrival time would be the fitted value of the intercept,  $\tau_{fit}$ . This method would not assume a speed of light propagation.

Figure 6.31 shows the velocity estimations as a function of the particle momentum, for the earliest hit time (left panel), average hit time (middle panel), and fitted hit time (right panel). The two bands correspond to the  $\pi^\pm$  and the  $p$  particles.  $\Delta\tau = 0.1$  ns. Notice how, for the earliest hit time method, the velocities are significantly biased towards larger values. For the multi-hit methods, the  $\tau_{fit}$  estimate appears to produce a larger variance than when using the  $\tau_{average}$  method.

## CHAPTER 6. PARTICLE ID IN ND-GAr

### 3663 6.4.2 Proton and pion separation

3664 Once we have the velocities of the particles, one can estimate their masses through  
3665 Eq. (6.19). The resulting mass spectra are shown in Fig. 6.32. I computed the masses  
3666 for the three arrival time estimates discussed above, and three different values of the  
3667 time resolution:  $\Delta\tau = 0.00$  (perfect time resolution),  $\Delta\tau = 0.10$  ns, and  $\Delta\tau = 0.70$  ns.  
3668 Although in all cases we have the same number of events, it appears as if the entries  
3669 in the histograms decrease as the time resolution increases. Sometimes, the particles  
3670 get unphysical values of  $\beta > 1$ , and in turn they do not contribute to the mass spectra.  
3671 This is more likely to happen for higher values of  $\Delta\tau$ .

3672 As noted before, the average hit time method produces the most robust estimates  
3673 when increasing  $\Delta\tau$ . Intuitively this makes sense, as by taking the mean one averages  
3674 out the effect of the Gaussian smearing. Going forward, I will use this arrival time  
3675 estimator, as it appears to be the best performing one.

3676 It is possible to use the velocity estimations to select a sample of protons. In this  
3677 case, I do so by dividing the relevant momentum range in bins of  $0.1 \text{ GeV}/c$ . For each  
3678 momentum bin, I compute the expected velocity for the protons via the inverse of Eq.  
3679 (6.19), and then take the fractional residuals of the measured velocities. Using that  
3680 distribution, I choose the cut that maximises the  $F_1$ -score of the proton selection.

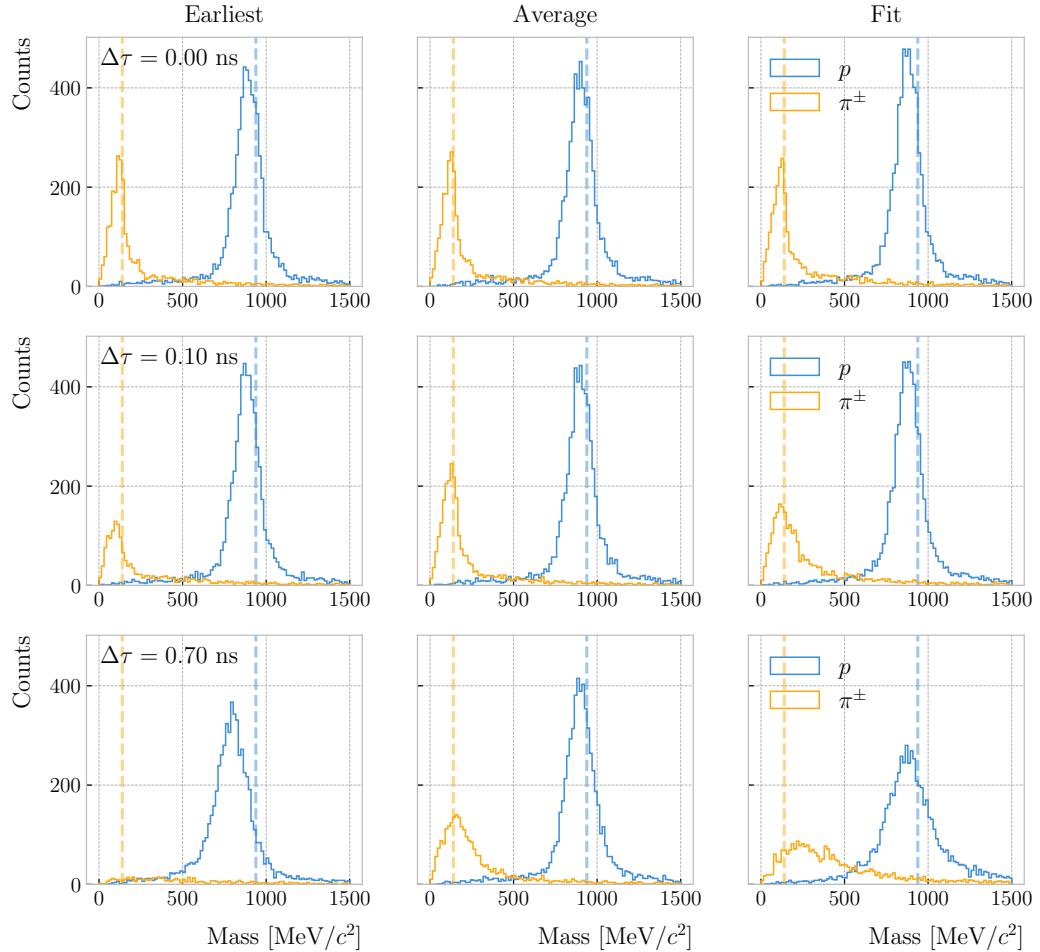
3681 The results can be seen in Fig. 6.33, for the case  $\Delta\tau = 0.10$  ns. As expected from  
3682 Fig. 6.31, the performance of the selection degrades rapidly with increasing momentum.  
3683 However, the purity is still around 75% at  $3.0 \text{ GeV}/c$ . This is likely to be sufficient, as  
3684 we do not expect protons or charged pions with higher energies from the beam neutrino  
3685 interactions.

3686 Figure 6.34

### 3687 6.5 Charged pion decay in flight

3688 As discussed previously, in GArSoft the TPC tracks are formed after a pattern recognition  
3689 algorithm and a Kalman filter are applied to the TPC clusters. These two steps can

## 6.5. CHARGED PION DECAY IN FLIGHT

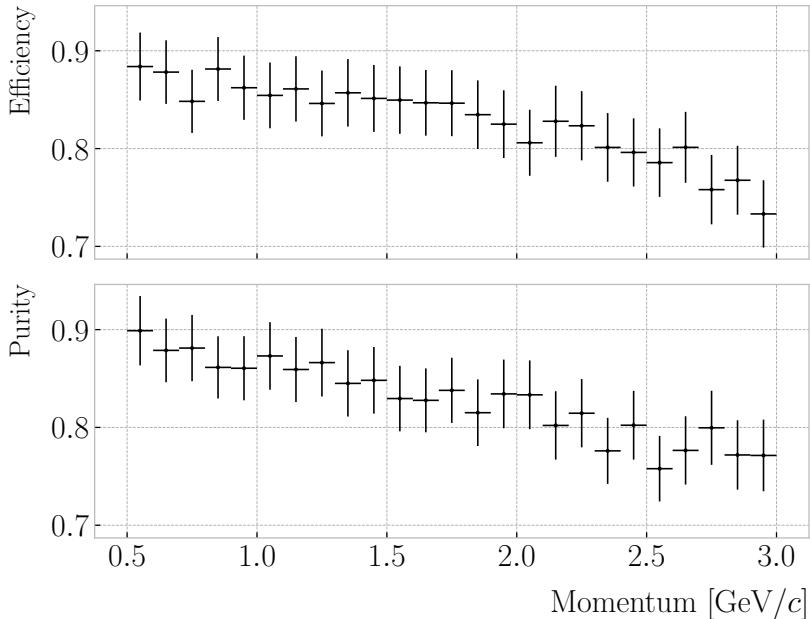


**Figure 6.32:** Mass spectra for  $p$  (blue) and  $\pi^\pm$  (yellow) particles, using different ECal time resolution values (from top to bottom, in ascending order), and arrival time estimates. From left to right: earliest hit time, average hit time, and fitted hit time. The dashed lines indicate the true masses of the particles.

3690 find discontinuities in the track candidates (e.g. due to a particle decay) when these  
 3691 so-called breakpoints are large enough. However, for some, more subtle, cases they may  
 3692 miss them and form a single reconstructed track. It has been noted in the literature  
 3693 that Kalman filters offer, as a by-product, additional information to form test statistics  
 3694 to identify these breakpoints [163, 164].

3695 Considering the mean life of the charged pion,  $\tau = (2.6033 \pm 0.0005) \times 10^{-8}$  s, one  
 3696 can estimate that about 12% of the pions with momentum  $p \sim \mathcal{O}(500 \text{ MeV}/c)$  (roughly  
 3697 the peak of the pion momentum distribution in  $\nu_\mu$  CC interactions off argon) decay

## CHAPTER 6. PARTICLE ID IN ND-GAr



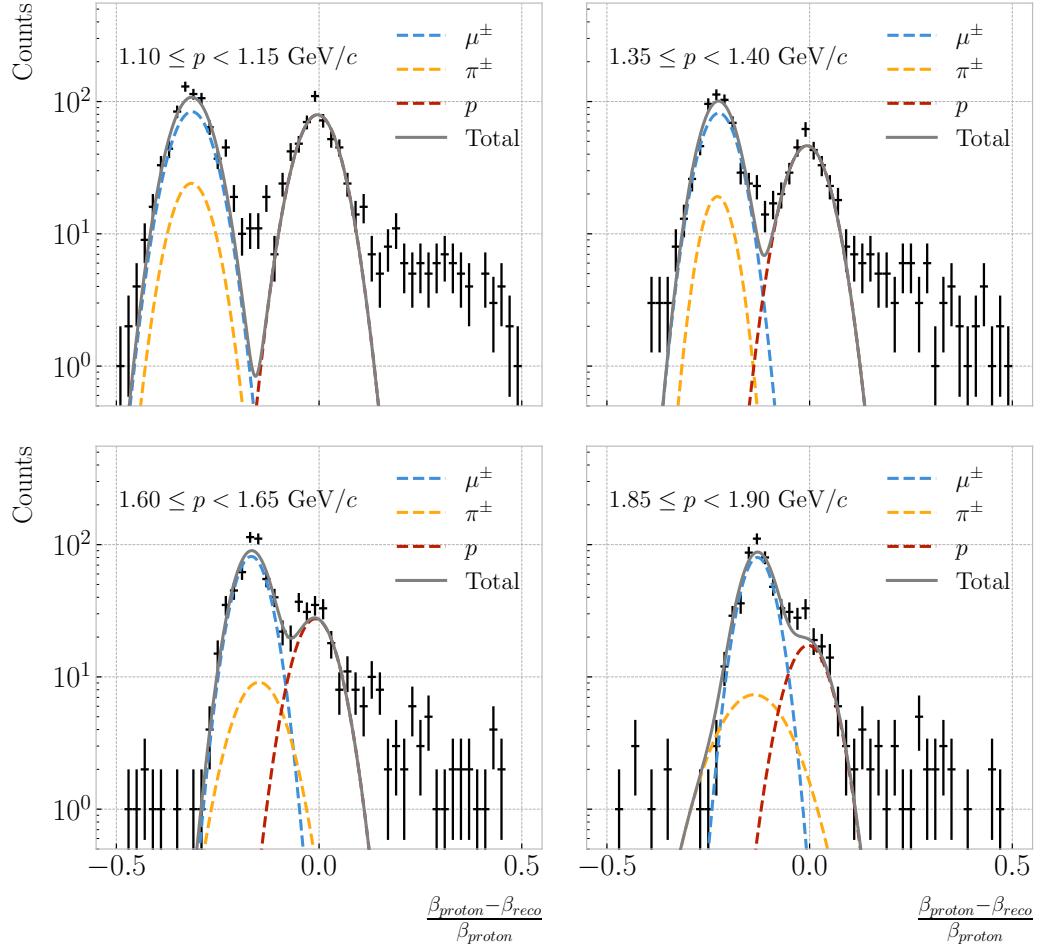
**Figure 6.33:** Efficiency (top panel) and purity (bottom panel) for the proton selection as a function of the momentum, for  $\Delta\tau = 0.10$  ns.

3698     inside the TPC. Figure 6.35 (left panel) shows the amount of charged pions decaying in  
 3699     the full TPC and fiducial volumes from an isotropic, monoenergetic sample of 100000  
 3700     negatively charged pions with  $p = 500$  MeV/ $c$ . We see that about 10% of those decayed,  
 3701     with more than half of them decaying inside the TPC fiducial volume.

3702       Figure 6.35 (right panel) shows an example event display of a charged pion (magenta  
 3703       line) decays in flight inside the TPC, but because the angle of the muon (blue line) is  
 3704       small both were reconstructed as one single track (black line). In this case, the composite  
 3705       track reaches the ECal, where it undergoes a muon-like interaction, thus being classified  
 3706       as a muon.

3707       A way to understand what decaying pion tracks were totally or partially reconstructed  
 3708       together with the daughter muon is looking at the relative energy contributions to the  
 3709       reconstructed track. In order to select a sample of such events, I require that a minimum  
 3710       50% of the total energy comes from the pion and at least 20% from the muon.

## 6.5. CHARGED PION DECAY IN FLIGHT



**Figure 6.34:** Distributions of the velocities measured by ToF with the inner ECal, for different momentum bins, in a FHC neutrino interaction sample. The Gaussian fits are performed around the maxima for each particle species.

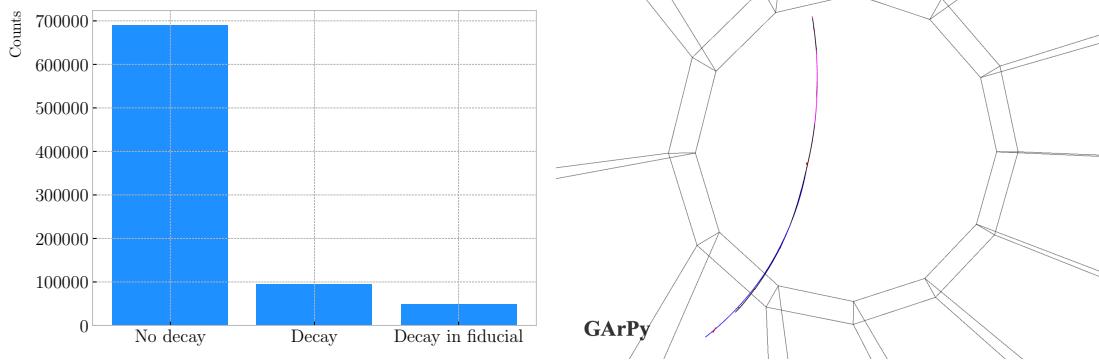
### 3711 6.5.1 Track breakpoints

3712 To identify potential decays we can use the information we obtain from the Kalman  
 3713 filter at each step of the fitted track. The simplest test we can think about is computing  
 3714 the  $\chi^2$  of the mismatch between all the parameters in the forward and the backward fits:

$$\chi_k^{2 (FB)} = (\hat{x}_k^B - \hat{x}_k^F)^T [V^{(\hat{x}_k, B)} + V^{(\hat{x}_k, F)}]^{-1} (\hat{x}_k^B - \hat{x}_k^F), \quad (6.23)$$

3715 where  $\hat{x}_k^F$ ,  $\hat{x}_k^B$  are the Kalman filter state vector estimates at step  $k$  in the forward and  
 3716 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.

## CHAPTER 6. PARTICLE ID IN ND-GAr



**Figure 6.35:** Left panel: number of non-decaying, decaying and decaying in the fiducial volume pions for a MC sample of 100000,  $p = 500$  MeV/c isotropic positively charged pions inside the TPC. Right panel: event display for a positive pion decaying inside the fiducial volume, with a single reconstructed track for the pion and muon system.

3717 Using the values of the  $\chi^2$  at measurement  $k$  for the forward and backward fits we can  
3718 compute another  $\chi^2$  value that characterises the overall track fit:

$$\chi_{track}^2 = \chi_k^{2(F)} + \chi_k^{2(B)} + \chi_k^{2(FB)}, \quad (6.24)$$

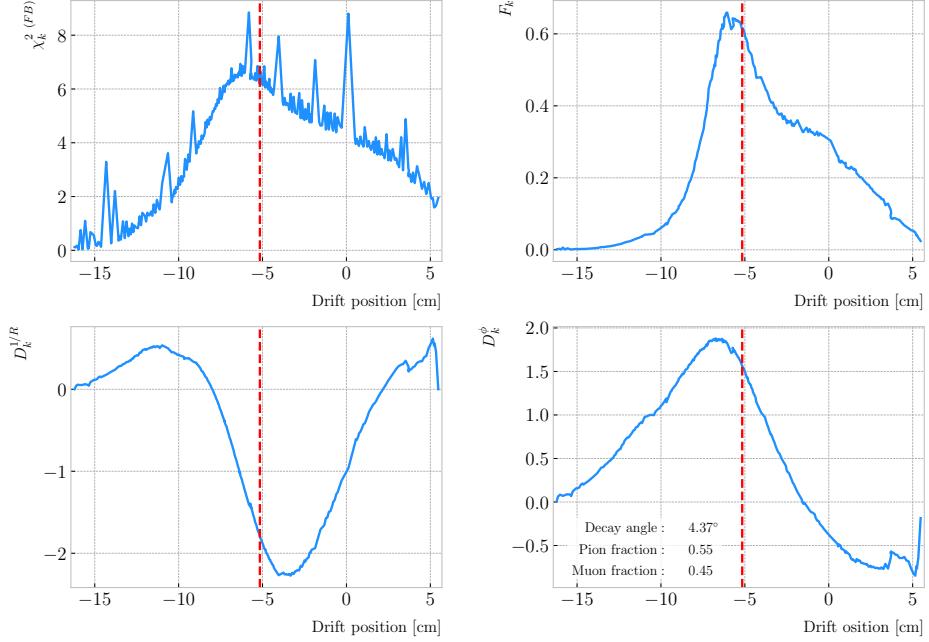
3719 which remains approximately constant for all  $k$ .

3720 An alternative approach proposed in the context of the NOMAD experiment was  
3721 using a fit with a more elaborate breakpoint hypothesis, so we can perform a comparison  
3722 of the  $\chi^2$  with and without breakpoints. This can be achieved by using some alternative  
3723 parametrisation with extra parameters, which allows some of the track parameters to  
3724 be discontinuous at certain points. A decay changes the momentum magnitude and  
3725 direction, so we can use the new state vector:

$$\alpha = (y, z, 1/R_F, 1/R_B, \phi_F, \phi_B, \tan\lambda_F, \tan\lambda_B)^T. \quad (6.25)$$

3726 As we already have the estimates from the standard Kalman filter and their  
3727 covariance matrices at each point, we do not need to repeat the Kalman fit for the new  
3728 parametrisation. Instead, I can compute the values of  $\alpha$  at each point  $k$  that minimise

## 6.5. CHARGED PION DECAY IN FLIGHT



**Figure 6.36:** Values of  $\chi_k^2(FB)$  (top left panel),  $F_k$  (top right panel),  $D_k^{1/R}$  (bottom left panel) and  $D_k^\phi$  (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.

3729 the  $\chi^2$  resulting from comparing them to  $\{\hat{x}_k^B, \hat{x}_k^F\}$ . Introducing the two  $5 \times 8$  matrices:

$$H^F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad H^B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (6.26)$$

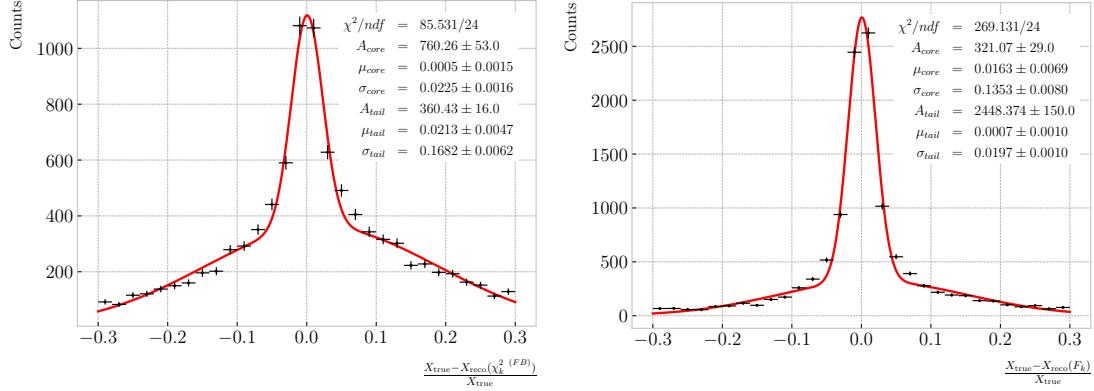
3730 we can write this as:

$$\begin{aligned} \chi_k^2(FB)(\alpha) &= (\hat{x}_k^F - H^F \alpha)^T \left[ V^{(\hat{x}_k, F)} \right]^{-1} (\hat{x}_k^F - H^F \alpha) \\ &\quad + (\hat{x}_k^B - H^B \alpha)^T \left[ V^{(\hat{x}_k, B)} \right]^{-1} (\hat{x}_k^B - H^B \alpha). \end{aligned} \quad (6.27)$$

3731 The minimum of  $\chi_k^2(FB)(\alpha)$  is found when the measured new state vector takes the  
3732 value:

$$\hat{\alpha}_k = V^{(\hat{\alpha}_k)} H^T (V^{(\hat{x}_k)})^{-1} \hat{X}, \quad (6.28)$$

## CHAPTER 6. PARTICLE ID IN ND-GAr



**Figure 6.37:** Fractional residual distributions of the true and reconstructed decay position along the drift coordinate, using the position of the maximum of  $\chi_k^2(FB)$  (left panel) and  $F_k$  (right panel) as estimates of the decay position. Also shown are double Gaussian fits to these points (red lines).

3733 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  $V^{(\hat{x}_k, B)}$   
 3734 and  $V^{(\hat{\alpha}_k)}$  is the covariance matrix of  $\hat{\alpha}_k$ , given by:

$$V^{(\hat{\alpha}_k)} = \left( H^T (V^{(\hat{x}_k)})^{-1} H \right)^{-1}. \quad (6.29)$$

3735 From these new fit estimates we can compute the  $F$  statistic, which tells us whether  
 3736 the model with breakpoint provides a statistically significant better fit:

$$F_k = \left( \frac{\chi_{track,k}^2 - \chi_{full,k}^2}{8 - 5} \right) / \left( \frac{\chi_{full,k}^2}{N - 8} \right). \quad (6.30)$$

3737 One can also compute the signed difference of the duplicated variables divided by  
 3738 their standard deviation at each point. These represent how significant the discontinuity  
 3739 in each variable is. For any variable  $\eta$  we can write it as:

$$D_k^\eta = \frac{\hat{\eta}_k^B - \hat{\eta}_k^F}{\sqrt{\text{Var}[\hat{\eta}_k^F] + \text{Var}[\hat{\eta}_k^B] - 2\text{Cov}[\hat{\eta}_k^F, \hat{\eta}_k^B]}}. \quad (6.31)$$

3740 In our case, the relevant ones to look at are  $D_k^{1/R}$  and  $D_k^\phi$ .

3741 Figure 6.36 shows the values of  $\chi_k^2(FB)$ ,  $F_k$ ,  $D_k^{1/R}$  and  $D_k^\phi$  as functions of the position  
 3742 along the drift direction, for an example reconstructed track with 55.5% of the energy

## 6.5. CHARGED PION DECAY IN FLIGHT

3743 coming from the charged pion and 45.5% from the daughter muon. The true position of  
 3744 the decay is indicated (dashed red lines). Notice how  $\chi_k^2(FB)$  and  $F_k$ ,  $D_k^{1/R}$  reach their  
 3745 maxima near the decay point. In the former case this indicates a large forward-backward  
 3746 difference in the track fit. In the later it represents that the extended state vector  
 3747 improves the fit particularly around that point.

3748 I can estimate the decay position finding resolution by computing the difference  
 3749 between the  $X$  position of the maxima of  $\chi_k^2(FB)$  and  $F_k$  and the  $X$  position of the  
 3750 true decay. Figure 6.37 represent the the fractional residual distributions for both  
 3751 cases, from the sample of tracks containing pion decays. Fitting a double Gaussian to  
 3752 the distributions (red lines) I find a resolution of  $(3.31 \pm 0.15)\%$  and  $(6.94 \pm 0.31)\%$   
 3753 respectively.

3754 In principle, the  $F$ -statistic should follow a Fisher distribution with  $(8 - 5)$  and  
 3755  $(N - 8)$  degrees of freedom under the null hypothesis. In most of our cases  $N \sim \mathcal{O}(100)$ ,  
 3756 so the probability density functions will look very similar. In this case, it is safe to take  
 3757 the limit  $N \rightarrow \infty$  in the Fisher PDF:

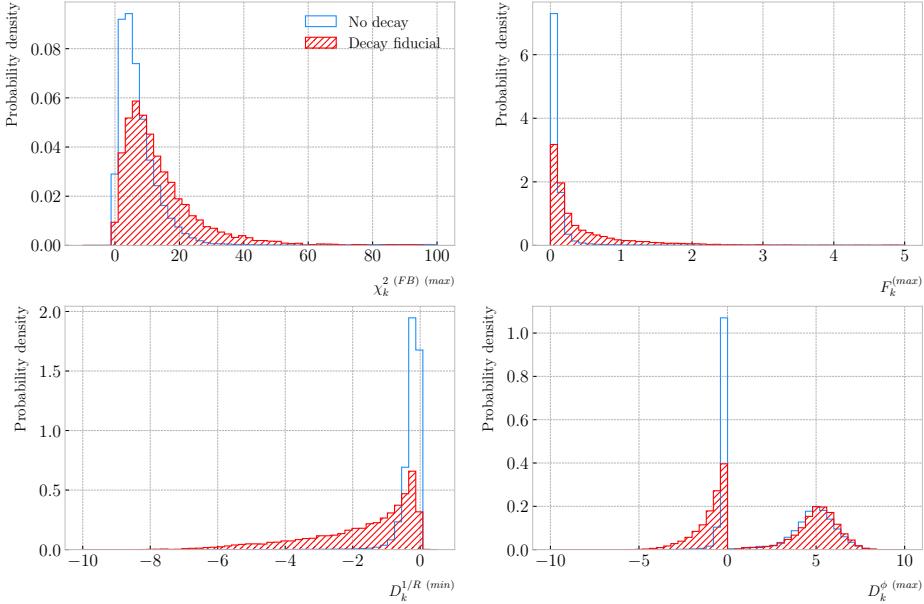
$$\begin{aligned}\tilde{f}(x; a - b) &= \lim_{N \rightarrow \infty} f(x; a - b, N - a) \\ &= \frac{2^{-\frac{a-b}{2}}}{\Gamma\left(\frac{a-b}{2}\right)} (a - b)^{\frac{a-b}{2}} x^{\frac{a-b}{2}-1} e^{-\frac{a-b}{2}x}.\end{aligned}\tag{6.32}$$

3758 In our case  $a - b = 8 - 5 = 3$ , so we would obtain a p-value of 0.05 at  $x = 2.60$ .

3759 Figure 6.38 contains the distributions of the maxima of  $\chi_k^2(FB)$ ,  $F_k$  and  $D_k^\phi$  and the  
 3760 minima of  $D_k^{1/R}$  for a sample of non-decaying pion tracks (blue) and another sample of  
 3761 reconstructed tracks containing part of the pion and the daughter muon from a decay  
 3762 inside the fiducial volume (red). Notice that, even though the values of  $F_k^{(max)}$  for the  
 3763 decay sample are typically larger than for the non-decaying one, just a small fraction of  
 3764 the events go beyond the aforementioned value of  $F = 2.60$ . Therefore, from a practical  
 3765 point of view, it is not the most efficient variable to use for selecting the decay events.

3766 However, looking at the  $D_k^{1/R \text{ (min)}}$  distribution we can see there is a big difference  
 3767 between non-decaying and decaying events in this variable. One can use a combination

## CHAPTER 6. PARTICLE ID IN ND-GAr



**Figure 6.38:** Distributions of the extreme values of  $\chi_k^2(FB)$  (top left panel),  $F_k$  (top right panel),  $D_k^{1/R}$  (bottom left panel) and  $D_k^\phi$  (bottom right panel) for non-decaying reconstructed pion tracks (blue) and tracks which include the decay inside the fiducial volume (red).

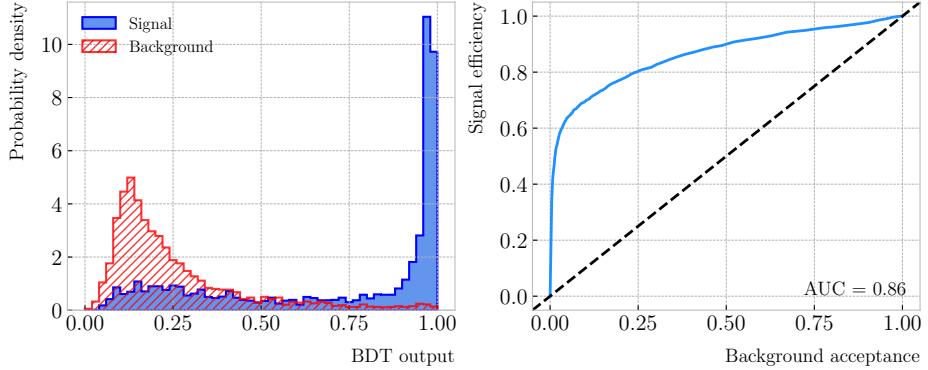
3768 of these four variables to distinguish between the pion decay events (signal) and the  
 3769 non-decaying pions (background).

3770 An approach to this classification could be using a boosted decision tree (BDT). One  
 3771 of the advantages of BDTs is that they are easy to interpret and identify the relative  
 3772 importance of the different input variables. Training a BDT with 400 estimators and a  
 3773 maximum depth of 4 I can obtain an efficient classification without overtraining. Figure  
 3774 6.39 (left panel) shows the distribution of probabilities predicted by the BDT for a test  
 3775 sample. The signal efficiency as a function of background acceptance, the so-called ROC  
 3776 curve, is shown in Fig. 6.39 (right panel). With a relative importance of 0.83, the most  
 3777 important variable turned out to be  $D_k^{1/R} \text{ (min)}$ .

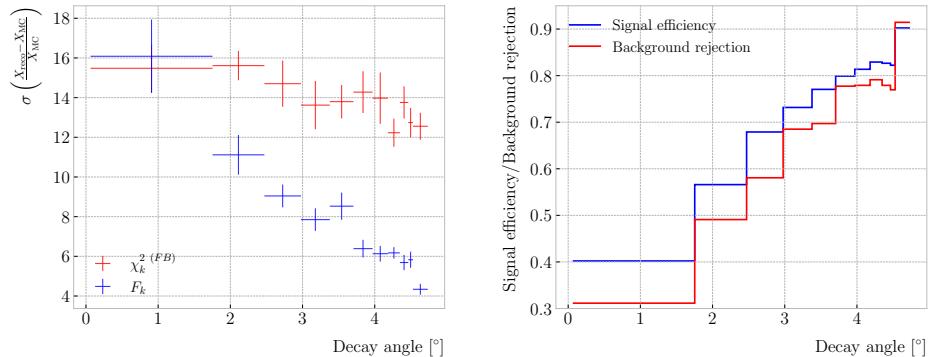
3778 One thing we can check is how the resolution to the decay and the signal efficiency in  
 3779 the classification changes with the true decay angle. Using an equal-frequency binning  
 3780 for the decay angles, we can repeat the previous steps for each bin.

3781 Figure 6.40 (left panel) shows the dependence on the decay angle of the decay finding

## 6.5. CHARGED PION DECAY IN FLIGHT



**Figure 6.39:** Left panel: distributions of the predicted probabilities assigned by the BDT classifier to a test sample of decaying pion+muon tracks (blue) and non-decaying pion tracks (red). Left: signal efficiency versus background acceptance (ROC curve) obtained from the BDT for the test sample.



**Figure 6.40:** Left panel: dependence of the decay position finding resolution on the true value of the decay angle for the  $\chi_k^2(FB)$  (red) and  $F_k$  (blue) methods. Right panel: signal efficiency (blue line) and background rejection (red line) from the BDT classifier versus true decay angle.

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 6.40 (right panel) represents the change in signal efficiency (blue)

## CHAPTER 6. PARTICLE ID IN ND-GAr

3790 and background rejection (red) with the value of the true decay angles.

## 3791 6.6 Neutral particle identification

### 3792 6.6.1 ECal clustering

3793 Another important reconstruction item is the clustering algorithm of ECal hits in  
3794 GArSoft. The default module features a NN algorithm that treats all hits in the same  
3795 way, independently of the layer each hit comes from. However, the current ECal design  
3796 of ND-GAr has two very different types of scintillator layers. The inner layers are made  
3797 out of tiles, which provide excellent angular and timing resolutions. On the other hand,  
3798 the outer layers are cross scintillator strips. That way, an algorithm that treats hits  
3799 from both kinds of layers differently may be able to improve the current performance.

3800 Inspired by the reconstruction of T2K’s ND280 downstream ECal [165], the idea  
3801 was to put together a clustering module that first builds clusters for the different ECal  
3802 views (tiles, strips segmented in the  $X$  direction and strips segmented in  $Y$  direction),  
3803 and then tries to match them together to form the final clusters.

3804 Working on a module-by-module basis, the algorithm first separates the hits depending  
3805 on the layer type they come from. Then, it performs a NN clustering for the 3 sets of  
3806 hits separately. For the tile hits it clusters together all the hits which are in nearest-  
3807 neighbouring tiles and nearest-neighbouring layers, for strip hits it looks at nearest-  
3808 neighbouring strips and next-to-nearest-neighbouring layers (as the layers with strips  
3809 along the two directions are alternated). For strip clusters an additional cut in the  
3810 direction along the strip length is needed.

3811 After this first clustering I then apply a recursive re-clustering for each collection  
3812 of strip clusters based on a PCA method. In each case, we loop over the clusters with  
3813  $N_{hits} \geq 2$ , computing the centre of mass and three principal components. Propagating  
3814 these axes up to the layers of the rest of the clusters, we check if the propagated point  
3815 and the centre of mass of the second cluster are within next-to-nearest-neighbouring  
3816 strips. An additional cut in the direction along the strip length is also needed. Moreover,

## 6.6. NEUTRAL PARTICLE IDENTIFICATION

3817 I require that the two closest hits across the two clusters are at most in next-to-nearest-  
3818 neighbouring strips. I merge the clusters if these three conditions are satisfied. The  
3819 re-clustering is repeated until no more cluster pairs pass the cuts.

3820 The clusters in each strip view are combined if their centres of mass are close enough  
3821 and they point in the same direction. An alternative approach for the strip cluster  
3822 merging could be to compute the overlap between the ellipsoids defined by the principal  
3823 axes of the clusters, and then merge the pair if the overlap exceeds some threshold.  
3824 Further study is needed to understand if this change would have an impact in the overall  
3825 clustering performance.

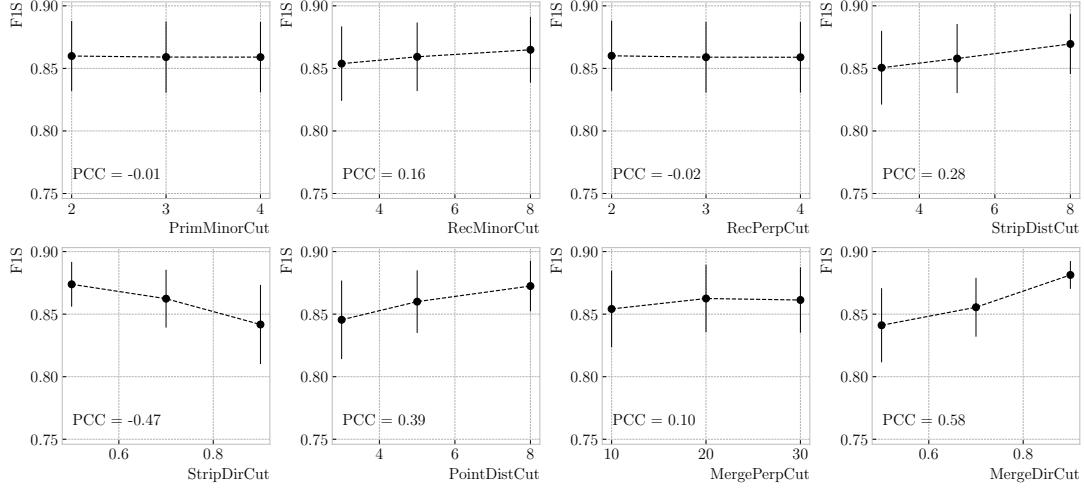
3826 To merge the tile clusters to the combined strip clusters I propagate the principal  
3827 axis of the strip cluster towards the inner layers, up to the centre of mass layer of the  
3828 tile cluster. I merge the clusters if the distance between the propagated point and the  
3829 centre of mass is below a certain cut.

3830 The last step is to check if clusters in neighbouring modules should be merged  
3831 together, both across two barrel modules, across end cap modules and between barrel  
3832 end cap modules. I check the distance between the two closest hits in the pair of clusters  
3833 and merge them if it passes this and an additional direction cut.

3834 Figure ?? presents an example of the clustering steps relevant for strip layer hits, from  
3835 the input hits (top left panel) to the NN clustering (top right panel) and re-clustering  
3836 (bottom left panel) for each strip view and the final merging strip clusters (bottom  
3837 right panel). It shows the hits from a single ECal barrel module in a  $\nu_\mu$  CC interaction  
3838 event with a neutral pion and a proton in the final state. The two clusters on the left  
3839 correspond to the photon pair from the  $\pi^0$  decay and the one on the upper right corner  
3840 is associated to the proton.

3841 This algorithm has a total number of eight free parameters that need to be optimised.  
3842 I used a sample of 1000  $\nu_\mu$  CC interactions in order to obtain the optimal configuration of  
3843 clustering parameters. This sample was generated up to the default ECal hit clustering  
3844 level, so then I could run the new clustering algorithm each time with a different  
3845 configuration of parameters. As the number of parameters is relatively large, I only

## CHAPTER 6. PARTICLE ID IN ND-GAr



**Figure 6.41:** Mean values of the  $F_1$ -score marginal distributions for the different free parameters of the new clustering algorithm, with the error bars representing one standard deviation around the mean. The  $F_1$ -score values were computed for the 6561 possible parameter configurations using 1000  $\nu_\mu$  CC interaction events.

**Table 6.5:** Summary of parameters and sampled values used in the optimisation of the clustering algorithm.

Name	Units	Sampled values	Description
PrimMinorCut	strips	2, 3, 4	Distance along strip length in NN clustering
RecMinorCut	strips	3, 5, 8	Distance between propagated point and CM along strip length in re-clustering
RecPerpCut	strips	2, 3, 4	Closest hit pair distance in re-clustering
StripDistCut	strips	3, 5, 8	Distance between CMs in strip cluster merging
StripDirCut	cos	0.5, 0.7, 0.9	Main axes direction cut in strip cluster merging
PointDistCut	tiles	3, 5, 8	Distance between propagated point and CM in strip-tile matching
MergePerpCut	cm	10, 20, 30	Closest hit pair distance in module merging
MergeDirCut	cos	0.5, 0.7, 0.9	Main axes direction cut in module merging

3846 performed a coarse-grained scan of the parameter space. Sampling each of the eight  
 3847 parameters at three different points each I obtain 6561 different configurations. These  
 3848 parameters, together with the used values, are summarised in Tab. 6.5.

3849 In order to measure the performance of the clustering, I use a binary classification  
 3850 approach. For each formed cluster, I identify the Geant4 Track ID of the matching MC  
 3851 particle and the energy fraction of each hit. Then, I assign to each cluster the Track ID  
 3852 with the highest total energy fraction. For each of the different Track IDs associated to

## 6.6. NEUTRAL PARTICLE IDENTIFICATION

3853 the clusters, I select the cluster with the highest energy (only from the hits with the  
3854 same Track ID). I identify such a cluster as the main cluster for that Track ID. I count  
3855 as true positives (TPs) the hits with the correct Track ID in each main cluster. False  
3856 positives (FPs) are the hits with the incorrect Track ID for the cluster they are in, not  
3857 only main clusters. The false negatives (FNs) are the hits with the correct Track ID in  
3858 clusters other than the main.

3859 Figure 6.41 shows the computed  $F_1$ -score values for the different cuts. In each case,  
3860 the central value represents the mean of the  $F_1$ -score distribution for the specified value  
3861 of the corresponding variable and the vertical error bar represents one standard deviation  
3862 around the mean. Also shown are the Pearson correlation coefficients of these central  
3863 values. We can see that five of the variables have a sizeable effect on the  $F_1$ -score, with  
3864 an absolute difference between the last and first values as big as 4%.

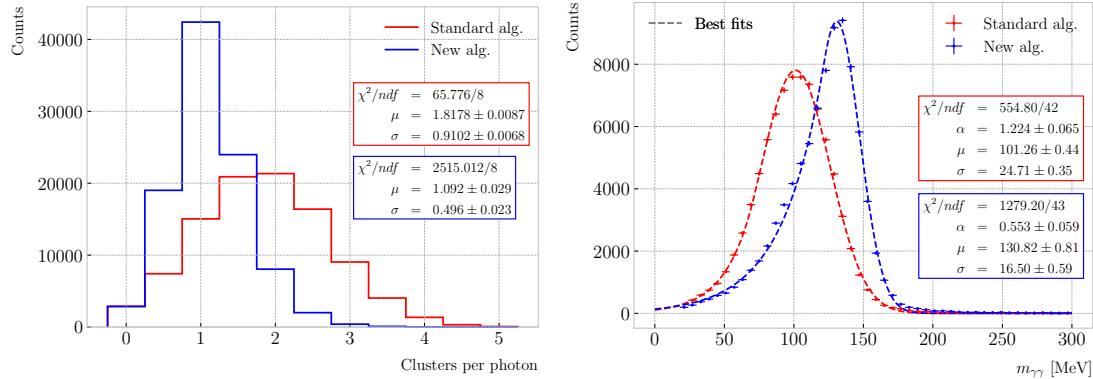
3865 The working configuration is obtained as follows. I first select all configurations  
3866 with purity  $\geq 90\%$ . Among those, I choose the combinations that yield the maximum  
3867  $F_1$ -score. If more than one configuration remains I select the one with the highest  
3868 sensitivity. Doing so, I end up with a parameter configuration with an efficiency of 88%  
3869 and a 90% purity. Compared with the default algorithm, which gives an efficiency of  
3870 76% and a purity of 91% for the same sample, I have managed to improve the efficiency  
3871 by a factor of 1.16.

### 3872 6.6.2 $\pi^0$ reconstruction

3873 One of the potential applications of the new ECal hit clustering is the reconstruction of  
3874 neutral particles, in particular pions. Neutral pions decay promptly after being produced,  
3875 through the  $\pi^0 \rightarrow \gamma\gamma$  channel ( $98.823 \pm 0.034\%$ ) of the time. The photon pair does  
3876 not leave any traces in the HPgTPC (unless one or both of them converts into an  
3877 electron-positron pair), but each of them will produce an electromagnetic shower in  
3878 the ECal.

3879 To test the potential impact of the new algorithm in  $\pi^0$  reconstruction, I generated  
3880 a MC sample of single, isotropic neutral pions inside the HPgTPC. All pions were

## CHAPTER 6. PARTICLE ID IN ND-GAr



**Figure 6.42:** Left panel: distributions of the number of ECal clusters per photon from  $\pi^0$  decays for the standard (red) and new (blue) clustering algorithms. Right panel: reconstructed invariant mass distributions for photon pairs from single  $\pi^0$  events using the standard (red) and new (blue) ECal clustering algorithms.

3881 generated with a momentum  $p = 500$  MeV and their initial positions were uniformly  
 3882 sampled inside a  $2 \times 2 \times 2$  m box aligned with the centre of the TPC. I ran both the  
 3883 default and the new clustering algorithms, using for the latter the optimised configuration  
 3884 discussed above.

3885 The first thing to notice is that the number of clusters produced per photon has  
 3886 decreased. Figure 6.42 (left panel) shows these distributions for the default (red) and  
 3887 new (blue) algorithms. Using a simple Gaussian fit, we see that the mean number of  
 3888 ECal clusters per photon went from  $1.82 \pm 0.01$  to  $1.09 \pm 0.03$ . This effectively means that  
 3889 with the new algorithm the ECal activity of one true particle is typically reconstructed  
 3890 as a single object. From the reconstruction point of view this can be an advantage. As  
 3891 now most of the photon energy ends up in a single ECal cluster, I can simply use cluster  
 3892 pairs to identify the  $\pi^0$  decay.

3893 In general, one calculates the invariant mass of the photon pair as:

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos \theta)}, \quad (6.33)$$

3894 where  $E_i$  are the energies of the photons and  $\theta$  the opening angle between them. In this  
 3895 case I can use the energies deposited in the ECal and their incident directions. This  
 3896 quantity is computed for all possible pairs of clusters, using their position together with

## 6.6. NEUTRAL PARTICLE IDENTIFICATION

3897 the true decay point. In a more realistic scenario, e.g.  $\nu_\mu$  CC interaction, one could use  
 3898 the position of the reconstructed primary vertex instead. I also tried to use the principal  
 3899 direction of the clusters, but that approach gave considerably worse results. For each  
 3900 event I only keep the pair with an invariant mass closer to the true  $\pi^0$  mass value.

3901 Figure 6.42 (right panel) shows the invariant mass distributions for the photon pairs  
 3902 we get using the default (red) and the new (blue) ECal clustering algorithms. For the fit  
 3903 I used a modified version of the Crystal Ball function [166], obtained by taking the limit  
 3904 where the parameter controlling the power-law tail goes to infinity:

$$f(x; N, \mu, \sigma, \alpha) = N \cdot \begin{cases} e^{\frac{\alpha(2x-2\mu+\alpha\sigma)}{2\sigma}}; & x \leq \mu - \alpha\sigma, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}; & x > \mu - \alpha\sigma. \end{cases} \quad (6.34)$$

3905 Comparing the fitted mean and standard deviation values for the Gaussian cores, we  
 3906 see that the distribution for the new algorithm is a 67% narrower and also peaks much  
 3907 closer to the true  $m_{\pi^0}$  value, going from  $101.3 \pm 0.4$  MeV to  $130.8 \pm 0.6$  MeV.



# Event selection in ND-GAr

## 3910 7.1 Data sample

## 3911 7.2 $\nu_\mu$ CC selection

3912 In a  $\nu_\mu$  CC inclusive selection, the signal topology we look for is a neutrino-induced  
 3913 muon with or without other final state particles. Here, I also require the neutrino vertex  
 3914 to be located inside the fiducial volume (FV) of ND-GAr.

3915 The FV is defined as a smaller cylinder within the cylindrical volume of the HPgTPC.  
 3916 The FV has a radius  $R_{\text{FV}}$  and a half-length  $L_{\text{FV}}$ . For a particle position to lie within  
 3917 the FV it must satisfy:

$$\vec{x}_i \in \left\{ \vec{x} \in \mathbb{R}^3 \mid |x_0| \leq L_{\text{FV}} \text{ \& } \sqrt{x_1^2 + x_2^2} \leq R_{\text{FV}} \right\}, \quad (7.1)$$

3918 in the reference frame of the HPgTPC. For convenience, I define:

$$\begin{aligned} \Delta R_{\text{FV}} &= R_{\text{HPgTPC}} - R_{\text{FV}}, \\ \Delta L_{\text{FV}} &= L_{\text{HPgTPC}} - L_{\text{FV}}, \end{aligned} \quad (7.2)$$

3919 where  $R_{\text{HPgTPC}}$  and  $L_{\text{HPgTPC}}$  refer to the radius and the half-length of the HPgTPC,  
 3920 respectively. Figure 7.1 shows the HPgTPC volume with the FV inside of it. In that  
 3921 representation, the FV is defined as  $\Delta L_{\text{FV}} = 30.0$  cm and  $\Delta R_{\text{FV}} = 30.0$  cm. Also shown  
 3922 is the HPgTPC reference frame, with  $x$  being the drift direction and  $z$  aligned along the

## CHAPTER 7. EVENT SELECTION IN ND-GAr

**Table 7.1:** Event rates in ND-GAr.

Process	Events/ton/year	
	$1.1 \times 10^{21}$ POT/year	$1.9 \times 10^{21}$ POT/year
All $\nu_\mu$ -CC	$1.60 \times 10^6$	$2.83 \times 10^6$
CC $0\pi$	$5.28 \times 10^5$	$9.35 \times 10^5$
CC $1\pi^\pm$	$3.02 \times 10^5$	$5.34 \times 10^5$
CC $1\pi^0$	$1.65 \times 10^5$	$2.92 \times 10^5$
CC $2\pi$	$3.18 \times 10^5$	$5.63 \times 10^5$
CC $3\pi$	$1.36 \times 10^5$	$2.41 \times 10^5$
CC other	$1.52 \times 10^5$	$2.69 \times 10^5$
All $\bar{\nu}_\mu$ -CC	$7.54 \times 10^4$	$1.33 \times 10^5$
All NC	$5.50 \times 10^5$	$9.73 \times 10^5$
All $\nu_e$ -CC	$2.70 \times 10^4$	$4.78 \times 10^4$

3923 beam direction.

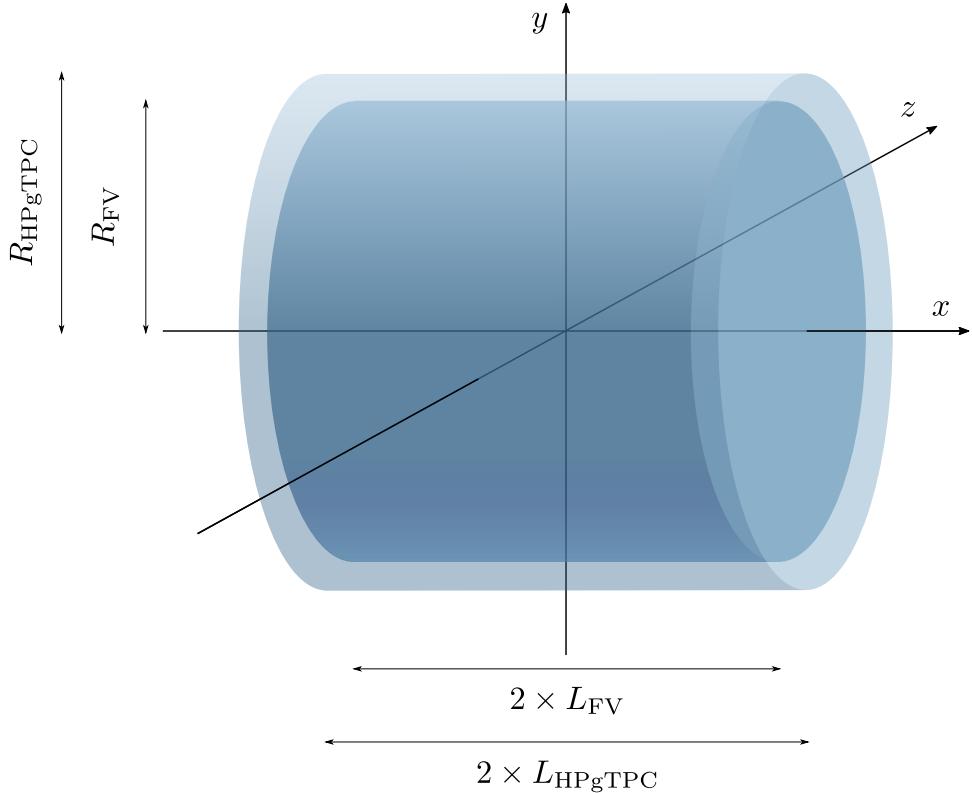
3924 In some cases, it is interesting to divide the signal events in different categories  
3925 based on their true interaction mode. In this work, I will distinguish between charged-  
3926 current quasi-elastic (CCQE), coherent (CCCOH), resonant (CCRES), and deep-inelastic  
3927 (CCDIS) interactions. I also use a separate category for the interactions not included in  
3928 any of the other categories (CCOther).

3929 Any other events are considered backgrounds. For this selection, I use the following  
3930 categorisation of background events:

- 3931 • Out of FV: if the true neutrino vertex lies outside the defined FV.
- 3932 • NC: if the event is a true neutral-current event.
- 3933 •  $\bar{\nu}_\mu$  CC: if the true neutrino candidate is of muon antineutrino flavour.
- 3934 • Other: if the event is not signal nor falls in any of the other background categories.

3935 The key to the CC selection is the identification of a primary muon candidate.  
3936 Typically, this is the longest track in the event. However, sometimes protons and pions  
3937 leave tracks longer than that of the muon. This is particularly important in the GAr

## 7.2. $\nu_\mu$ CC SELECTION



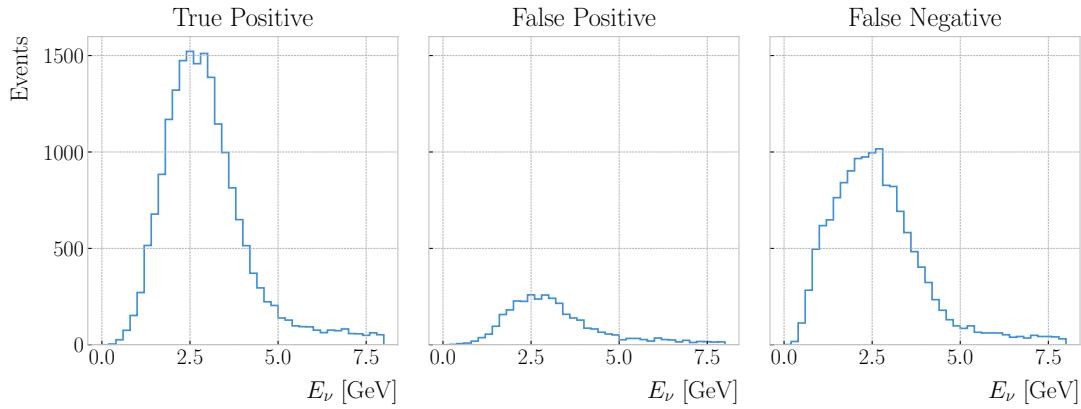
**Figure 7.1:** Schematic diagram of the HPgTPC including the fiducial volume (FV). In this case the FV is given by  $\Delta L_{\text{FV}} = 30.0$  cm and  $\Delta R_{\text{FV}} = 30.0$  cm.

3938 medium, considerably less dense than the LAr. For this reason, the muon identification  
 3939 in ND-GAr relies heavily on the capabilities of the ECal.

3940 The selection strategy proposed combines the information coming from the three  
 3941 main detection systems of ND-GAr: the HPgTPC charge readout, and the ECal and  
 3942  $\mu$ ID detectors. It consists of five steps:

- 3943 1. Event contains reconstructed particles.  
 3944 2. Select particles with reconstructed negative charge,  $q_{\text{reco}} = -1$ .  
 3945 3. Select particles passing the muon score cut,  $\mu_{\text{score}} \geq \mu_{\text{score}}^{\text{cut}}$ .  
 3946 4. Keep reconstructed particle with the highest momentum,  $\max [p_{\text{reco}}]$ .  
 3947 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  $\nu_\mu$  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.

3948 All the events passing these cuts are classified as signal, and the selected particle is  
 3949 regarded as the primary muon candidate.

3950 **7.2.1 Selection optimisation**

3951 I performed an optimisation of this selection, comparing the performance of a number of  
 3952 configurations. For the muon selection, I varied the value of  $\mu_{\text{score}}^{\text{cut}}$  from 0.05 to 0.95,  
 3953 using a step size of 0.05. Additionally, to optimise the FV, I systematically explored a  
 3954 number of different parameter configurations, moving within the 10.0 – 70.0 cm range for  
 3955  $\Delta L_{\text{FV}}$  and 25.0 – 75.0 cm for  $\Delta R_{\text{FV}}$ , in increments of 10.0 cm and 5.0 cm respectively.

3956 For each parameter configuration, I extract three different true neutrino energy  
 3957 distributions. These are built combining the results of the selection described previously,  
 3958 which we can refer to as the “reco” selection, and a “true” selection. The later identifies  
 3959 the true  $\nu_\mu$  CC events using the GENIE event records, and checks that the true neutrino  
 3960 vertices are contained in the FV.

3961 The first distribution consists of the events passing both selections, i.e., these are  
 3962 the true  $\nu_\mu$  CC events which pass the “reco” selection. The second distribution contains  
 3963 the events passing the “reco” selection but failing the “true” selection. These are  
 3964 the background events that the selection misidentifies. Finally, the third distribution

## 7.2. $\nu_\mu$ CC SELECTION

3965 corresponds to the events picked by the “true” selection but not by the “reco” one. In  
 3966 other words, these are the true  $\nu_\mu$  CC events that our selection misses. In analogy to  
 3967 the machine learning jargon, I refer to these distributions as the true positive (TP),  
 3968 false positive (FP), and false negative (FN) spectra, respectively. Figure 7.2 shows an  
 3969 example of these three distributions for the case  $\mu_{\text{score}}^{\text{cut}} = 0.75$ ,  $\Delta L_{\text{FV}} = 30.0$  cm, and  
 3970  $\Delta R_{\text{FV}} = 30.0$  cm.

3971 By making different combinations of these distributions one can compute a series of  
 3972 performance metrics. Using the full information from the spectra allows to obtain the  
 3973 scores as a function of the true neutrino energy, whereas the totals can be obtained by  
 3974 integrating the histograms. This way, the efficiency of the selection is given by:

$$\begin{aligned} \text{Efficiency} &= \frac{\text{Selected true } \nu_\mu \text{ CC events}}{\text{Total true } \nu_\mu \text{ CC events}} \\ &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \end{aligned} \quad (7.3)$$

3975 while the purity can be written as:

$$\begin{aligned} \text{Purity} &= \frac{\text{Selected true } \nu_\mu \text{ CC events}}{\text{Total selected events}} \\ &= \frac{\text{TP}}{\text{TP} + \text{FP}}. \end{aligned} \quad (7.4)$$

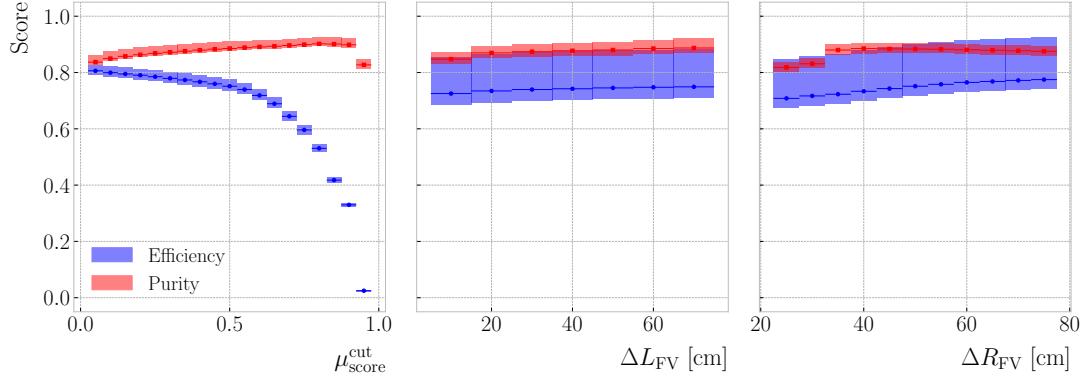
3976 Another scoring metric typically used when quantifying the performance of a selection  
 3977 is the significance. It is defined as:

$$\text{Significance} = \frac{S}{\sqrt{S+B}} = \frac{\text{TP}}{\sqrt{\text{TP}+\text{FP}}}. \quad (7.5)$$

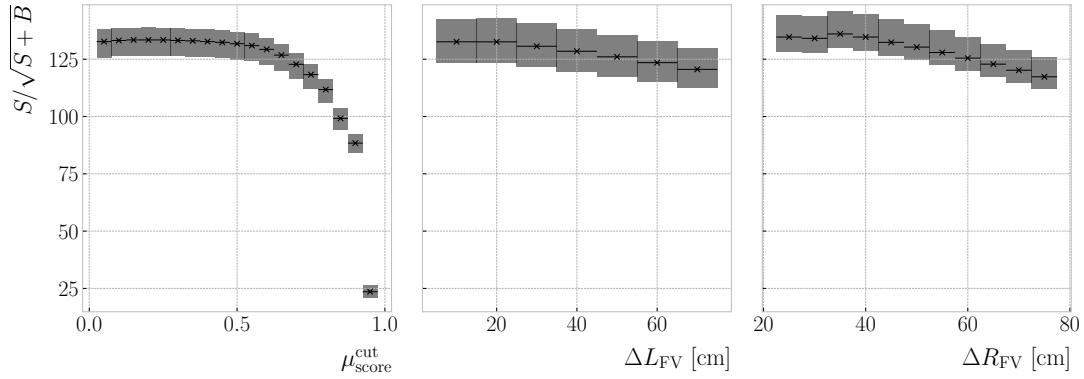
3978 The significance measures the relative size of the true signal within the selection,  $S = \text{TP}$   
 3979 with respect to one standard deviation of the counting experiment. Assuming Poisson  
 3980 statistics, the variance is equal to the number of observations, and therefore the standard  
 3981 deviation equals to  $\sqrt{N} = \sqrt{S+B} = \sqrt{\text{TP}+\text{FP}}$ . I use this metric to

3982 Figure 7.3 shows the change in efficiency (blue) and purity (red) of the  $\nu_\mu$  CC  
 3983 selection as a function of the different cuts. From left to right, I vary  $\mu_{\text{score}}^{\text{cut}}$ ,  $\Delta L_{\text{FV}}$ ,

## CHAPTER 7. EVENT SELECTION IN ND-GAR



**Figure 7.3:** Efficiency (blue) and purity (red) for the  $\nu_\mu$  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 mean.

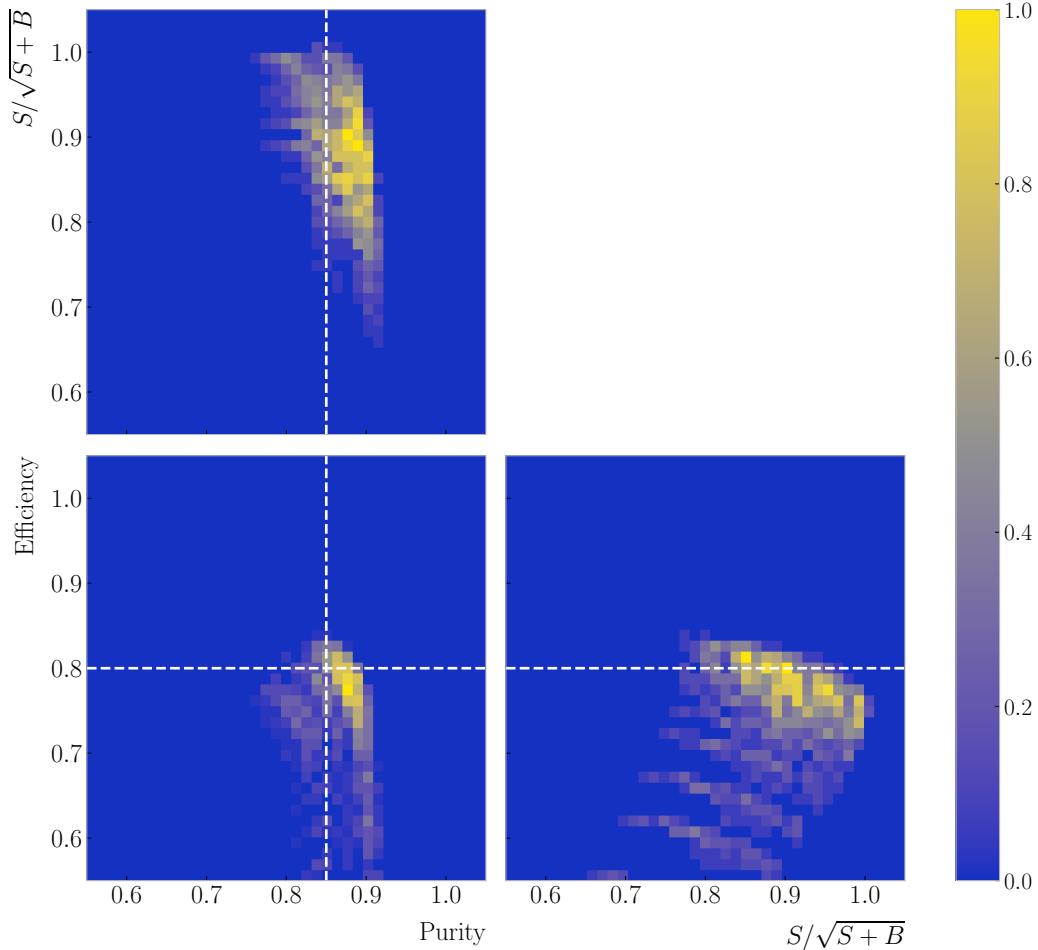


**Figure 7.4:** Significance for the  $\nu_\mu$  CC selection as a function of the muon score cut (left panel), FV half-length cut (middle panel), and radial cut (right panel). The height of the boxes represents the IQR of the conditional distributions, whereas the horizontal line corresponds to the mean.

3984 and  $\Delta R_{\text{FV}}$ . For each value of the cuts, I compute the median and IQR (represented  
 3985 by the horizontal lines and the heights of the boxes, respectively) of the corresponding  
 3986 conditional distributions of efficiency and purity. This representation is useful to get  
 3987 an idea of the general trend the scores follow with the cuts, as well as the spread. It  
 3988 is clear that the muon score cut has the biggest impact on the efficiency, which ranges  
 3989 between 0.05 to 0.80, whereas the purity remains stable with values around 0.85.

3990 A similar depiction of the significance can be found in Fig. 7.4. In this case, one can  
 3991 see that the  $S/\sqrt{S+B}$  decreases as the cuts grow tighter. However, there are hints of

## 7.2. $\nu_\mu$ CC SELECTION

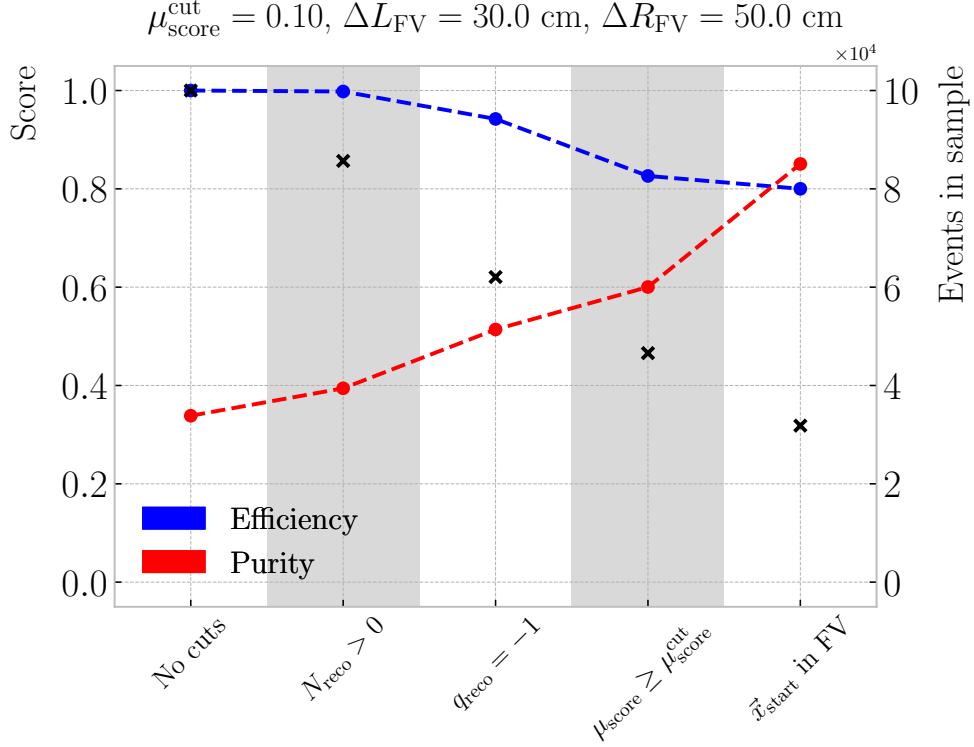


**Figure 7.5:** Normalised 2D distributions of efficiency, purity and significance for the  $\nu_\mu$  CC selection. The  $S/\sqrt{S+B}$  is normalised to the highest value achieved. The vertical dashed line indicates a purity value of 0.85, whereas the horizontal one corresponds to an efficiency of 0.80.

3992 local maxima at intermediate values.

3993 Selecting the cut configuration with the highest significance,  $147 \pm 11$  for the parameter  
 3994 values explored here, results in an efficiency and purity of  $0.754 \pm 0.006$  and  $0.833 \pm 0.007$ ,  
 3995 respectively. Figure 7.5 shows the 2D distributions resulting when combining pairs of  
 3996 efficiency, purity and significance, obtained for the cut configurations explored. The  
 3997 significance is normalised to the highest value obtained in the parameter scan. Looking  
 3998 at this, it is clear that a selection with highest efficiency and purity can be achieved,  
 3999 maintaining a similar significance level.

## CHAPTER 7. EVENT SELECTION IN ND-GAR



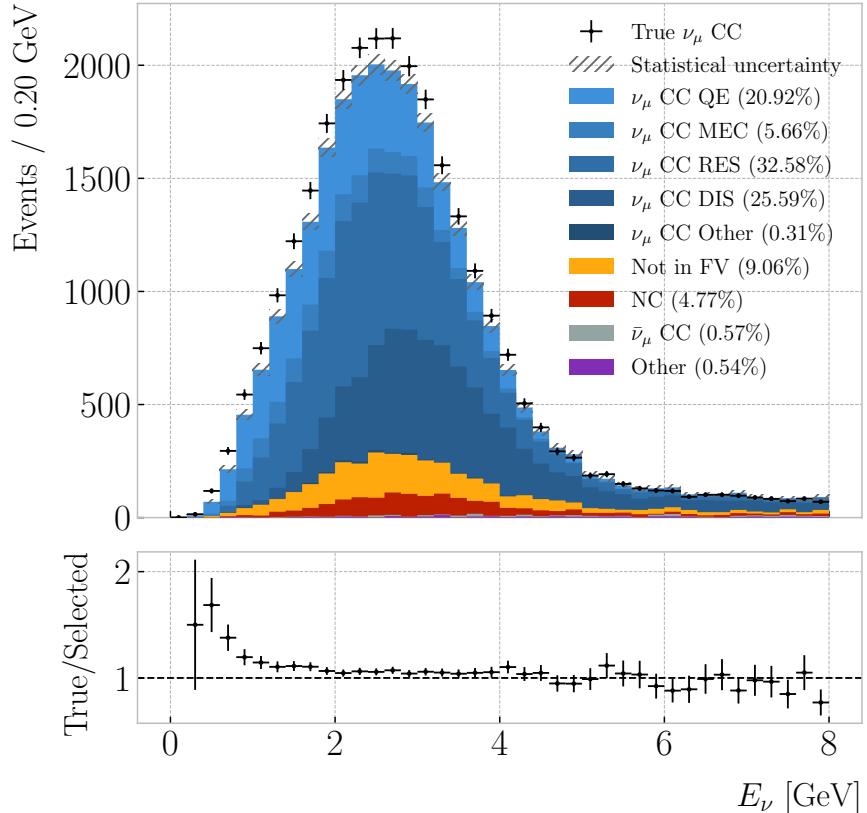
**Figure 7.6:** Cumulative efficiency (blue) and purity (red) of the  $\nu_\mu$  CC selection. The secondary axis indicates the number of events in the sample after each cut (black crosses).

**Table 7.2:** Step-by-step  $\nu_\mu$  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 $\vec{x}_{\text{start}}$ 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  $0.800 \pm 0.007$  and purity of  $0.851 \pm 0.008$ ,

## 7.2. $\nu_\mu$ CC SELECTION



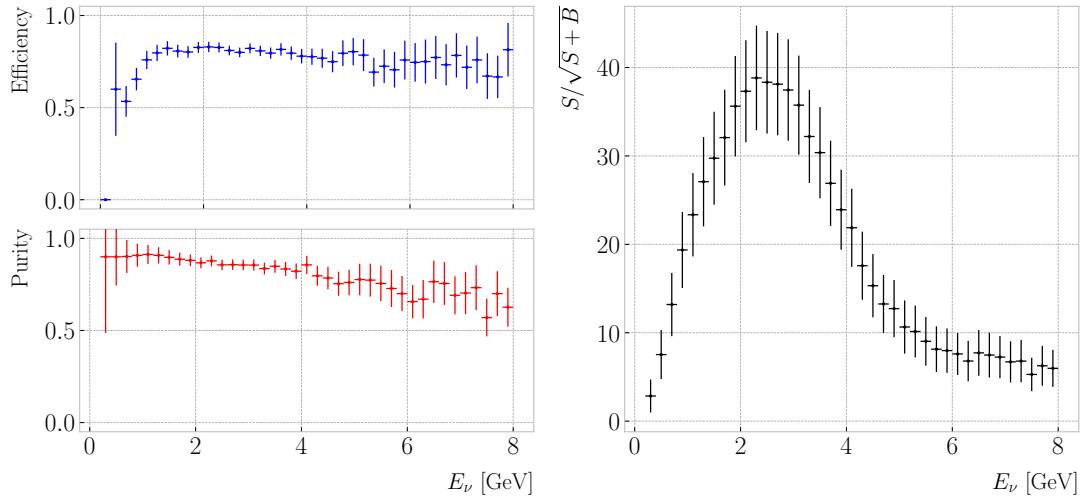
**Figure 7.7:** True neutrino energy spectra for the  $\nu_\mu$  CC selection. The selected events correspond to the coloured stacked histogram, broken down by signal and background subcategories. The statistical uncertainty is drawn in hatched gray. The true distribution is also shown with the black data points. The bottom panel shows the ratio between the number of true and selected  $\nu_\mu$  CC events per bin.

with a significance of  $138 \pm 11$ . Hereafter, I use this optimised selection cuts, unless specified otherwise.

A summary of the selection can be found in Tab. 7.2. It shows the number of events in the selected sample after each selection cut, as well as the absolute and relative passing rates. Figure 7.6 shows the overall efficiencies (blue) and purities (red) after each cut in the event selection is applied. As expected, the efficiency drops while the purity increases with the successive cuts.

Notice how, out of the cuts prior to the FV constraint, the sign selection produces the highest increase in purity. This is one of the advantages of having a magnetised TPC, and can also be used for a  $\bar{\nu}_\mu$  CC selection when running in RHC mode.

## CHAPTER 7. EVENT SELECTION IN ND-GAR



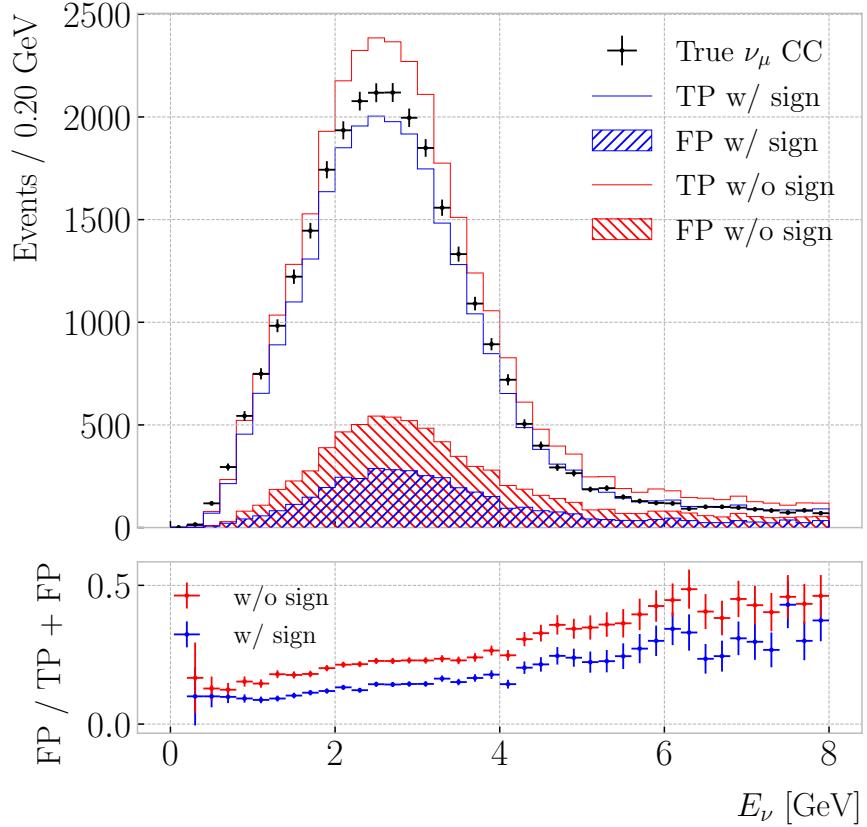
**Figure 7.8:** Left panel: efficiency (top panel) and purity (bottom panel) for the  $\nu_\mu$  CC selection as a function of the true neutrino energy. Right panel: significance for the  $\nu_\mu$  CC selection as a function of the true neutrino energy

### 4015 7.2.2 Selection performance

4016 Using the stored spectra discussed above, the true neutrino energy distribution for the  
 4017 selected events can be recovered doing TP + FP. Similarly, the combination TP + FN  
 4018 gives the true spectrum. Figure 7.7 shows the true (black data points) and selected  
 4019 (coloured stacked histogram)  $E_\nu$  distributions for the optimised  $\nu_\mu$  CC selection. The  
 4020 colours in the selected spectrum indicate the different signal categories and backgrounds,  
 4021 with the overall statistical uncertainty represented by the gray hatched mess. The ratio  
 4022 between the true and selected events is also shown. One can see that it sits around 1 for  
 4023 most of the energy range. However, for energies  $\leq 1$  GeV there is a significant deficit of  
 4024 selected events.

4025 These spectra also allow to compute the efficiency and purity of the selection as  
 4026 a function of the true neutrino energy, as shown in Fig. 7.8 (left panel). As it could  
 4027 be expected from the previous ratio plot, the efficiency is low at low neutrino energies.  
 4028 Nonetheless, it raises quickly with the energy, until it stabilises around a value of 0.80.  
 4029 Looking at the purity, one may notice that, although it starts at around 0.90, there is a  
 4030 significant decrease towards the high end of the spectrum. Figure 7.8 (right panel) also

## 7.2. $\nu_\mu$ CC SELECTION

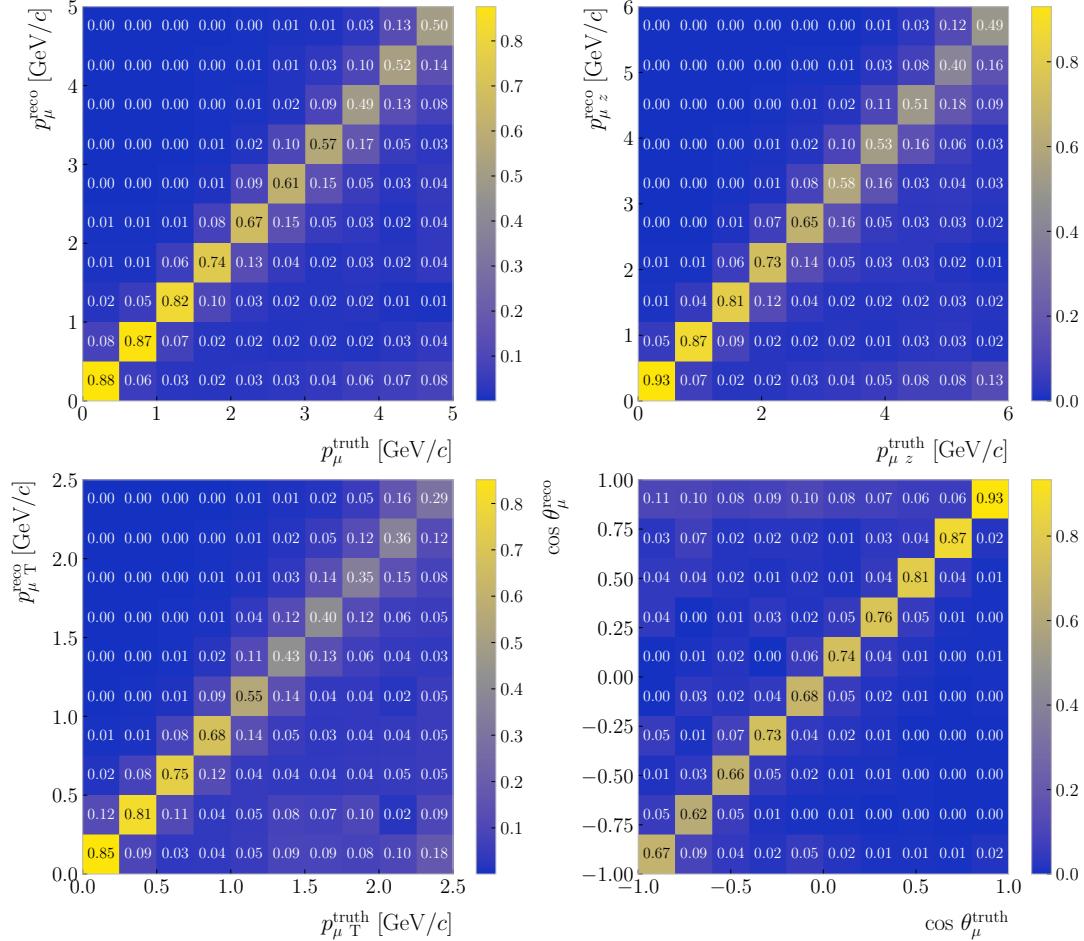


**Figure 7.9:** True neutrino energy spectra for the  $\nu_\mu$  CC selection with (blue) and without (red) sign selection. The selected events are broken down by true positives (signal) and false positives (background). The true distribution is also shown (black data points). The bottom panel shows the ratios between the number of false positives and total selected events per bin.

4031 shows the significance as a function of the energy. In this case, the highest  $S/\sqrt{S+B}$  is  
 4032 achieved around the energies where the spectrum peaks.

4033 A variation of the  $\nu_\mu$  CC selection one can try is to apply it without the reconstructed  
 4034 charge cut. Figure 7.9 (top panel) shows the  $E_\nu$  distributions corresponding to the  
 4035 selection with (blue stacked histogram) and without (red stacked histogram) the sign  
 4036 selection. In the former case, the out of FV contamination amounts to 9.06% of the  
 4037 total, while the NC contamination results 4.77% and the wrong-sign contamination  
 4038 0.57%. For the later, these backgrounds account for the 10.01%, 10.82%, and 2.18%  
 4039 of the selected events, respectively. As expected, removing the positive particles does  
 4040 not change the FV-related effects noticeably. However, the sign selection proves its

## CHAPTER 7. EVENT SELECTION IN ND-GAR



**Figure 7.10:** 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.

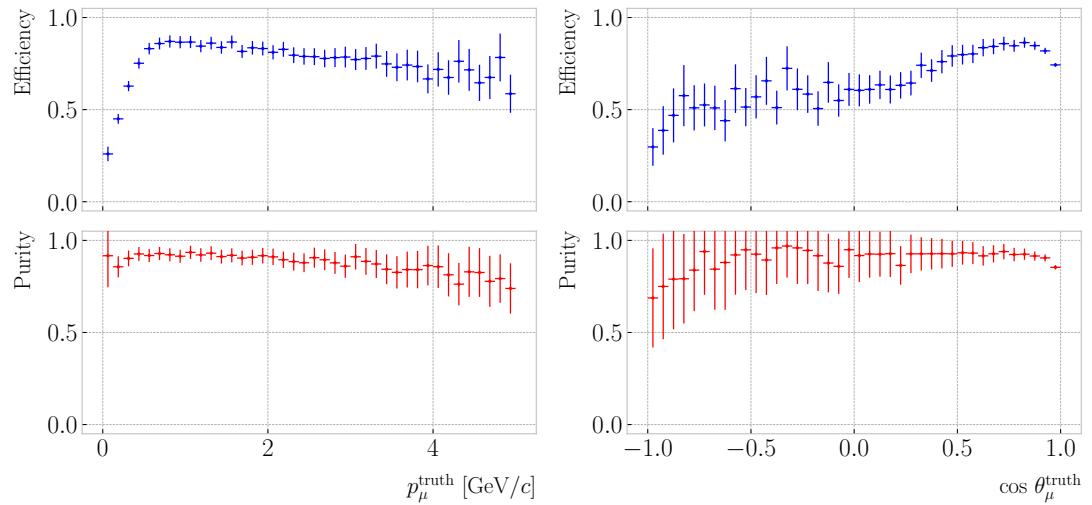
worth in the rejection of  $\bar{\nu}_{\mu}$  CC events, which drop almost by one order of magnitude.

Additionally, the charge selection cuts the NC events in half, as it reduces the chances of misidentifying a positively charged hadron for a muon.

### 7.2.3 Primary muon kinematics

Figure 7.10 shows a comparison between some of the reconstructed and truth primary muon kinematic variables. Notice that, for the reconstructed values, the . That means that, in some cases, we are comparing the kinematic, i.e. the account for both

### 7.3. CHARGED PION IDENTIFICATION



**Figure 7.11:** Efficiency (blue) and purity (red) of the  $\nu_\mu$  CC selection as a function of the primary muon true momentum (left panel) and beam angle (right panel).

4048 reconstruction and selection deficiencies.

4049 study the performance of the  $\nu_\mu$  CC selection as a function of the kinematic variables  
4050 of the primary muon.

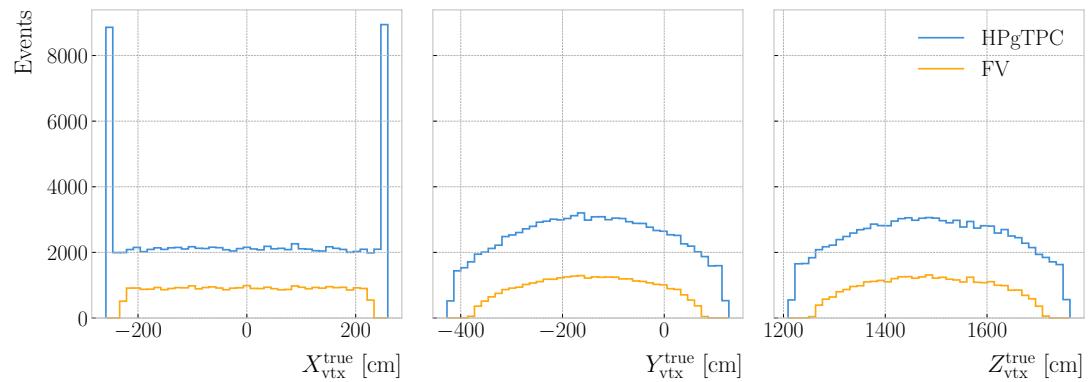
## 4051 7.3 Charged pion identification

### 4052 7.3.1 $\nu_\mu$ CC $1\pi^\pm$ selection

## 4053 7.4 Neutral pion identification

## 4054 7.5 Detector systematic uncertainties

**CHAPTER 7.** EVENT SELECTION IN ND-GAR



**Figure 7.12:** Distributions of the true  $\nu_\mu$  CC vertex positions for the full HPgTPC volume (blue) and the optimised FV (yellow), given by  $\Delta L_{\text{FV}} = 30.0$  cm and  $\Delta R_{\text{FV}} = 50.0$  cm.

8

4055

4056

## Conclusion and outlook



4057

A [REDACTED]

4058

An appendix



## Bibliography

- 4060 [1] J.N. Bahcall, A.M. Serenelli and S. Basu, *New solar opacities, abundances,*  
 4061 *helioseismology, and neutrino fluxes, Astrophys. J. Lett.* **621** (2005) L85  
 4062 [[astro-ph/0412440](#)]. 22, 44, 126, 127, 128
- 4063 [2] R. Garani and S. Palomares-Ruiz, *Dark matter in the Sun: scattering off electrons*  
 4064 *vs nucleons, JCAP* **05** (2017) 007 [[1702.02768](#)]. 22, 124, 125, 128, 129, 152
- 4065 [3] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara and S. Midorikawa,  
 4066 *Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model,*  
 4067 *Phys. Rev. D* **92** (2015) 023004 [[1502.03916](#)]. 22, 133
- 4068 [4] M. Colom i Bernadich and C. Pérez de los Heros, *Limits on Kaluza-Klein dark*  
 4069 *matter annihilation in the Sun from recent IceCube results, Eur. Phys. J. C,* **80** 2  
 4070 (*2020*) 129 **80** (2019) [[1912.04585](#)]. 23, 137, 138
- 4071 [5] ANTARES collaboration, *Search for Dark Matter in the Sun with the ANTARES*  
 4072 *Neutrino Telescope in the CMSSM and mUED frameworks, Nucl. Instrum. Meth.*  
 4073 *A* **725** (2013) 76 [[1204.5290](#)]. 23, 137, 138
- 4074 [6] N. Deutschmann, T. Flacke and J.S. Kim, *Current LHC Constraints on Minimal*  
 4075 *Universal Extra Dimensions, Phys. Lett. B* **771** (2017) 515 [[1702.00410](#)]. 23, 137,  
 4076 138
- 4077 [7] ICECUBE collaboration, *Search for GeV-scale dark matter annihilation in the Sun*  
 4078 *with IceCube DeepCore, Phys. Rev. D* **105** (2022) 062004 [[2111.09970](#)]. 24, 149,  
 4079 150

## BIBLIOGRAPHY

- 4080 [8] C.-S. Chen, F.-F. Lee, G.-L. Lin and Y.-H. Lin, *Probing Dark Matter*  
4081 *Self-Interaction in the Sun with IceCube-PINGU*, *JCAP* **10** (2014) 049  
4082 [1408.5471]. 24, 149, 150
- 4083 [9] N.F. Bell, M.J. Dolan and S. Robles, *Searching for dark matter in the Sun using*  
4084 *Hyper-Kamiokande*, *JCAP* **11** (2021) 004 [2107.04216]. 24, 149, 150
- 4085 [10] E. Behnke et al., *Final Results of the PICASSO Dark Matter Search Experiment*,  
4086 *Astropart. Phys.* **90** (2017) 85 [1611.01499]. 24, 149, 150
- 4087 [11] PICO collaboration, *Dark Matter Search Results from the Complete Exposure of*  
4088 *the PICO-60 C<sub>3</sub>F<sub>8</sub> Bubble Chamber*, *Phys. Rev. D* **100** (2019) 022001  
4089 [1902.04031]. 24, 149, 150
- 4090 [12] DARKSIDE collaboration, *Constraints on Sub-GeV Dark-Matter–Electron*  
4091 *Scattering from the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018) 111303  
4092 [1802.06998]. 24, 154, 155
- 4093 [13] XENON collaboration, *Light Dark Matter Search with Ionization Signals in*  
4094 *XENON1T*, *Phys. Rev. Lett.* **123** (2019) 251801 [1907.11485]. 24, 154, 155
- 4095 [14] P.F. de Salas, D.V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena,  
4096 C.A. Ternes et al., *2020 global reassessment of the neutrino oscillation picture*,  
4097 *JHEP* **02** (2021) 071 [2006.11237]. 33, 57, 65
- 4098 [15] W. Pauli, *Dear radioactive ladies and gentlemen*, *Phys. Today* **31N9** (1978) 27. 39
- 4099 [16] F. Reines and C.L. Cowan, *Detection of the free neutrino*, *Phys. Rev.* **92** (1953)  
4100 830. 39
- 4101 [17] S.L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** 579. 39
- 4102 [18] S. Weinberg, *A model of leptons*, *Physical Review Letters* **19** 1264. 39
- 4103 [19] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968)  
4104 367. 39

## BIBLIOGRAPHY

- 4105 [20] A. Pich, *The Standard Model of Electroweak Interactions*, in *2010 European  
4106 School of High Energy Physics*, pp. 1–50, 1, 2012 [[1201.0537](#)]. 40
- 4107 [21] ALEPH, DELPHI, L3, OPAL, SLD, LEP ELECTROWEAK WORKING GROUP,  
4108 SLD ELECTROWEAK GROUP, SLD HEAVY FLAVOUR GROUP collaboration,  
4109 *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006)  
4110 257 [[hep-ex/0509008](#)]. 43
- 4111 [22] R. Davis, D.S. Harmer and K.C. Hoffman, *Search for neutrinos from the sun*,  
4112 *Physical Review Letters* **20** 1205. 44
- 4113 [23] J.N. Bahcall, N.A. Bahcall and G. Shaviv, *Present Status of the Theoretical  
4114 Predictions for the  $^{37}\text{Cl}$  Solar-Neutrino Experiment*, . 44
- 4115 [24] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,  
4116 *Measurement of the solar electron neutrino flux with the Homestake chlorine  
4117 detector*, *Astrophys. J.* **496** (1998) 505. 44
- 4118 [25] SAGE collaboration, *Measurement of the solar neutrino capture rate with gallium  
4119 metal. III: Results for the 2002–2007 data-taking period*, *Phys. Rev. C* **80** (2009)  
4120 015807 [[0901.2200](#)]. 44, 56
- 4121 [26] F. Kaether, W. Hampel, G. Heusser, J. Kiko and T. Kirsten, *Reanalysis of the  
4122 GALLEX solar neutrino flux and source experiments*, *Phys. Lett. B* **685** (2010) 47  
4123 [[1001.2731](#)]. 44, 56
- 4124 [27] Q.R. Ahmad, R.C. Allen, T.C. Andersen, J.D. Anglin, G. Bühler, J.C. Barton  
4125 et al., *Measurement of the Rate of  $\nu_e + d \rightarrow p + p + e^-$  Interactions Produced by  
4126  $^8\text{B}$  Solar Neutrinos at the Sudbury Neutrino Observatory*, . 45
- 4127 [28] Q.R. Ahmad, R.C. Allen, T.C. Andersen, J. D. Anglin, J.C. Barton, E.W. Beier  
4128 et al., *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current  
4129 Interactions in the Sudbury Neutrino Observatory*, . 45
- 4130 [29] T.K. Gaisser and M. Honda, *Flux of atmospheric neutrinos*, . 45

## BIBLIOGRAPHY

- 4131 [30] K. Hirata, T. Kajita, M. Koshiba, M. Nakahata, S. Ohara, Y. Oyama et al.,  
4132 *Experimental study of the atmospheric neutrino flux*, . 46
- 4133 [31] D. Casper, R. Becker-Szendy, C.B. Bratton, D.R. Cady, R. Claus, S.T. Dye et al.,  
4134 *Measurement of atmospheric neutrino composition with the imb-3 detector*, . 46
- 4135 [32] M. Ambrosio, R. Antolini, C. Aramo, G. Auriemma, A. Baldini, G. C. Barbarino  
4136 et al., *Measurement of the atmospheric neutrino-induced upgoing muon flux using*  
4137 *macro*, . 46
- 4138 [33] W. Allison, G. Alner, D. Ayres, W. Barrett, C. Bode, P. Border et al.,  
4139 *Measurement of the atmospheric neutrino flavour composition in soudan 2*, . 46
- 4140 [34] SUPER-KAMIOKANDE collaboration, *Evidence for oscillation of atmospheric*  
4141 *neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [[hep-ex/9807003](#)]. 46
- 4142 [35] P. Minkowski,  $\mu \rightarrow e\gamma$  at a Rate of One Out of  $10^9$  Muon Decays?, *Phys. Lett. B*  
4143 **67** (1977) 421. 49
- 4144 [36] M. Gell-Mann, P. Ramond and R. Slansky, *Complex Spinors and Unified Theories*,  
4145 *Conf. Proc. C* **790927** (1979) 315 [[1306.4669](#)]. 49
- 4146 [37] T. Yanagida, *Horizontal gauge symmetry and masses of neutrinos*, *Conf. Proc. C*  
4147 **7902131** (1979) 95. 49
- 4148 [38] R.N. Mohapatra and G. Senjanovic, *Neutrino Mass and Spontaneous Parity*  
4149 *Nonconservation*, *Phys. Rev. Lett.* **44** (1980) 912. 49
- 4150 [39] J. Schechter and J.W.F. Valle, *Neutrino Masses in  $SU(2) \times U(1)$  Theories*, *Phys.*  
4151 *Rev. D* **22** (1980) 2227. 49
- 4152 [40] B. Pontecorvo, *Mesonium and anti-mesonium*, *Sov. Phys. JETP* **6** (1957) 429. 49
- 4153 [41] M. Gell-Mann and A. Pais, *Behavior of neutral particles under charge conjugation*,  
4154 . 49

## BIBLIOGRAPHY

- 4155 [42] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of*  
 4156 *Leptonic Charge, Zh. Eksp. Teor. Fiz.* **53** (1967) 1717. 50
- 4157 [43] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge, Zh.*  
 4158 *Eksp. Teor. Fiz.* **34** (1957) 247. 51
- 4159 [44] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary*  
 4160 *particles, Prog. Theor. Phys.* **28** (1962) 870. 51
- 4161 [45] PARTICLE DATA GROUP collaboration, *Review of particle physics, Phys. Rev. D*  
 4162 **110** (2024) 030001. 53
- 4163 [46] L. Wolfenstein, *Neutrino Oscillations in Matter, Phys. Rev. D* **17** (1978) 2369. 54
- 4164 [47] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,  
 4165 *Measurement of the solar electron neutrino flux with the Homestake chlorine*  
 4166 *detector, Astrophys. J.* **496** (1998) 505. 56
- 4167 [48] G. Bellini et al., *Precision measurement of the  $^{7}\text{Be}$  solar neutrino interaction rate*  
 4168 *in Borexino, Phys. Rev. Lett.* **107** (2011) 141302 [1104.1816]. 56
- 4169 [49] SUPER-KAMIOKANDE collaboration, *Solar neutrino measurements in*  
 4170 *super-Kamiokande-I, Phys. Rev. D* **73** (2006) 112001 [hep-ex/0508053]. 56
- 4171 [50] SNO collaboration, *Combined Analysis of all Three Phases of Solar Neutrino*  
 4172 *Data from the Sudbury Neutrino Observatory, Phys. Rev. C* **88** (2013) 025501  
 4173 [1109.0763]. 56
- 4174 [51] SUPER-KAMIOKANDE collaboration, *Atmospheric neutrino oscillation analysis*  
 4175 *with external constraints in Super-Kamiokande I-IV, Phys. Rev. D* **97** (2018)  
 4176 072001 [1710.09126]. 56, 157
- 4177 [52] ICECUBE collaboration, *Measurement of Atmospheric Neutrino Oscillations at*  
 4178 *6–56 GeV with IceCube DeepCore, Phys. Rev. Lett.* **120** (2018) 071801  
 4179 [1707.07081]. 56

## BIBLIOGRAPHY

- 4180 [53] KamLAND collaboration, *Reactor On-Off Antineutrino Measurement with  
4181 KamLAND*, *Phys. Rev. D* **88** (2013) 033001 [[1303.4667](#)]. 57
- 4182 [54] RENO collaboration, *Measurement of Reactor Antineutrino Oscillation Amplitude  
4183 and Frequency at RENO*, *Phys. Rev. Lett.* **121** (2018) 201801 [[1806.00248](#)]. 57
- 4184 [55] DAYA BAY collaboration, *Measurement of the Electron Antineutrino Oscillation  
4185 with 1958 Days of Operation at Daya Bay*, *Phys. Rev. Lett.* **121** (2018) 241805  
4186 [[1809.02261](#)]. 57
- 4187 [56] K.J. Kelly, P.A.N. Machado, S.J. Parke, Y.F. Perez-Gonzalez and R.Z. Funchal,  
4188 *Neutrino mass ordering in light of recent data*, *Phys. Rev. D* **103** (2021) 013004.  
4189 57
- 4190 [57] P. Dunne, “Latest Neutrino Oscillation Results from T2K.” Jul, 2020. 57
- 4191 [58] MINOS collaboration, *Combined analysis of  $\nu_\mu$  disappearance and  $\nu_\mu \rightarrow \nu_e$   
4192 appearance in MINOS using accelerator and atmospheric neutrinos*, *Phys. Rev.  
4193 Lett.* **112** (2014) 191801 [[1403.0867](#)]. 57
- 4194 [59] OPERA collaboration, *Final Results of the OPERA Experiment on  $\nu_\tau$   
4195 Appearance in the CNGS Neutrino Beam*, *Phys. Rev. Lett.* **120** (2018) 211801  
4196 [[1804.04912](#)]. 57
- 4197 [60] K2K collaboration, *Measurement of Neutrino Oscillation by the K2K Experiment*,  
4198 *Phys. Rev. D* **74** (2006) 072003 [[hep-ex/0606032](#)]. 57
- 4199 [61] DUNE collaboration, *Long-baseline neutrino oscillation physics potential of the  
4200 DUNE experiment*, *Eur. Phys. J. C* **80** (2020) 978 [[2006.16043](#)]. 57
- 4201 [62] HYPER-KAMIOKANDE collaboration, *Physics potential of Hyper-Kamiokande for  
4202 neutrino oscillation measurements*, *PoS NuFact2019* (2019) 040. 57
- 4203 [63] SUPERNEMO collaboration, *Probing New Physics Models of Neutrinoless Double  
4204 Beta Decay with SuperNEMO*, *Eur. Phys. J. C* **70** (2010) 927 [[1005.1241](#)]. 58

## BIBLIOGRAPHY

- 4205 [64] SNO+ collaboration, *Current Status and Future Prospects of the SNO+*  
 4206 *Experiment, Adv. High Energy Phys.* **2016** (2016) 6194250 [[1508.05759](#)]. 58
- 4207 [65] NEXT collaboration, *Sensitivity of a tonne-scale NEXT detector for neutrinoless*  
 4208 *double beta decay searches, JHEP* **2021** (2021) 164 [[2005.06467](#)]. 58
- 4209 [66] P. Coloma and P. Huber, *Impact of nuclear effects on the extraction of neutrino*  
 4210 *oscillation parameters, Phys. Rev. Lett.* **111** (2013) 221802 [[1307.1243](#)]. 59
- 4211 [67] P. Coloma, P. Huber, C.-M. Jen and C. Mariani, *Neutrino-nucleus interaction*  
 4212 *models and their impact on oscillation analyses, Phys. Rev. D* **89** (2014) 073015  
 4213 [[1311.4506](#)]. 59
- 4214 [68] U. Mosel, *Neutrino Interactions with Nucleons and Nuclei: Importance for*  
 4215 *Long-Baseline Experiments, Ann. Rev. Nucl. Part. Sci.* **66** (2016) 171  
 4216 [[1602.00696](#)]. 59
- 4217 [69] J.A. Formaggio and G.P. Zeller, *From ev to eev: Neutrino cross sections across*  
 4218 *energy scales, Rev. Mod. Phys.* **84** (2012) 1307 [[1305.7513](#)]. 60
- 4219 [70] L. Bathe-Peters, S. Gardiner and R. Guenette, *Comparing generator predictions*  
 4220 *of transverse kinematic imbalance in neutrino-argon scattering,* [2201.04664](#). 61
- 4221 [71] R.A. Smith and E.J. Moniz, *Neutrino reactions on nuclear targets, Nucl. Phys. B*  
 4222 **43** (1972) 605. 61
- 4223 [72] H. Nakamura and R. Seki, *Quasi-elastic neutrino-nucleus scattering and spectral*  
 4224 *function, Nuclear Physics B - Proceedings Supplements* **112** 197. 61
- 4225 [73] V. Pandey, N. Jachowicz, T. Van Cuyck, J. Ryckebusch and M. Martini,  
 4226 *Low-energy excitations and quasielastic contribution to electron-nucleus and*  
 4227 *neutrino-nucleus scattering in the continuum random-phase approximation, Phys.*  
 4228 *Rev. C* **92** (2015) 024606 [[1412.4624](#)]. 61

## BIBLIOGRAPHY

- 4229 [74] A. Nikolakopoulos, R. González-Jiménez, N. Jachowicz, K. Niewczas, F. Sánchez  
4230 and J.M. Udías, *Benchmarking intranuclear cascade models for neutrino*  
4231 *scattering with relativistic optical potentials*, *Phys. Rev. C* **105** (2022) 054603  
4232 [2202.01689]. 61
- 4233 [75] MINIBooNE collaboration, *First Measurement of the Muon Neutrino Charged*  
4234 *Current Quasielastic Double Differential Cross Section*, *Phys. Rev. D* **81** (2010)  
4235 092005 [1002.2680]. 62
- 4236 [76] MINERvA collaboration, *Measurements of the Inclusive Neutrino and*  
4237 *Antineutrino Charged Current Cross Sections in MINERvA Using the Low- $\nu$  Flux*  
4238 *Method*, *Phys. Rev. D* **94** (2016) 112007 [1610.04746]. 62
- 4239 [77] MICROBooNE collaboration, *First Measurement of Energy-Dependent Inclusive*  
4240 *Muon Neutrino Charged-Current Cross Sections on Argon with the MicroBooNE*  
4241 *Detector*, *Phys. Rev. Lett.* **128** (2022) 151801 [2110.14023]. 62
- 4242 [78] SBND collaboration, *Neutrino cross-section measurement prospects with SBND*,  
4243 *PoS NuFact2017* (2018) 067. 62
- 4244 [79] NOvA collaboration, *Measurement of the double-differential muon-neutrino*  
4245 *charged-current inclusive cross section in the NOvA near detector*, *Phys. Rev. D*  
4246 **107** (2023) 052011 [2109.12220]. 62
- 4247 [80] T2K collaboration, *Measurement of the  $\nu_\mu$  charged-current cross sections on*  
4248 *water, hydrocarbon, iron, and their ratios with the T2K on-axis detectors*, *PTEP*  
4249 **2019** (2019) 093C02 [1904.09611]. 62
- 4250 [81] \$NU\$PRISM collaboration,  *$\nu$ PRISM: A new way of probing neutrino*  
4251 *interactions*, *PoS NUFACT2014* (2015) 046. 62
- 4252 [82] C. Hasnip, *DUNE-PRISM – A New Method to Measure Neutrino Oscillations*,  
4253 Ph.D. thesis, Oxford U., 2023. 62

## BIBLIOGRAPHY

- 4254 [83] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*  
 4255 *Detector Technical Design Report, Volume I Introduction to DUNE*, JINST **15**  
 4256 (2020) T08008 [2002.02967]. 63, 64, 65, 70, 71, 81, 82, 83, 84
- 4257 [84] DUNE collaboration, *Snowmass Neutrino Frontier: DUNE Physics Summary*,  
 4258 2203.06100. 65
- 4259 [85] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*  
 4260 *Detector Technical Design Report, Volume II: DUNE Physics*, 2002.03005. 66, 67,  
 4261 68, 75, 133, 142, 146
- 4262 [86] SUPER-KAMIOKANDE collaboration, *Search for Proton Decay via  $p \rightarrow e^+ \pi_0$  and*  
 4263  *$p \rightarrow \mu^+ \pi_0$  in a Large Water Cherenkov Detector*, Phys. Rev. Lett. **102** (2009)  
 4264 141801 [0903.0676]. 66
- 4265 [87] S. Raby, *Grand Unified Theories*, in *2nd World Summit: Physics Beyond the*  
 4266 *Standard Model*, 8, 2006 [[hep-ph/0608183](#)]. 66
- 4267 [88] KAMIOKANDE-II collaboration, *Observation of a Neutrino Burst from the*  
 4268 *Supernova SN 1987a*, Phys. Rev. Lett. **58** (1987) 1490. 66
- 4269 [89] R.M. Bionta et al., *Observation of a Neutrino Burst in Coincidence with*  
 4270 *Supernova SN 1987a in the Large Magellanic Cloud*, Phys. Rev. Lett. **58** (1987)  
 4271 1494. 66
- 4272 [90] J. Strait, E. McCluskey, T. Lundin, J. Willhite, T. Hamernik, V. Papadimitriou  
 4273 et al., *Long-baseline neutrino facility (lbnf) and deep underground neutrino*  
 4274 *experiment (dune) conceptual design report volume 3: Long-baseline neutrino*  
 4275 *facility for dune june 24, 2015*, 1601.05823. 67
- 4276 [91] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE) Near*  
 4277 *Detector Conceptual Design Report, Instruments* **5** (2021) 31 [2103.13910]. 69,  
 4278 70, 73, 74, 76, 77, 159

## BIBLIOGRAPHY

- 4279 [92] DUNE collaboration, *DUNE Phase II: Scientific Opportunities, Detector*  
4280 *Concepts, Technological Solutions*, 2408.12725. 72, 78
- 4281 [93] DUNE collaboration, *A Gaseous Argon-Based Near Detector to Enhance the*  
4282 *Physics Capabilities of DUNE*, 2203.06281. 74
- 4283 [94] F. Sauli, *Gem: A new concept for electron amplification in gas detectors*, *Nuclear*  
4284 *Instruments and Methods in Physics Research Section A: Accelerators,*  
4285 *Spectrometers, Detectors and Associated Equipment* **386** 531. 76, 80
- 4286 [95] A. Ritchie-Yates et al., *First operation of an ALICE OROC operated in high*  
4287 *pressure Ar-CO<sub>2</sub> and Ar-CH<sub>4</sub>*, *Eur. Phys. J. C* **83** (2023) 1139 [2305.08822]. 79
- 4288 [96] ALICE TPC collaboration, *The upgrade of the ALICE TPC with GEMs and*  
4289 *continuous readout*, *JINST* **16** (2021) P03022 [2012.09518]. 79
- 4290 [97] F. Sauli, *The gas electron multiplier (gem): Operating principles and applications*,  
4291 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,*  
4292 *Spectrometers, Detectors and Associated Equipment* **805** 2. 80
- 4293 [98] C. Lippmann, *A continuous read-out tpc for the alice upgrade*, *Nuclear*  
4294 *Instruments and Methods in Physics Research Section A: Accelerators,*  
4295 *Spectrometers, Detectors and Associated Equipment* **824** 543. 80
- 4296 [99] C. Calabria, *Large-size triple gem detectors for the cms forward muon upgrade*,  
4297 *Nuclear and Particle Physics Proceedings* **273–275** 1042. 80
- 4298 [100] DUNE collaboration, *The DUNE Far Detector Vertical Drift Technology*,  
4299 *Technical Design Report*, 2312.03130. 85, 86
- 4300 [101] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*  
4301 *Detector Technical Design Report, Volume IV: Far Detector Single-phase*  
4302 *Technology*, *JINST* **15** (2020) T08010 [2002.03010]. 87

## BIBLIOGRAPHY

- 4303 [102] DUNE DAQ, “dtp-firmware.”  
4304 <https://gitlab.cern.ch/dune-daq/readout/dtp-firmware>, 2020. 89
- 4305 [103] DUNE DAQ, “dtp-simulation.”  
4306 <https://gitlab.cern.ch/dune-daq/readout/dtp-simulation>, 2020. 92
- 4307 [104] DUNE DAQ, “dtpe emulator.”  
4308 [https://github.com/DUNE-DAQ/dtpe emulator/tree/fmlopez/filter\\_ana](https://github.com/DUNE-DAQ/dtpe emulator/tree/fmlopez/filter_ana),  
4309 2022. 92
- 4310 [105] J. McClellan and T. Parks, *A personal history of the Parks-McClellan algorithm*,  
4311 *IEEE Signal Processing Magazine* **22** (2005) 82. 93
- 4312 [106] G. Turin, *An introduction to matched filters*, *IEEE Transactions on Information*  
4313 *Theory* **6** (1960) 311. 96
- 4314 [107] J.W. Goodman, *Statistical Optics*, Wiley (1985). 98
- 4315 [108] B. Dwork, *Detection of a pulse superimposed on fluctuation noise*, *Proceedings of*  
4316 *the IRE* **38** (1950) 771. 99
- 4317 [109] L. Wainstein and V. Zubakov, *Extraction of Signals from Noise*, Prentice-Hall  
4318 (1962). 99
- 4319 [110] E.D. Church, *Larsoft: A software package for liquid argon time projection drift*  
4320 *chambers*, **1311.6774**. 102, 103
- 4321 [111] S.V. Stehman, *Selecting and interpreting measures of thematic classification*  
4322 *accuracy*, *Remote Sensing of Environment* **62** (1997) 77. 116
- 4323 [112] A.A. Taha and A. Hanbury, *Metrics for evaluating 3d medical image*  
4324 *segmentation: analysis, selection, and tool*, *BMC Medical Imaging* **15** (2015) . 117
- 4325 [113] J. Silk, K.A. Olive and M. Srednicki, *The Photino, the Sun and High-Energy*  
4326 *Neutrinos*, *Phys. Rev. Lett.* **55** (1985) 257. 123

## BIBLIOGRAPHY

- 4327 [114] M. Srednicki, K.A. Olive and J. Silk, *High-Energy Neutrinos from the Sun and*  
4328 *Cold Dark Matter*, *Nucl. Phys. B* **279** (1987) 804. 123
- 4329 [115] J.S. Hagelin, K.W. Ng and K.A. Olive, *A High-energy Neutrino Signature From*  
4330 *Supersymmetric Relics*, *Phys. Lett. B* **180** (1986) 375. 123
- 4331 [116] T.K. Gaisser, G. Steigman and S. Tilav, *Limits on Cold Dark Matter Candidates*  
4332 *from Deep Underground Detectors*, *Phys. Rev. D* **34** (1986) 2206. 123
- 4333 [117] N. Bernal, J. Martín-Albo and S. Palomares-Ruiz, *A novel way of constraining*  
4334 *WIMPs annihilations in the Sun: MeV neutrinos*, *JCAP* **08** (2013) 011  
4335 [1208.0834]. 123, 124, 130
- 4336 [118] C. Rott, J. Siegal-Gaskins and J.F. Beacom, *New Sensitivity to Solar WIMP*  
4337 *Annihilation using Low-Energy Neutrinos*, *Phys. Rev. D* **88** (2013) 055005  
4338 [1208.0827]. 123, 130
- 4339 [119] C. Rott, S. In, J. Kumar and D. Yaylali, *Dark Matter Searches for Monoenergetic*  
4340 *Neutrinos Arising from Stopped Meson Decay in the Sun*, *JCAP* **11** (2015) 039  
4341 [1510.00170]. 123
- 4342 [120] DUNE collaboration, *Searching for solar KDAR with DUNE*, *JCAP* **10** (2021)  
4343 065 [2107.09109]. 123
- 4344 [121] G. Busoni, A. De Simone and W.-C. Huang, *On the Minimum Dark Matter Mass*  
4345 *Testable by Neutrinos from the Sun*, *JCAP* **07** (2013) 010 [1305.1817]. 124
- 4346 [122] V.A. Bednyakov and F. Simkovic, *Nuclear spin structure in dark matter search:*  
4347 *The Zero momentum transfer limit*, *Phys. Part. Nucl.* **36** (2005) 131  
4348 [hep-ph/0406218]. 125
- 4349 [123] A. Gould, *WIMP Distribution in and Evaporation From the Sun*, *Astrophys. J.*  
4350 **321** (1987) 560. 125

## BIBLIOGRAPHY

- 4351 [124] J.N. Bahcall, M.H. Pinsonneault and S. Basu, *Solar models: Current epoch and*  
 4352 *time dependences, neutrinos, and helioseismological properties*, *Astrophys. J.* **555**  
 4353 (2001) 990 [[astro-ph/0010346](#)]. 128
- 4354 [125] T. Golan, J.T. Sobczyk and J. Zmuda, *NuWro: the Wroclaw Monte Carlo*  
 4355 *Generator of Neutrino Interactions*, *Nucl. Phys. B Proc. Suppl.* **229-232** (2012)  
 4356 499. 132
- 4357 [126] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for*  
 4358 *likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [[1007.1727](#)].  
 4359 134
- 4360 [127] T. Kaluza, *Zum Unitätsproblem der Physik, Sitzungsber. Preuss. Akad. Wiss.*  
 4361 *Berlin (Math. Phys.)* **1921** (1921) 966 [[1803.08616](#)]. 135
- 4362 [128] O. Klein, *Quantum Theory and Five-Dimensional Theory of Relativity. (In*  
 4363 *German and English)*, *Z. Phys.* **37** (1926) 895. 135
- 4364 [129] T. Appelquist, H.-C. Cheng and B.A. Dobrescu, *Bounds on universal extra*  
 4365 *dimensions*, *Phys. Rev. D* **64** (2001) 035002 [[hep-ph/0012100](#)]. 135
- 4366 [130] N. Arkani-Hamed, S. Dimopoulos and G.R. Dvali, *The Hierarchy problem and new*  
 4367 *dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263 [[hep-ph/9803315](#)]. 135
- 4368 [131] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra*  
 4369 *dimension*, *Phys. Rev. Lett.* **83** (1999) 3370 [[hep-ph/9905221](#)]. 135
- 4370 [132] G. Servant and T.M.P. Tait, *Is the lightest Kaluza-Klein particle a viable dark*  
 4371 *matter candidate?*, *Nucl. Phys. B* **650** (2003) 391 [[hep-ph/0206071](#)]. 136
- 4372 [133] H.-C. Cheng, K.T. Matchev and M. Schmaltz, *Radiative corrections to*  
 4373 *Kaluza-Klein masses*, *Phys. Rev. D* **66** (2002) 036005 [[hep-ph/0204342](#)]. 136
- 4374 [134] M. Blennow, J. Edsjö and T. Ohlsson, *Neutrinos from WIMP annihilations using*  
 4375 *a full three-flavor Monte Carlo*, *JCAP* **01** (2008) 021 [[0709.3898](#)]. 136, 139

## BIBLIOGRAPHY

- 4376 [135] J. Edsjö, J. Elevant and C. Niblaeus, *WimpSim Neutrino Monte Carlo.* 136, 139
- 4377 [136] U. Haisch and A. Weiler, *Bound on minimal universal extra dimensions from*  
4378 *anti- $B \rightarrow X(s)\gamma$ ,* *Phys. Rev. D* **76** (2007) 034014 [[hep-ph/0703064](#)]. 138
- 4379 [137] A. Freitas and U. Haisch, *Anti- $B \rightarrow X(s)\gamma$  in two universal extra*  
4380 *dimensions,* *Phys. Rev. D* **77** (2008) 093008 [[0801.4346](#)]. 138
- 4381 [138] C. Rott, D. Jeong, J. Kumar and D. Yaylali, *Neutrino Topology Reconstruction at*  
4382 *DUNE and Applications to Searches for Dark Matter Annihilation in the Sun,*  
4383 *JCAP* **07** (2019) 006 [[1903.04175](#)]. 141
- 4384 [139] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al.,  
4385 *Scikit-learn: Machine learning in Python,* *Journal of Machine Learning Research*  
4386 **12** (2011) 2825. 147
- 4387 [140] J. Kopp, V. Niro, T. Schwetz and J. Zupan, *DAMA/LIBRA and leptonically*  
4388 *interacting Dark Matter,* *Phys. Rev. D* **80** (2009) 083502 [[0907.3159](#)]. 151
- 4389 [141] PARTICLE DATA GROUP collaboration, *Review of Particle Physics,* *PTEP* **2020**  
4390 (2020) 083C01. 152
- 4391 [142] M. Beltran, D. Hooper, E.W. Kolb and Z.C. Krusberg, *Deducing the nature of*  
4392 *dark matter from direct and indirect detection experiments in the absence of*  
4393 *collider signatures of new physics,* *Phys. Rev. D* **80** (2009) 043509 [[0808.3384](#)].  
4394 153
- 4395 [143] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters,* *Astron.*  
4396 *Astrophys.* **641** (2020) A6 [[1807.06209](#)]. 153
- 4397 [144] C. Principato, *An Indirect Search for Weakly Interacting Massive Particles in the*  
4398 *Sun Using Upward-going Muons in NOvA*, Ph.D. thesis, Virginia U., 2021.  
4399 10.18130/x5z2-1466. 156

## BIBLIOGRAPHY

- 4400 [145] G. Wikström and J. Edsjö, *Limits on the wimp-nucleon scattering cross-section*  
 4401 *from neutrino telescopes*, 0903.2986. 156
- 4402 [146] SUPER-KAMIOKANDE collaboration, *Search for neutrinos from annihilation of*  
 4403 *captured low-mass dark matter particles in the Sun by Super-Kamiokande*, *Phys.*  
 4404 *Rev. Lett.* **114** (2015) 141301 [1503.04858]. 156
- 4405 [147] C. Rott, T. Tanaka and Y. Itow, *Enhanced sensitivity to dark matter*  
 4406 *self-annihilations in the sun using neutrino spectral information*, 1107.3182. 156
- 4407 [148] M.M. Boliev, S.V. Demidov, S.P. Mikheyev and O.V. Suvorova, *Search for muon*  
 4408 *signal from dark matter annihilations inthe Sun with the Baksan Underground*  
 4409 *Scintillator Telescope for 24.12 years*, *JCAP* **09** (2013) 019 [1301.1138]. 156, 157
- 4410 [149] M. Honda, T. Kajita, K. Kasahara, S. Midorikawa and T. Sanuki, *Calculation of*  
 4411 *atmospheric neutrino flux using the interaction model calibrated with atmospheric*  
 4412 *muon data*, *Phys. Rev. D* **75** (2007) 043006 [astro-ph/0611418]. 157
- 4413 [150] C. Green, J. Kowalkowski, M. Paterno, M. Fischler, L. Garren and Q. Lu, *The*  
 4414 *Art Framework*, *J. Phys. Conf. Ser.* **396** (2012) 022020. 160
- 4415 [151] H. Bethe, *Zur theorie des durchgangs schneller korpuskularstrahlen durch materie*,  
 4416 *Annalen der Physik* **397** (1930) 325. 165
- 4417 [152] E. Fermi, *The ionization loss of energy in gases and in condensed materials*,  
 4418 *Physical Review* **57** (1940) 485. 166
- 4419 [153] R. Sternheimer, M. Berger and S. Seltzer, *Density effect for the ionization loss of*  
 4420 *charged particles in various substances*, *Atomic Data and Nuclear Data Tables* **30**  
 4421 (1984) 261. 166
- 4422 [154] W.W.M. Allison and J.H. Cobb, *Relativistic Charged Particle Identification by*  
 4423 *Energy Loss*, *Ann. Rev. Nucl. Part. Sci.* **30** (1980) 253. 166

## BIBLIOGRAPHY

- 4424 [155] W. Blum, L. Rolandi and W. Riegler, *Particle detection with drift chambers*,  
4425 Particle Acceleration and Detection (2008), 10.1007/978-3-540-76684-1. 166
- 4426 [156] ALICE TPC collaboration, *Particle identification of the ALICE TPC via dE/dx*,  
4427 *Nucl. Instrum. Meth. A* **706** (2013) 55. 166
- 4428 [157] L. Landau, *On the energy loss of fast particles by ionization*, *J. Phys. (USSR)* **8**  
4429 (1944) 201. 169
- 4430 [158] W. Ulmer and E. Matsinos, *Theoretical methods for the calculation of Bragg*  
4431 *curves and 3D distributions of proton beams*, *The European Physical Journal*  
4432 *Special Topics* **190** (2010) 1. 170
- 4433 [159] E. Aprile, A.E. Bolotnikov, A.L. Bolozdynya and T. Doke, *Noble Gas Detectors*,  
4434 Wiley (Oct., 2008), 10.1002/9783527610020. 173
- 4435 [160] D.S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press. 210
- 4436 [161] ALICE collaboration, *Production of pions, kaons and protons in pp collisions at*  
4437  $\sqrt{s} = 900 \text{ gev with alice at the lhc}$ , **1101.4110**. 212
- 4438 [162] U. Einhaus, *Charged hadron identification with de/dx and time-of-flight at future*  
4439 *higgs factories*, **2110.15115**. 212
- 4440 [163] R. Frühwirth, *Application of filter methods to the reconstruction of tracks and*  
4441 *vertices in events of experimental high energy physics*, Ph.D. thesis, Technischen  
4442 Universität Wien, 1988. 217
- 4443 [164] P. Astier, A. Cardini, R.D. Cousins, A. Letessier-Selvon, B.A. Popov and  
4444 T. Vinogradova, *Kalman filter track fits and track breakpoint analysis*, *Nuclear*  
4445 *Instruments and Methods in Physics Research Section A: Accelerators*,  
4446 *Spectrometers, Detectors and Associated Equipment* **450** (2000) 138. 217
- 4447 [165] T2K UK collaboration, *The Electromagnetic Calorimeter for the T2K Near*  
4448 *Detector ND280*, *JINST* **8** (2013) P10019 [1308.3445]. 226

## BIBLIOGRAPHY

- 4449 [166] J.E. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the  $J/\psi$  and  $\psi'$ ,*  
4450 Ph.D. thesis, Stanford University, 1982. 231