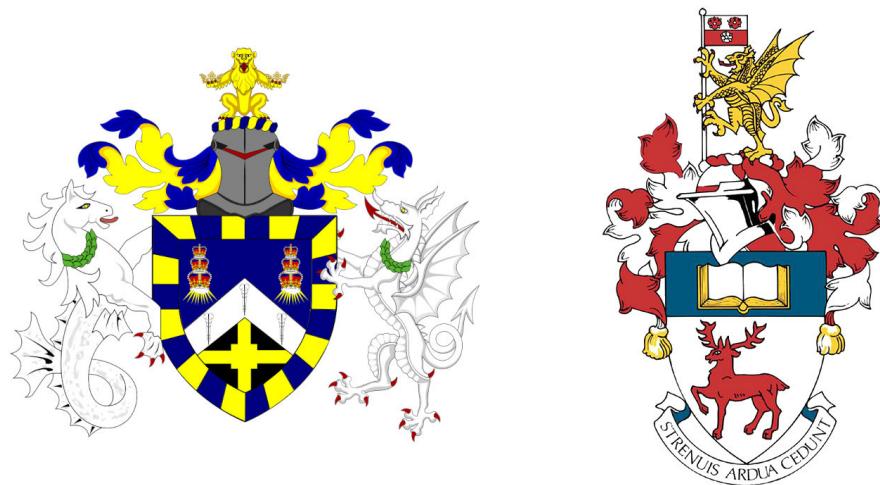


<sup>1</sup> ADVANCING NEUTRINO  
<sup>2</sup> DETECTION AND TRIGGERING IN  
<sup>3</sup> DUNE



<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



## <sup>13</sup> Statement of originality

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



# <sup>31</sup> Abstract

<sup>32</sup> Work in progress . . .



*¡Oh memoria, enemiga mortal de mi descanso!*

---

*El ingenioso hidalgo don Quijote de la Mancha*

MIGUEL DE CERVANTES SAAVEDRA



## <sup>33</sup> Acknowledgements

<sup>34</sup> Work in progress . . .



# <sup>35</sup> Contents

<sup>36</sup>	Statement of originality . . . . .	3
<sup>37</sup>	Abstract . . . . .	5
<sup>38</sup>	Acknowledgements . . . . .	9
<sup>39</sup>	<b>List of Figures</b>	<b>15</b>
<sup>40</sup>	<b>List of Tables</b>	<b>31</b>
<sup>41</sup>	<b>List of Abbreviations</b>	<b>33</b>
<sup>42</sup>	<b>1 Introduction</b>	<b>35</b>
<sup>43</sup>	<b>2 Neutrino physics</b>	<b>37</b>
<sup>44</sup>	2.1 Historical remarks . . . . .	38
<sup>45</sup>	2.2 Neutrinos in the SM . . . . .	38
<sup>46</sup>	2.3 Massive neutrinos . . . . .	41
<sup>47</sup>	2.4 Neutrino oscillations . . . . .	42
<sup>48</sup>	2.4.1 Oscillations in vacuum . . . . .	43
<sup>49</sup>	2.4.2 Oscillations in matter . . . . .	44
<sup>50</sup>	2.4.3 Current status of neutrino oscillations . . . . .	45
<sup>51</sup>	2.5 Open questions in the neutrino sector . . . . .	46
<sup>52</sup>	2.6 Neutrino interactions . . . . .	47
<sup>53</sup>	<b>3 The Deep Underground Neutrino Experiment</b>	<b>49</b>

## CONTENTS

54	3.1 Overview . . . . .	49
55	3.2 Physics goals of DUNE . . . . .	51
56	3.3 Far Detector . . . . .	53
57	3.3.1 Horizontal Drift . . . . .	54
58	3.3.2 Vertical Drift . . . . .	57
59	3.3.3 FD Data Acquisition System . . . . .	59
60	3.4 Near Detector . . . . .	60
61	3.4.1 ND-LAr . . . . .	62
62	3.4.2 TMS/ND-GAr . . . . .	63
63	3.4.3 PRISM . . . . .	64
64	3.4.4 SAND . . . . .	66
65	3.5 LBNF beamline . . . . .	66
66	<b>4 ND-GAr</b>	<b>69</b>
67	4.1 Requirements . . . . .	69
68	4.2 Reference design . . . . .	70
69	4.2.1 HPgTPC . . . . .	70
70	4.2.2 ECal . . . . .	71
71	4.2.3 Magnet . . . . .	72
72	4.2.4 Muon system . . . . .	73
73	4.3 GArSoft . . . . .	73
74	4.3.1 Event generation . . . . .	73
75	4.3.2 Detector simulation . . . . .	74
76	4.3.3 Reconstruction . . . . .	75
77	<b>5 FWTPG offline software</b>	<b>79</b>
78	<b>6 Matched Filter approach to induction wire Trigger Primitives</b>	<b>81</b>
79	6.1 Motivation . . . . .	81
80	6.2 Signal-to-noise ratio definition . . . . .	83

## CONTENTS

81	6.3	Low-pass FIR filter design . . . . .	85
82	6.4	Matched filters . . . . .	88
83	6.5	Using simulated samples . . . . .	94
84	6.5.1	Angular dependence . . . . .	100
85	6.5.2	Distortion and peak asymmetry . . . . .	102
86	6.5.3	Hit sensitivity . . . . .	105
87	<b>7</b>	<b>DM searches with neutrinos from the Sun</b>	<b>115</b>
88	7.1	Motivation . . . . .	115
89	7.2	Gravitational capture of DM by the Sun . . . . .	115
90	7.3	Neutrino flux from DM annihilations . . . . .	122
91	7.4	Computing limits from solar neutrino fluxes . . . . .	123
92	7.5	Example: Kaluza-Klein Dark Matter . . . . .	127
93	7.6	High energy DM neutrino signals . . . . .	131
94	7.6.1	DIS events . . . . .	133
95	7.6.2	Single proton QEL events . . . . .	138
96	7.6.3	Results . . . . .	141
97	7.7	Example: Leptophilic Dark Matter . . . . .	143
98	<b>8</b>	<b>Particle ID in GArSoft</b>	<b>149</b>
99	8.1	$dE/dx$ measurement in the TPC . . . . .	150
100	8.1.1	Energy calibration . . . . .	152
101	8.1.2	Truncated $dE/dx$ mean . . . . .	162
102	8.1.3	Mean $dE/dx$ parametrisation . . . . .	165
103	8.1.4	Particle identification . . . . .	169
104	8.2	Muon and pion separation in the ECal and MuID . . . . .	169
105	8.2.1	Track-ECal matching . . . . .	172
106	8.2.2	Classification strategy . . . . .	176
107	8.2.3	Feature selection and importance . . . . .	180
108	8.2.4	Hyperparameter optimisation . . . . .	191

## CONTENTS

109	8.2.5 Probability calibration . . . . .	195
110	8.2.6 Performance . . . . .	197
111	8.3 ECal time-of-flight . . . . .	197
112	8.3.1 Arrival time estimations . . . . .	199
113	8.3.2 Proton and pion separation . . . . .	200
114	8.4 Charged pion decay in flight . . . . .	203
115	8.4.1 Track breakpoints . . . . .	204
116	8.5 Neutral particle identification . . . . .	211
117	8.5.1 ECal clustering . . . . .	211
118	8.5.2 $\pi^0$ reconstruction . . . . .	214
119	<b>9 Event selection in ND-GAr</b>	<b>217</b>
120	9.1 CAFs and CAFAna . . . . .	217
121	9.2 Event selection . . . . .	217
122	9.2.1 $\nu_\mu$ CC selection . . . . .	217
123	9.2.2 Charged pion multiplicity . . . . .	217
124	<b>10 Conclusions</b>	<b>219</b>
125	<b>A An appendix</b>	<b>221</b>
126	<b>Bibliography</b>	<b>223</b>

# <sup>127</sup> List of Figures

<sup>128</sup> 3.1	Schematic diagram of the DUNE experiment and the LBNF beamline [1].	50
<sup>129</sup> 3.2	Schematic diagram showing the operating principle of a LArTPC with wire readout . . . . .	54
<sup>131</sup> 3.3	Proposed design for the FD-1 and FD-2 modules following the HD principle.	55
<sup>132</sup> 3.4	Schematic representation of an APA frames showing the U, V, X and G wires. . . . .	56
<sup>134</sup> 3.5	A PDS module containing 24 X-ARAPUCAs and the location of the modules on the APAs. . . . .	56
<sup>136</sup> 3.6	Proposed design for the FD-3 module following the VD principle. . . . .	57
<sup>137</sup> 3.7	Schematic representation of the electrode strip configuration for a top and bottom CRU. . . . .	58
<sup>139</sup> 3.8	Detailed diagram of the DUNE FD DAQ system. Figure taken from Ref. [2]. . . . .	59
<sup>141</sup> 3.9	Representation of the ND hall in Phase II, showing the different subcomponents.	61
<sup>142</sup> 3.10	Schematic representation of the external components of ND-LAr, including the cryostat and the PRISM movable system and detailed drawing of one ArgonCube module. . . . .	62
<sup>145</sup> 3.11	Schematic view of the TMS detector, highlighting its main parts. . . . .	63
<sup>146</sup> 3.12	Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet. . . . .	64

## LIST OF FIGURES

148	3.13 Predicted beam muon neutrino flux at the ND location for different off-axis	
149	positions. . . . .	65
150	3.14 Schematic longitudinal section of the LBNF beamline at Fermilab. . . . .	66
151	3.15 Predicted neutrino fluxes at the FD in FHC mode and RHC mode. . . . .	67
152	4.1 Diagram of the ALICE TPC, showing the two drift chambers, inner and	
153	outer field cages and readout chambers. . . . .	71
154	4.2 Diagram of the ALICE TPC, showing the two drift chambers, inner and	
155	outer field cages and readout chambers. . . . .	72
156	6.1 <i>Schematic representation of an APA. The black lines represent the APA</i>	
157	<i>steel frame. The green and magenta lines correspond to the direction of the</i>	
158	<i>U and V induction wires respectively. The blue lines indicate the direction</i>	
159	<i>of the X collection wires and the wire shielding G.</i> . . . . .	82
160	6.2 <i>Left panel: Zoomed unfiltered waveform corresponding to channel 7840</i>	
161	<i>from the ProtoDUNE-SP raw data capture felix-2020-07-17-21:31:44</i>	
162	<i>(blue line). The green dashed lines mark the region <math>\pm 3\sigma_{raw}</math>. The resulting</i>	
163	<i>noise waveform is also shown (red line). Top right panel: ADC distribution</i>	
164	<i>for channel 7840, where the green shaded region represents <math>\pm \sigma_{raw}</math>. Bottom</i>	
165	<i>right panel: noise ADC distribution for channel 7840, where the green</i>	
166	<i>shaded region represents <math>\pm \sigma_{noise}</math>.</i> . . . . .	83
167	6.3 <i>Left panel: Zoomed filtered waveform corresponding to channel 7840 from</i>	
168	<i>the ProtoDUNE-SP raw data capture felix-2020-07-17-21:31:44 (blue</i>	
169	<i>line). The filter used was the current implementation of the low-pass FIR</i>	
170	<i>filter in <code>dtp-firmware</code>. The green dashed lines mark the region <math>\pm 3\sigma_{raw}</math>.</i>	
171	<i>The resulting noise waveform is also shown (red line). Top right panel:</i>	
172	<i>ADC distribution for channel 7840 after filtering, where the green shaded</i>	
173	<i>region represents <math>\pm \sigma_{raw}</math>. Bottom right panel: noise ADC distribution for</i>	
174	<i>channel 7840 after filtering, where the green shaded region represents <math>\pm \sigma_{noise}</math></i>	85

## LIST OF FIGURES

175	6.4 Power spectrum in decibels for the current implementation of the low-pass	
176	FIR filter in <i>dtp-firmware</i> (blue line), compared to the response of an	
177	optimal filter obtained using the Parks-McClellan algorithm for the same	
178	pass-band (red line). Also for comparison I include the spectrum of the	
179	optimal filter when taking only the integer part of the coefficients (red	
180	dashed line) . . . . .	86
181	6.5 Relative change in the S/N for the ProtoDUNE-SP raw data capture	
182	<i>felix-2020-07-17-21:31:44</i> , using different values of the cutoff frequency	
183	$f_c$ and the transition width $\delta f$ . The optimal Chebyshev filters were applied	
184	using just the integer part of the coefficients given by the Parks-McClellan	
185	algorithm. . . . .	87
186	6.6 Distribution of the relative change of the S/N on the different wire planes	
187	from the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i>	
188	after the optimal Chebyshev filter was applied. The filter was computed	
189	with the Parks-McClellan algorithm using a cutoff of $f_c = 0.068 \text{ ticks}^{-1}$	
190	and a transition width $\delta f = 0.010 \text{ ticks}^{-1}$ . . . . .	88
191	6.7 Left panel: Zoomed match filtered waveform corresponding to channel 7840	
192	from the ProtoDUNE-SP raw data capture <i>felix-2020-07-17-21:31:44</i>	
193	(blue line). The filter used was directly extracted from the data, being	
194	the 32 values around the first peak in the original waveform. The green	
195	dashed lines mark the region $\pm 3\sigma_{\text{raw}}$ . The resulting noise waveform is	
196	also shown (red line). Top right panel: ADC distribution for channel 7840	
197	after match filtering, where the green shaded region represents $\pm \sigma_{\text{raw}}$ .	
198	Bottom right panel: noise ADC distribution for channel 7840 after match	
199	filtering, where the green shaded region represents $\pm \sigma_{\text{noise}}$ . . . . .	89
200	6.8 Relative improvement in the S/N for the raw data capture <i>felix-2020-07-17-21:31:44</i> ,	
201	using the matched filter following the parametrisation in Eq. (6.17). The	
202	black crosses in both panels denote the location of the maximum ratio value.	92

## LIST OF FIGURES

203	6.9 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. (6.17) 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 <i>felix-2020-07-17-21:31:44</i> after their respective optimal matched filters were applied. . . . .	95
210	6.10 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 $\pi^0 \rightarrow \gamma\gamma$ . . . . .	96
215	6.11 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. . . . .	98
223	6.12 Distributions of the mean $S/N$ improvement per event for the corresponding sample after applying the matched filters. Here I separated the change in the $U$ plane (blue) and the $V$ plane (red) channels. From top left to the right: muon, electron, proton and neutral pion. All the events have a fixed kinetic energy of $E_k = 100$ MeV. . . . .	99

## LIST OF FIGURES

## LIST OF FIGURES

248	6.16 Left panel: peak asymmetry distribution for the case of the monoenergetic	
249	$E_k = 100$ MeV muon sample. Each value corresponds to a single bipolar	
250	signal peak from a channel in any event. The blue distribution represents	
251	the peaks on $U$ plane channels, whereas the red corresponds to signal peaks	
252	in $V$ wires. Right panel: relation between the mean peak asymmetry per	
253	event with the $S/N$ for $U$ channel waveforms from the $E_k = 100$ MeV	
254	muon sample. The top subplot shows the decimal logarithm of the mean	
255	$S/N$ for the raw (red) and the matched filtered (blue) waveforms. The	
256	bottom subplot contains the mean $S/N$ improvement ratio after the matched	
257	filter was applied. . . . .	104
258	6.17 Raw data display in the plane time (in firmware ticks) vs. offline channel	
259	number for an $E_k = 100$ MeV electron event. The produced true hits are	
260	superimposed (black boxes) as well as the hits comming from the standard	
261	hit finder chain (blue circles) and the hit finder using the matched filter	
262	(green triangles). . . . .	106
263	6.18 Dependence of the precision (blue), sensitivity (red) and $F_1$ (green) scores	
264	on the threshold values used in the hit finder, for the FIR (left panel)	
265	and matched filter (right panel) cases. The results were obtained after	
266	matching the hits to the true hits in the case of the isotropic muon sample	
267	with kinetic energy in the range 5 to 100 MeV, taking only into account	
268	the induction plane channels. The points represent the mean value while	
269	the error bars indicate one standard deviation around that mean value. .	108

## LIST OF FIGURES

- 270    6.19 *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.* . . . . 110  
 271  
 272  
 273  
 274  
 275  
 276  
 277
- 278    6.20 *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.* . . . . . 111  
 279  
 280  
 281
- 282    6.21 *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.* 112  
 283  
 284  
 285  
 286
- 287    7.1 *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 [3].* . . . . . 119  
 288  
 289  
 290  
 291
- 292    7.2 *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. [4]. All the rates are shown for a choice of scattering cross section of  $\sigma_i = 10^{-40} \text{ cm}^2$ .* . . . 120  
 293  
 294  
 295  
 296

## LIST OF FIGURES

297	7.3 <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> . . . . .	124
302	7.4 <i>Expected atmospheric neutrino flux as a function of the neutrino energy <math>E_\nu</math> at Homestake at solar minimum, taken from Ref. [5]. The blue solid (dashed) line correspond to muon neutrinos (antineutrinos) and the red solid (dashed) line correspond to electron neutrinos (antineutrinos)</i> . . . .	126
306	7.5 Feynman diagrams for $B^1B^1$ annihilation into SM fermions. . . . .	128
307	7.6 Feynman diagrams for $B^1B^1$ annihilation into a Higgs boson pair. . . .	128
308	7.7 <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> . . . . .	129
311	7.8 <i>Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin- dependent <math>B^1</math>-proton scattering cross section as a function of <math>M_{\text{LKP}}</math> (green dots). I also show the previous limits from IceCube [6] (blue line) and Antares [7] (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 [8]</i> . . . . .	130
317	7.9 <i>Computed spectra of muon neutrinos at the DUNE FD site from <math>\tau^+\tau^-</math> (left panel) and <math>b\bar{b}</math> (right panel) annihilations in the Sun for the DM masses <math>m_{\text{DM}} = 10</math> GeV (red line), 50 GeV (green line) and 100 GeV (blue line), plotted in relative energy units.</i> . . . . .	131
321	7.10 <i>Distribution of the muon neutrino energies from the <math>\tau^+\tau^-</math> (left panel) and <math>b\bar{b}</math> (right panel) annihilation channels, for <math>m_{\text{DM}} = 10</math> GeV, separated by CC interaction type: QEL (blue), MEC (orange), RES (green) and DIS (red)</i> . . . . .	133

## LIST OF FIGURES

<p>325    7.11 <i>Distributions of <math>\theta_\mu</math> (left panel), <math>\theta_j</math> (central panel) and <math>\theta_{plane}</math> (right panel)</i>            326    <i>for the <math>b\bar{b}</math> sample with <math>m_{DM} = 10</math> GeV (blue) and the atmospheric</i>            327    <i>background (red).</i> . . . . .</p> <p>328    7.12 <i>Left panel: signal efficiencies (blue lines) and background rejections (red</i>            329    <i>lines) for events passing the cuts <math>\theta &lt; \theta_{cut}</math> for the jet (solid lines) and</i>            330    <i>muon (dashed lines) angles. Right panel: signal efficiency (blue line) and</i>            331    <i>background rejection (red line) for events passing the cut <math>\theta_{plane} &lt; \theta_{cut}</math> for</i>            332    <i>the momentum conservation plane deviation.</i> . . . . .</p> <p>333    7.13 <i>Signal efficiencies for the <math>\tau^+\tau^-</math> (blue line) and <math>b\bar{b}</math> (red line) DIS samples</i>            334    <i>as functions of the DM mass, <math>m_{DM}</math>, obtained by applying the optimal</i>            335    <i>angular cuts <math>\theta_\mu &lt; 27^\circ</math>, <math>4^\circ &lt; \theta_j &lt; 26^\circ</math> and <math>\theta_{plane} &lt; 3.5^\circ</math>.</i> . . . . .</p> <p>336    7.14 <i>Distributions of <math>\cos \theta_\mu</math> (left panel), <math>\cos \theta_p</math> (central panel) and <math>\cos \theta_N</math></i>            337    <i>(right panel) for the <math>\tau^+\tau^-</math> QEL sample with <math>m_{DM} = 5</math> GeV (blue) and</i>            338    <i>the atmospheric background (red).</i> . . . . .</p> <p>339    7.15 <i>Left panel: value of the loss function for the training sample (blue line)</i>            340    <i>and accuracy for the validation sample (red line) versus the number of</i>            341    <i>iterations for the MLP classifier training. Right panel: distributions of the</i>            342    <i>predicted probabilities assigned by the MLP classifier to the test sample</i>            343    <i>for the <math>\tau^+\tau^-</math> QEL signal with <math>m_{DM} = 5</math> GeV (blue) and the atmospheric</i>            344    <i>background (red).</i> . . . . .</p> <p>345    7.16 <i>Signal efficiencies for the <math>\tau^+\tau^-</math> (blue line) and <math>b\bar{b}</math> (red line) single proton</i>            346    <i>QEL samples as functions of the DM mass, <math>m_{DM}</math>, obtained by requiring a</i>            347    <i>minimum predicted probability from the MLP classifier of 0.97 in order to</i>            348    <i>achieve a background rejection greater than 99.8%.</i> . . . . .</p>	<p>134</p> <p>135</p> <p>137</p> <p>138</p> <p>139</p> <p>140</p>
---	---

## LIST OF FIGURES

349	7.17 Projected 90% confidence level upper limit for DUNE (400 kT yr) on 350 the spin-dependent DM-nucleon scattering cross section as a function of 351 $m_{\text{DM}}$ , for the annihilation channels $\tau^+\tau^-$ (blue) and $b\bar{b}$ (red) separated 352 by interaction type (up triangles denote DIS interactions whereas down 353 triangles represent QEL interactions). I also show the previous limits 354 from IceCube [9] (solid lines) and the projected sensitivities for Pingu [10] 355 (dashed lines) and Hyper-Kamiokande [11] (dash-dotted lines), as well 356 as the direct detection limits from PICASSO [12] (solid green line) and 357 PICO-60 C <sub>3</sub> F <sub>8</sub> [13] (dashed green line). . . . .	142
358	7.18 Left panel: Projected 90% confidence level sensitivity of DUNE (400 kT 359 yr) to the scale $\Lambda$ of an EFT containing only leptophilic DM axial-axial 360 interactions (blue line). Right panel: . In both cases the corresponding 361 limits from DarkSide-50 [14] (dotted green line) and XENON1T [15] 362 (dashed red line) are also shown, together with the configurations for which 363 the correct relic density is achieved (black line), all for the coupling values 364 $c_A^e = 10^3$ and $c_A^\nu = 10^{-2}$ . . . . .	146
365	8.1 Distribution of the fraction of energy deposits with residual range less 366 than 20% of the total track length, and distribution of the ionisation per 367 unit length after removing the tracks with less than 30% of their energy 368 deposits in the last 20% of the track. . . . .	153
369	8.2 Distribution of the reconstructed ionisation charge per unit length for 370 different reclustering values, and distribution of the median change in 371 $dQ/dx$ per track for the $N_{\text{group}} = 4$ reclustering. . . . .	154
372	8.3 Distribution of the Geant4-simulated energy losses per unit length versus 373 residual range for the stopping proton sample. . . . .	156
374	8.4 Fitted most probable $dQ/dx$ values for each $dE/dx$ bin, together with 375 best fit to the logarithmic calibration function. . . . .	157

## LIST OF FIGURES

376	8.5 Fitted most probable $dQ/dx$ values for each $dE/dx$ bin for three different	
377	ADC bit limits. . . . .	159
378	8.6 Area normalised $dE/dx$ distributions for the true and the reconstructed	
379	energy deposits in the stopping proton sample, both after applying the	
380	calibration and the calibration and the normalisation correction. . . . .	160
381	8.7 Fractional residuals between the true and the corrected $dE/dx$ means	
382	and the 60% truncated means, and fractional residuals between the true	
383	and the uncorrected, corrected and uncalibrated $dE/dx$ 60% truncated	
384	means. . . . .	162
385	8.8 Estimated values of the mean $dE/dx$ bias and resolution for the stopping	
386	proton sample at different values of the truncation factor. . . . .	163
387	8.9 Examples of the truncated mean $dE/dx$ LanGauss fits for various $\beta\gamma$	
388	bins, from a simulated FHC neutrino sample. . . . .	165
389	8.10 Resulting one and two dimensional projections of the posterior probability	
390	distributions of the ALEPH $\langle dE/dx \rangle$ parameters obtained by fitting the	
391	60% truncated mean $dE/dx$ values from a FHC neutrino sample. . . . .	166
392	8.11 Truncated mean $dE/dx$ obtained for the FHC neutrino sample as a	
393	function of the $\beta\gamma$ product, together with the fitted most probable values	
394	for each $\beta\gamma$ bin and the best fit obtained using the ALEPH parametrisation.	167
395	8.12 Distribution of the 60% truncated mean $dE/dx$ versus reconstructed	
396	momentum for the FHC neutrino sample. . . . .	168
397	8.13 Estimated values of the mean $dE/dx$ bias and resolution obtained for the	
398	true protons in a FHC neutrino sample. . . . .	169
399	8.14 True momentum distribution for the primary muon in $\nu_\mu$ CC $N\pi^\pm$	
400	interactions inside the fiducial volume of ND-GAr, compared to the	
401	post FSI charged pion spectrum. . . . .	170
402	8.15 Distributions of energy deposits in the ECal for a muon and a charged	
403	pion with similar momenta. . . . .	171

## LIST OF FIGURES

404	8.16 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). . . . .	173
410	8.17 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, $t_{cluster}$ . . . . .	175
414	8.18 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. . . . .	177
419	8.19 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. . . . .	178
423	8.20 Example ECal feature distributions for muons and charged pions in the five different momentum ranges considered. . . . .	182
425	8.21 Example MuID feature distributions for muons and charged pions in the three different momentum ranges considered. . . . .	183
427	8.22 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 $0.3 \leq p < 0.8 \text{ GeV}/c$ . . . . .	185

## LIST OF FIGURES

433	8.23 Left panel: cumulative explained variance for the first three principal	
434	components (top panel) and contribution of the different features to the	
435	principal axes in feature space (bottom panel). Right panel: Shapley	
436	(blue) and Gini (red) feature importances for the different input features.	
437	Both figures correspond to the samples in the momentum range $0.8 \leq$	
438	$p < 1.5 \text{ GeV}/c$ . . . . .	186
439	8.24 Evolution of the SHAP importance for the top six most important features	
440	across all five momentum ranges. . . . .	188
441	8.25 Permutation importances for the ten most important features in the	
442	different momentum ranges (from left to right, top to bottom, in increasing	
443	momentum order). The bars indicate the effect that permutations of	
444	each feature have on the purity (blue) and the sensitivity (yellow), the	
445	translucent regions representing one standard deviation around the central	
446	value. . . . .	190
447	8.26 Values of the precision and sensitivity obtained for 10000 BDT hyperparameter	
448	configurations, for the momentum regions I, III and V. . . . .	193
449	8.27 Reliability diagrams for the BDT classifier used in the momentum range	
450	$0.3 \leq p < 0.8 \text{ GeV}/c$ , both for the original (blue circles) and calibrated	
451	(yellow squares) responses. For reference, the response of a perfectly	
452	calibrated classifier is also shown (black dashed line). . . . .	195
453	8.28 Uncalibrated (left panel) and calibrated (right panel) predicted probabilities	
454	assigned by the BDT classifiers for true muons (blue) and charged pions	
455	(red) in the momentum range $0.3 \leq p < 0.8 \text{ GeV}/c$ . . . . .	196
456	8.29 Schematic of the hit selection used for the ToF measurement. The grid	
457	represents the layers of the inner ECal, with coloured squares indicating	
458	the tiles with hits. Green squares indicate the selected hits. . . . .	198
459	8.30 Particle velocity versus momentum measured with different ECal arrival	
460	time estimations. From left to right: earliest hit time, average hit time,	
461	and fitted hit time. In all cases the time resolution is $\Delta\tau = 0.1 \text{ ns}$ . . . . .	200

## LIST OF FIGURES

462	8.31 Mass spectra for $p$ (blue) and $\pi^\pm$ (yellow) particles, using different ECal 463 time resolution values (from top to bottom, in ascending order), and 464 arrival time estimates. From left to right: earliest hit time, average hit 465 time, and fitted hit time. The dashed lines indicate the true masses of 466 the particles. . . . .	201
467	8.32 Efficiency (top panel) and purity (bottom panel) for the proton selection 468 as a function of the momentum, for $\Delta\tau = 0.10$ ns. . . . .	202
469	8.33 Distributions of the velocities measured by ToF with the inner ECal, for 470 different momentum bins, in a FHC neutrino interaction sample. The 471 Gaussian fits are performed around the maxima for each particle species.	203
472	8.34 Left panel: number of non-decaying, decaying and decaying in the fiducial 473 volume pions for a MC sample of 100000, $p = 500$ MeV/c isotropic 474 positively charged pions inside the TPC. Right panel: event display 475 for a positive pion decaying inside the fiducial volume, with a single 476 reconstructed track for the pion and muon system. . . . .	204
477	8.35 Values of $\chi_k^{2(FB)}$ (top left panel), $F_k$ (top right panel), $D_k^{1/R}$ (bottom 478 left panel) and $D_k^\phi$ (bottom right panel) versus position along the drift 479 direction for a reconstructed track in a positive pion decay event. The 480 vertical red dashed line indicates the true location of the decay point. . .	206
481	8.36 Fractional residual distributions of the true and reconstructed decay 482 position along the drift coordinate, using the position of the maximum 483 of $\chi_k^{2(FB)}$ (left panel) and $F_k$ (right panel) as estimates of the decay 484 position. Also shown are double Gaussian fits to these points (red lines).	207
485	8.37 Distributions of the extreme values of $\chi_k^{2(FB)}$ (top left panel), $F_k$ (top 486 right panel), $D_k^{1/R}$ (bottom left panel) and $D_k^\phi$ (bottom right panel) for 487 non-decaying reconstructed pion tracks (blue) and tracks which include 488 the decay inside the fiducial volume (red). . . . .	209

## LIST OF FIGURES

489	8.38 Left panel: distributions of the predicted probabilities assigned by the	
490	BDT classifier to a test sample of decaying pion+muon tracks (blue) and	
491	non-decaying pion tracks (red). Left: signal efficiency versus background	
492	acceptance (ROC curve) obtained from the BDT for the test sample. . .	210
493	8.39 Left panel: dependence of the decay position finding resolution on the	
494	true value of the decay angle for the $\chi_k^2(FB)$ (red) and $F_k$ (blue) methods.	
495	Right panel: signal efficiency (blue line) and background rejection (red	
496	line) from the BDT classifier versus true decay angle. . . . .	210
497	8.40 Mean values of the $F_1$ -score marginal distributions for the different	
498	free parameters of the new clustering algorithm, with the error bars	
499	representing one standard deviation around the mean. The $F_1$ -score	
500	values were computed for the 6561 possible parameter configurations	
501	using 1000 $\nu_\mu$ CC interaction events. . . . .	213
502	8.41 Left panel: distributions of the number of ECal clusters per photon from	
503	$\pi^0$ decays for the standard (red) and new (blue) clustering algorithms.	
504	Right panel: reconstructed invariant mass distributions for photon pairs	
505	from single $\pi^0$ events using the standard (red) and new (blue) ECal	
506	clustering algorithms. . . . .	215



# List of Tables

508	2.1	Values of $T_3$ and $Y/2$ assigned to the first generation of fermions. . . . .	39
509	2.2	Neutral current couplings. . . . .	40
510	2.3	Summary of neutrino oscillation parameters determined in the Neutrino	
511		Global Fit of 2020 [16]. . . . .	46
512	3.1	Summary of the two-phased plan for DUNE . . . . .	51
513	3.2	Exposure and time required to achieve the different physics milestones of	
514		the two phases . . . . .	52
515	6.1	<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> . . . . .	103
519	8.1	Calibration parameters obtained from the fit of the ND-GAr simulated stopping proton sample to the calibration function, for different ADC limits. . . . .	159
520			
521	8.2	Momentum ranges and description of the PID approach assumed for the muon and pion classification task. . . . .	179
522			
523	8.3	Optimal values of the hyperparameters used by the BDT, for each momentum range. . . . .	194
524			
525	8.4	Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges. . . . .	194
526			

## LIST OF TABLES

527	8.5 Summary of parameters and sampled values used in the optimisation of	
528	the clustering algorithm. . . . .	213

529

<sup>530</sup> **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.



<sup>531</sup> Chapter 1

<sup>532</sup> Introduction



533 Chapter 2

534 Neutrino physics

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

538                          – Terry Pratchett, *Sourcery*

539        Ever since they were postulated in 1930 by Wolfgang Pauli to explain the continuous  
540         $\beta$  decay spectrum [17] and later found by F. Reines and C. Cowan at the Savannah  
541        River reactor in 1953 [18], neutrinos have had a special place among all other elementary  
542        particles. They provide a unique way to probe a wide range of quite different physics,  
543        from nuclear physics to cosmology, from astrophysics to colliders. Moreover, there is  
544        compelling evidence to believe that the study of neutrinos may be key to unveil different  
545        aspects of physics beyond the SM, difficult to test elsewhere.

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

## Chapter 2. Neutrino physics

### 549 2.1 Historical remarks

### 550 2.2 Neutrinos in the SM

551 The SM of fundamental interactions was initially proposed in 1967 by S. Glashow, S.  
552 Weinberg and A. Salam[19–21]. This theoretical framework describes the dynamics of  
553 leptons and quarks, by introducing one scalar particle, the Higgs boson. It assumes  
554 that the local  $SU(3) \times SU(2)_L \times U(1)_Y$  gauge symmetry is an internal symmetry of the  
555 system, with  $SU(3)$  describing quantum chromodynamics, and  $SU(2)_L \times U(1)_Y$  being the  
556 gauge groups of the electroweak sector. For a detailed overview of the SM of electroweak  
557 interactions, see Ref. [22].

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

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

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

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

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

570 Left and right-handed fermions transform differently under  $SU(2)_L \times U(1)_Y$  rotations,

## 2.2. 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

as the right-handed particles are singlets under  $SU(2)_L$ . Applying a local transformation,  
they change as:

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

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

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

$$Q = T_3 + \frac{Y}{2},\tag{2.4}$$

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

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

## 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

given by:

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

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, respectively, and  $g$  and  $g'$  are the corresponding gauge couplings. It can be shown that these fields transform in the adjoint representation of the gauge group.

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:

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

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

### 2.3. Massive neutrinos

604 and the electric charge:

$$e = g' \cos \theta_W = g \sin \theta_W. \quad (2.7)$$

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

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

608 with the currents defined as:

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

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

612 As seen in Eq. (7.8), in the electroweak theory neutrinos are coupled to the  $Z$  boson  
 613 in a universal way. Therefore, by measuring the so-called invisible decay width of the  $Z$   
 614 boson we have an estimate of the number of light (i.e. lighter than the  $Z$  boson) neutrino  
 615 flavours. This number was measured by LEP in a combined analysis of  $e^+ e^- \rightarrow \mu^+ \mu^-$   
 616 and  $e^+ e^- \rightarrow \text{hadrons}$  to be  $N_\nu = 2.9840 \pm 0.0082$  [23].

### 617 2.3 Massive neutrinos

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

## Chapter 2. Neutrino physics

introducing an arbitrary number of sterile neutrinos,  $N_i$ ,  $i = 1, \dots, m$ .

$$-\mathcal{L}_{M_\nu} = M_D^{ij} \bar{N}_i \nu_{Lj} + \frac{1}{2} M_N^{ij} \bar{N}_i N_j^c + \text{h.c.}, \quad (2.10)$$

where  $M_D$  is a complex  $m \times 3$  matrix and  $M_N$  a complex, 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, the other fermions. The second term, Majorana mass term, is allowed in the Lagrangian,singlet of the gauge group. However, it violates lepton number conservation by two units.

$$-\mathcal{L}_{M_\nu} = \frac{1}{2} (\bar{\nu}_L^c \quad \bar{N}) \begin{pmatrix} 0 & M_D^T \\ M_D & M_N \end{pmatrix} \begin{pmatrix} \nu_L \\ N^c \end{pmatrix} + \text{h.c.} = \bar{\nu}^c M_\nu \nu + \text{h.c.}, \quad (2.11)$$

## 2.4 Neutrino oscillations

The evidence for neutrino oscillation [24], and therefore the existence of non-zero neutrino masses, constitutes one of the groundbreaking discoveries of modern Physics and has acted as driving force for Beyond the Standard Model (BSM) Physics. The minimal extension of the Standard Model (SM) we can do to address these phenomena is introducing distinct 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  $\nu_\tau$ .

The way to relate these two sets of neutrino eigenstates is via a  $3 \times 3$  unitary matrix, called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [25, 26], as:

$$|\nu_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle, \quad (2.12)$$

where the Greek index  $\alpha$  denotes the flavour  $\{e, \mu, \tau\}$  and the Latin index  $i$  the associated masses  $\{1, 2, 3\}$ . This leptonic mixing matrix may be parametrized in terms of 6 parameters, 3 of which are mixing angles  $\theta_{12}$ ,  $\theta_{13}$  and  $\theta_{23}$ , one CP-violating phase  $\delta_{CP}$

## 2.4. Neutrino oscillations

and 2 Majorana phases  $\alpha$  and  $\beta$ :

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha} & 0 \\ 0 & 0 & e^{i\beta} \end{pmatrix}, \quad (2.13)$$

where  $c_{ij} \equiv \cos \theta_{ij}$  and  $s_{ij} \equiv \sin \theta_{ij}$ . This matrix is analogous to the Cabibbo-Kobayashi-Maskawa (CKM) matrix in the quark sector. If neutrinos are Dirac fermions, we can drop the Majorana phases in the PMNS matrix. But, in any case, these phases play no role on the neutrino oscillations.

### 2.4.1 Oscillations in vacuum

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

$$|\nu_\alpha(t)\rangle = \sum_{i=1}^3 U_{\alpha i}^* e^{-iE_i t} |\nu_i(t=0)\rangle, \quad (2.14)$$

as the mass eigenstates are also eigenstates of the free Hamiltonian. Now, if we express the mass eigenstates as a superposition of flavour eigenstates, the last expression can be rewritten as:

$$|\nu_\alpha(t)\rangle = \sum_{i=1}^3 U_{\beta i} e^{-iE_i t} U_{\alpha i}^* |\nu_\beta\rangle. \quad (2.15)$$

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

$$P(\nu_\alpha \rightarrow \nu_\beta) = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 = \left| \sum_{i=1}^3 U_{\beta i} e^{-iE_i t} U_{\alpha i}^* \right|^2. \quad (2.16)$$

A usual approximation to take at this point is to consider ultra-relativistic neutrinos, i.e.  $E \approx |\vec{p}|$ , so we can write the dispersion relations as:

$$E_i = \sqrt{p^2 + m_i^2} \approx E + \frac{m_i^2}{2E}, \quad (2.17)$$

## Chapter 2. Neutrino physics

655 so we can write the oscillation probability as:

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* e^{-i \frac{\Delta m_{ij}^2}{2E} t} \\ &= \delta_{\alpha\beta} - 4 \sum_{i < j} \Re [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin^2 \left( \frac{\Delta m_{ij}^2}{4E} t \right) \\ &\quad + 2 \sum_{i < j} \Im [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin \left( \frac{\Delta m_{ij}^2}{2E} t \right), \end{aligned} \quad (2.18)$$

656 where  $\Delta m_{ij}^2$  is the difference of the squared masses of the  $j$ th and  $i$ th neutrino mass  
 657 eigenvalues. At this point, it is usual to write the phase responsible for the oscillations  
 658 as (under the approximate assumption  $t \approx L$ ):

$$\Delta_{ij} \equiv \frac{\Delta m_{ij}^2}{4E} L \simeq 1.27 \frac{\Delta m_{ij}^2}{(\text{eV}^2)} \frac{L}{(\text{km})} \frac{(\text{GeV})}{E}. \quad (2.19)$$

659 Notice that, in the case of antineutrinos the only difference would be the sign of the  
 660 last term in the oscillation probability. This way, one can write the CP asymmetry as:

$$\begin{aligned} A_{CP}^{\alpha\beta} &= P(\nu_\alpha \rightarrow \nu_\beta) - P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) \\ &= 4 \sum_{i < j} \Im [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin 2\Delta_{ij}. \end{aligned} \quad (2.20)$$

### 661 2.4.2 Oscillations in matter

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

668 The first proposed model to account for neutrino oscillations in matter was proposed  
 669 by Mikhaev, Smirnov and Wolfenstein (MSW) [27]. It relies on the fact that, as the  
 670 only charged lepton present in ordinary matter is the electron, electron neutrinos can

## 2.4. Neutrino oscillations

undergo both charged and neutral-current interactions with matter whereas for muon and tau neutrinos just neutral currents are possible.

### 2.4.3 Current status of neutrino oscillations

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

**Solar neutrino experiments** detect neutrinos produced in thermonuclear reactions inside the Sun, mainly from the so-called *pp* chain and the CNO cycle. These neutrinos have a typical energy in the range from 0.1 to 20 MeV. These experiments (Homestake [28], GALLEX [29], SAGE [30], Borexino [31], Super-Kamiokande [32] and SNO [33]) provide the best sensitivities to  $\theta_{12}$  and  $\Delta m_{21}^2$ .

**Atmospheric neutrino experiments** detect the neutrino flux produced when cosmic rays scatter with particles in Earth's atmosphere. These collisions generate particle showers that eventually produce electron and muon neutrinos (and antineutrinos). Their energies range from few MeV to about  $10^9$  GeV. Experiments, like Super-Kamiokande [34] and IceCube [35] use atmospheric neutrinos to measure oscillations and are specially sensitive to  $\theta_{23}$  and  $\Delta m_{32}^2$ .

**Reactor neutrino experiments** look for the  $\bar{\nu}_e$  spectrum produced by nuclear reactors, with energies in the MeV scale. Depending on the distance to the source, long-baseline experiments like KamLAND [36] are sensitive to the solar mass splitting  $\Delta m_{21}^2$  whereas much shorter baseline experiment such as RENO [37] or DayaBay [38] 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

---

<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.

## Chapter 2. Neutrino physics

**Table 2.3:** Summary of neutrino oscillation parameters determined in the Neutrino Global Fit of 2020 [16].

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

697 configuration one can obtain a beam made of mostly neutrinos or antineutrinos. The  
 698 typical energies of these neutrinos are in the GeV range. Experiments such as NOvA  
 699 [39], T2K [40], MINOS [41], OPERA [42] and K2K [43] (and in the future DUNE [44])  
 700 are primarily sensitive to  $\theta_{13}$ ,  $\theta_{23}$  and  $\Delta m_{32}^2$ . Also, in the coming years DUNE [44] and  
 701 Hyper-Kamiokande [45] will be sensitive to  $\delta_{CP}$ .

## 702 2.5 Open questions in the neutrino sector

703 A crucial question that remains open these days, and is of vital importance for oscillation  
 704 phenomena, is whether the mass eigenvalue  $\nu_3$  is the heaviest (what we call normal  
 705 ordering) or the lightest (referred to as inverted ordering) of the mass eigenstates. In  
 706 other words, this means that we do not know the sign of  $\Delta m_{32}^2$ , so we can either have  
 707  $m_1 < m_2 < m_3$  (NO) or  $m_3 < m_1 < m_2$  (IO).

708 Another big puzzle is related to the value of  $\delta_{CP}$ . Nowadays it is poorly constrained,  
 709 with all values between  $\pi$  and  $2\pi$  being consistent with data. A prospective measurement  
 710 different from  $\delta_{CP} = 0, \pi$  will predict CP-violation in the leptonic sector, and thus  
 711 contribute along with the one measured in the quark sector to the total amount of

## 2.6. Neutrino interactions

712 CP-violation. Although it is true that these two contributions by themselves are not  
713 enough to explain the matter anti-matter asymmetry in our universe, the amount of  
714 CP-violation in the leptonic sector can be key to explain such imbalance.

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

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

723 Other open question concerns the nature itself of the neutrinos. If neutrinos are Dirac  
724 particles then their mass term can be generated through the usual Higgs mechanism  
725 by adding right-handed neutrino fields. However, if they are Majorana particles and  
726 therefore their own antiparticles, there is no need to add extra fields to have the mass  
727 term in the Lagrangian. Experiments like SuperNEMO [46], SNO+ [47] and NEXT  
728 [48], which search for neutrino-less double beta decay, will be able to determine whether  
729 neutrinos are Dirac or Majorana.

## 730 2.6 Neutrino interactions



731 **Chapter 3**

732 **The Deep Underground Neutrino  
733 Experiment**

734 The Deep Underground Neutrino Experiment (DUNE) is a next generation long-baseline  
735 neutrino experiment [1]. It will aim to address several questions in neutrino physics,  
736 study neutrinos from astrophysical sources and search for beyond the standard model  
737 physics.

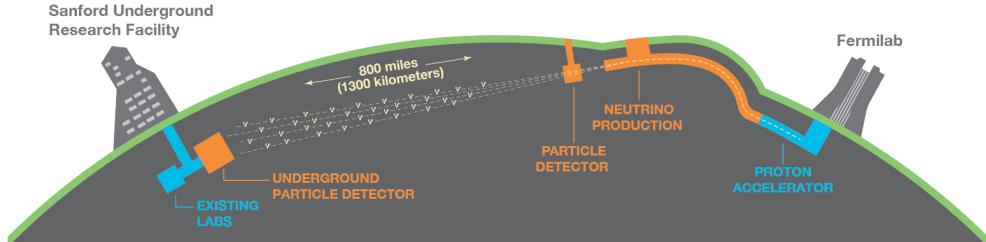
738 This chapter reviews the main goals of the DUNE experiment, the design of the far  
739 detector modules and their data acquisition (DAQ) system, and the role that the near  
740 detector plays in the physics program of DUNE.

741 **3.1 Overview**

742 The main physics goals of DUNE are:

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

## Chapter 3. The Deep Underground Neutrino Experiment



**Figure 3.1:** Schematic diagram of the DUNE experiment and the LBNF beamline [1].

747        The design of DUNE has bee tailored with these goals in mind. It will consist  
748        of two neutrino detectors. A near detector (ND) complex will be placed in Fermilab,  
749        574 m downstream of the neutrino production point, whereas a larger far detector (FD)  
750        will be built in the Sandford Underground Research Facility (SURF), South Dakota,  
751        approximately 1300 km away. Fig. 3.1 shows a simplified view of the various components  
752        of DUNE (not to scale).

753        The beam neutrinos to be used in DUNE will be provided by the LBNF beamline,  
754        the multi-megawatt wide-band neutrino beam planned for Fermilab. First, an intense  
755        proton beam is extracted from the Fermilab Main Injector. Then, these protons with  
756        energies between 60 GeV and 120 GeV collide with a high-power production target and  
757        produce charged mesons. Two magnetic horns allow to focus the mesons and perform a  
758        sign selection (thus having the capability to switch between neutrino and antineutrino  
759        mode). Soon after that, the mesons decay and produce neutrinos (or antineutrinos)  
760        which are then aimed to SURF.

761        Before arriving to the FD, the neutrino beam meets the ND complex, which serves as  
762        the experiment's control. Its role is to measure the unoscillated neutrino energy spectra.  
763        From these we can predict the unoscillated spectra at the FD, which can be compared to  
764        the spectra measured at the FD in order to extract the oscillation parameters. Therefore,  
765        the design of the DUNE ND is mainly driven by the needs of the oscillation physics  
766        program.

767        The liquid Argon time projection chamber (LArTPC) technology has been chosen for

## 3.2. Physics goals of DUNE

**Table 3.1:** Summary of the two-phased plan for DUNE. Adapted from Ref. [49].

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

the FD modules of DUNE. Its four modules will record neutrino interactions from the accelerator-produced beam arriving at predictable times. As it also aims at recording rare events, the FD requires trigger schemes which can deal with both kinds of physics, and also maximum uptime.

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

A summary of the DUNE science program can be found in the DUNE FD Technical Design Report (TDR) Volume I [1]. For a detailed discussion on the two-phased approach the reader is referred to the DUNE Snowmass 2021 report [49].

## 3.2 Physics goals of DUNE

As noted in the literature (see for instance Ref. [16] for a review), the parameter space of the neutrino oscillation phenomena within the three-flavour picture is quite constrained by current experimental data. However, there are still crucial open questions, like the mass ordering, the value of  $\delta_{CP}$  or the  $\theta_{13}$  octant. One of the main goals of DUNE is to shed some light on the values of these parameters [50].

To address these questions DUNE can look to the subdominant oscillation channel  $\nu_\mu \rightarrow \nu_e$  ( $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ) and study the energy dependence of the  $\nu_e$  ( $\bar{\nu}_e$ ) appearance probability.

## Chapter 3. The Deep Underground Neutrino Experiment

**Table 3.2:** Exposure and time required to achieve the different physics milestones of the two phases. The predictions assume a Phase II staging scenario where FD modules 3 and 4 are deployed in years 4 and 6 and both the beam and ND are upgraded after 6 years. Adapted from Ref. [49].

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

791 When we focus on the antineutrino channel  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  there is a change in the sign of  $\delta_{CP}$ ,  
 792 thus introducing CP-violation. Moreover, due to the fact that there are no positrons in  
 793 the composition of Earth, there is a sign difference for the matter effect contribution  
 794 when looking to the antineutrino channel. This asymmetry is proportional to the baseline  
 795 length  $L$  and is sensitive to the sign of  $\Delta_{31}$ , and thus to the neutrino mass ordering.

796 Another of the main physics goals of DUNE is the search for baryon-number violating  
 797 processes. Specifically, it will try to answer the question of whether protons are stable  
 798 or not. There is no symmetry argument that forbids protons from decaying, but its  
 799 apparent stability seems to suggest that baryon number is conserved [51]. However,  
 800 proton decay is a usual feature of grand-unified theories, where electromagnetic, weak  
 801 and strong interactions are unified above a certain energy scale [52].

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

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

### 3.3. Far Detector

811 events from the 1987A supernova located in the Magellanic Cloud, 50 kpc away from  
812 Earth [53, 54].

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

819 Moreover, due to the stringent requirements that the main physics goals set for  
820 DUNE, it will allow also to perform searches for all kind of BSM physics. Among  
821 others, DUNE will be able to look for: active-sterile neutrino mixing, non-unitarity of  
822 the PMNS matrix, non-standard interactions, Lorentz and CPT violations, neutrino  
823 trident production, light-mass DM, boosted DM and heavy neutral leptons. The reader  
824 is referred to the DUNE FD TDR Volume II [50] for a full discussion of the physics  
825 scope of DUNE.

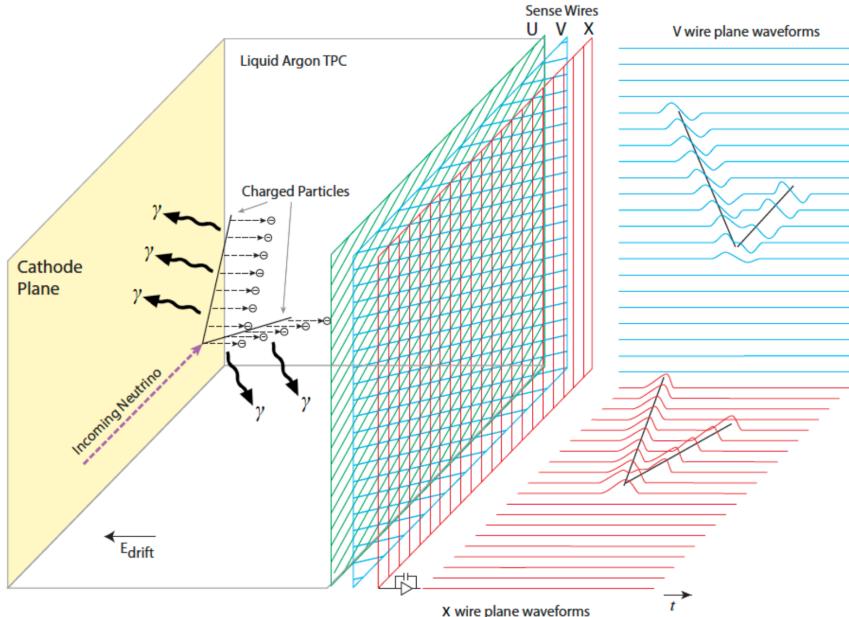
### 826 3.3 Far Detector

827 The so-called DUNE FD complex will sit 1.5 km underground at SURF, South Dakota.  
828 Two caverns will host the four FD modules, two of them per cavern, each embedded in  
829 cryostats of dimensions 18.9 m (w)  $\times$  17.8 m (h)  $\times$  65.8 m (l). A central, smaller cavern  
830 will host the cryogenic system.

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

836 For each event, with energies ranging from a few MeV to several GeV, these detectors  
837 collect both the scintillation light and the ionisation electrons created when the charged

## Chapter 3. The Deep Underground Neutrino Experiment



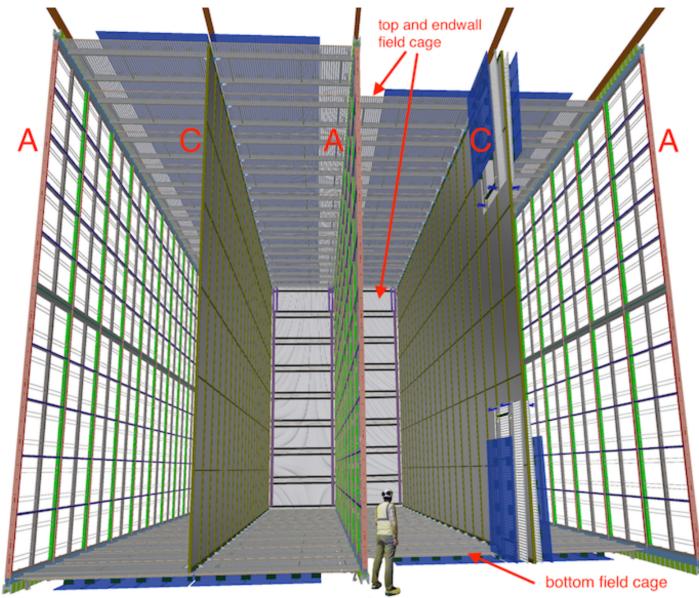
**Figure 3.2:** Schematic diagram showing the operating principle of a LArTPC with wire readout. Figure taken from Ref. [1].

838 particles produced in neutrino-nucleus interactions ionise the argon nuclei. In both HD  
 839 and VD designs the characteristic 128 nm scintillation light of argon is collected by a  
 840 photon detection system (PDS). This light will indicate the time at which electrons  
 841 start to drift, thus enabling reconstruction over the drift coordinate when compared  
 842 to the time when the first ionisation electron arrives to the anode. Reconstruction of  
 843 the topology in the transverse direction is achieved using the charge readout. Fig. 3.2  
 844 illustrates the detection principle described, for the case of a HD detector with a wire  
 845 readout.

### 846 3.3.1 Horizontal Drift

847 Within the HD design the ionisation electrons produced as charged particles traverse the  
 848 LAr drift horizontally towards the anode planes, made out of three layers of wire readout,  
 849 due to the effect of an electric field. This design, previously known as single-phase (SP),  
 850 was tested by the ProtoDUNE-SP detector at CERN. The prototype collected data from  
 851 a hadron beam and cosmic rays, providing high-quality data sets for calibration studies  
 852 and proving the excellent performance of this design.

### 3.3. Far Detector



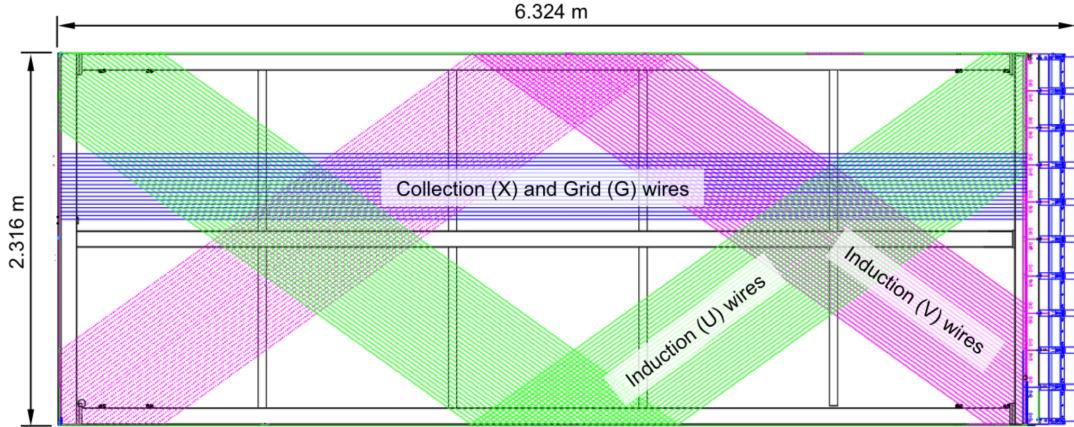
**Figure 3.3:** Proposed design for the FD-1 and FD-2 modules following the HD principle. Figure taken from Ref. [1].

853     Each FD HD detector module is divided in four drift regions, with a maximum drift  
 854    length of 3.5 m, by alternating anode and cathode walls. The surrounding field cage  
 855    ensures the uniformity of the 500 V/cm horizontal electric field across the drift volumes.  
 856    The three anode walls, which constitute the charge readout of the detector, are built by  
 857    stacking anode plane assemblies (APAs), 2 high times 25 wide. The design of the HD  
 858    modules is shown in Fig. 3.3.

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

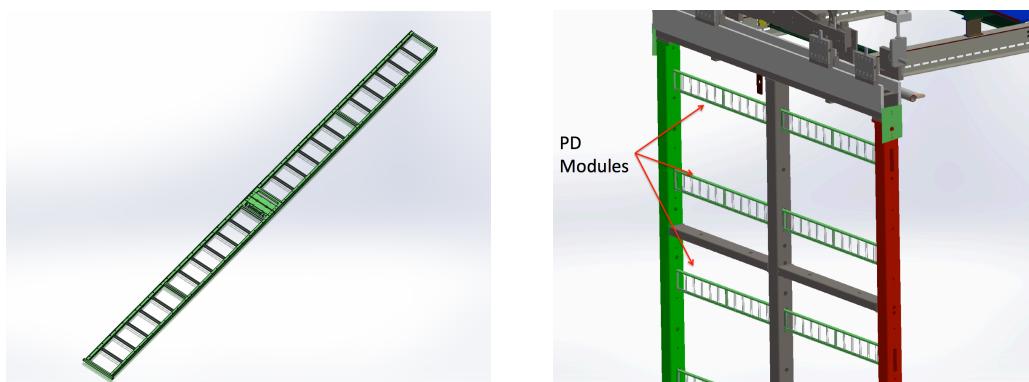
866     The front-end readout electronics, or cold electronics as they are immerse in the LAr,  
 867    are attached to the top of the up APAs and the bottom of the down APAs. Mounted on  
 868    the front-end mother boards we have a series of ASICs that digitize the signals from the

## Chapter 3. The Deep Underground Neutrino Experiment



**Figure 3.4:** 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. [1].

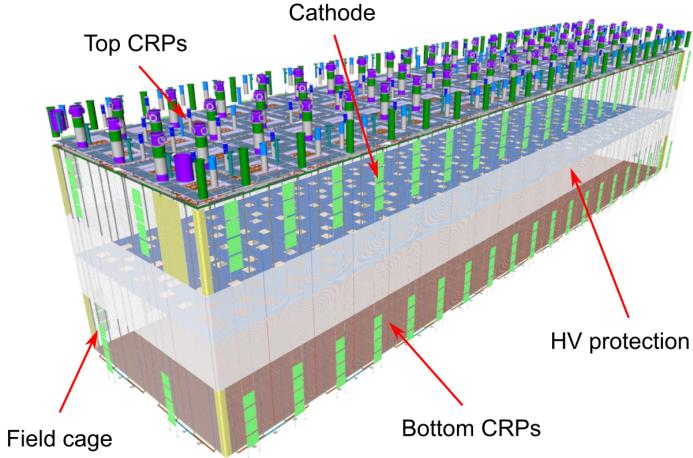
869 collection and induction planes. Each wire signal goes to a charge-sensitive amplifier,  
 870 then there is a pulse-shaping circuit and this is followed by the analogue-to-digital  
 871 converter. This part of the process happens inside the LAr to minimise the number of  
 872 cables penetrating the cryostat. The digitised signals come out finally via a series of  
 873 high-speed serial links to the warm interface boards (WIBs), from where the data is sent  
 874 to the back-end DAQ through optical fibers.



**Figure 3.5:** A PDS module containing 24 X-ARAPUCAs (left) and the location of the modules on the APAs (right). Figure taken from Ref. [1].

875 The PDS uses modules of X-ARAPUCA devices, mounted on the APA frames  
 876 between the wire planes. Each X-ARAPUCA consists of layers of dichroic filter and

### 3.3. Far Detector



**Figure 3.6:** Proposed design for the FD-3 module following the VD principle. Figure adapted from Ref. [55].

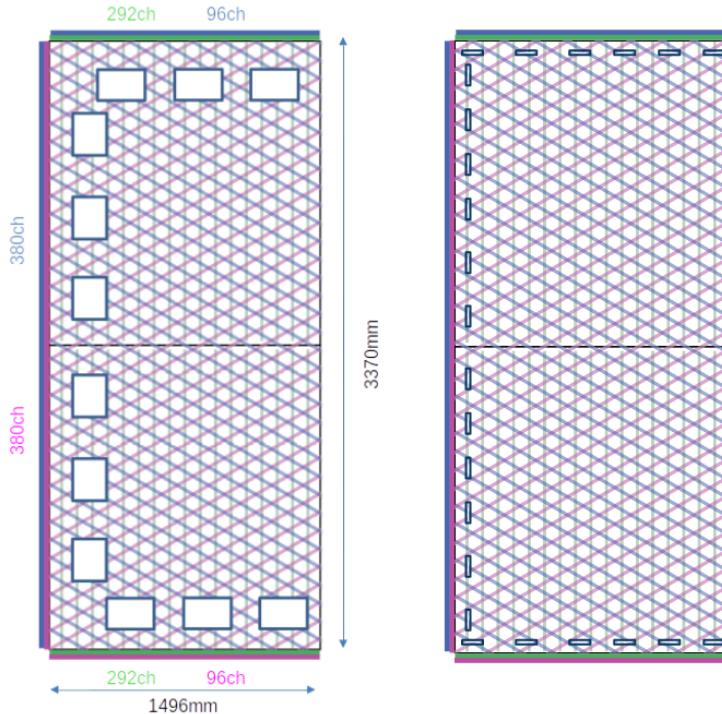
wavelength-shifter. They shift the VUV scintillation light into the visible spectrum, sending then the visible photons to silicon photomultiplier (SiPM) devices. The PDS modules are  $209\text{ cm} \times 12\text{ cm} \times 2\text{ cm}$  bars, containing 24 X-ARAPUCAs. There are 10 of these PDS modules per APA. Fig. 3.5 shows a PDS module (left) and the placement of the modules on the APAs (right).

#### 3.3.2 Vertical Drift

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

The current design of the FD VD module counts with two drift chambers with a maximum drift distance of 6.5 cm. A cathode plane splits the detector volume along the drift direction while the two anode planes are connected to the bottom and top walls of the detector. The layout of the VD module is shown in Fig. 3.6. Compared with

## Chapter 3. The Deep Underground Neutrino Experiment



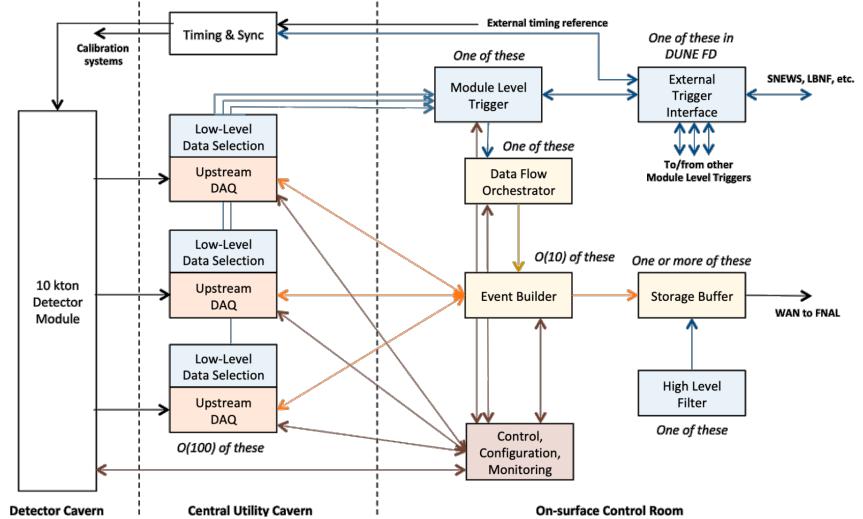
**Figure 3.7:** Schematic representation of the electrode strip configuration for a top (left) and bottom (right) CRU. Figure taken from Ref. [55].

the HD design, the VD option offers a slightly larger instrumented volume and a more cost-effective solution for the charge readout.

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

The PCB face opposite to the cathode has a copper guard plane which acts as shielding, while its reverse face is etched with electrode strips forming the first induction plane. The outer PCB has electrode strips on both faces, the ones facing the inner PCB form the second induction plane while the outermost ones form the collection plane. Fig. 3.7 shows the layout of the electrode strips for the top (left) and bottom (right) CRUs. The magenta and blue lines represent the first and second induction planes respectively, and the green lines correspond to the collection plane.

### 3.3. Far Detector



**Figure 3.8:** Detailed diagram of the DUNE FD DAQ system. Figure taken from Ref. [2].

907        The PDS in the VD module will use the same X-ARAPUCA technology developed  
 908        for the HD design. The plan is to place the PDS modules on the cryostat walls and on  
 909        the cathode, in order to maximise the photon yield.

910        **3.3.3 FD Data Acquisition System**

911        The task of the data acquisition (DAQ) system is to receive, process and store data from  
 912        the detector modules. In the case of DUNE the DAQ architecture is designed to work  
 913        for all FD modules interchangeably, except some aspects of the upstream part which  
 914        may depend on the specific module technology.

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

921        The DAQ system of the DUNE FD is composed of five different subsystems. The  
 922        first one is the upstream DAQ, which receives the raw data from the detector, buffers it

## Chapter 3. The Deep Underground Neutrino Experiment

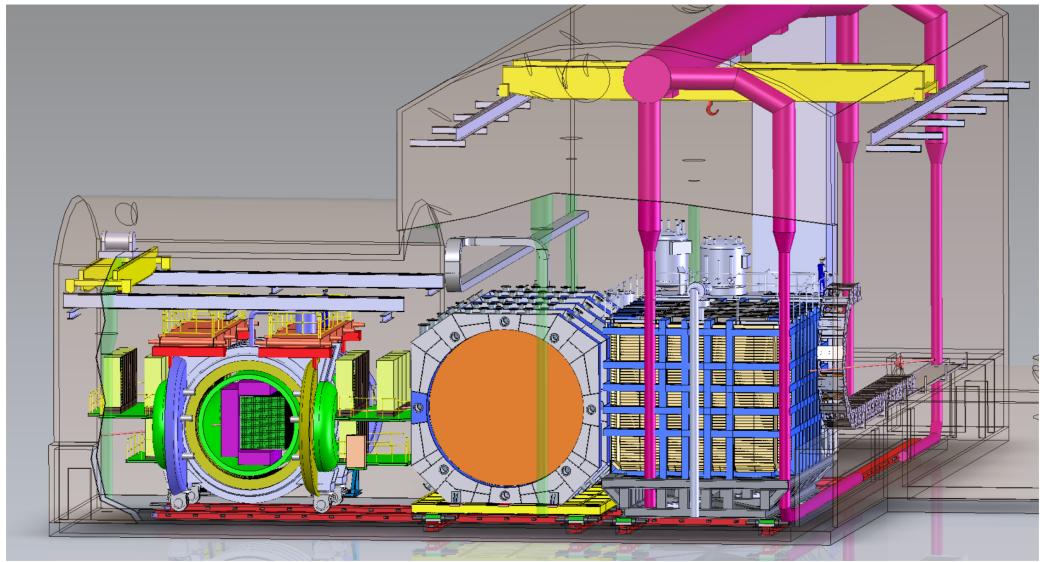
and perform some low-level pre-processing. The minimally processed data is then fed into a hierarchical data selection system, which then performs a module level trigger decision. In case of a positive decision a trigger command is produced and executed by the data flow orchestrator, located in the back-end (BE) DAQ subsystem. Subsequently the DAQ BE retrieves the relevant data from the buffers located in the upstream DAQ, adds all the data into a cohesive record and saves it to permanent storage. Watching over all the other subsystems we also have the control, configuration and monitoring subsystem and the time and synchronization subsystem. Fig. 3.8 shows a schematic diagram of the DAQ system, showing the different subsystems and their relations.

A notorious challenge for the DUNE DAQ system comes from its broad physics goals. We must be prepared to process events spanning a wide range of time windows (from 5 ms in the case of beam and cosmic neutrinos and nucleon decay to 100 s in the case of SNBs) and therefore this requires a continuous readout of the detector modules. Moreover, because of the off-beam measurements we need to ensure the capabilities of online data processing and self-triggering. Having this into account, together with the technical constraints, the DUNE FD DAQ faces a series of challenges: it needs to be fault tolerant and redundant to reduce downtime, accommodate new components while it keeps serving the operational modules, have large upstream buffers to handle SNB physics, be able to support a wide range of readout windows and last reduce the throughput of data to permanent storage to be at most 30 PB/year.

### 3.4 Near Detector

In order 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 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.

### 3.4. Near Detector



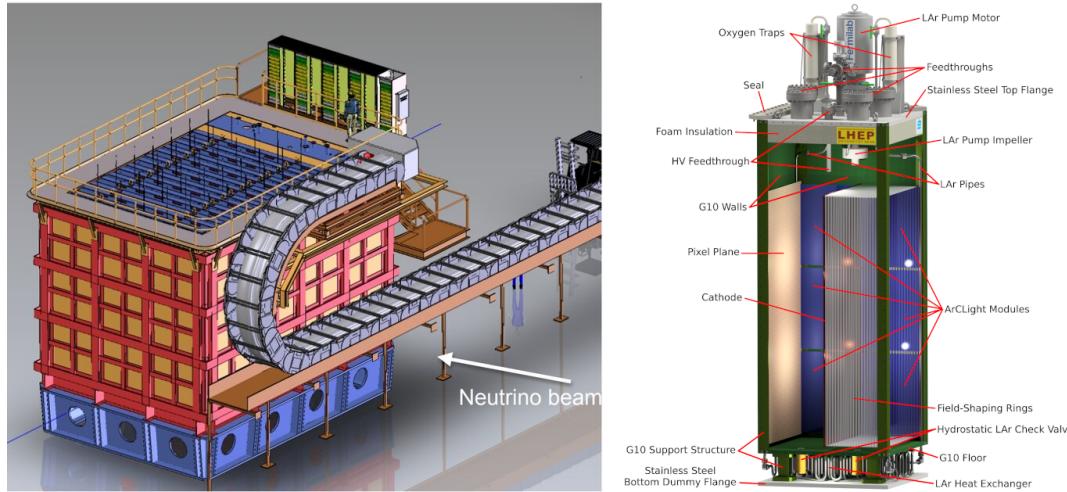
**Figure 3.9:** Representation of the ND hall in Phase II, showing the different subcomponents. From right to left, in the direction of the beam, we have ND-LAr, ND-GAr and SAND. Figure taken from Ref. [56].

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

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

965 Additionally, the ND will have a physics program of its own. In particular, it will

## Chapter 3. The Deep Underground Neutrino Experiment



**Figure 3.10:** 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. [1].

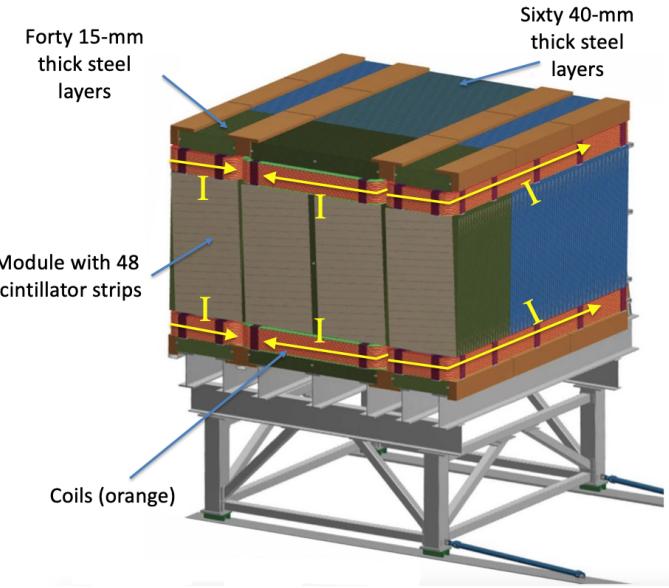
measure neutrino cross sections that will then be used to constrain the model used in the long-baseline oscillation analysis. It will also be used to search for BSM phenomena such as heavy neutral leptons, dark photons, millicharged particles, etc.

The DUNE ND can be divided in three main components, a LArTPC known as ND-LAr, a magnetised muon spectrometer, which will be the Temporary Muon Spectrometer (TMS) in Phase I and ND-GAr in Phase II, and the System for on-Axis Neutrino Detection (SAND). The layout of the Phase II DUNE ND can be seen in Fig. 3.9. The first two components of the ND will be able to move off-axis, in what is called the Precision Reaction-Independent Spectrum Measurement (PRISM) concept. More details on the purpose and design of the ND can be found in the DUNE ND Conceptual Design Report (CDR) [56].

### 3.4.1 ND-LAr

ND-LAr is a LArTPC, as the ND needs a LAr component in order to reduce cross section and detector systematic uncertainties in the oscillation analysis. However, its design differs significantly from those proposed for the FD modules. Because of the high event rates at the ND, approximately 55 neutrino interaction events per 10  $\mu\text{s}$  spill,

### 3.4. Near Detector



**Figure 3.11:** Schematic view of the TMS detector, highlighting its main parts. Figure adapted from Ref. [1].

ND-LAr will be built in a modular way. Each of the modules, based on the ArgonCube technology, is a fully instrumented, optically isolated TPC with a pixelated readout. The pixelisation allows for a fully 3D reconstruction and the optical isolation reduces the problems due to overlapping interactions. Fig. 3.10 shows a representation of the external parts of ND-LAr (left) and a detailed diagram of an ArgonCube module (right).

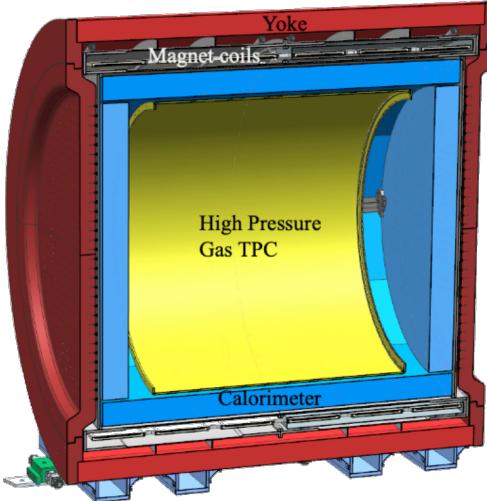
With a fiducial mass of 67 t and dimensions 7 m (w)  $\times$  3 m (h)  $\times$  5 m (l), ND-LAr will be able to provide high statistics and contain the hadronic systems from the beam neutrino interactions, but muons with a momentum higher than 0.7 GeV will exit the detector.

#### 3.4.2 TMS/ND-GAr

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

In Phase I that role will be fulfilled by TMS. It is a magnetised sampling calorimeter, with alternating steel and plastic scintillator layers. Fig. 3.11 shows a schematic view of

## Chapter 3. The Deep Underground Neutrino Experiment



**Figure 3.12:** Cross section of the ND-GAr geometry, showing the HPgTPC, ECal and magnet. Figure adapted from Ref. [1].

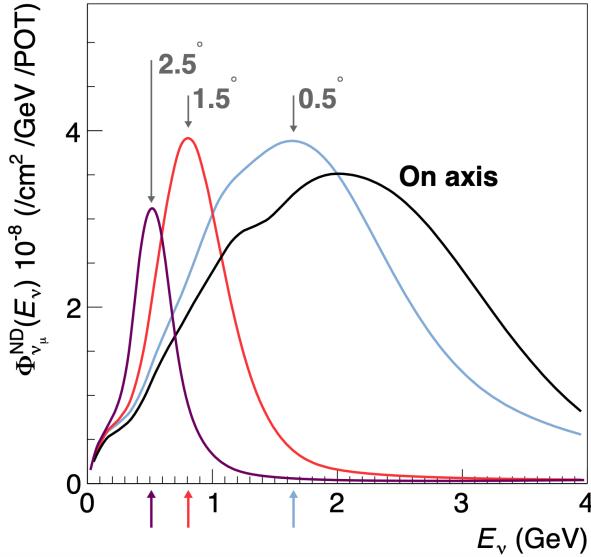
997 the TMS detector. The magnetic field allows a precise measurement of the sign of the  
998 muon, so one can distinguish between neutrino and antineutrino interactions.

999 After the Phase II upgrade, TMS will be replaced with ND-GAr. This detector is  
1000 a magnetised, high-pressure GAr TPC (often denoted as HPgTPC) surrounded by an  
1001 electromagnetic calorimeter (ECal) and a muon tagger. A cross section of its geometry  
1002 can be seen in Fig. 3.12. ND-GAr will be able to measure the momenta of the outgoing  
1003 muons while also detect neutrino interactions inside the GAr volume. This allows  
1004 ND-GAr to constrain the systematic uncertainties even further, as it will be able to  
1005 accurately measure neutrino interactions at low energies thanks to the lower tracking  
1006 thresholds of GAr.

### 1007 3.4.3 PRISM

1008 In general, the observed peak neutrino energy of a neutrino beam decreases as the  
1009 observation angle with respect to the beam direction increases. This feature has been  
1010 used in other long-baseline neutrino experiments, like T2K ( $2.5^\circ$  off-axis) and NOvA  
1011 ( $0.8^\circ$  off-axis), in order to achieve narrower energy distributions. The DUNE PRISM  
1012 concept exploits this effect using a movable ND. Within PRISM both ND-LAr and the

### 3.4. Near Detector



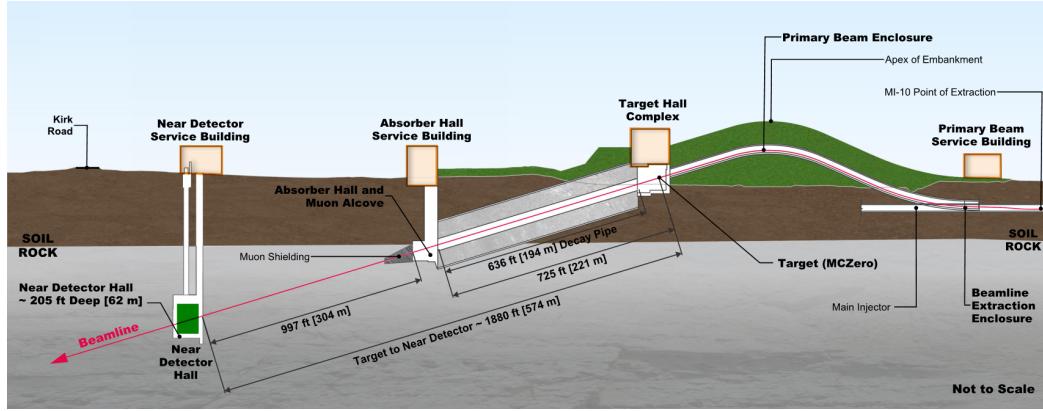
**Figure 3.13:** Predicted beam muon neutrino flux at the ND location for different off-axis positions. Figure taken from Ref. [56].

1013 muon spectrometer (TMS in Phase I and ND-GAr in Phase II) can be moved up to 3.2°  
 1014 off-axis, equivalent to move the detectors 30.5 m laterally through the ND hall.

1015 This allows to record additional data samples with different energy compositions.  
 1016 Fig. 3.13 compares the on-axis muon neutrino flux at the ND with the fluxes at different  
 1017 off-axis positions. As the off-axis position increases the neutrino flux becomes closer to  
 1018 a monoenergetic beam with a lower peak energy. These samples can be used to perform  
 1019 a data-driven determination of the relation between true and reconstructed neutrino  
 1020 energy, in order to reduce the dependence on the interaction model. The off-axis samples  
 1021 are linearly combined to produce a narrow Gaussian energy distribution centered on  
 1022 a target true energy. From the combination coefficients one can build a sample of  
 1023 reconstructed neutrino events that will determine the energy mapping.

1024 The PRISM samples can also be used to form a flux at the ND location similar in  
 1025 shape to the oscillated flux measured by the FD. This method can be used to extract  
 1026 the oscillation parameters with minimal input from the neutrino interaction model.

## Chapter 3. The Deep Underground Neutrino Experiment



**Figure 3.14:** Schematic longitudinal section of the LBNF beamline at Fermilab (not to scale). Figure taken from Ref. [57].

### 1027 3.4.4 SAND

1028 The role of SAND is to monitor the beam stability by measuring the on-axis neutrino  
 1029 energy spectra. As the PRISM program requires that ND-LAr and its downstream  
 1030 muon spectrometer spend about half of the time in off-axis positions, it is not possible  
 1031 to monitor the stability with the movable detectors. Moreover, for the success of PRISM  
 1032 it is essential to have a stable beam configuration, or, at least, a quick assessment and  
 1033 modeling of the distortions.

1034 The SAND detector is magnetised, and it counts with an inner low density tracker,  
 1035 a LAr target with optical readout and surrounding sampling calorimeter.

## 1036 3.5 LBNF beamline

1037 The Long-Baseline Neutrino Facility (LBNF) project is responsible for producing the  
 1038 neutrino beam for the DUNE detectors. A detailed discussion of the LBNF program  
 1039 can be found in the DUNE/LBNF CDR Volume III [57].

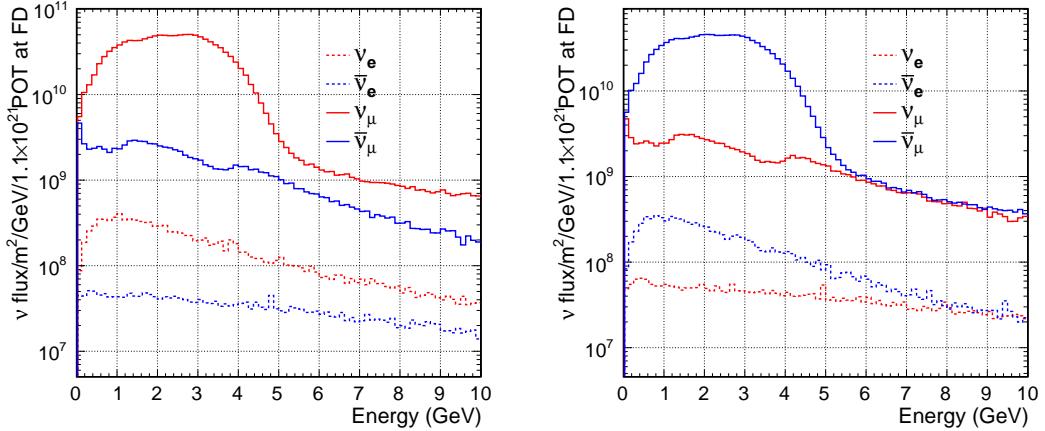
1040 The LBNF beamline will provide a high-intensity neutrino beam within the adequate  
 1041 energy range in order to meet the long-baseline oscillation physics goals of DUNE. A  
 1042 schematic diagram of the longitudinal section of the LBNF beamline is shown in Fig.  
 1043 3.14. First, a beam of  $60 - 120$  GeV protons is extracted from the Fermilab Main

### 3.5. LBNF beamline

1044 Injector. This beam is aimed towards the target area, where it collides with a cylindrical  
 1045 graphite target to produce pions and kaons.

1046 The diffuse, secondary beam of particles is focused by a pair of magnetic horns.  
 1047 These select the positively charged particles when operated in Forward Horn Current  
 1048 (FHC) mode, or the negatively charged ones when the current is reversed, also known as  
 1049 Reverse Horn Current (RHC) mode. The focused secondary beam then enters a 194 m  
 1050 decay pipe where the pions and kaons will predominantly produce  $\mu^+\nu_\mu$  pairs when in  
 1051 FHC mode (or  $\mu^-\bar{\nu}_\mu$  in RHC mode).

1052 At the end of the decay pipe a hadron absorber removes the undecayed hadrons and  
 1053 muons from the beam, which reduces the  $\nu_e$  ( $\bar{\nu}_e$ ) and  $\bar{\nu}_\mu$  ( $\nu_\mu$ ) contamination coming  
 1054 from the  $\mu^+$  ( $\mu^-$ ) decays. The resulting neutrino flux at the FD is shown in Fig. 3.15,  
 1055 both for FHC (left) and RHC (right) modes. These predictions show the intrinsic  $(\bar{\nu}_e)$   
 1056 contamination and wrong sign component from wrong sign and neutral meson decays,  
 1057 as well as muons decaying before reaching the absorber.



**Figure 3.15:** Predicted neutrino fluxes at the FD in FHC mode (left panel) and RHC mode (right panel). Figures taken from Ref. [50].



1058 **Chapter 4**

1059 **ND-GAr**

1060 ND-GAr is a magnetised, high-pressure gaseous argon TPC (HPgTPC), surrounded by  
1061 an electromagnetic calorimeter (ECal) and a muon detector (commonly refer to as  $\mu$ ID).  
1062 A detailed discussion on the requirements, design, performance and physics of ND-GAr  
1063 can be found in the DUNE ND CDR [56] and the ND-GAr whitepaper (cite).

1064 In DUNE Phase II ND-GAr will fulfill the role of TMS, measuring the momentum  
1065 and sign of the charged particles exiting ND-LAr. Additionally, it will be able to measure  
1066 neutrino interactions inside the HPgTPC, achieving lower energy thresholds than those  
1067 of the ND and FD LArTPCs. By doing so ND-GAr will allow to constrain the relevant  
1068 systematic uncertainties for the LBL analysis even further.

1069 The goal of the present chapter is to review the requirements that the physics program  
1070 of DUNE impose on ND-GAr, present the current status of its design and describe the  
1071 GArSoft package, its simulation and reconstruction software.

1072 **4.1 Requirements**

1073 The primary requirement for ND-GAr is to the measure the momentum and charge of  
1074 muons from  $\nu_\mu$  and  $\bar{\nu}_\mu$  CC interactions in ND-LAr, in order to measure their energy  
1075 spectrum. To achieve the sensitivity to the neutrino oscillation parameters described  
1076 in the DUNE FD TDR Volume II [50] ND-GAr should be able to constrain the muon

## Chapter 4. ND-GAr

1077 energy within a 1% uncertainty or better. The main constraint will come from the  
1078 calibration of the magnetic field, performed using neutral kaon decays in the HPgTPC.

1079 Another requirement for ND-GAr is the precise measurement of neutrino interactions  
1080 on argon for the energies relevant to the neutrino oscillation program. The goal is to  
1081 constrain the cross section systematic uncertainties in the regions of phase space that  
1082 are not accessible to ND-LAr. This requires the kinematic acceptance for muons in  
1083 ND-GAr to exceed that of ND-LAr, being comparable to the one observed in the FD.

1084 ND-GAr should also be able to the relationship between true and reconstructed energy  
1085 from neutrino interactions on argon with low thresholds, being sensitive to particles that  
1086 are not observed or may be misidentified in ND-LAr. In particular, ND-GAr needs to  
1087 have low tracking thresholds in order to measure the spectrum of pions and protons  
1088 produced in final-state interactions (FSI). It also must be able to accurately measure  
1089 the pion multiplicity in 1, 2 and 3 pions final states, to inform the pion mass correction  
1090 in the LArTPCs.

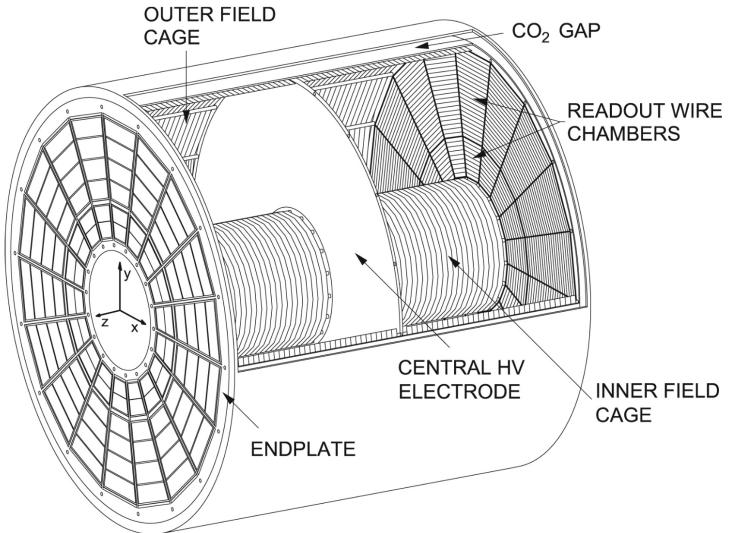
### 1091 4.2 Reference design

1092 The final design of ND-GAr is still under preparation. However, a preliminary baseline  
1093 design was in place at the time of the ND CDR. This section summarises the main  
1094 features of that design, as it is also the one used for the default geometry in our simulation.  
1095 A DUNE Phase II whitepaper, discussing the different options under consideration for  
1096 the ND-GAr design, is in progress.

#### 1097 4.2.1 HPgTPC

1098 The reference design for the ND-GAr HPgTPC follow closely that of the ALICE TPC.  
1099 It is a cylinder with a central high-voltage cathode, generating the electric field for  
1100 the two drift volumes, with a maximum drift distance of 2.5 m each. The anodes will  
1101 be instrumented with charge readout chambers. The original design repurposed the  
1102 multi-wire proportional readout chambers of ALICE, however the current R&D efforts

## 4.2. Reference design



**Figure 4.1:** Diagram of the ALICE TPC, showing the two drift chambers, inner and outer field cages and readout chambers. Figure taken from Ref. [1].

1103 focus on a gas electron multiplier option instead. Fig. 4.1 shows a schematic diagram of  
 1104 the ALICE TPC design. The basic ND-GAr geometry will resemble this, except for the  
 1105 inner field cage.

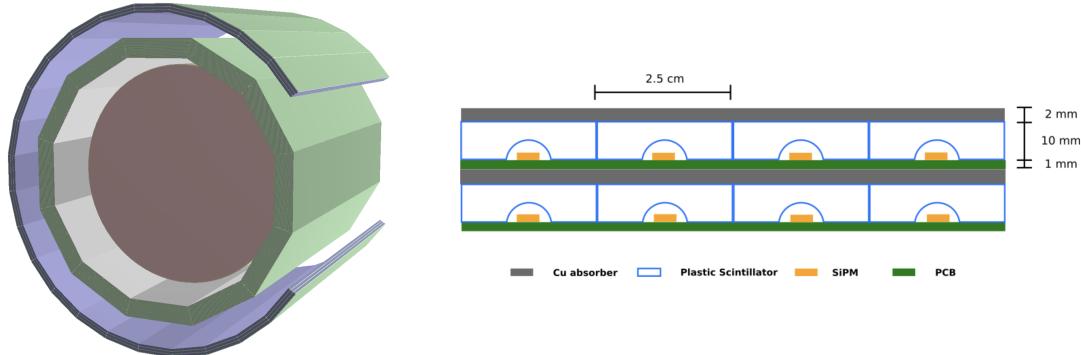
1106 It will use a 90-10 molar fraction argon-CH<sub>4</sub> mixture at 10 bar. With this baseline  
 1107 gas mixture light collection is not possible, as the quenching gas absorbs most of the  
 1108 VUV photons. Additional R&D efforts are underway, to understand if different mixtures  
 1109 allow for the light signal to be used to provide a  $t_0$  while maintaining stable charge gain.

1110 **4.2.2 ECal**

1111 The main role of the ND-GAr ECal is the calorimetric measurement of the electron  
 1112 energies and the reconstruction of photons, in particular those from neutral pion decays.  
 1113 Also, the ECal is able to provide a  $t_0$  timestamp for neutrino interactions, by associating  
 1114 its activity to the tracks in the HPgTPC. The ECal will also be able to perform  
 1115 neutron reconstruction using time of flight and reject external backgrounds, thanks to  
 1116 its sub-nanosecond time resolution.

1117 The ECal design features three independent subdetectors, two end caps at each side  
 1118 and a barrel surrounding the HPgTPC. Each of the detectors is divided in modules,

## Chapter 4. ND-GAr



**Figure 4.2:** 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. [1].

which combine alternating layers of plastic scintillator and absorber material readout by SiPMs. The inner scintillator layers consist of  $2.5 \times 2.5 \text{ cm}^2$  high-granularity tiles, whereas the outer ones are made out of 4 cm wide cross-strips spanning the whole module length. The current barrel geometry consists of 8 tile layers and 34 strip layers, while the end caps feature 6 and 36 respectively. The thickness of the scintillator layers is 7 mm and 5 mm for the Pb absorber layers. The 12-sided geometry of the ECal barrel (left) and the layout of the tile layers (left)<sup>1</sup> can be seen in Fig. 4.2.

### 4.2.3 Magnet

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

The solenoid is a single layer coil, based on niobium titanium superconducting

<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.

### 4.3. GArSoft

1135 Rutherford cable. The total length of the coil is 7.5 m. The bobbin will be split in four  
1136 segments grouped in pairs with two identical cryostats, connected in series. The iron  
1137 yoke features an aperture in the upstream side to allow the muons coming from ND-LAr.  
1138 Still, its material will be enough to reduce the magnetic field reaching SAND, and also  
1139 stop the charged pions produced inside the HPgTPC.

#### 1140 4.2.4 Muon system

1141 The design of the ND-GAr muon system is still in a preliminary stage. Its role is to  
1142 distinguish between muons and pions punching through the ECal. This is especially  
1143 important for wrong-sign determination, to separate these from neutral current events.

1144 In its current form, the muon system consists of three layers of longitudinal sampling  
1145 structures. It alternates 10 cm Fe absorber slabs with 2 cm plastic scintillator strips.  
1146 The transverse granularity required is still under study.

### 1147 4.3 GArSoft

1148 GArSoft is a software package developed for the simulation and reconstruction of events  
1149 in ND-GAr. It is inspired by the LArSoft toolkit used for the simulation of LArTPC  
1150 experiments, like the DUNE FD modules. It is based on `art`, the framework for event  
1151 processing in particle physics experiments [58]. Other of its main dependencies are `ROOT`,  
1152 `NuTools`, `GENIE` and `Geant4`. It allows the user to run all the steps of a generation-  
1153 simulation-reconstruction workflow using FHiCL configuration files.

#### 1154 4.3.1 Event generation

1155 The standard generator FHiCLs in GArSoft run the event generation and particle  
1156 propagation simulation (i.e. Geant4) in the same job by default. However, it is possible  
1157 to split them up if needed. The current version of GArSoft provides five different event  
1158 generators, each of them producing `simb::MCTruth` products defined in `NuTools`. The  
1159 available modules are:

## Chapter 4. ND-GAr

- 1160     • **SingleGen**: particle gun generator. It produces the specified particles with a given  
1161       distribution of momenta, initial positions and angles.
- 1162     • **TextGen**: text file generator. The input file must follow the `hepevt` format<sup>2</sup>, the  
1163       module simply copies this to `simb::MCTruth` data products.
- 1164     • **GENIEGen**: GENIE neutrino event generator. The module runs the neutrino-nucleus  
1165       interaction generator using the options specified in the driver FHiCL file (flux file,  
1166       flavour composition, number of interactions per event,  $t_0$  distribution, ...). Current  
1167       default version is `v3_04_00`.
- 1168     • **RadioGen**: radiological generator. It produces a set list of particles to model  
1169       radiological decays. Not tested.
- 1170     • **CRYGen**: cosmic ray generator. The module runs the CRY event generator with a  
1171       configuration specified in the FHiCL file (latitude and altitude of detector, energy  
1172       threshold, ...). Not tested.

1173       The module `GArG4` searches for all the generated `simb::MCTruth` data products, using  
1174       them as inputs to the Geant4 simulation with the specified detector geometry. A constant  
1175       0.5 T magnetic field along the drift coordinate is assumed. The main outputs of this step  
1176       are `simb::MCParticle` objects for the generated Geant4 particles, `gar::EnergyDeposit`  
1177       data products for the energy deposits in the HPgTPC and `gar::CaloDeposit` data  
1178       products for the energy deposits in the ECal and muon system.

### 1179 4.3.2 Detector simulation

1180       The standard detector simulation step in GArSoft is all run with a single FHiCL, but  
1181       the different modules can be run independently as well. First the `IonizationReadout`

<sup>2</sup>In brief, each event contains at least two lines. The first line contains two entries, the event number and the number of particles in the event. Each following line contains 15 entries to describe each particle. The entries are: status code, pdg code for the particle, entry of the first mother for this particle, entry of the second mother for this particle, entry of the first daughter for this particle, entry of the second daughter for this particle, x component of the particle momentum, y component of the particle momentum, z component of the particle momentum, energy of the particle, mass of the particle, x component of the particle initial position, y component of the particle initial position, z component of the particle initial position and time of the particle production.

### 4.3. GArSoft

1182 module simulates the charge readout of the HPgTPC, and later the `SiPMReadout` module  
1183 runs twice, once for the ECal and then for the muon system, with different configurations.

1184 The `IonizationAndScintillation` module collects all the `gar::EnergyDeposit`  
1185 data products, to compute the equivalent number of ionization electrons for each energy  
1186 deposit. The `ElectronDriftAlg` module simulates the electron diffusion numerically  
1187 both in the longitudinal and transverse directions and applies an electron lifetime  
1188 correction factor. The induced charge on the nearest and neighbouring readout pads  
1189 is modeled using the provided pad response functions. The digitisation of the data is  
1190 then simulated with the `TPCReadoutSimAlg` module. By default, the ADC sampling  
1191 rate used is 50.505 MHz. The resulting raw waveforms for each channel are stored with  
1192 zero-suppression, in order to save memory and CPU time. The algorithms keep blocks  
1193 of ADC values above a certain threshold, plus some adjustable additional early and late  
1194 tick counts. The results of these three steps are `gar::raw::RawDigit` data products.

1195 For the ECal and the muon system the `SiPMReadout` module calls either the  
1196 `ECALReadoutSimStandardAlg` or `MuIDReadoutSimStandardAlg` modules. These take  
1197 all the `gar::CaloDeposit` data products in the corresponding detector and do the  
1198 digitisation depending on whether the hit was in a tile or strip layer. They include single  
1199 photon statistics, electronic noise, SiPM saturation and time smearing. The resulting  
1200 objects are `gar::raw::CaloRawDigit` data products.

#### 1201 4.3.3 Reconstruction

1202 The reconstruction in GArSoft is also run as a single job by default. It first runs the hit  
1203 finding, clustering, track fitting and vertex identification in the HPgTPC, followed by  
1204 the hit finding and clustering in the ECal and muon system. After those it produces the  
1205 associations between the associations between the tracks and the ECal clusters.

1206 Focusing first on the HPgTPC reconstruction, the `CompressedHitFinder` module  
1207 takes the zero-suppressed ADCs from the `gar::raw::RawDigit` data products. The  
1208 reconstructed hits largely correspond to the above threshold blocks, however the hit  
1209 finder identifies waveforms with more than one maximum, diving them in multiple hits

## Chapter 4. ND-GAr

1210 if they dip below a certain threshold. The data products produced are of the form  
1211 `gar::rec::Hit`. These are the inputs to the clustering of hits in the `TPCHitCluster`  
1212 module. Hits close in space and time are merged, and the resulting centroids are found.  
1213 This module outputs `gar::rec::TPCClusters` objects and associations to the input  
1214 hits.

1215 The following step prior to the track fitting is pattern recognition. The module  
1216 called `tpcvecchitfinder2` uses the `gar::rec::TPCClusters` data products to find track  
1217 segments, typically called vector hits. They are identified by performing linear 2D fits  
1218 to the positions of the clusters in a 10 cm radius, one fit for each coordinate pair. A  
1219 3D fit defines the line segment of the vector hit, using as independent variable the one  
1220 whose sum of (absolute value) slopes in the 2D fits is the smallest. The clusters are  
1221 merged to a given vector hit if they are less than 2 cm away from the line segment. The  
1222 outputs are `gar::rec::VecHit` data products, as well as associations to the clusters. The  
1223 `tpcpatrec2` module takes the `gar::rec::VecHit` objects to form the track candidates.  
1224 The vector hits are merged together if their direction matches, their centers are within  
1225 60 cm and their direction vectors point roughly to their respective centers. Once  
1226 the clusters of vector hits are formed they are used to make a first estimation of the  
1227 track parameters, simply taking three clusters along the track. The module produces  
1228 `gar::rec::Track` data products and associations between these tracks and the clusters  
1229 and vector hits.

1230 The track is fitted by means of a Kalman filter in the `tpctrackfit2` module, using  
1231 the position along the drift direction as the independent variable. Two different fits are  
1232 performed per track, a forward and a backwards fit, each starting from one of the track  
1233 ends. The Kalman filter state vector ( $y, z, R, \phi, \tan\lambda$ ) is estimated at each point along  
1234 the track using a Bayesian update. The track parameters reported in the forward and  
1235 backwards fits are the ones computed at the opposite end where the fit started. The  
1236 main outputs of the track fit are the `gar::rec::Track` objects. Additionally, the module  
1237 stores the fitted 3D positions along the track in the `gar::rec::TrackTrajectory` data  
1238 products and the total charge and step sizes for each point also get stored in the form of

### 4.3. GArSoft

1239    `gar::rec::TrackIonization` objects.

1240    After the tracking step, the `vertexfinder1` module looks at the reconstructed  
1241    `gar::rec::Track` products, creating vertex candidates with the track ends that are  
1242    within 12 cm of each other. The vertices are then fitted using linear extrapolations from  
1243    the different track ends associated. The results are `gar::rec::Vertex` data products,  
1244    and associations to the tracks and corresponding track ends.

1245    For the ECal and muon tagger, the `SiPMHitFinder` module runs twice with different  
1246    configurations, adapted to the particular capabilities of both. The module simply takes  
1247    the `gar::raw::CaloRawDigit` products, applies a calibration factor to convert the ADC  
1248    counts to MeV and for the strip layer hits it calculates the position along the strip using  
1249    the times recorded of both SiPMs. This module produces `gar::rec::CaloHit` data  
1250    products. Next, these objects are used as inputs to the `CaloClustering` module. It  
1251    merges the hits based on a simple nearest neighbours (NN) algorithm. For the resulting  
1252    clusters it also computes the total energy and position of the centroid. The results are  
1253    stored as `gar::rec::Cluster` data products, with associations to the hits.

1254    The last step in the reconstruction is associating the reconstructed tracks in the  
1255    HPgTPC to the clusters formed in the ECal and muon system. The `TPCECALAssociation`  
1256    module checks first the position of the track end points, considering only the points  
1257    that are at least 215 cm away from the cathode or have a radial distance to the center  
1258    greater than 230 cm. The candidates are propagated up to the radial position, in the  
1259    case of clusters in the barrel, or the drift coordinate position, for the end cap cluster, of  
1260    the different clusters in the collection using the track parameters computed at the end  
1261    point. The end point is associated to the cluster if certain proximity criteria are met.  
1262    This module creates associations between the tracks, the end points and the clusters.  
1263    The criteria for the associations are slightly different for the ECal and the muon tagger.



<sub>1264</sub> Chapter 5

<sub>1265</sub> FWTPG offline software



1266 Chapter 6

1267 Matched Filter approach to  
1268 induction wire Trigger Primitives

1269 6.1 Motivation

1270 The filter implemented in the firmware of the upstream DUNE FD DAQ is a 32nd-order  
1271 low-pass finite impulse-response (FIR) filter. The output of such filter for a discrete  
1272 system can be written as:

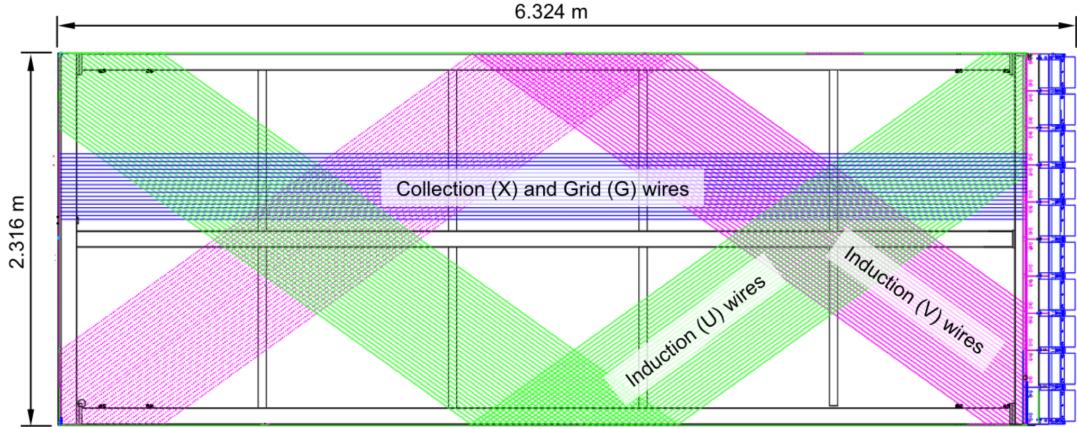
$$y[i] = \sum_{j=0}^N h[j]x[i-j], \quad (6.1)$$

1273 where  $N$  is the order of the filter,  $y$  is the output sequence,  $x$  is the input sequence and  $h$   
1274 is the set of coefficients of the filter. The current implementation within `dtp-firmware`  
1275 [59] uses a set of 16 non-zero integer coefficients.

1276 Filtering is a vital step in the hit finder chain. It helps to suppress the noise and  
1277 enhance the signal peaks with respect to the noiseless baseline. A good filtering strategy  
1278 allows us to use lower thresholds when forming the trigger primitives (TPs) and thus  
1279 increasing the sensitivity of our detector to low energy physics events. In such events,  
1280 the hits produced by the ionisation electrons tend to have lower amplitudes than those  
1281 of interest to the baseline physics programme of the DUNE experiment.

1282 This is particularly important for the induction planes. In general, signal peaks in

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



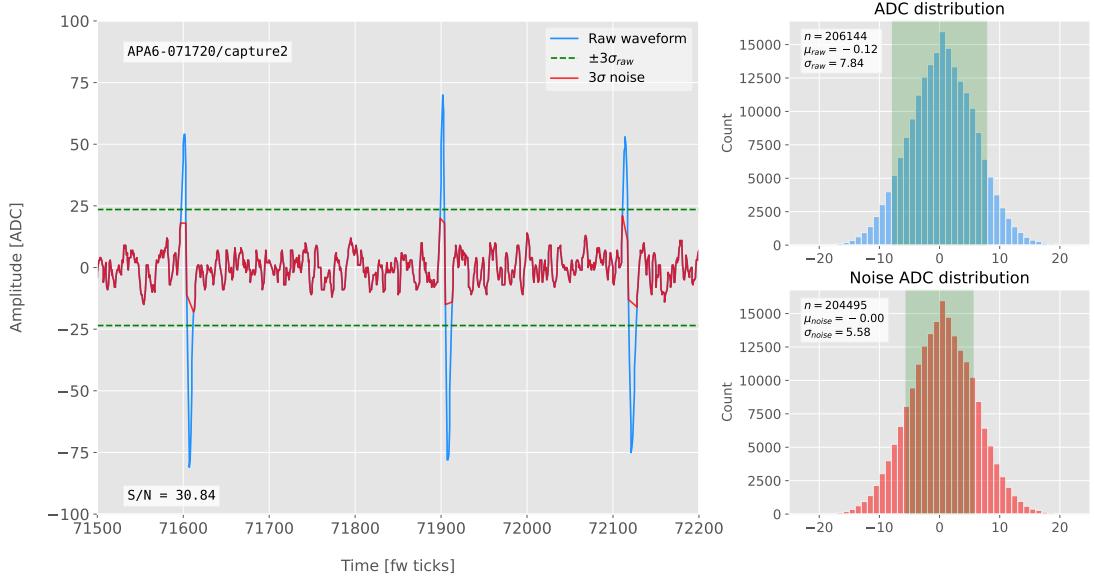
**Figure 6.1:** Schematic representation of an APA. The black lines represent the APA steel frame. The green and magenta lines correspond to the direction of the U and V induction wires respectively. The blue lines indicate the direction of the X collection wires and the wire shielding G.

the induction wires have smaller amplitude than the ones in the induction plane. This, together with the fact that the pulse shapes are bipolar, reduces our capacity to detect the hits on these channels. The inefficiency of detecting TPs in the induction planes (denoted as U and V planes) lead trigger algorithms to focus mainly on the TPs from the collection plane (so-called X plane). As a result, the possibility of making trigger decisions based on the coincidence of TPs across the three wire planes remains nowadays unexploited in DUNE. Fig. 6.1 shows a schematic view of an anode plane assembly (APA), with the different wire plane orientations highlighted.

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

The goal is to implement a better finite-impulse response filter design and to evaluate its performance relative to the current filter. To do so, we need to take into account the limitations of the firmware: the FIR filter shall have maximum 32 coefficients (so-called taps) whose values are 12-bit unsigned integers. Although it is technically possible to include non-integer coefficients, it would be a technical challenge as we have 40 FIR

## 6.2. Signal-to-noise ratio definition



**Figure 6.2:** 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  $\pm \sigma_{\text{noise}}$ .

1300 instances per APA, as there are 4 FIR per optical link and 10 optical links per APA.  
 1301 With these restrictions, the task is to provide a set of 32 coefficients which yield an  
 1302 optimal filter performance for the induction wires.

## 1303 6.2 Signal-to-noise ratio definition

1304 I introduce the signal to noise ratio (S/N) as a measure of the FIR filter performance  
 1305 and demonstrate how to extract its value for a set of ProtoDUNE-SP data. The S/N  
 1306 metrics allow us to compare different filter implementations and serve as a basis for more  
 1307 detailed studies presented later in this document. Specifically, I use the ADC capture  
 1308 `felix-2020-07-17-21:31:44` (data capture taken for firmware validation purposes). I  
 1309 defined S/N as the height of the signal peaks relative to the size of the noise peaks.  
 1310 To quantify this quantity channel by channel one first need to estimate the standard  
 1311 deviation of the ADC data for each channel,  $\sigma_{\text{ADC}}$ . Then, I define the corresponding

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives

noise waveform to be the ADC values in the range  $\pm 3\sigma_{ADC}$ . From this new noise data one can estimate again the mean and standard deviation,  $\mu_{noise}$  and  $\sigma_{noise}$ , so I can write the S/N for any given channel as:

$$S/N = \frac{\max [ADC] - \mu_{noise}}{\sigma_{noise}}, \quad (6.2)$$

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

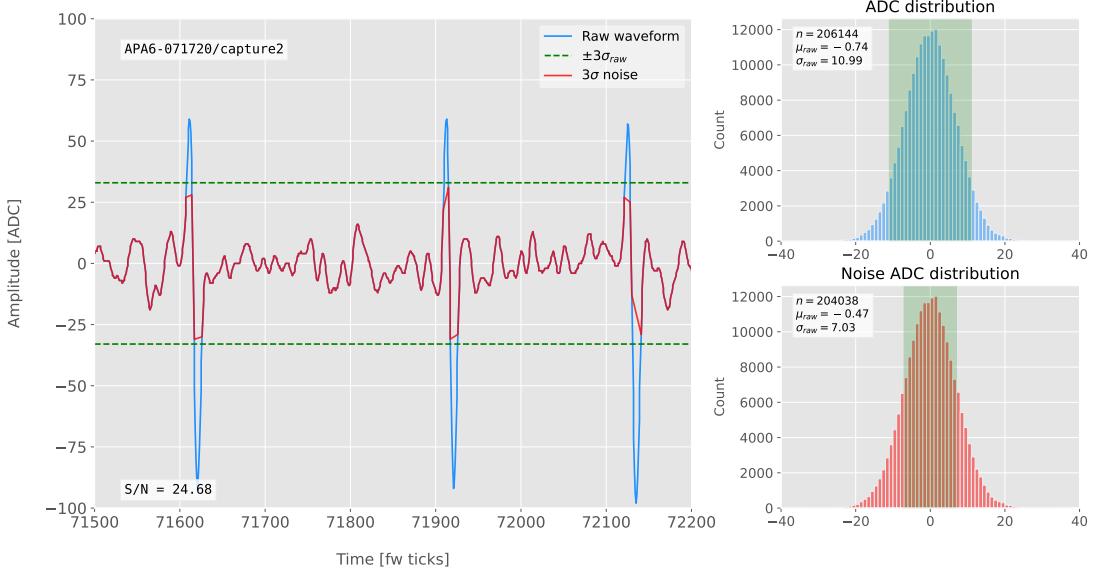
One can apply this definition of the S/N with a waveform from one of the channels of the data capture<sup>1</sup>. Fig. 6.2 shows a zoomed region of the waveform corresponding to channel 7840 (blue line), where one can clearly see three signal peaks and continuous additive noise (we actually see 6 peaks, 3 positive and 3 negative, but, because by design for induction channels the expected signal pulse shapes are bipolar, I treat them as a collection of 3 individual signal peaks). I estimated the standard deviation of this raw waveform to be  $\sigma_{raw} = 7.84$  ADC, so I am able to define the noise waveform (red line) as the ADC values in the range  $\pm 23.52$  ADC. This way one obtains  $\mu_{noise} = 0$  and  $\sigma_{noise} = 5.58$  ADC, which gives  $S/N = 30.84$ .

We can repeat this calculation now for the corresponding filtered waveform (using the current firmware FIR filter). In Fig. 6.3 I plotted the same time window for the filtered waveform from channel 7840 (blue line). In this case, the standard deviation of the waveform is larger than before, giving  $\sigma_{raw} = 10.99$  ADC. The resulting noise waveform (red line) results from selection the ADC values in the range  $\pm 32.91$  ADC, giving now  $\mu_{noise} = -0.47$  ADC and  $\sigma_{noise} = 7.03$  ADC. Finally, one obtains  $S/N = 24.68$ . Notice that the value of S/N decreases after the filtering. Clearly, one can see that the noise baseline has increased by a factor of 1.35 when we applied the FIR filter and at the same time the amplitude of the signal peaks has remained almost unchanged, leading to this poorer S/N value.

---

<sup>1</sup>All the original work was done within the `dtp-simulation` package [60], 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 [61]. 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.

### 6.3. Low-pass FIR filter design



**Figure 6.3:** 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  $\pm \sigma_{\text{noise}}$

### 1335 6.3 Low-pass FIR filter design

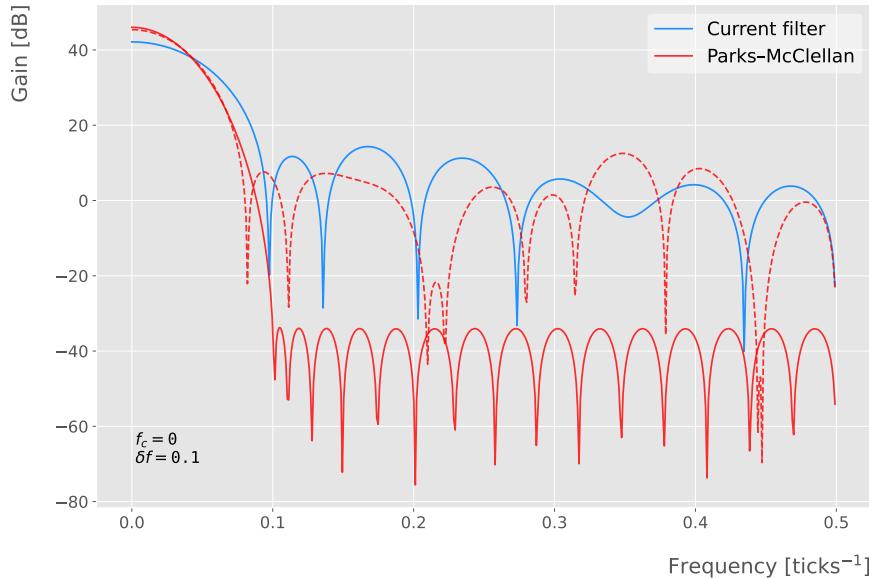
1336 In general, when one uses a method to optimize the frequency response of a digital filter,  
 1337 such as the Parks-McClellan algorithm, one finds a set of  $N$  real coefficients that give  
 1338 the best response for the specified pass-band and order of the filter [62].

1339 In our case, as the sampling frequency is defined as  $1 \text{ ticks}^{-1}$ , the Nyquist frequency  
 1340 will simply be  $1/2 \text{ ticks}^{-1}$ . The current implementation of the filter seems to have as  
 1341 pass-band the range  $[0, 0.1] \text{ ticks}^{-1}$ . This can be seen in Fig. 6.4, where I show the  
 1342 power spectrum, in decibels, of such filter implementation (blue solid line). For instance,  
 1343 the Park-McClellan algorithm finds the optimal Chebyshev FIR filter taking as input  
 1344 the boundaries of the target pass-band and stop-band, which can be written in the form:

$$\left\{ \begin{array}{l} [0, f_c] \\ [f_c + \delta f, f_N] \end{array} \right. , \quad (6.3)$$

## Chapter 6. Matched Filter approach to induction wire 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. 6.4 (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.

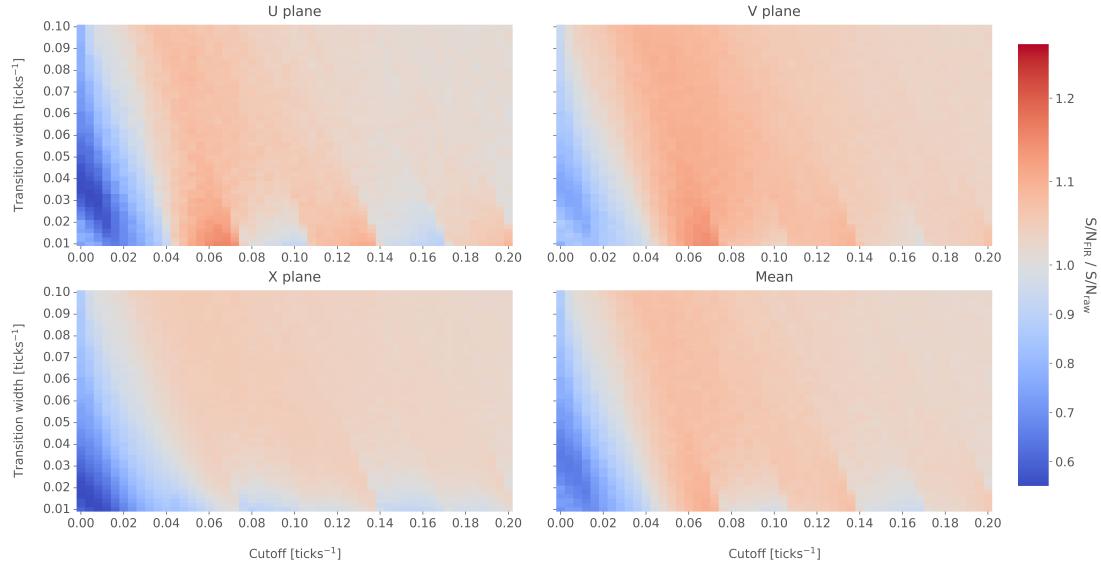


**Figure 6.4:** Power spectrum in decibels for the current implementation of the low-pass FIR filter in `dtp-firmware` (blue line), compared to the response of an optimal filter obtained using the Parks-McClellan algorithm for the same pass-band (red line). Also for comparison I include the spectrum 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. 6.5 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`,

### 6.3. Low-pass FIR filter design

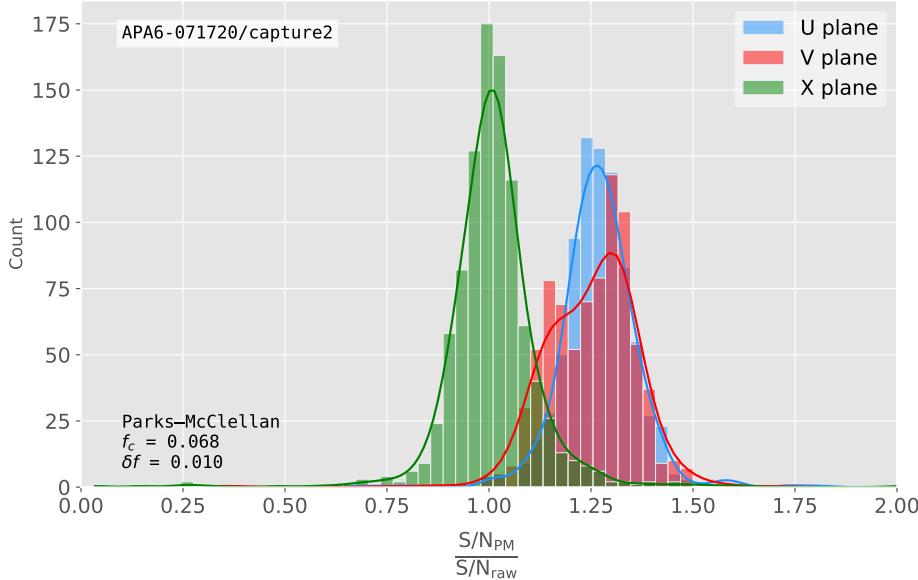


**Figure 6.5:** Relative change in the S/N for the ProtoDUNE-SP raw data capture `felix-2020-07-17-21:31:44`, using different values of the cutoff frequency  $f_c$  and the transition width  $\delta f$ . The optimal Chebyshev filters were applied using just the integer part of the coefficients given 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. 6.5), 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. 6.6 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 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.6:** 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.

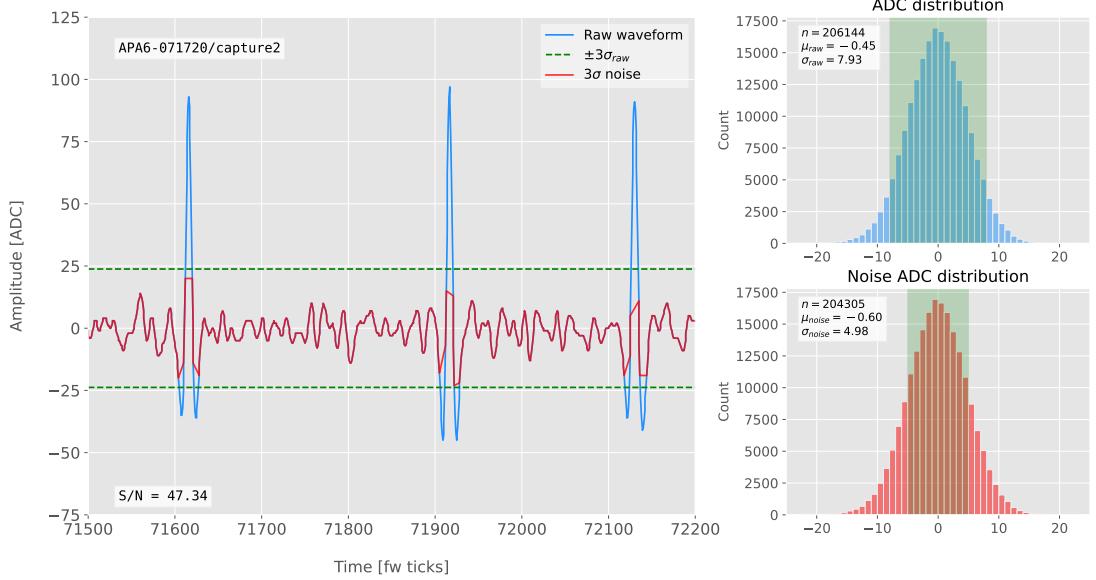
The results I obtained optimising the low pass filter with the Parks-McClellan method are promising. Nonetheless, the improvement found is rather marginal so I wondered if there could be an alternative approach to the filtering problem which yields better outputs. At this point, I found a possible alternative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

## 6.4 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 [63].

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

## 6.4. Matched filters



**Figure 6.7:** 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  $\pm \sigma_{\text{noise}}$

1387  $n(t)$ , the input signal can be written as:

$$x(t) = s(t) + n(t). \quad (6.4)$$

1388 Now, considering a linear time-invariant filter, whose impulse-response function I  
1389 will refer to as  $h(t)$ , one can write the output signal as:

$$\begin{aligned} y(t) &= x(t) * h(t) \\ &= (s(t) + n(t)) * h(t) \\ &= y_s(t) + y_n(t), \end{aligned} \quad (6.5)$$

1390 where  $y_s(t)$  and  $y_n(t)$  are simply the outputs of the filter due to the signal and the noise  
1391 components respectively.

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives

1392 The goal of the matched filter is to detect the presence of the signal  $s(t)$  in the input  
 1393 sample  $x(t)$  at a certain time  $t_0$ , which effectively means we need to maximise the S/N.  
 1394 This way, what one wants is to have a filter which gives a much bigger output when the  
 1395 known signal is present than when it is not. Putting it in other words, the instantaneous  
 1396 power of the signal output  $y_s(t)$  should be much larger than the average power of the  
 1397 noise output  $y_n(t)$  at some time  $t_0$ .

1398 For the case of the filtered signal, one can easily re-write it as an inverse Fourier  
 1399 transform:

$$y_s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega H(\omega)S(\omega)e^{i\omega t}, \quad (6.6)$$

1400 where  $H(\omega)$  and  $S(\omega)$  are the Fourier transforms of the impulse-response function (i.e.  
 1401 the transfer function of the filter) and of the input signal, respectively.

1402 Now focusing on the noise, we can use the Wiener-Khinchin theorem [64] to write  
 1403 the mean power of the noise after filtering as:

$$E|y_n(t)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega |H(\omega)|^2 S_n(\omega), \quad (6.7)$$

1404 where  $S_n(\omega)$  is the power spectral density of the noise.

1405 Having these, one can write the instantaneous S/N at time  $t_0$  as:

$$\begin{aligned} \left( \frac{S}{N} \right)_{t_0} &= \frac{|y_s|^2}{E|y_n(t)|^2} \\ &= \frac{1}{2\pi} \frac{\left| \int_{-\infty}^{\infty} d\omega H(\omega)S(\omega)e^{i\omega t_0} \right|^2}{\int_{-\infty}^{\infty} d\omega |H(\omega)|^2 S_n(\omega)}. \end{aligned} \quad (6.8)$$

1406 Once we have this expression, we need to find the upper limit of it to determine what  
 1407 would be the optimal choice for the transfer function. One can use the Cauchy-Schwarz  
 1408 inequality, which in the present case takes the form:

$$\left| \int_{-\infty}^{\infty} dx f(x)g(x) \right|^2 \leq \int_{-\infty}^{\infty} dx |f(x)|^2 + \int_{-\infty}^{\infty} dx |g(x)|^2, \quad (6.9)$$

## 6.4. Matched filters

1409 for any two analytical functions  $f(x)$  and  $g(x)$ . One can prove that making the choice:

$$\begin{aligned} f(x) &= H(\omega) \sqrt{S_n(\omega)} e^{i\omega t_0}, \\ g(x) &= \frac{S(\omega)}{\sqrt{S_n(\omega)}}, \end{aligned} \quad (6.10)$$

1410 leads to the following upper bound for the S/N:

$$\left( \frac{S}{N} \right)_{t_0} \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{|S(\omega)|^2}{S_n(\omega)}. \quad (6.11)$$

1411 From Eqs. (6.8), (6.9) and (6.10) one can also derive the form of the transfer function

1412 such that the upper bound is exactly reached [65]:

$$H(\omega) \propto \frac{S^*(\omega) e^{-i\omega t_0}}{S_n(\omega)}. \quad (6.12)$$

1413 From this last expression we can clearly see the way the matched filter acts. As the

1414 transfer function is proportional to the Fourier transform of the signal it will try to only

1415 pick the frequencies present in the signal [66].

1416 The matched filter transfer function can be greatly simplified if the input noise is

1417 Gaussian. In that case, the power spectral density of the noise is a constant, so it can be

1418 re-absorbed in the overall normalisation of the transfer function. Moreover, considering

1419 that the input signal is a real function, one can simply set  $S^*(\omega) = S(-\omega)$ , which gives:

$$H(\omega) \propto S(-\omega) e^{-i\omega t_0}. \quad (6.13)$$

1420 For a discrete signal, one can think of the input and impulse-response sequences as

1421 vectors of  $\mathbb{R}^N$ . Then, the matched filter tries to maximise the inner product of the signal

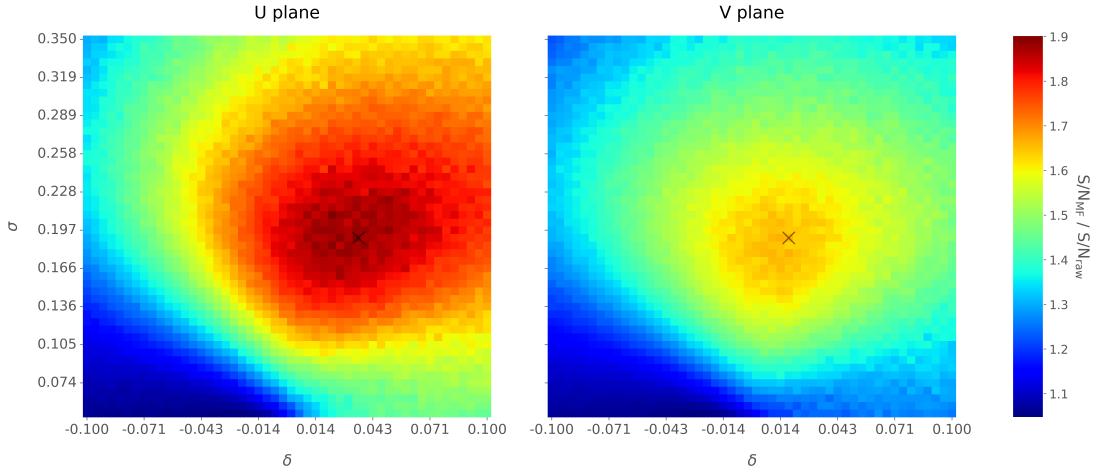
1422 and the filter while minimising the output due to the noise by choosing a filter vector

1423 orthogonal to the later. In the case of additive noise, that leads to the impulse-response

1424 vector:

$$h = \frac{1}{\sqrt{s^\dagger R_n^{-1} s}} R_n^{-1} s, \quad (6.14)$$

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.8:** 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. (6.17). The black crosses in both panels denote the location of the maximum ratio value.

1425 where  $s$  is a reversed signal template sequence of length  $N$  equal to the order of the filter  
 1426 and  $R_n$  is the covariance matrix associated with the noise sequence  $n$ . For the Gaussian  
 1427 noise case, the covariance matrix is simply the unit matrix, so the above expression  
 1428 simplifies again to:

$$h = \frac{s}{|s|}. \quad (6.15)$$

1429 For this first stage of the study, I use a definition of the S/N per channel given by:

$$\text{S/N} = \frac{\max [ADC] - \mu_{noise}}{\sigma_{noise}}, \quad (6.16)$$

1430 where the subscript *noise* refers to a subset of the data obtained by only taking into  
 1431 account waveform values within a  $\pm 3\sigma$  range around the mean of the data and  $\max [ADC]$   
 1432 is the maximum of the original waveform. This definition is further discussed in App.  
 1433 6.2, where I also show examples of its application to raw data and to a waveform filtered  
 1434 with the current low-pass FIR filter.

1435 To test whether this choice of filter is appropriate one needs to choose a signal  
 1436 template. As an example of how a matched filter would affect our signal, I simply took  
 1437 the filter coefficients to be the 32 ADC values around a signal peak present in the data.

## 6.4. Matched filters

1438 In Fig. 6.7 (left panel) I plotted a zoomed region for channel 7840 in the raw data capture  
1439 `felix-2020-07-17-21:31:44`, after applying the matched filter described before (blue  
1440 line). When compared to the raw and FIR filtered case (see App. 6.2), after applying  
1441 the match filter the standard deviation of the noise waveform (red line) decreases and at  
1442 the same time the signal peaks are enhanced. This leads to an improvement of the S/N  
1443 by a factor of 1.92 when compared to the raw waveform.

1444 In order to obtain the matched filter that is more suitable for our data, I explored  
1445 different configurations of signal templates. In order to perform this exploration, I  
1446 parametrised the signal using the bipolar function:

$$f(x) = -A(x + \delta) e^{-x^2/\sigma^2}, \quad (6.17)$$

1447 where the parameter  $\delta$  controls the asymmetry between the positive and negative peaks  
1448 and  $\sigma$  controls their width. The amplitude parameter  $A$  is set such that it keeps the  
1449 height of the biggest peak to be less than 200 ADC in absolute value.

1450 As this parametrisation is only adequate for bipolar signals I will focus exclusively  
1451 on the induction channels. Also, the optimal configurations I found for the U and V  
1452 plane will be kept separate, i.e. I will have two sets of coefficients that will be applied to  
1453 either the U and V planes of wires. I do so as I found this was the choice giving the  
1454 best performance. Even so, as I will discuss, the differences are not very pronounced. In  
1455 case it is not technically possible to separate channels in the firmware according to the  
1456 wire plane they come from and use different sets of filter coefficients for them, we can  
1457 just find a common unique set of coefficients. In such case, I do not expect our results  
1458 to change dramatically.

1459 In Fig. 6.8 I present the results of our parameter scan, for channels in the induction  
1460 planes U (left panel) and V (right panel). For each configuration of  $\sigma$  and  $\delta$  the resulting  
1461 matched filter was applied to all channels in the corresponding plane within the data  
1462 capture `felix-2020-07-17-21:31:44`, the S/N improvement was computed with respect  
1463 to the raw waveforms and then the S/N mean value was kept as a score for such filter.

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives

1464 One can see that the improvement obtained for the U plane is in general higher than the  
1465 one for the V plane. In any case, I got substantially higher ratios than the ones obtained  
1466 for the low-pass FIR filters. For the optimal configurations I attained improvements up  
1467 to a factor of 1.85 for the U plane and 1.65 for the V plane.

1468 The sets of optimal matched filter coefficients were obtained for the parameters  
1469  $\delta = 0.035$ ,  $\sigma = 0.191$  for the U plane and  $\delta = 0.018$ ,  $\sigma = 0.191$  for the V plane. I  
1470 show these two sets of coefficients in Fig. 6.9 (left panel). Also in Fig. 6.9 (right  
1471 panel) I plot the distribution of the S/N improvement after the optimal match filters  
1472 for the U and V were applied to the corresponding channels in the raw data capture  
1473 `felix-2020-07-17-21:31:44`. As mentioned before, the mean improvement achieved  
1474 for the U plane channels is slightly bigger than the one for the V channels. Note, however,  
1475 that the spread of the distribution for the V plane is also smaller than the one for the U  
1476 plane.

1477 I also performed a similar scan for the case of a low-pass FIR filter using the Parks-  
1478 McClellan algorithm. In that case, the parameters to check were the cutoff frequency  
1479 and the transition width of the filter. A summary of the results is given in App. 6.3.

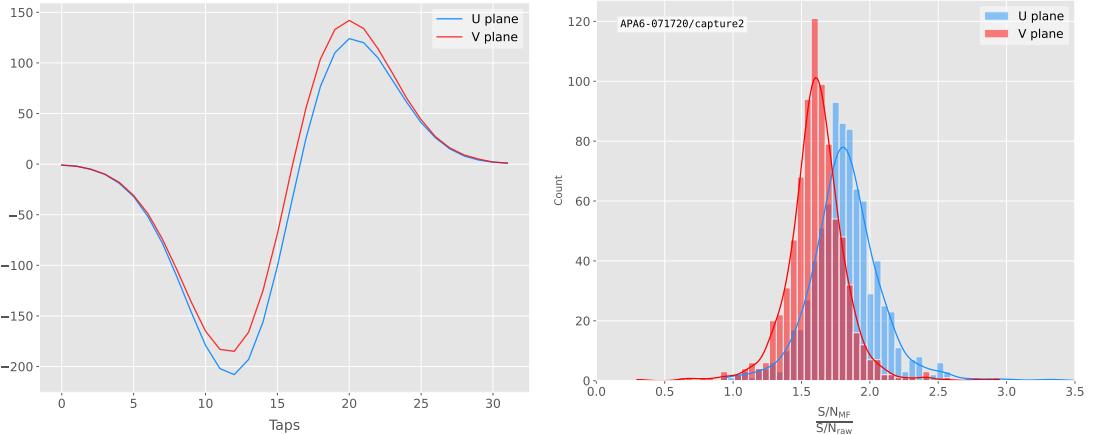
1480 Overall, one can see that the improvements on the S/N are much more significant in  
1481 the case of the matched filter than it is for the low-pass FIR filters. The analysis of this  
1482 and other raw data captures from ProtoDUNE-SP suggest that matched filters increase  
1483 the S/N of induction channels by a factor of 1.5 more than the optimal low-pass FIR  
1484 filters.

1485 Although these results are by themselves great points in favour of the matched  
1486 filter, more studies are needed to completely assess the robustness of this approach. I  
1487 proceeded then to test the matched filter with simulated data samples.

### 1488 6.5 Using simulated samples

1489 In order to further test the matched filter, the next step was to generate and process  
1490 data samples using *LArSoft* [67]. In this way, one can control the particle content of

## 6.5. Using simulated samples



**Figure 6.9:** 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. (6.17) 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.

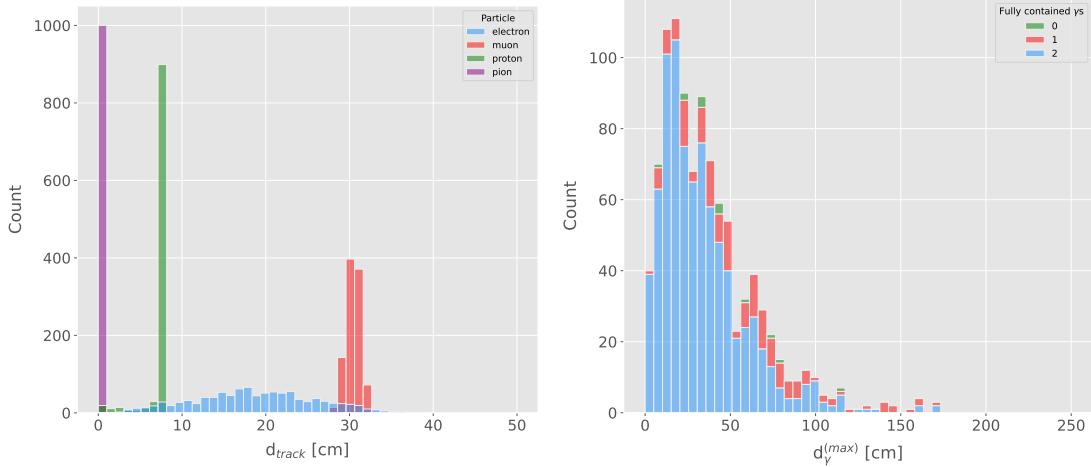
1491 the samples, the orientation of the tracks and their energy, and therefore see how the  
1492 matched filter behaves in various situations.

1493 To begin with, I prepared different monoenergetic and isotropic samples containing  
1494 a single particle per event. Each sample contains a different particle species, namely  
1495 electrons, muons, protons and neutral pions all with a kinetic energy of  $E_k = 100$  MeV.  
1496 I chose these because of the fairly different topologies they generate in the liquid argon,  
1497 ranging from shower-like to track-like. The procedure I followed to generate the samples  
1498 and process them is discussed in detail in App. ??.

1499 These were generated with the single particle gun and the Geant4 stage of the  
1500 *LArSoft* simulation [67] was performed with the standard configuration for the DUNE  
1501 FD 10kt module.

1502 For simplicity, I restricted the particles to start drifting in a single TPC volume  
1503 (in this case TPC 0), so I can focus exclusively on the signals coming from one APA.  
1504 The chosen kinetic energy for all the particles in my first trial is  $E_k = 100$  MeV, so a  
1505 necessary check is to see if all our tracks will be typically contained in one TPC volume.  
1506 Fig. 6.10 (left panel) shows the distributions of the track lengths in the liquid argon

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.10:** 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. 6.10 (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

## 6.5. 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.

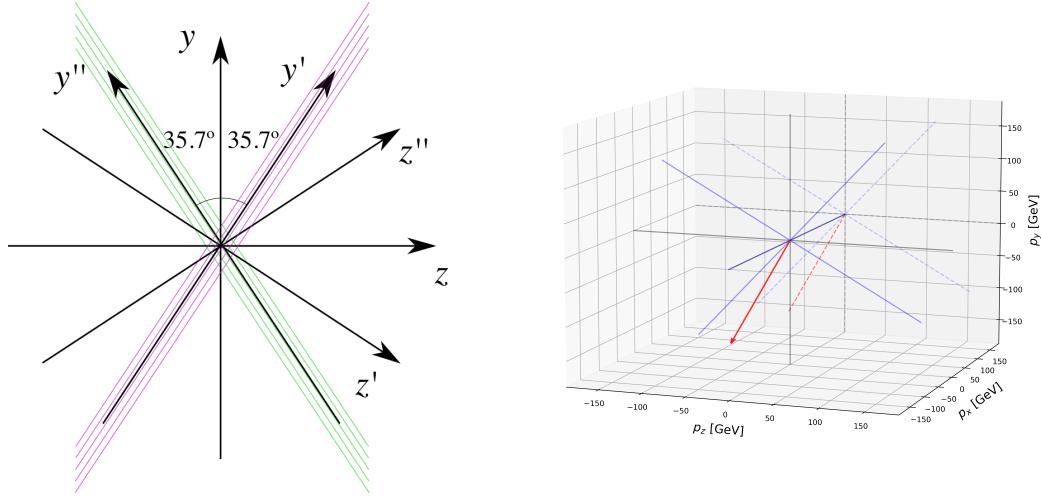
After the detector simulation stage, one needs to extract the no zero-suppressed noisy waveforms, along with their offline channel numbers, and store them in a certain format to be analysed later. To reduce the amount of data that will go for processing, I used the information from the Geant4 step of 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 6. Matched Filter approach to induction wire Trigger Primitives

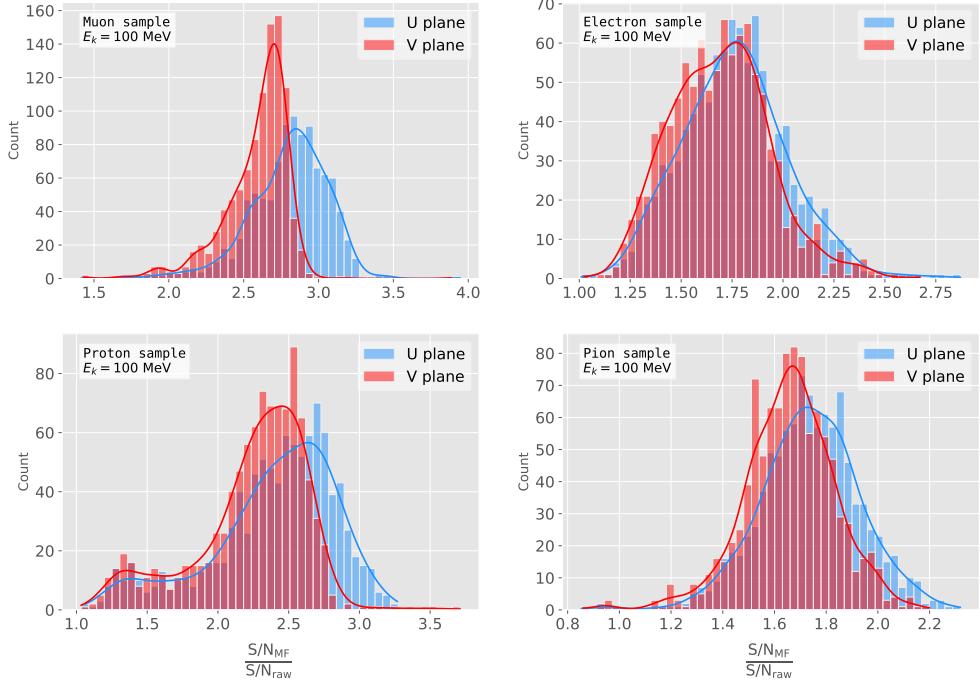


**Figure 6.11:** Left panel: schematic representation of the two new rotated reference frames used in this analysis (denoted as prime and double prime), viewed from the  $yz$  plane. The magenta stack of lines represent the wires in the  $U$  plane, whereas the green lines correspond to the wires in 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.

1553 U and V induction wires. Fig. 6.11 (left panel) shows a schematic representation of  
 1554 the original reference frame together with the two rotated ones (denoted by primed and  
 1555 double primed). This way, one can easily understand how parallel was a track to the  
 1556 wires in the two induction planes. Fig. 6.11 (right panel) shows a 3D representation of  
 1557 the momentum of a track (red arrow) in the original reference frame (black lines), along  
 1558 with the new reference frame for  $U$  wires (blue lines). I added the projection in the  $yz$   
 1559 plane of this three, to show the usefulness of the new reference frame to tell whether a  
 1560 track is parallel or normal to the wires in the induction plane.

1561 Fig. 6.12 shows the distribution of the average S/N improvement per event when one  
 1562 applies the optimal matched filters. I produced separate distributions for the channels  
 1563 in the  $U$  (red) and  $V$  (blue) induction wire planes. Notice that the S/N distributions  
 1564 for the track-like particles, i.e. muons (top left panel) and protons (bottom left panel),  
 1565 have significantly larger mean values than the distributions of the shower like particles,  
 1566 i.e. electrons (top right panel) and neutral pions (bottom right panel). An important

## 6.5. Using simulated samples



**Figure 6.12:** 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  $E_k = 100 \text{ MeV}$ .

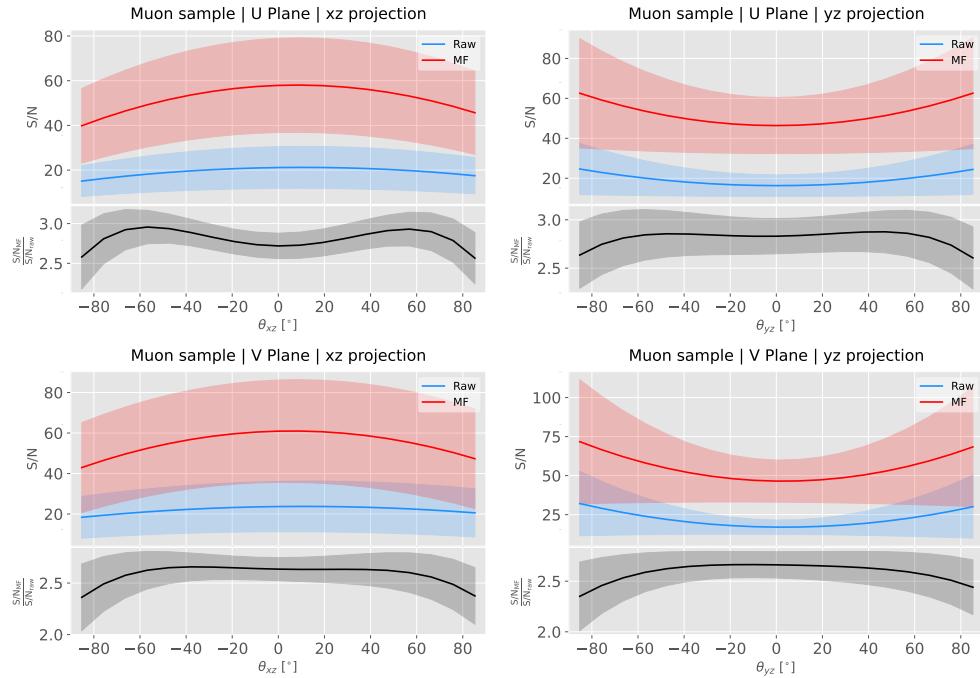
difference between these results and the ones seen before for ProtoDUNE data is that, overall, the improvements that I get for simulated data are bigger. This could be due either to the default noise model used in the *LArSoft* simulation or to the simulated hits having higher energy than the ones in 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:

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

$$(S/N_{raw})_{event} = \frac{\sum_{i=0}^{N_{chan}} (S/N_{raw})_i}{N_{chan}}.$$
(6.18)

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.13:** 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).

1577 However, for the ratio of the raw and filtered S/N (what I called the S/N improvement)  
 1578 per event I am not just taking the ratio of the previous two quantities but computing  
 1579 the average of the individual ratios per channel in the event:

$$\left( \frac{S/N_{fir}}{S/N_{raw}} \right)_{event} = \frac{\sum_{i=0}^{N_{chan}} \left( \frac{S/N_{fir}}{S/N_{raw}} \right)_i}{N_{chan}}, \quad (6.19)$$

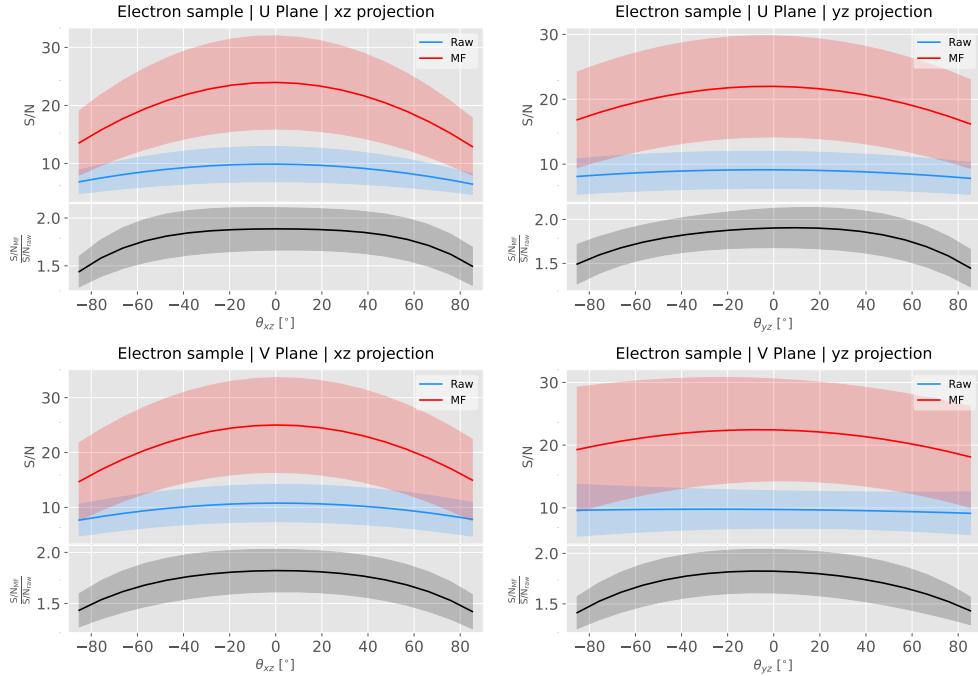
1580 and so:

$$\left( \frac{S/N_{fir}}{S/N_{raw}} \right)_{event} \neq \frac{(S/N_{fir})_{event}}{(S/N_{raw})_{event}}. \quad (6.20)$$

### 1581 6.5.1 Angular dependence

1582 Having these monoenergetic samples, one can also study the angular dependence of the  
 1583 performance of the matched filter. This is an important point, as it is a well established  
 1584 fact that for certain configurations (an extreme case configuration being signals normal

## 6.5. Using simulated samples



**Figure 6.14:** 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).

1585 to the wire plane and perpendicular to the induction wires at the same time) the S/N is  
 1586 much lower than average as the corresponding waveforms are severely distorted. In this  
 1587 sense, I am interested to see how the matched filter behaves for these cases and how the  
 1588 S/N improvement on those compare to the average.

1589 Fig. 6.13 shows the angular dependence of the S/N for the monoenergetic  $E_k =$   
 1590 100 MeV isotropic muons, for the different induction wire planes and projections. The  
 1591 angles for each event are given by the components of the initial value of the momentum  
 1592 of the particles, taking the angles of the projections on the  $xz$  and  $yz$  planes with respect  
 1593 to the  $z$  axis (more accurately, one needs to compute these angles twice for each event, a  
 1594 pair for the  $xy'z'$  coordinate system and the other for the  $xy''z''$ ). The top row shows the  
 1595 dependence on the angles corresponding to the U plane, i.e.  $\theta_{xz'}$  and  $\theta_{y'z'}$ , whereas the  
 1596 bottom row shows the angular dependence viewed from the V plane,  $\theta_{xz''}$  and  $\theta_{y''z''}$ . In  
 1597 each plot, the top subplot represents the mean values of the S/N for the raw (blue) and

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives

1598 matched filtered (red) signals, and the bottom subplot the averaged S/N improvement  
1599 (black). The solid lines represent the mean value obtained for the corresponding angular  
1600 value, whereas the semitransparent bands represent one standard deviation around the  
1601 mean at each point.

1602 As expected, the S/N is in general higher when tracks are parallel to the APA (i.e.  
1603  $\theta_{xz} \sim 0$ ) and lower when it is normal to the plane ( $\theta_{xz} \sim \pm 90^\circ$ ). In the same way, tracks  
1604 parallel to the wires ( $\theta_{yz} \sim \pm 90^\circ$ ) tend to have higher S/N than those perpendicular to  
1605 these ( $\theta_{yz} \sim \pm 0$ ).

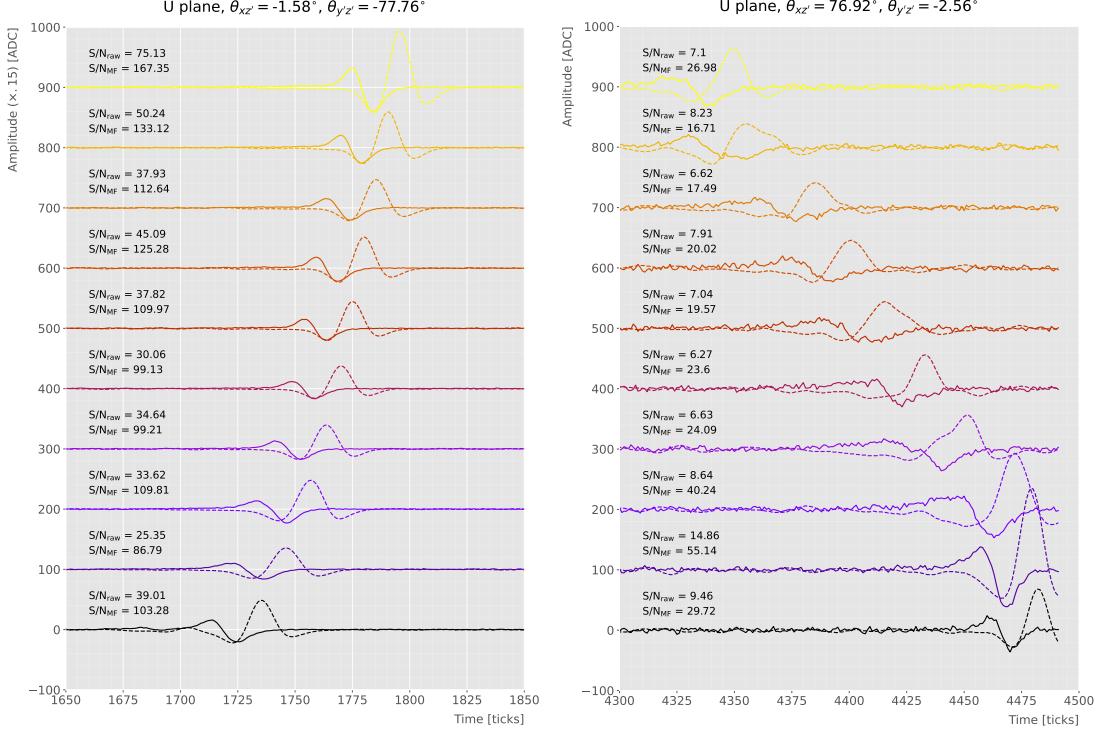
1606 Fig. 6.14 shows the corresponding angular dependence information for the  $E_k =$   
1607 100 MeV electrons sample. Notice that, in this case, the S/N behaviour discussed above  
1608 does not hold. A possible explanation can be that, because most hits in these events  
1609 are produced by the secondary particles generated in the EM shower, the signal peaks  
1610 whose S/N ratios were computed do not correspond to the directional information of  
1611 the primary electron.

### 1612 6.5.2 Distortion and peak asymmetry

1613 As a little case of study, I selected two of the simulated  $E_k = 100$  MeV monoenergetic  
1614 muon events. With respect to the U induction plane, one is parallel to the APA (low  
1615  $\theta_{xz'}$ ) and to the wires (high  $\theta_{y'z'}$ ) and the other is normal to the APA plane (high  $\theta_{xz'}$ )  
1616 and perpendicular to the wires (low  $\theta_{y'z'}$ ). As expected from the results on the angular  
1617 dependence discussed above, the former has a higher S/N (before and after the filtering)  
1618 when compared to the latter. An interesting thing to notice about these two samples  
1619 is that, even though one has a much bigger S/N than the other, it is the one with the  
1620 smallest S/N the one that got the biggest averaged S/N improvement. In Table 6.1  
1621 I included all the relevant parameters of these two  $E_k = 100$  MeV muon events I am  
1622 considering, namely, the angles with respect to the  $xy'z'$  reference frame, the values of  
1623 the S/N, the S/N improvement and also the so-called peak asymmetry  $\Delta_{peak}$  that I will  
1624 discuss next.

1625 One can try to understand better what is going on with these two events by looking

## 6.5. Using simulated samples



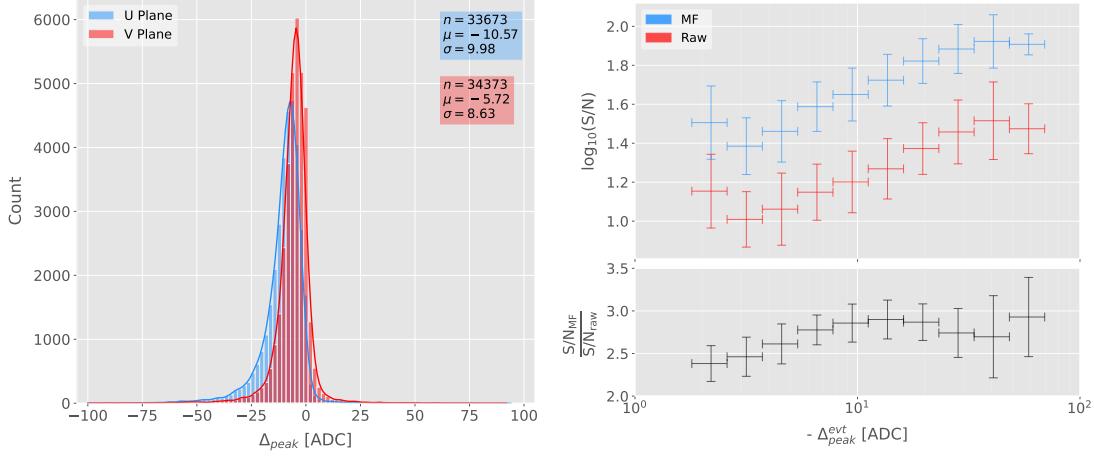
**Figure 6.15:** Selected consecutive waveforms corresponding to two monoenergetic  $E_k = 100$  MeV muon events, one is parallel to the APA and to the wires in the U plane (left panel) and the other is normal to the APA plane and perpendicular to the U plane wires (right panel). The solid lines represent 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 6.1:** 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.

	$\theta_{xz'} (\circ)$	$\theta_{yz'} (\circ)$	$S/N_{\text{raw}}$	$S/N_{\text{MF}}$	$\frac{S/N_{\text{MF}}}{S/N_{\text{raw}}}$	$\Delta_{\text{peak}} (\text{ADC})$
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

at the raw and filtered data from some of their active channels. Fig. 6.15 shows a selection of consecutive raw and filtered U plane waveforms from the event with high S/N (left panel) and the one with low S/N (right panel). Notice that to show both collections of waveforms at a similar scale I had to apply a factor of 0.15 to the waveforms with high S/N. Additionally, next to each waveform I included the values of the raw and matched filtered S/N for the corresponding channel. The first thing to notice in this plot

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.16:** 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.

is that the amplitude of the signal peaks from the normal track have a much smaller  
 amplitude, and also appear quite distorted when compared to the others. On the other  
 hand, although the matched filtered S/N is still smaller, the relative improvement is  
 bigger than in the parallel case.

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

$$\Delta_{peak} \equiv h_+ - h_-, \quad (6.21)$$

where both heights  $h_+$  and  $h_-$  are positive defined. Fig. 6.16 (left panel) shows the distribution of this peak asymmetry for all the waveforms corresponding to channels in the U (blue) and V (red) planes for the monoenergetic muon sample. One can see that these distributions are clearly shifted to negative values (with mean values  $\mu_\Delta^U = -10.57$  ADC and  $\mu_\Delta^V = -5.72$  ADC respectively). It is interesting to notice

## 6.5. Using simulated samples

that the peak asymmetry value of the sample with high S/N sits at the left tail of the distribution whereas the corresponding value of the sample with low S/N lies around the mean.

Now, one can try to correlate the peak asymmetry with the S/N and the S/N change per event. Fig. 6.16 (right panel) shows the result of comparing (minus) the mean peak asymmetry per event to the averaged raw (red) and matched filtered (blue) S/N per event (top subplot). The horizontal lines sit at the mean value obtained in the fit and represent the width of the  $-\Delta_{peak}$  bins used, while the vertical lines indicate one standard deviation around that mean value. Notice that, when taking decimal logarithm on both, there is an approximate linear relation between these quantities, except for peak asymmetry values bigger than  $-5$  ADC where the S/N remains constant.

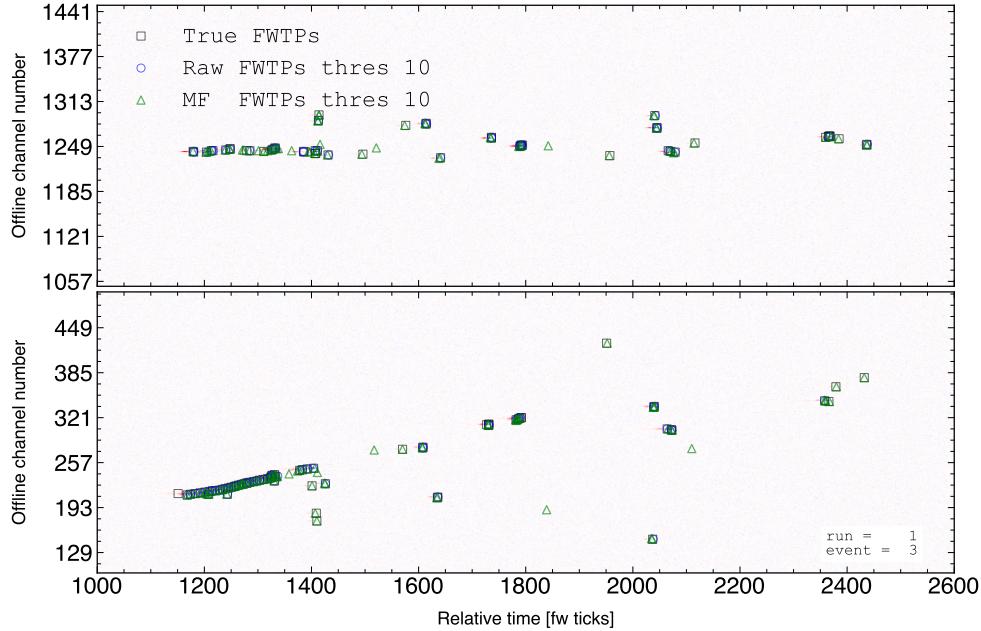
Also, in the bottom subplot of Fig. 6.16 (right panel) I show the relation between the peak asymmetry and the mean S/N improvement. In this case, one see that there is a maximum at  $\Delta_{peak} \sim -10$  ADC. As mentioned previously, this is also the value of the mean of the peak asymmetry distribution. In fact, it is expected that our filter favours the signal peaks with the most common values of the peak asymmetry, as this was one of the features I target in our filter coefficient optimisation through the parameter  $\delta$ .

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

### 6.5.3 Hit sensitivity

One of the advantages of the matched filter, directly related to increasing the S/N, is the capability of picking hits that before fell below the threshold. For instance, Fig. 6.17 shows the raw ADC data from an example event (electron,  $E_k = 100$  MeV) with the produced true hits superimposed (black boxes), together with the hits produced by the standard hit finder chain (blue circles), i.e. using the current FIR filter, and the hits

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.17:** 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).

1672 obtained using the matched filters (green triangles). Both the standard and the matched  
 1673 filter hit finders were run with a threshold of 10 ADC. Notice that the standard hits  
 1674 match well the true ones at the initial part of the event (where we have a track-like  
 1675 object), but they miss most of the hits produced by the EM shower at later times. On  
 1676 the other hand, the hits produced with the matched filter have a better agreement with  
 1677 the true hits even for the more diffuse shower activity.

1678 Notwithstanding that now I get more hits with this combination of matched filter  
 1679 and low threshold as a results of the enhancement of the signal peaks relative to the  
 1680 noise level, it is also true that I pick some spurious hits not related to any real activity  
 1681 if one lowers the thresholds too much. Therefore, some optimisation of the threshold is  
 1682 needed. Basically one will need to make a trade-off between precision and sensitivity.

1683 Having this in mind, I tried to compare the produced hits one gets from the standard  
 1684 hit finder and the ones resulting from applying the matched filter with the true hits.  
 1685 By running the hit finders on our samples with different values of the threshold one

## 6.5. Using simulated samples

1686 can understand, for instance, how low one can set the threshold without getting mostly  
1687 spurious hits and then evaluate the gains obtained from this.

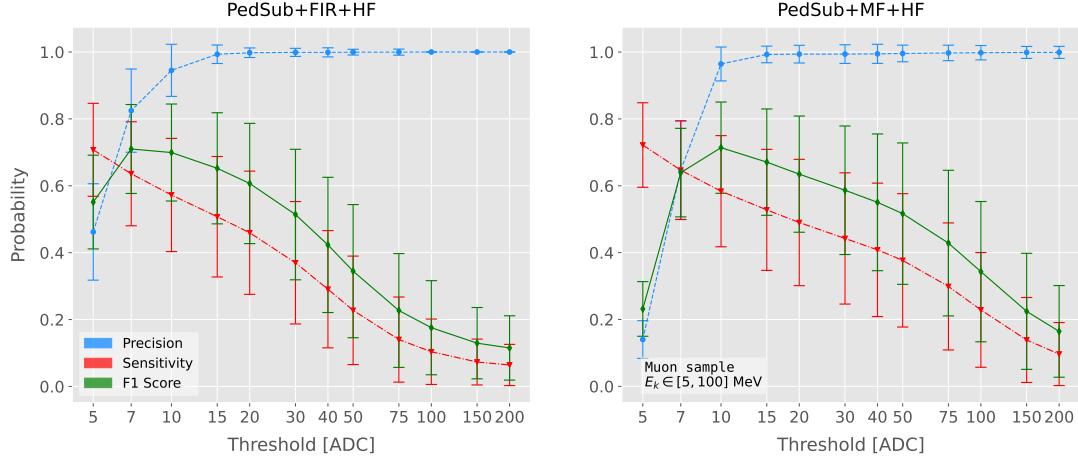
1688 Because now I am also interested in seeing how the hit sensitivity changes with the  
1689 energy, I prepared new isotropic samples with the same types of particles as before  
1690 (muons, electrons, protons and neutral pions) but with a flat kinetic energy distribution  
1691 ranging from 5 to 100 MeV.

1692 In order to estimate the hit sensitivity, given a certain sample, one needs to recover  
1693 the set of true hits to be able to compare these with the ones produced. To do so,  
1694 a modification in the procedure I was using to extract the raw waveforms is needed.  
1695 For this kind of study I run the detector simulation in two steps, first I produce the  
1696 waveforms without noise and extract them in the same format I used for the raw data,  
1697 then the noise is added and the noisy waveforms are then written to a file as well.

1698 To have a better comparison between the true hits and the ones produced from  
1699 the raw waveforms after applying the two filters, I applied also the FIR filter and the  
1700 matched filters to the noiseless waveforms and then I run the hit finder with a minimal  
1701 threshold (in this case I used 1 ADC) on these noiseless filtered waveforms. In this way  
1702 I generated two sets of true hits, I will refer to them as standard true hits (with the  
1703 current/default FIR filter) and matched filter true hits respectively. This allows a more  
1704 precise matching between the different groups of hits produced, as it will account for  
1705 any delays and distortions introduced by the FIR and the matched filters.

1706 In the case of the raw waveforms (with noise), I run the hit finder on them, with  
1707 different values of the threshold, after applying either the FIR or the matched filters. I  
1708 will name them simply standard hits and matched filter hits respectively. Then, I match  
1709 the generated hits to the true hits (the standard hits with the standard true hits and  
1710 the matched filter hits with the matched filter true hits). The matching is performed by  
1711 comparing the channel number and the timestamp of the hits. To count as a match,  
1712 I require that all hits with the same channel number and timestamp have overlapping  
1713 hit windows, i.e. the time windows between their hit end and hit start times need to  
1714 overlap. If more than one hit in one of the groups have hit overlap with the same hit in

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.18:** 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.

1715 the other group I only count the hit with closer hit peak time value.

1716 The generation of the samples, the procedure to produce the standard hits (with the  
1717 default FIR filter) and matched filter hits and the matching of these with the true hits  
1718 is described in detail in App. ??.

1719 To quantify the performance of the two hit finder approaches, I use a classical method  
1720 from statistical classification known as confusion matrix [68]. This is basically a way of  
1721 sorting the outputs of a binary classifier, considering the true values of the classification  
1722 and the predicted values. It divides the outputs in four categories: true positive (TP,  
1723 both true and predicted values are 1), false negative (FN, true value is 1 but predicted  
1724 is 0), false positive (FP, true value is 0 but predicted is 1) and true negative (TN, both  
1725 true and predicted values are 0)).

1726 The contents of the confusion matrix allow us to compute other derived scores to  
1727 judge the performance of our classifiers. In this study, I will make use of three of these  
1728 metrics, namely the precision or positive predictive value:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (6.22)$$

## 6.5. Using simulated samples

1729 the sensitivity or true positive rate:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (6.23)$$

1730 and the  $F_1$  score [69]:

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (6.24)$$

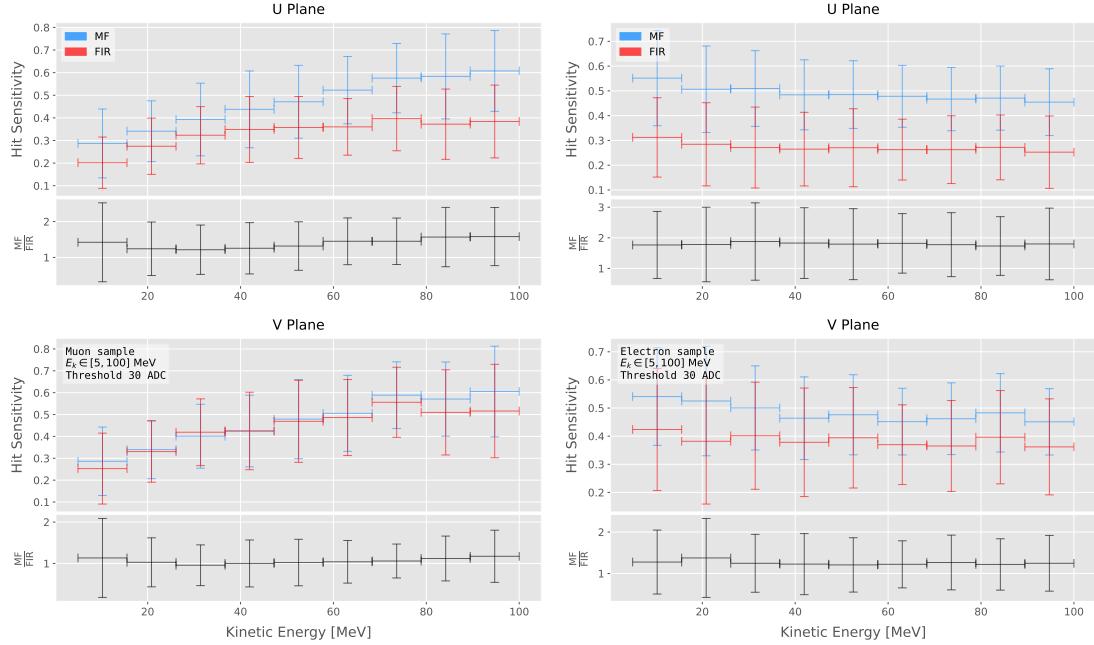
1731 which is the harmonic mean of the precision and the sensitivity.

1732 In our specific case I am not going to make use of the true negative value, as its  
1733 definition in this context can be ambiguous because one does not have clear instances in  
1734 the classification process. This way, I will only count the number of true positives as the  
1735 total amount of hits I can match between true and raw populations, the number of false  
1736 negatives will be the number of missing true hits and the false positive the number of  
1737 hits which do not match any true hit.

1738 In Fig. 6.18 I show the precision (blue), sensitivity (red) and  $F_1$  (green) scores I  
1739 obtained for different values of the threshold used in the hit finder for the case of the  
1740 muon sample. Because the matched filters are only applied to induction channels, I only  
1741 consider here hits coming from the U and V planes. The panel on the left corresponds  
1742 to the scores I got when I ran the hit finder on the FIR filtered waveforms, whereas the  
1743 right panel contains the scores for the matched filter case. The points are centered at  
1744 the threshold value used and represent the mean value obtained for each score using all  
1745 the generated events, while the error bars indicate one standard deviation around the  
1746 mean value.

1747 One can see that the precision for the matched filter case is lower when the thresholds  
1748 are very low, as the noise baseline is slightly amplified, but then rises to high values  
1749 quicker than for the FIR case. The other difference one can spot is that the sensitivity  
1750 in the FIR case starts dropping faster at around the same threshold values where the  
1751 precision stabilizes around 1, while in contrast for the matched filter this rapid decrease  
1752 starts at higher threshold values. A similar scan for the same thresholds was performed  
1753 for the electron sample in the same energy range, yielding similar results.

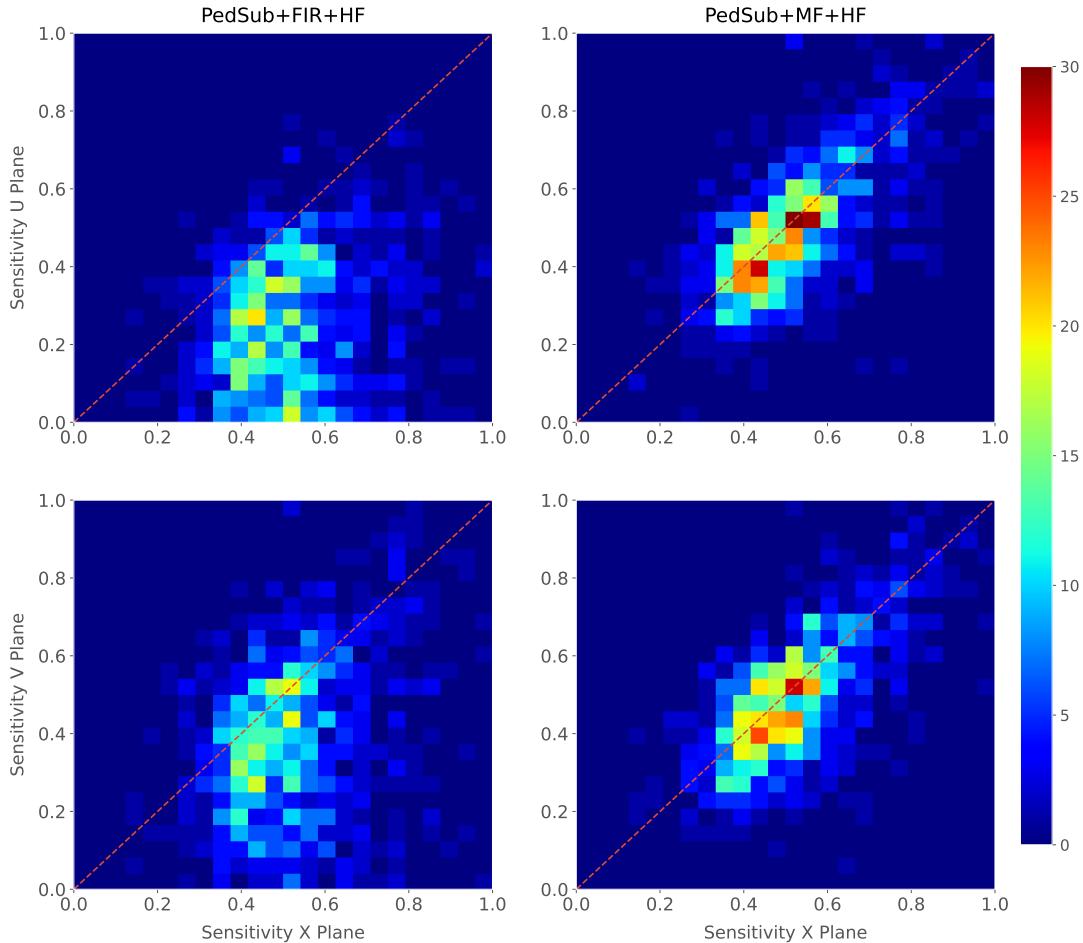
## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.19:** 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.

1754 In Fig. 6.19 I show the averaged hit sensitivity versus the kinetic energy of the  
 1755 events, both for the matched filter hits (blue) and the standard hits (red). The left  
 1756 panel corresponds to the muon sample, whereas the one on the right corresponds to the  
 1757 electron sample, both with kinetic energies between 5 and 100 MeV. In each panel the  
 1758 top plot corresponds to hits in the U plane, while the bottom plot contains the same  
 1759 information for the V plane. Each plot contains two subplots, the one on the top shows  
 1760 the hit sensitivity values for the matched filter and standard hits separate, while the  
 1761 bottom subplot depicts the ratio between the matched filter and standard sensitivities.  
 1762 The horizontal lines are placed at the mean value obtained in the fit and represent the  
 1763 width of the  $E_k$  bins used, while the vertical error bars indicate one standard deviation  
 1764 around that mean value. In both cases the threshold used was 30 ADC, as I required  
 1765 the precision to be higher than 0.99 for both matched filter and standard cases.

## 6.5. Using simulated samples

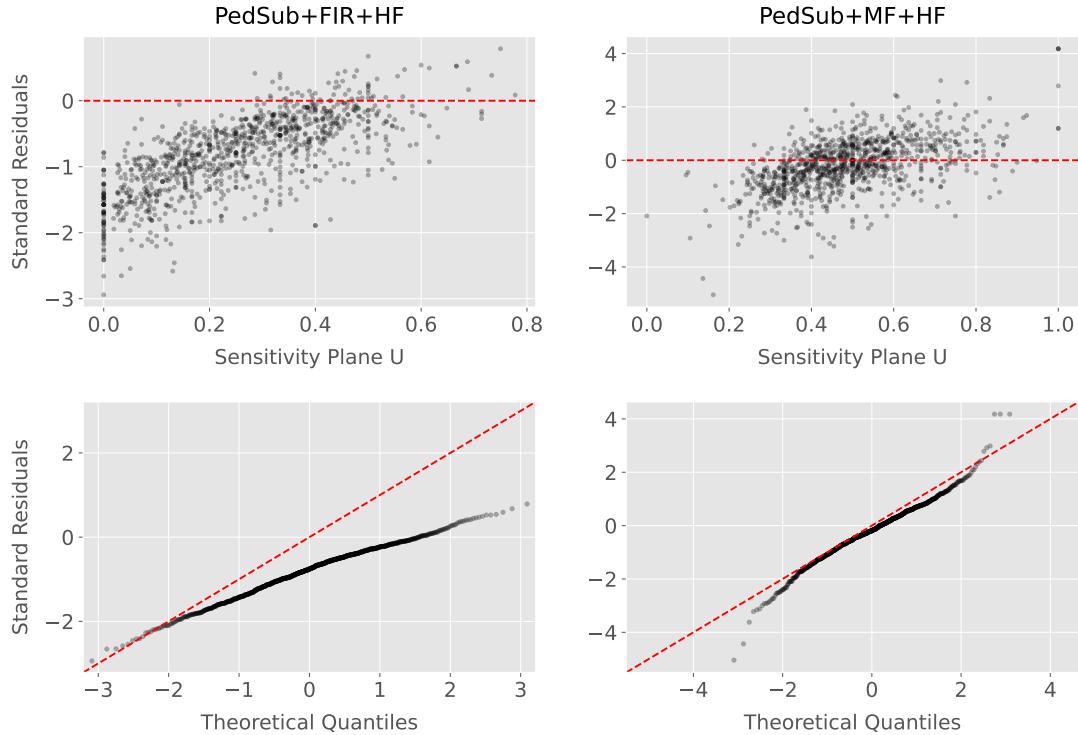


**Figure 6.20:** Distributions of the hit sensitivity in the U (top panels) and V (bottom panels) planes versus the hit sensitivity in the X plane, both for the standard hits (left panels) and the matched filter hits (right panels), in the case of the electron sample and a threshold of 30 ADC.

1766 One can see that, in general, the improvements are better for the U than for the V  
 1767 plane. While for the U channels I achieved a mean improvement of 50% and 80% for  
 1768 muons and electrons respectively, the improvement in the V plane is stalled at 10% and  
 1769 25%. Nevertheless, if I look at the sensitivities for the matched filter hits in both planes  
 1770 one can see these have similar mean values for each energy bin, while on the contrary  
 1771 for the standard hits the sensitivity remains relatively high for the V plane. This way, it  
 1772 looks there was a less significant gain because the hit sensitivity was already high.

1773 Another interesting observation is the different behaviors for muons and electrons.  
 1774 While hit sensitivity for muons grows significantly with energy, in the case of electrons

## Chapter 6. Matched Filter approach to induction wire Trigger Primitives



**Figure 6.21:** 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.

1775 this slightly decreases the higher the kinetic energy of the event is. In any case, when it  
 1776 comes to the improvement on the sensitivities, this remains almost constant in all cases.

1777 Furthermore, we can look at how the concurrence of hits between the different wire  
 1778 planes has changed. For any given event, I expect to have a similar number of hits in the  
 1779 three planes. As the ionisation electrons need to cross the  $U$  and  $V$  planes prior to reach  
 1780 the collection plane  $X$  they will induce current in those wire planes. A way to check the  
 1781 concurrence of hits across planes is looking at the relation between the hit sensitivities  
 1782 for each individual event. One cannot expect the sensitivities to be exactly equal across  
 1783 planes, but ideally they should be normally distributed around the diagonal.

1784 Fig. 6.20 shows the hit sensitivity in the  $U$  (top panels) and  $V$  (bottom panels)  
 1785 planes versus the hit sensitivity in the  $X$  plane, for the case of the standard hits (left  
 1786 panels) and the matched filter hits (right panels). All plots were generated for the

## 6.5. Using simulated samples

1787 electron sample and a threshold of 30 ADC. From these one can see a clear trend,  
1788 when I use the standard hit finder chain the sensitivities in the induction planes are  
1789 systematically lower than the hit sensitivity in the X plane, i.e. most of the points sit  
1790 below the diagonal (red dashed line). In contrast, when the matched filters are applied,  
1791 the majority of the events are distributed around the diagonal. This points out that the  
1792 concurrence of hits across planes has improved.

1793 To exemplify the improvement I obtained, one can consider the residuals of the hit  
1794 sensitivities for the X and U planes. Assuming the diagonal hypothesis, i.e. given a  
1795 dataset of the form  $(x, y)$  for any  $x$  I take the predicted  $y$  value to be equal to the value  
1796 of  $x$ , I can compute the standard residuals for the hit sensitivities in U given the ones for  
1797 X. In Fig. 6.21 (top panels) I show these standard residuals against the corresponding  
1798 values of the hit sensitivity in the U plane, for our electron sample with kinetic energy  
1799 between 5 and 100 MeV. If I compare the scatter points in the case of the standard  
1800 hits (left panel) and the matched filter hits (right panel), I see that the residuals of the  
1801 standard hit finder case follow a certain pattern and their mean deviates from 0.

1802 To see clearly if the residuals are normally distributed, in Fig. 6.21 (bottom panels)  
1803 I plot the corresponding quantile-quantile plot for both the standard (left panel) and  
1804 matched filter (right panel) standard residuals. One can clearly see that the points for  
1805 the standard case follow a strongly non-linear pattern, suggesting that the residuals  
1806 do not follow a normal distribution. In contrast, for the matched filter hits the points  
1807 conform to a roughly linear path, implying that in this case the normality condition is  
1808 fulfilled.

1809 All these results hint at the fact that the concurrence of hits across the wire planes  
1810 can be strengthened by applying the matched filters.



1811 Chapter 7

1812 DM searches with neutrinos from  
1813 the Sun

1814 7.1 Motivation

1815 The idea of detecting neutrino signals coming from the Sun’s core to probe DM is not new.  
1816 The main focus of these searches has usually been high-energy neutrinos originated from  
1817 DM annihilations into heavy particles [70–73], although recent studies have proposed to  
1818 look at the low-energy neutrino flux arising from the decay of light mesons at rest in the  
1819 Sun [74–77] previously thought undetectable.

1820 In this chapter I try to demonstrate the capability of DUNE to constrain different  
1821 DM scenarios. I used the neutrino fluxes arising from DM annihilations in the core  
1822 of the Sun to compute the projected limits that DUNE would be able to set on the  
1823 annihilation rates in the Sun and the DM scattering cross sections.

1824 7.2 Gravitational capture of DM by the Sun

1825 The Sun and the centre of the Earth are possible sources of DM annihilations, specially  
1826 interesting because of their proximity. Their gravitational attraction ensured the capture  
1827 of DM from the local halo through repeated scatterings of DM particles crossing them.

## Chapter 7. DM searches with neutrinos from the Sun

1828 Only neutrinos produced from DM annihilations can escape the dense interior of these  
1829 objects. Therefore, neutrino telescopes are the most useful experimental layouts to  
1830 pursue DM searches from their cores.

1831 The neutrino flux from DM annihilations inside the Sun depends on the DM capture  
1832 rate, which is proportional to the DM scattering cross section, and the annihilation rate,  
1833 which is proportional to the velocity-averaged DM annihilation cross-section. The total  
1834 number of DM particles inside the Sun follows the Boltzmann equation [74]:

$$\frac{dN_{DM}}{dt} = C_{\odot} - A_{\odot} N_{DM}^2, \quad (7.1)$$

1835 where  $C_{\odot}$  and  $A_{\odot}$  are the total Sun DM capture and annihilation rates respectively.  
1836 In this expression I neglected the evaporation term, proportional to  $N_{DM}$ , which only  
1837 contribute for  $m_{DM} \lesssim 4$  GeV [78]. As the current threshold of neutrino telescopes is  
1838 a few GeV, this region falls below the probed range but can be important in future  
1839 low-energy projects.

1840 This equation has an equilibrium solution:

$$N_{DM}^{eq} = \sqrt{\frac{C_{\odot}}{A_{\odot}}}, \quad (7.2)$$

1841 which represents the amount of DM inside the Sun if the capture and annihilation have  
1842 reached equilibrium. As the Sun is approximately 4.6 Gyr old, it is usually assumed that  
1843 equilibrium has been achieved. Therefore, the anomalous neutrino flux from the Sun  
1844 would only depend on the DM scattering cross section, enabling us to set limits on this  
1845 quantity. If one does not assume equilibrium, some assumptions on the DM annihilation  
1846 cross section are necessary to extract predictions from neutrino signals.

1847 Here, I am going to consider three possible scenarios for the DM interactions: DM  
1848 scattering off electrons, spin-dependent (SD) and spin-independent interactions off nuclei.  
1849 For the case of these last two, the cross sections will be given in terms of the SD and  
1850 SI elastic scattering DM cross section off protons (assuming that DM interactions off

## 7.2. Gravitational capture of DM by the Sun

1851 protons and neutrons are identical),  $\sigma_p^{\text{SD}}$  and  $\sigma_p^{\text{SI}}$ , as [4, 74]:

$$\sigma_i^{\text{SD}} = \left( \frac{\tilde{\mu}_{A_i}}{\tilde{\mu}_p} \right)^2 \frac{4(J_i + 1)}{3J_i} |\langle S_{p,i} \rangle + \langle S_{n,i} \rangle|^2 \sigma_p^{\text{SD}}, \quad (7.3)$$

$$\sigma_i^{\text{SI}} = \left( \frac{\tilde{\mu}_{A_i}}{\tilde{\mu}_p} \right)^2 A_i^2 \sigma_p^{\text{SI}}, \quad (7.4)$$

1852 where  $\tilde{\mu}_{A_i}$  is the reduced mass of the DM-nucleus  $i$  system,  $\tilde{\mu}_p$  is the reduced mass of  
 1853 the DM-proton system,  $A_i$  and  $J_i$  the mass number and total angular momentum of  
 1854 nucleus  $i$  and  $\langle S_{p,i} \rangle$  and  $\langle S_{n,i} \rangle$  the expectation value of the spins of protons and neutrons  
 1855 averaged over all nucleons, respectively (see Ref. [79] for a review on spin expectation  
 1856 values).

1857 Since the Sun is mainly composed of Hydrogen, the capture of DM from the halo  
 1858 is expected to occur mainly through spin-dependent scattering. However, since the  
 1859 spin-independent cross section is proportional to the square of the atomic mass, heavy  
 1860 elements can contribute to the capture rate (even though they constitute less than 2%  
 1861 of the mass of the Sun). Heavy elements can also contribute to the spin-dependent cross  
 1862 section if the DM has also momentum-dependent interactions.

1863 DM particles can get captured by the Sun if after repeated scatterings off solar  
 1864 targets their final velocity is lower than the escape velocity of the Sun. In the limit of  
 1865 weak cross sections, this capture rate can be approximately written as [4]:

$$C_{\odot}^{\text{weak}} = \sum_i \int_0^{R_{\odot}} dr 4\pi r^2 \int_0^{\infty} du_{\chi} \frac{\rho_{\chi}}{m_{\chi}} \frac{f_{v_{\odot}}(u_{\chi})}{u_{\chi}} \omega(r) \int_0^{v_e(r)} dv R_i^-(\omega \rightarrow v) |F_i(q)|^2, \quad (7.5)$$

1866 where the summation extends over all possible nuclear targets. In this expression,  $R_{\odot}$   
 1867 is the radius of the Sun,  $\rho_{\chi}$  is the local DM density,  $m_{\chi}$  the mass of the DM particle,  
 1868  $f_{v_{\odot}}(u_{\chi})$  the DM velocity distribution seen from the Sun's reference frame,  $R_i^-(\omega \rightarrow v)$   
 1869 is the differential rate at which a DM particle with velocity  $v$  scatters a solar target of  
 1870 mass  $m_i$  to end up with a velocity  $\omega$  and  $|F_i(q)|$  is the nuclear form factor of target  $i$ .

1871 The differential scattering rate takes a rather simple form when considering velocity-

## Chapter 7. DM searches with neutrinos from the Sun

1872 independent and isotropic cross sections. In that case, this quantity is given by [4, 80]:

$$R_i^-(\omega \rightarrow v) = \frac{2}{\sqrt{\pi}} \frac{\mu_{i,+}^2}{\mu_i} \frac{v}{\omega} n_i(r) \sigma_i \left[ \chi(-\alpha_-, \alpha_+) + \chi(-\beta_-, \beta_+) e^{\mu_i(\omega^2 - v^2)/u_i^2(r)} \right], \quad (7.6)$$

1873 where  $\mu_i$  is the ratio between the DM mass and the mass of target  $i$ ,  $\mu_{i,\pm}$  is defined as:

$$\mu_{i,\pm} \equiv \frac{\mu_i \pm 1}{2}, \quad (7.7)$$

1874  $n_i(r)$  is the density profile of target  $i$  in the solar medium,  $u_i(r)$  is the most probable

1875 velocity of target  $i$  given by:

$$u_i(r) = \sqrt{\frac{2T_\odot(r)}{m_i}}, \quad (7.8)$$

1876 where  $T_\odot(r)$  is the temperature of the Sun, the quantities  $\alpha_\pm$  and  $\beta_\pm$  are defined as:

$$\alpha_\pm \equiv \frac{\mu_{i,+}v \pm \mu_{i,-}\omega}{u_i(r)}, \quad (7.9)$$

$$\beta_\pm \equiv \frac{\mu_{i,-}v \pm \mu_{i,+}\omega}{u_i(r)}, \quad (7.10)$$

1877 and the function  $\chi(a, b)$  is a Gaussian integral of the form:

$$\chi(a, b) \equiv \int_a^b dx e^{-x^2}. \quad (7.11)$$

1878 Finally, if one assumes the DM halo velocity distribution in the galactic rest frame

1879 to be a Maxwell-Boltzmann distribution, one can write the halo velocity distribution for

1880 an observer moving at the speed of the Sun with respect to the DM rest frame as:

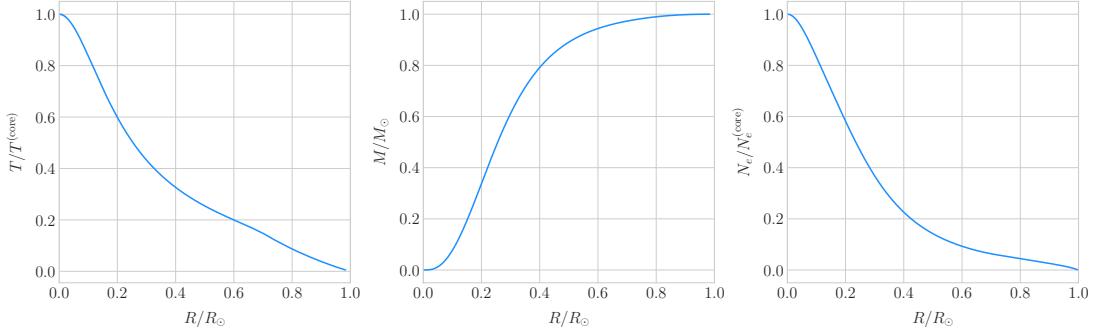
$$f_{v_\odot}(u_\chi) = \sqrt{\frac{3}{2\pi}} \frac{u_\chi}{v_\odot v_d} \left( e^{-\frac{3(u_\chi - v_\odot)^2}{2v_d^2}} - e^{-\frac{3(u_\chi + v_\odot)^2}{2v_d^2}} \right), \quad (7.12)$$

1881 where:

$$\omega^2 = u_\chi + v_e(r)^2, \quad (7.13)$$

1882 is the DM velocity squared,  $v_\odot$  the relative velocity of the Sun from the DM rest frame

## 7.2. Gravitational capture of DM by the Sun



**Figure 7.1:** 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 [3].

1883 and  $v_d \simeq \sqrt{3/2}v_\odot$  the velocity dispersion.

1884 For the case of strong scattering cross section, Eq. (7.5) ceases to be valid, as it  
 1885 escalates indefinitely with the cross section. In that limit, the capture rate saturates to  
 1886 the case where the probability of interaction is equal to one, which can be written as:

$$C_\odot^{\text{geom}} = \pi R_\odot^2 \left( \frac{\rho_\chi}{m_\chi} \right) \langle v \rangle \left( 1 + \frac{3}{2} \frac{v_e^2(R_\odot)}{v_d^2} \right) \xi(v_\odot, v_d), \quad (7.14)$$

1887 where  $v_d = \sqrt{8/3\pi}v_d$  is the mean velocity in the DM rest frame and the factor  $\xi(v_\odot, v_d)$   
 1888 accounts for the suppression due to the motion of the Sun:

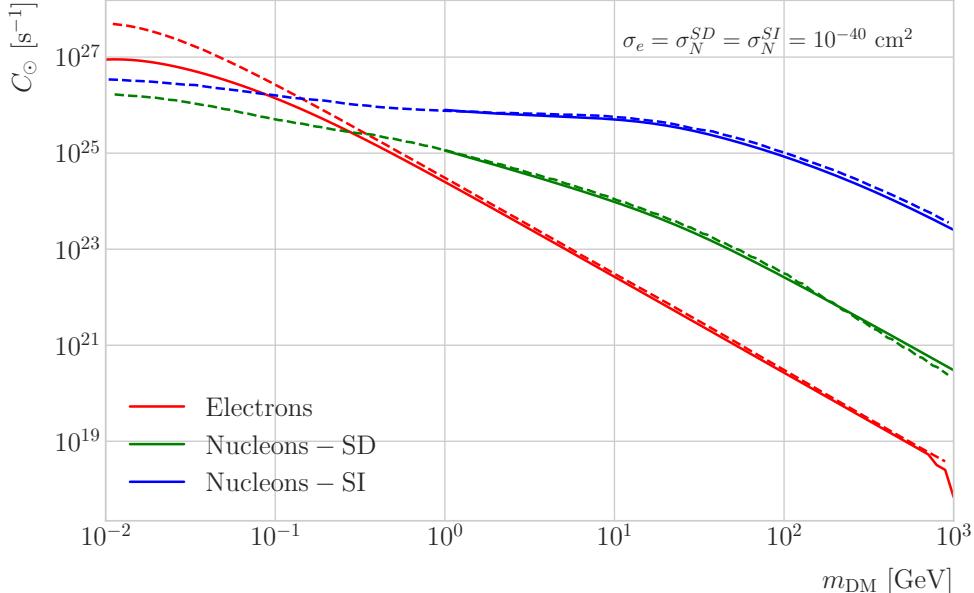
$$\xi(v_\odot, v_d) = \frac{v_d^2 e^{-\frac{3v_\odot^2}{2v_d^2}} + \sqrt{\frac{\pi}{6}} \frac{v_d}{v_\odot} (v_d^2 + 3v_e^2(R_\odot) + 3v_\odot^2) \text{Erf} \left( \sqrt{\frac{3}{2}} \frac{v_\odot}{v_d} \right)}{2v_d^2 + 3v_e^2(R_\odot)}. \quad (7.15)$$

1889 Having these into account, one can write the total capture rate as a combination of  
 1890 both contributions, allowing a smooth transition between the two, as:

$$C_\odot = C_\odot^{\text{weak}} \left( 1 - e^{C_\odot^{\text{geom}}/C_\odot^{\text{weak}}} \right). \quad (7.16)$$

1891 I computed the capture rate from Eq. (7.16) in the case of interactions with  
 1892 electrons. To do so, I used the standard solar model BS2005-OP [3]. Fig. 7.1 shows the

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.2:** 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. [4]. All the rates are shown for a choice of scattering cross section of  $\sigma_i = 10^{-40} \text{ cm}^2$ .

1893 three parameters from the solar model that are needed for the computation, the solar

1894 temperature (left panel), mass (central panel) and electron density (right panel) profiles.

1895 For the case of the interactions off nuclei, the computations are more convoluted

1896 as one needs to add up the contributions of the different most abundant nuclei in

1897 the Sun. Also, in contrast to the electron scenario where the form factor is trivially

1898  $|F_e(q)|^2 = 1$ , for any nucleus  $i$  one would need to consider some appropriate nuclear

1899 density distribution (either a Gaussian approximation, a Woods-Saxon distribution, etc)

1900 which would complicate the calculations even further.

1901 That is the reason why, at this stage of our study, I decided to take an alternative

1902 approach to the computation of the DM-nucleus capture rates. I used the **DarkSUSY**

1903 software, that allows us to compute these quantities performing a full numerical

1904 integration over the momentum transfer of the form factors. The default standard

1905 solar model used by **DarkSUSY** is BP2000<sup>1</sup> [81].

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

## 7.2. Gravitational capture of DM by the Sun

1906 In Fig. 7.2 I show the results I obtained for the capture rates, for the case of  
 1907 interactions off electrons (red solid line), SD (green solid line) and SI (blue solid line)  
 1908 interactions of nucleons. In all cases I used a value of the scattering cross sections of  
 1909  $\sigma_i = 10^{-40} \text{ cm}^2$ . Note here one of the limitations of the **DarkSUSY** approach, one can  
 1910 not extend the computation below  $m_{\text{DM}} = 1 \text{ GeV}$ . Nevertheless, this is not something  
 1911 to worry about in this case, as I will discuss next. As a comparison, I added also the  
 1912 values computed in Ref. [4] (same color scheme, dashed lines). One can see there is good  
 1913 agreement between these and the **DarkSUSY** computation of the SD and SI interactions  
 1914 for  $m_{\text{DM}} \geq 1 \text{ GeV}$ . In this regime their computations also matches quite well our  
 1915 result for the electron capture rate. However, these start to differ significantly below  
 1916  $m_{\text{DM}} = 1 \text{ GeV}$ , being their estimate up to a factor of 5 bigger than ours for low masses.

1917 Let us comment briefly about the assumption I made before about not including  
 1918 an evaporation term in the Boltzmann equation. If I include this term in the equation  
 1919 (which will be proportional to the number of DM particles) the equilibrium solution  
 1920 takes the form:

$$N_{\text{DM}}^{eq} = \sqrt{\frac{C_{\odot}}{A_{\odot}}} \frac{1}{\kappa + \frac{1}{2} E_{\odot} \tau_{eq}}, \quad (7.17)$$

1921 where  $E_{\odot}$  is the total evaporation rate,  $\tau_{eq}$  is the equilibrium time in the absence of  
 1922 evaporation:

$$\tau_{eq} = \frac{1}{\sqrt{C_{\odot} A_{\odot}}}, \quad (7.18)$$

1923 and  $\kappa$  is defined as:

$$\kappa \equiv \sqrt{1 + \left( \frac{E_{\odot} \tau_{eq}}{2} \right)^2}. \quad (7.19)$$

1924 Now, it is easy to proof that in case evaporation dominates  $\kappa \gg 1$  and therefore:

$$N_{\text{DM}}^{eq} \simeq \frac{C_{\odot}}{E_{\odot}}. \quad (7.20)$$

1925 In contrast, if evaporation is irrelevant  $\kappa \simeq 1$  and one recovers Eq. (7.2).

---

model is relatively old and do not see why they are not using others like [3]. Maybe one can double-check in the code to make sure.

## Chapter 7. DM searches with neutrinos from the Sun

1926 In this way, one can define the evaporation mass as the mass for which the number  
1927 of DM particles in equilibrium approaches Eq. (7.20) at 10% level:

$$\left| N_{DM}^{eq}(m_{\text{evap}}) - \frac{C_{\odot}(m_{\text{evap}})}{E_{\odot}(m_{\text{evap}})} \right| = 0.1 N_{DM}^{eq}(m_{\text{evap}}). \quad (7.21)$$

1928 This can be regarded as the minimum testable mass one can reach using the annihilation  
1929 products of the DM in the Sun.

1930 It was reported in Ref. [4] that, in the case of both SD and SI DM interactions  
1931 off nuclei, this value ranges from 2 to 4 GeV depending on the specific scattering  
1932 cross section value, compatible with the usual assumptions in the literature. What is  
1933 interesting is the case of the electron capture. It was found that, when one applies a  
1934 cutoff in the velocity distribution of the DM trapped in the Sun slightly below the escape  
1935 velocity, the evaporation mass for the DM-electron interaction decreases remarkably. For  
1936 a moderate choice of  $v_c(r) = 0.9v_e(r)$  one gets an evaporation mass of around 200 to  
1937 600 MeV. This possibility opens a region of the parameter space that could be tested  
1938 with neutrino detectors.

### 1939 7.3 Neutrino flux from DM annihilations

1940 When WIMPs annihilate inside the Sun a flux of high-energy neutrinos is expected from  
1941 heavy quarks, gauge bosons and  $\tau^+\tau^-$  final states, which decay before losing energy  
1942 in the dense solar medium, as they will produce a continuum spectra up to  $E_{\nu} \sim m_{\chi}$   
1943 (in the case of direct annihilation to neutrinos one would have a line at  $E_{\nu} = m_{\chi}$ ) [75].  
1944 This kind of signal has been extensively studied in the literature, allowing to put strong  
1945 limits on the SD WIMP-proton cross section for large  $m_{\chi}$ . However, the number of  
1946 high-energy neutrinos per WIMP annihilation is small and the spectrum depends on the  
1947 unknown final state. Moreover, background rejection is easier for large  $m_{\chi}$  but neutrinos  
1948 with  $E_{\nu} \gtrsim 100$  GeV are significantly attenuated by interactions in the Sun.

1949 Nevertheless, most WIMP annihilation final states eventually produce a low-energy  
1950 neutrino spectrum. In this case one does not just consider the more massive final

## 7.4. Computing limits from solar neutrino fluxes

1951 states but also annihilations into  $e^+e^-$ ,  $\mu^+\mu^-$  and light quarks [74]. In particular, light  
 1952 mesons would be produced and stopped in the dense medium, thus decaying at rest and  
 1953 producing a monoenergetic neutrino signal. The decay-at-rest of kaons will produce  
 1954 a  $E_\nu = 236$  MeV  $\nu_\mu$  while in the case of pions one would have a  $E_\nu = 29.8$  MeV  $\nu_\mu$ .  
 1955 In practice only  $K^+$  and  $\pi^+$  contribute to these signals, as  $K^-$  and  $\pi^-$  are usually  
 1956 Coulomb-captured in an atomic orbit and get absorbed by the nucleus. There is also a  
 1957 low-energy neutrino signal coming from muon decays, which are produced in kaon or  
 1958 pion decays, leptonic decays of other hadrons and heavy leptons or even directly from  
 1959 WIMP annihilations, which can decay at rest and contribute to the previous low-energy  
 1960 neutrino flux with a well known spectrum below 52.8 MeV.

1961 These monoenergetic MeV neutrinos were previously considered undetectable but,  
 1962 due to the large yield, the known spectra and the modern advances in the detector  
 1963 technology, these low-energy neutrino flux can be a good probe of the SD WIMP-proton  
 1964 cross-section in standard solar WIMP capture scenario, as it is sensitive to low WIMP  
 1965 masses and insensitive to the particular final state. A good place to look for these signals  
 1966 are next-generation neutrino experiments such as DUNE.

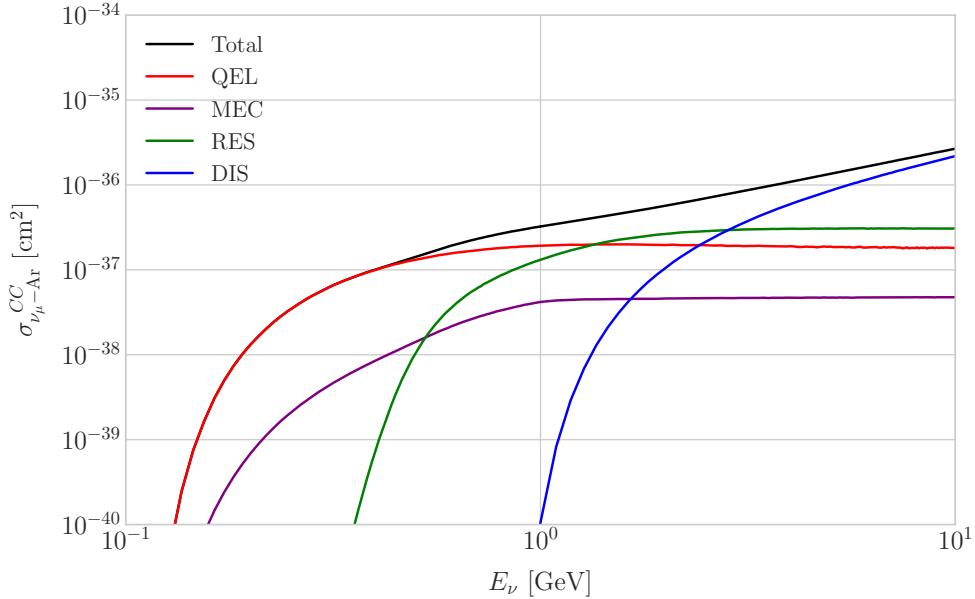
## 1967 7.4 Computing limits from solar neutrino fluxes

1968 In order to use the neutrino fluxes from DM annihilations in the Sun, the first thing I  
 1969 need to do is to determine the expected number of atmospheric background events, for  
 1970 a given exposure, after directionality selection has been applied. I can write this number  
 1971 as:

$$N_B = \eta_B \int d\Omega \int_{E_{min}}^{E_{max}} dE_\nu \frac{d^2\Phi_{atm}^\mu}{dE_\nu d\Omega} \times \left( A_{eff}^{(\mu)}(E_\nu) T \right), \quad (7.22)$$

1972 where  $\eta_B$  is the background efficiency,  $E_{min}$  and  $E_{max}$  the minimum and maximum  
 1973 energies to integrate over,  $d^2\Phi_{atm}^\mu/dE_\nu d\Omega$  the differential flux of atmospheric muon  
 1974 neutrinos,  $A_{eff}^{(\mu)}$  is the effective area of DUNE to muon neutrinos and  $T$  is the exposure  
 1975 time. The effective area can be expressed as the product of the neutrino-nucleus scattering  
 1976 cross section and the number of nuclei in the fiducial volume of the detector. This way

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.3:** *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).

1977 for DUNE I can write:

$$A_{eff}^{(\mu)}(E_\nu) = (6.0 \times 10^{-10} \text{ m}^2) \left( \frac{\sigma_{\nu - \text{Ar}}^{(\mu)}(E_\nu)}{10^{-38} \text{ cm}^2} \right) \left( \frac{M_{target}}{40 \text{ kT}} \right), \quad (7.23)$$

1978 where  $\sigma_{\nu - \text{Ar}}^{(\mu)}$  is the  $\nu_\mu - {}^{40}\text{Ar}$  charged-current scattering cross section. In Fig. 7.3 I  
 1979 show the computed value of this cross section as a function of the neutrino energy  $E_\nu$ ,  
 1980 in the range of interest both for the atmospheric background and signal events. It was  
 1981 computed using the NuWro Monte Carlo neutrino event generator [82], including the  
 1982 charged-current contributions of the quasi-elastic scattering (red line), resonant pion  
 1983 exchange (green line), deep inelastic scattering (blue line) and meson exchange current  
 1984 (purple line).

1985 The background rejection will depend on the resolution of the detector and the  
 1986 selection one applies on the events. A geometry argument can be used to estimate  
 1987 the maximum background rejection one can achieve in this case, considering one can

## 7.4. Computing limits from solar neutrino fluxes

1988 efficiently discriminate all events coming from a direction different from that of the  
 1989 Sun. In that case, the optimal background efficiency will simply be the relative angular  
 1990 coverage of the Sun. Taking the angular diameter of the Sun as seen from the Earth to  
 1991 be  $0.5^\circ$ , I have:

$$\eta_B^{(opt)} \approx \frac{\pi \left(\frac{0.5}{2}\right)^2}{360 \times 180} \simeq 3.03 \times 10^{-6}. \quad (7.24)$$

1992 This value will give a very optimistic estimate of the number of background events.  
 1993 However, it can be regarded as an lower limit, as it represents the best case scenario.

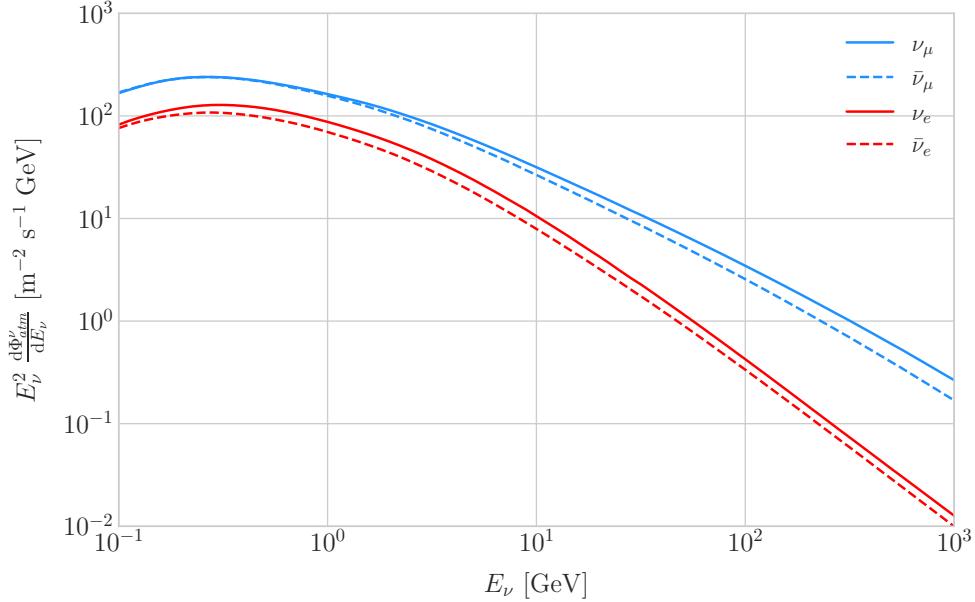
1994 In Fig. 7.4 I show the fluxes of atmospheric neutrinos at the Homestake mine during  
 1995 solar minimum, taken from Ref. [5]. The values are averaged over the two angular  
 1996 directions. In blue I have the flux of muon neutrinos while in red I indicate the flux  
 1997 of electron neutrinos. Additionally, the dashed lines correspond to both antineutrino  
 1998 species.

1999 Using these values for the muon neutrino and the corresponding total CC cross  
 2000 section, one can compute the number of expected background events by integrating over  
 2001 the given energy range (as in this case the angular integral is trivial). As for the energy  
 2002 range to integrate over, I choose the range for DUNE specified in [50],  $E_{min} = 10^{-1}$  GeV  
 2003 and  $E_{max} = 10$  GeV. Taking all these into account, I found the number of background  
 2004 events to be:

$$N_B \simeq \eta_B \times (3.827 \times 10^4) \times \left( \frac{\text{exposure}}{400 \text{ kT yr}} \right). \quad (7.25)$$

2005 In order to estimate the sensitivity of DUNE to this kind signal, one can consider a  
 2006 hypothetical data set where the number of observed neutrinos is taken to be the expected  
 2007 number of background events rounded to the nearest integer,  $N_{obs} = \text{round}(N_B)$  [83].  
 2008 Now, if I assume that the number of signal and background events seen by DUNE are  
 2009 given by Poisson distributions with means equal to the expected number of signal and  
 2010 background events,  $N_S$  and  $N_B$ , one can denote by  $N_S^{90}$  to the number of expected  
 2011 signal events such that the probability of having an experimental run with a number of  
 2012 events greater than  $N_{obs}$  is 90%. This number can be obtained as the numerical solution

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.4:** Expected atmospheric neutrino flux as a function of the neutrino energy  $E_\nu$  at Homestake at solar minimum, taken from Ref. [5]. The blue solid (dashed) line correspond to muon neutrinos (antineutrinos) and the red solid (dashed) line correspond to electron neutrinos (antineutrinos).

2013 to the equation:

$$1 - \frac{\Gamma(N_{obs} + 1, N_S^{90} + N_B)}{N_{obs}!} = 0.9, \quad (7.26)$$

2014 where  $\Gamma(x, y)$  is the upper incomplete gamma function.

2015 The number of signal events is related to the neutrino flux from DM annihilations in  
2016 a similar way as the background events to the atmospheric neutrino flux. In this case I  
2017 have:

$$N_S = \eta_S \Gamma_A^{eq} \int_{z_{min}}^{z_{max}} dz \frac{dN_\nu}{dAdN_Adz} \times (A_{eff}^\mu(z)T), \quad (7.27)$$

2018 where  $\eta_S$  is the signal efficiency,  $\Gamma_A^{eq}$  is the total annihilation rate of DM particles at  
2019 equilibrium,  $\Gamma_A^{eq} = A_\odot (N_{DM}^{eq})^2$ ,  $z_{min}$  and  $z_{max}$  the minimum and maximum relative  
2020 energies to integrate over (in such a way that  $z_{min,max} \leq E_{min,max}/m_{DM}$  for each  $m_{DM}$ )  
2021 and  $dN_\nu/dAdN_Adz$  the muon neutrino flux per DM annihilation in the Sun.

2022 Knowing  $N_S^{90}$  one can use the relation in Eq. (7.27) to obtain  $\Gamma_A^{eq,90}$  for different  
2023 values of the DM mass. From there I can directly translate those values into the

## 7.5. Example: Kaluza-Klein Dark Matter

2024 upper limits for DUNE on the DM scattering cross sections, for a given exposure. The  
2025 relation between the annihilation rate and the DM-nucleon cross section comes from the  
2026 equilibrium condition through the solar DM capture rate. The details of the evolution  
2027 of the number of DM particles inside the Sun and the computation of the capture rates  
2028 are discussed in App. 7.2.

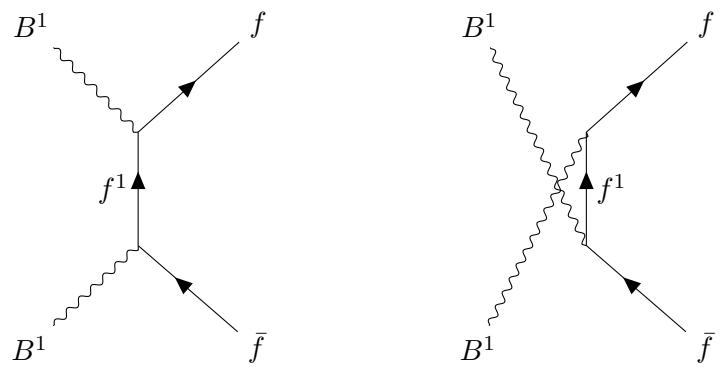
### 2029 7.5 Example: Kaluza-Klein Dark Matter

2030 Even though there are plenty of BSM theories which provide viable dark matter  
2031 candidates, Kaluza-Klein type of models [84, 85] within the universal extra dimensions  
2032 (UED) paradigm naturally predict the existence of a massive, stable particle that can  
2033 play the role of the dark matter. In the UED scenario all the SM fields can propagate  
2034 in one or more compact extra dimensions [86], as opposed to the idea of brane worlds  
2035 [87, 88], where just gravity can propagate in the bulk while SM particles live at fixed  
2036 points.

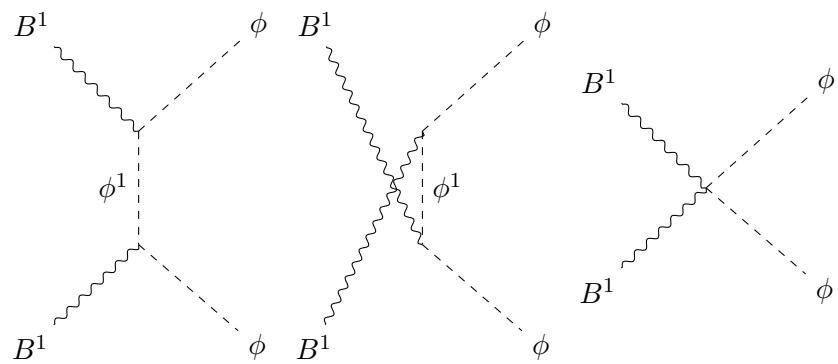
2037 Furthermore, in UED there is no violation of the translational invariance along the  
2038 extra dimensions, thus leading to degenerate KK modes masses and also the conservation  
2039 of the KK number in the effective four dimensional theory. At loop level, radiative  
2040 corrections and boundary terms shift the masses of the KK modes and break KK  
2041 number conservation into a KK parity. As a result, this theory only contains interactions  
2042 between an even number of odd KK modes and therefore the lightest among the first KK  
2043 excitations will be stable. This particle is usually denoted as the lightest Kaluza-Klein  
2044 particle (LKP) and its mass is proportional to  $1/R$ , being  $R$  the size of the extra  
2045 dimension.

2046 A viable DM candidate needs to be electrically neutral and non-baryonic, therefore  
2047 good candidates among the first Kaluza-Klein excitations would be the KK neutral  
2048 gauge bosons and the KK neutrinos [89]. Another possible candidate is the first KK  
2049 excitation of the graviton, which receives negligible radiate contributions and therefore  
2050 has a mass almost equal to  $1/R$ , but it has been shown that the lightest eigenstate from

## Chapter 7. DM searches with neutrinos from the Sun

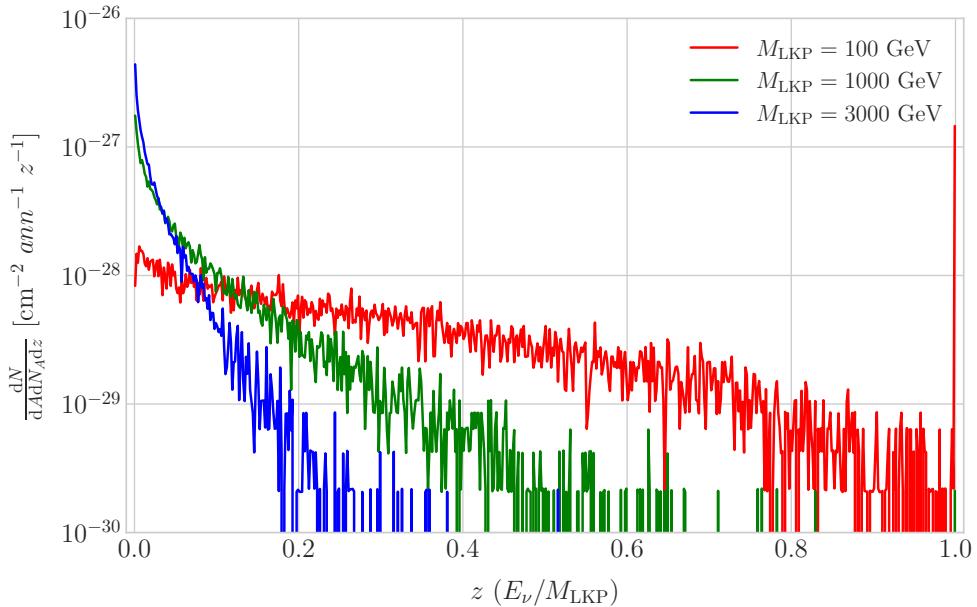


**Figure 7.5:** Feynman diagrams for  $B^1B^1$  annihilation into SM fermions.



**Figure 7.6:** Feynman diagrams for  $B^1B^1$  annihilation into a Higgs boson pair.

## 7.5. Example: Kaluza-Klein Dark Matter

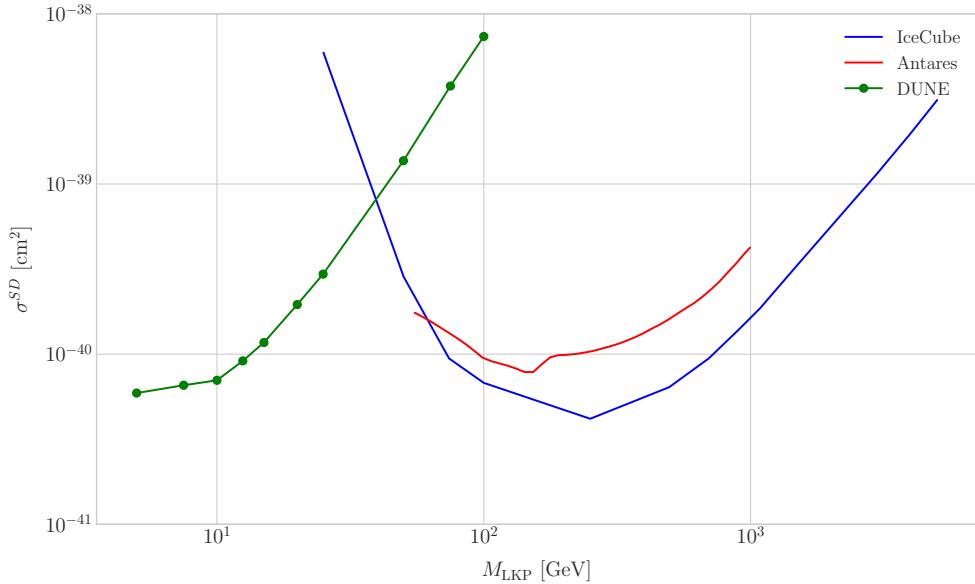


**Figure 7.7:** Computed spectra of muon neutrinos at the DUNE FD site from  $B^1$  annihilations in the Sun for three different values of  $M_{\text{LKP}}$ , plotted in relative energy units for legibility.

the mixing of the gauge mass states  $(B^1, W_3^1)$  would be lighter, as  $B^1$  and  $W_3^1$  receive negative radiative corrections [90]. 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 [90]. 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` [91, 92] 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. 7.7 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  $\chi\chi \rightarrow \nu\bar{\nu}$ .

## Chapter 7. DM searches with neutrinos from the Sun

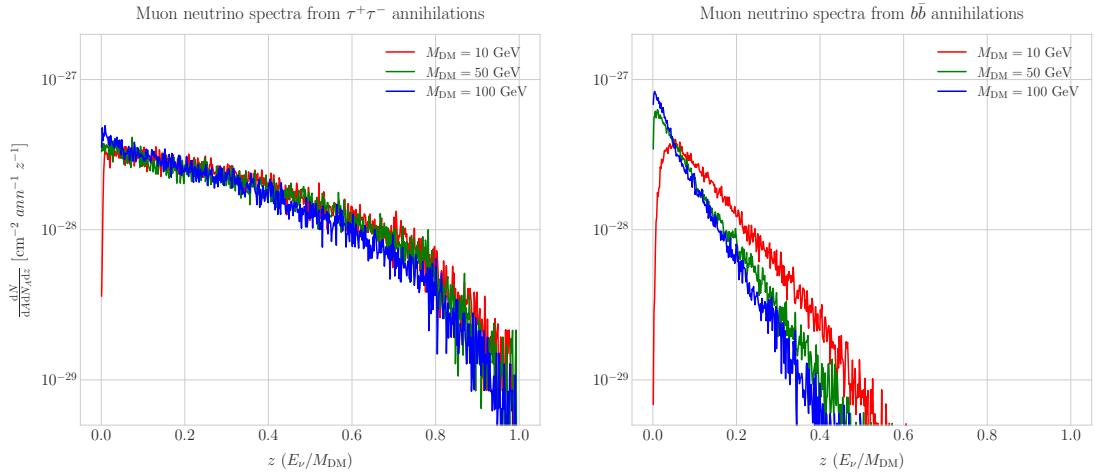


**Figure 7.8:** Projected 90% confidence level upper limit for DUNE (400 kT yr) on the spin-dependent  $B^1$ -proton scattering cross section as a function of  $M_{LKP}$  (green dots). I also show the previous limits from IceCube [6] (blue line) and Antares [7] (red line) on the LKP cross section. The shaded area represents the disfavoured region (at 95% confidence level) on the mass of the LKP from LHC data [8].

Now, one can estimate the sensitivity of DUNE to this particular model by using the methods I previously discussed. To begin with, I will use the optimistic estimation of the background efficiency in Eq. (7.24) 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. (7.26) 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. (7.2) and the capture rates I computed with DarkSUSY.

In Fig. 7.8 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 [6] (blue line) and Antares [7] (red line). The shaded area represents the disfavoured region from combined searches for UED by ATLAS and CMS [8].

## 7.6. High energy DM neutrino signals



**Figure 7.9:** 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.

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 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 [8] and other rare decay measurements [93, 94], it still constitutes an alternative indirect probe.

## 7.6 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 fluxes. Other channels, like  $W^+W^-$  and  $ZZ$ , are open for more massive WIMPs, but

## Chapter 7. DM searches with neutrinos from the Sun

2096 those will produce usually a higher energy neutrino flux that will be out of reach for  
2097 DUNE (usually the maximum neutrino energy is taken to be  $E_{max} = 10$  GeV).

2098 In Fig. 7.9 I show the `WimpSim` [91, 92] generated muon neutrino spectra at the  
2099 DUNE FD location ( $44^\circ 20' N, 103^\circ 45' W$ ) from  $\tau^+\tau^-$  (left panel) and  $b\bar{b}$  (right panel)  
2100 annihilations in the core of the Sun, for different DM masses. Here, one can clearly see  
2101 the meaning of the previous distinction between hard and soft channels. For the same  
2102 DM mass value, the muon neutrino spectrum from the  $\tau^+\tau^-$  channel is more flat and  
2103 reaches higher energies than the one from the  $b\bar{b}$  channel, which drops faster.

2104 In this case, I prepared two sets of files, one for  $\tau^+\tau^-$  and the other for  $b\bar{b}$ , for DM  
2105 masses in the range from 5 to 100 GeV (actually for  $b\bar{b}$  the first mass point I took is  
2106 7.5 GeV, as a WIMP with  $m_{DM} = 5$  GeV can not kinematically self annihilate into  $b\bar{b}$ ).  
2107 Then, I prepared the `WimpSim` output fluxes in a specific way to use them as inputs to  
2108 `NuWro`, which simulates the neutrino interaction with the argon.

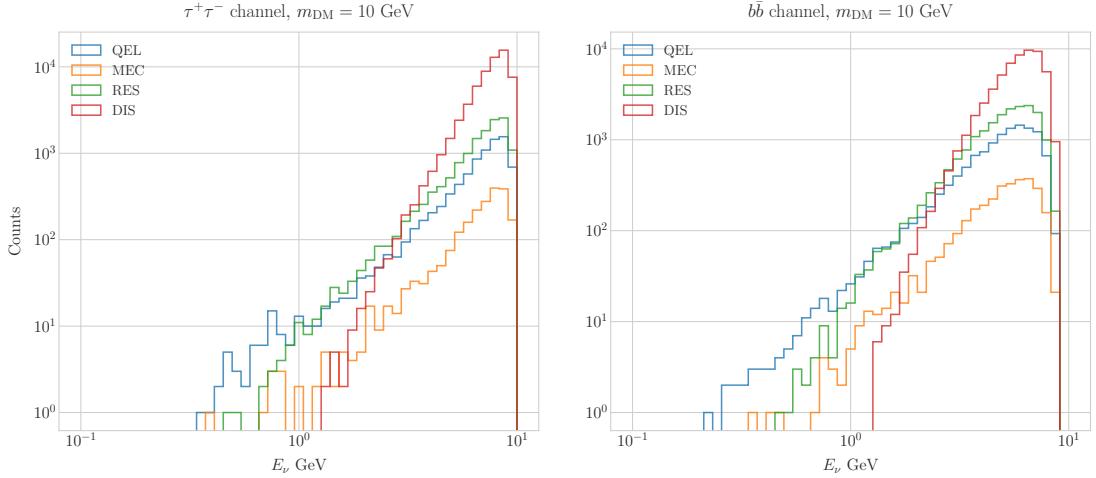
2109 Because `WimpSim` outputs an event list together with the fluxes, I can use the former  
2110 to generate the events. The direction of these is given in terms of the azimuth and  
2111 altitude angles viewed from the specified location, so first I need to convert these into the  
2112 DUNE FD coordinates. Once I have done it, each event can be processed with `NuWro`.  
2113 To increase the number of samples and optimise the computation time, I generate 100  
2114 interactions (i.e. `NuWro` events) for each `WimpSim` event<sup>2</sup>. I restrict the event generation  
2115 to charged current interactions, but I allow all the different contributions to the CC  
2116 cross section, i.e. quasielastic scattering (QEL), meson exchange current process (MEC),  
2117 resonant pion production (RES) and deep inelastic scattering (DIS). I just take into  
2118 account the CC contribution because I am only interested in final states with charged  
2119 leptons, as we have better chances of reconstructing the kinematics of CC events.

2120 For the atmospheric fluxes I follow a similar procedure, only that this time I do not  
2121 have a set of events but the fluxes binned in azimuth and altitude angles. This way, I  
2122 transform these to DUNE coordinates and process the fluxes for each bin separated with

---

<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.

## 7.6. High energy DM neutrino signals



**Figure 7.10:** Distribution of the muon neutrino energies from the  $\tau^+\tau^-$  (left panel) and  $b\bar{b}$  (right panel) annihilation channels, for  $m_{\text{DM}} = 10 \text{ GeV}$ , separated by CC interaction type: QEL (blue), MEC (orange), RES (green) and DIS (red).

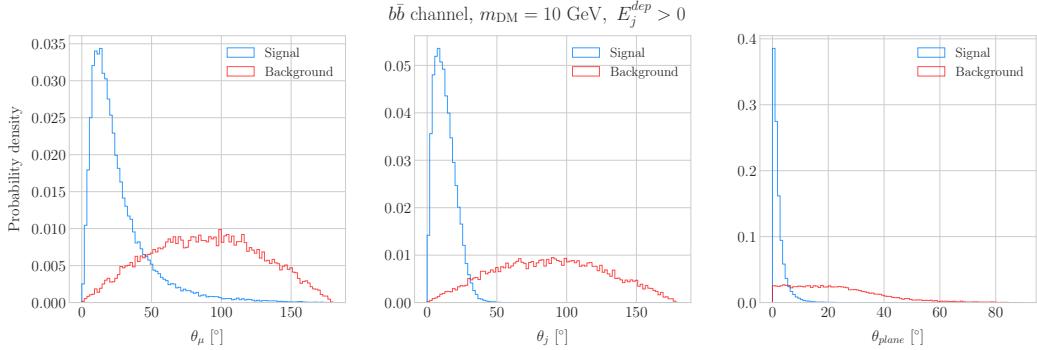
2123 NuWro.

2124 At this point, I have two sets of events with different energies and final states.  
 2125 In Fig. 7.10 one can see the distribution of the muon neutrino energies for the case  
 2126  $m_{\text{DM}} = 10 \text{ GeV}$ , both for the  $\tau^+\tau^-$  (left panel) and  $b\bar{b}$  (right panel) channels, separated  
 2127 by interaction. One can clearly see that there are different energy regimes where the  
 2128 primary interaction type is different. This leads to a plurality of event topologies,  
 2129 therefore making it difficult to implement a general approach to the selection of events  
 2130 in detriment of the background. As a way to proceed, I decided to split our samples,  
 2131 based on the different interaction modes and contents of the final state, into a CC DIS  
 2132 sample and a single proton CC QEL sample.

2133 **7.6.1 DIS events**

2134 To begin with, I consider the high energy part of the spectrum. In this region DIS events  
 2135 dominate, i.e. interactions of the form  $\nu_\mu + q_d(\bar{q}_u) \rightarrow \mu^- + q_u(\bar{q}_d)$ . Therefore, our final  
 2136 estates will contain a muon and a hadronic jet from the fragmentation of the outgoing  
 2137 quark. As all these events have  $E_\nu \gtrsim 1 \text{ GeV}$  the momentum transfer to the remnant  
 2138 nucleus is negligible, for this reason the neutrino energy can be effectively reconstructed

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.11:** Distributions of  $\theta_\mu$  (left panel),  $\theta_j$  (central panel) and  $\theta_{\text{plane}}$  (right panel) for the  $b\bar{b}$  sample with  $m_{\text{DM}} = 10 \text{ GeV}$  (blue) and the atmospheric background (red).

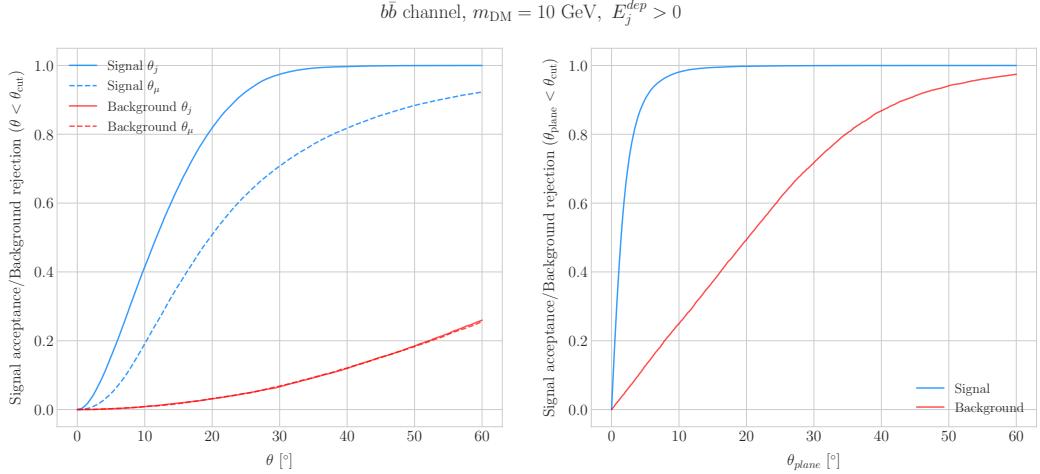
just taking into account the momenta of the muon and the jet. This technique was successfully used in Ref. [95] to select monoenergetic DM solar neutrino events from  $\nu\bar{\nu}$  annihilation channels.

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

To account for the limited angular resolution of the detector, I smeared the momenta of the muons and hadrons. In a liquid argon TPC muons are expected to be tracked with high precision, therefore I take the associated angular resolution to be  $1^\circ$ . In the case of jets, it is expected that for the hadrons dominating the cascade a detector like DUNE has an angular resolution between  $1^\circ$  to  $5^\circ$  [50], so I take the latter, more conservative, estimate.

As a first selection step, I will just take into account particles with kinetic energies above the detection threshold of DUNE. For muons and photons the specified threshold energy is 30 MeV, for charged pions 100 MeV and for other hadrons 50 MeV [50]. This way, if the outgoing muon in a certain event has an energy lower than the required

## 7.6. High energy DM neutrino signals



**Figure 7.12:** Left panel: signal efficiencies (blue lines) and background rejections (red lines) for events passing the cuts  $\theta < \theta_{\text{cut}}$  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  $\theta_{\text{plane}} < \theta_{\text{cut}}$  for the momentum conservation plane deviation.

threshold I will drop such event. For the case of hadrons and photons, I will only require to have at least one particle above the energy threshold, so then one can compute the jet momentum using the (smeared) momenta of the  $N$  particles above threshold as:

$$\vec{p}_j = \sum_{i=1}^N \vec{p}_i. \quad (7.28)$$

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

$$E_j^{\text{dep}} = m_{^{39}\text{Ar}} - m_{^{40}\text{Ar}} + \sum_{i=1}^N \sqrt{|\vec{p}_i|^2 + m_i^2}. \quad (7.29)$$

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

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

## Chapter 7. DM searches with neutrinos from the Sun

2169 incoming neutrino as:

$$\cos \theta_\mu = \hat{p}_\nu \cdot \hat{p}_\mu, \quad (7.30)$$

$$\cos \theta_j = \hat{p}_\nu \cdot \hat{p}_j, \quad (7.31)$$

2170 and the deviation from the momentum conservation plane as:

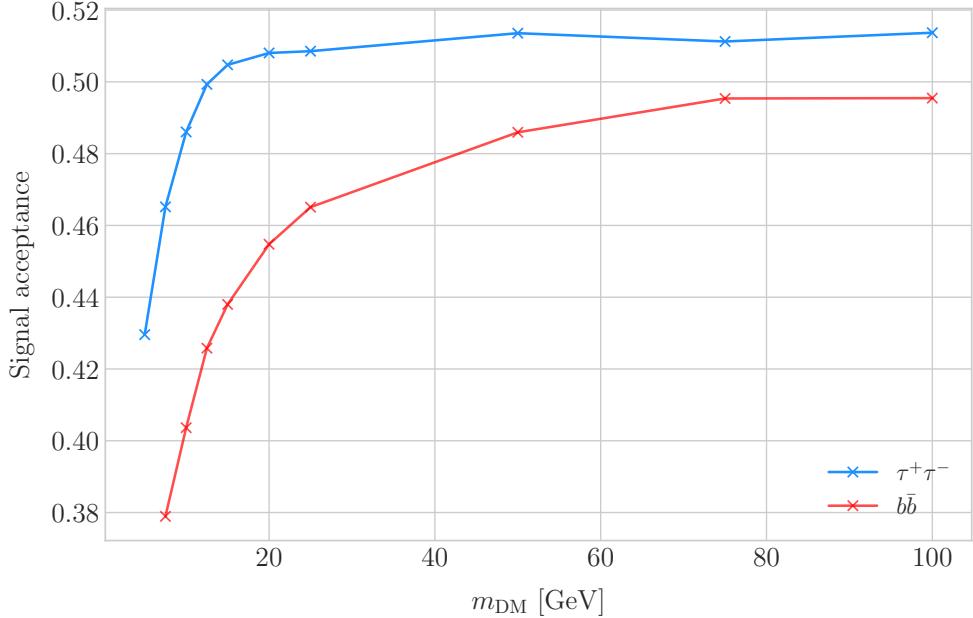
$$\sin \theta_{plane} = \left| \frac{\hat{p}_\mu \times \hat{p}_\nu}{|\hat{p}_\mu \times \hat{p}_\nu|} \cdot \hat{p}_j \right|. \quad (7.32)$$

2171 In Fig. 7.11 I show some distributions of these quantities for the case of the  $b\bar{b}$  sample  
2172 with  $m_{DM} = 10$  GeV (blue histograms) and for the atmospheric backgrounds (red).  
2173 In order to select the atmospheric events I followed the same criteria as for the signal  
2174 events. However, because in the signal case I used the true direction of the neutrino  
2175 as input, as it should be that of the Sun at that time and therefore known, in the  
2176 atmospheric case I used a set of solar positions as our ansatz for the neutrino direction.  
2177 From the distributions, one can see that the muon and the jet for the signal events are  
2178 predominantly forward and also that the deviations from the momentum conservation  
2179 plane are peaked at zero, as one should expect.

2180 Now, I can start applying cuts to maximise our signal selection efficiency while at  
2181 the same time I try to minimise the amount of atmospheric background events passing  
2182 the selection. To this end, I will need to find some lower and upper cuts for  $\theta_j$  and  
2183  $\theta_\mu$  and an upper bound for  $\theta_{plane}$ . In Fig. 7.12 I show how upper bound cuts in the  
2184 different angular variables affect the signal efficiency (blue lines) and the background  
2185 rejection (red lines). Notice that the signal efficiency behaves in a quite different way  
2186 when I apply cuts in the jet and the muon angles. On the contrary, the cuts on both  
2187 variables have a similar effect on the background rejection.

2188 In order to obtain the optimal set of cuts, I perform a multidimensional scan. I  
2189 do this separately for the  $\tau^+\tau^-$  and the  $b\bar{b}$  samples. For each case, I scan the possible  
2190 cuts for each mass point and then I take the mean value of the signal efficiency for

## 7.6. High energy DM neutrino signals

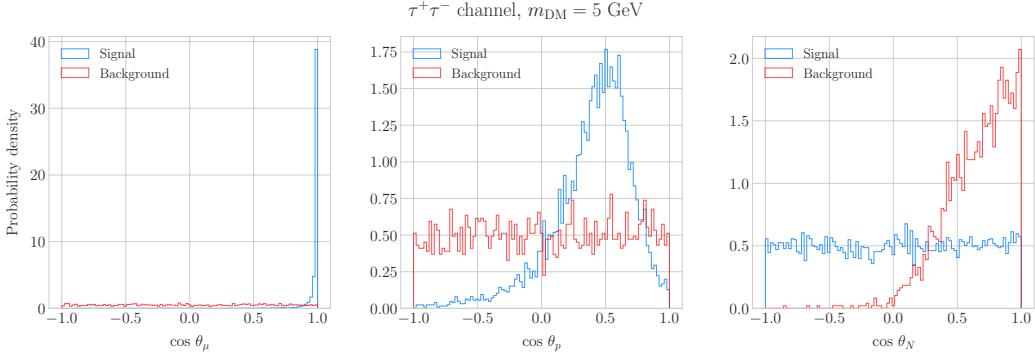


**Figure 7.13:** Signal efficiencies for the  $\tau^+\tau^-$  (blue line) and  $b\bar{b}$  (red line) DIS samples as functions of the DM mass,  $m_{\text{DM}}$ , obtained by applying the optimal angular cuts  $\theta_\mu < 27^\circ$ ,  $4^\circ < \theta_j < 26^\circ$  and  $\theta_{\text{plane}} < 3.5^\circ$ .

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

In Fig. 7.13 I show the signal efficiencies as a function of the DM mass for the  $\tau^+\tau^-$  (blue line) and the  $b\bar{b}$  (red line) DIS events, after applying the cuts discussed above, as well as the energy threshold and hadronic visible energy selections. One can see that the efficiency grows with the mass, as annihilations of more massive DM particles will produce a neutrino spectrum centered at higher energies, where DIS events dominate. Notice also that the efficiency is higher for the  $\tau^+\tau^-$  case at every mass point, as in general this channel produces neutrinos at higher energies than the corresponding  $b\bar{b}$  channel.

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.14:** 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).

### 2205 7.6.2 Single proton QEL events

2206 Now, one can try to explore the low energy tail of the neutrino energy distributions. This  
 2207 regime is dominated by the QEL interactions, i.e. events of the type  $\nu_\mu + n \rightarrow \mu^- + p$ .  
 2208 In this case, as the typical energies are  $E_\nu \lesssim 1 \text{ GeV}$ , the momentum transfer to the  
 2209 remnant nucleus is sizeable. Therefore, I can not make the approximation I did before  
 2210 and assume that the momentum of the muon and the proton will give an adequate  
 2211 estimation of the reconstructed neutrino energy.

2212 In any case, as before, I can take the direction of the incoming neutrino as known.  
 2213 That way, one can estimate the energy of the neutrino as:

$$E_\nu^{reco} = E_\mu + E_p + m_{^{39}\text{Ar}} - m_{^{40}\text{Ar}}, \quad (7.33)$$

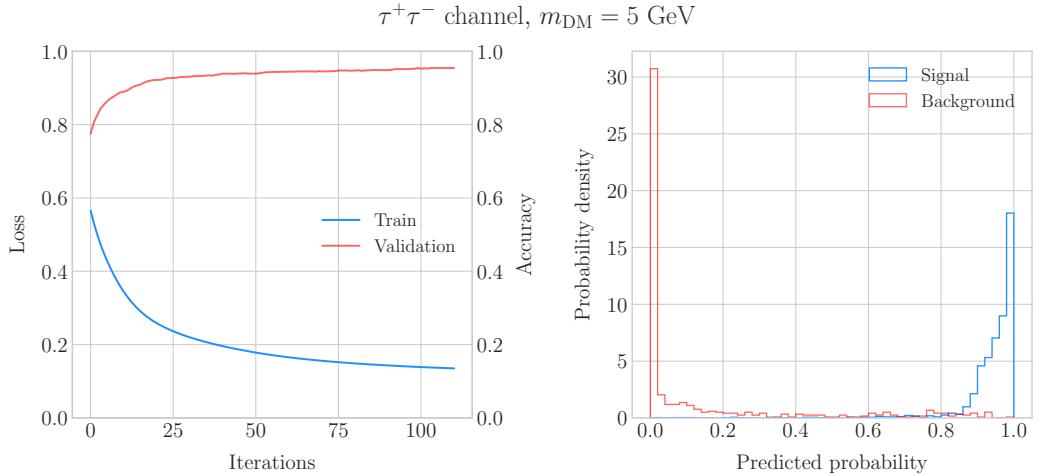
2214 and using momentum conservation I can write the momentum of the remnant nucleus  
 2215 as:

$$\vec{p}_N = \hat{p}_\nu (E_\mu + E_p + m_{^{39}\text{Ar}} - m_{^{40}\text{Ar}}) - \vec{p}_\mu - \vec{p}_p. \quad (7.34)$$

2216 As in the previous case, I need to drop the events where the muon or the proton fall  
 2217 below the kinetic energy detection threshold [50]. Also, I again apply a smearing to the  
 2218 momenta of the particles, a 1% for muons and 5% for protons.

2219 Having done that, one can compute the following angular variables for our selected

## 7.6. High energy DM neutrino signals



**Figure 7.15:** 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).

2220 events:

$$\cos \theta_\mu = \hat{p}_\nu \cdot \hat{p}_\mu, \quad (7.35)$$

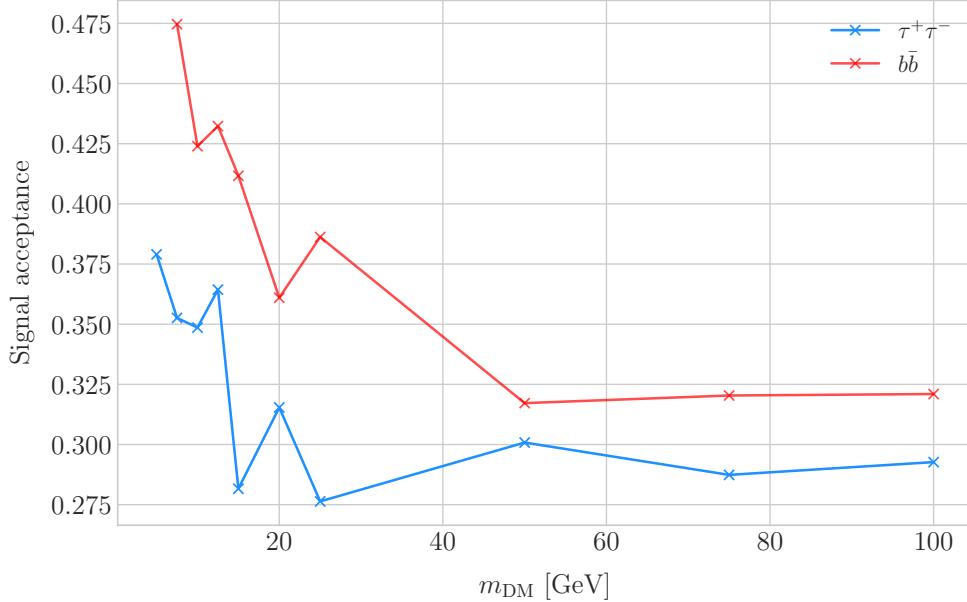
$$\cos \theta_p = \hat{p}_\nu \cdot \hat{p}_p, \quad (7.36)$$

$$\cos \theta_N = \hat{p}_\nu \cdot \hat{p}_N. \quad (7.37)$$

2221 Fig. 7.14 shows the distributions of these angular variables for the  $\tau^+\tau^-$  QEL  
 2222 sample with  $m_{\text{DM}} = 5 \text{ GeV}$  (blue) and the atmospheric background (red). Again, for  
 2223 the atmospheric events I used a random solar position as the ansatz for the incoming  
 2224 neutrino direction. Notice that now, opposed to the DIS case where the signal had very  
 2225 sharp distributions for the variables considered, the shapes of the angular distributions  
 2226 for signal and background are not that much different.

2227 This effectively means that the usual approach of applying simple angular cuts would  
 2228 not work as well as in the previous situation. Therefore, as a possible solution, I tried to  
 2229 use a multilayer perceptron (MLP) classifier to separate between signal and background  
 2230 events. Thus, the power of the hypothesis test will serve as an estimate of the signal

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.16:** 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%.

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

For each DM mass value and channel, as well as for the background sample, I divide our events into training, validation and test samples. The input variables for the classifier were the reconstructed neutrino energy from Eq. (7.33) and the angular variables defined in Eqs. (7.35 - 7.37). I used the MLP classifier implemented in `scikit-learn` [96], with a total of five hidden layers, the rectified linear unit activation function and adaptive learning rate. In order to account for fluctuations due to artifacts in the training process I repeated the training a thousand times for each sample, redefining each time the training, validation and test subsets, so one can take as our signal efficiency and background rejection the mean values of the powers and sizes of the tests.

The results of one of these training processes for the  $\tau^+\tau^-$  QEL signal with  $m_{\text{DM}} = 5$  GeV is shown in Fig. 7.15. On the left panel I show the loss function values (blue) and accuracy (red) at each iteration for the training and the validation samples respectively.

## 7.6. High energy DM neutrino signals

2245 The training stops either when the maximum number of iterations is reached (1000 in  
2246 this case) or when the accuracy for the validation sample reaches a certain tolerance  
2247 (I chose  $10^{-4}$  as our tolerance). On the right panel I have the distributions for the  
2248 predicted probability by the model, separated in true signal (blue) and background  
2249 (red) events, for the test sample. One can see that both populations are well separated,  
2250 obtaining a power of 44.97% and a size of 0.17% when I require a predicted probability  
2251 greater than 0.97.

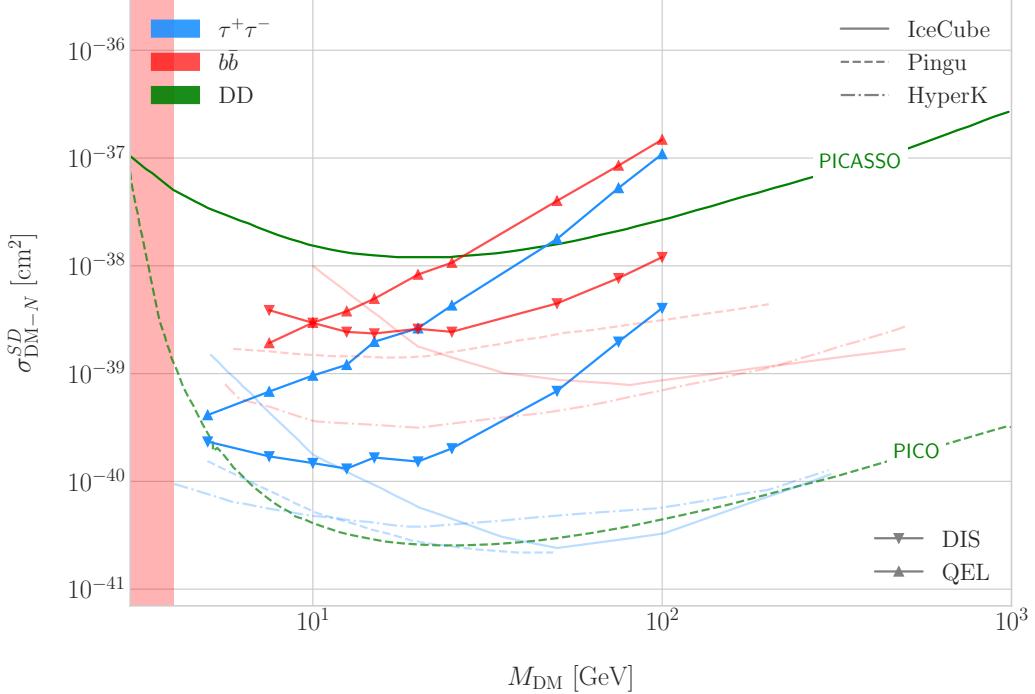
2252 Applying this criteria for each sample, I obtain the mean signal efficiencies shown in  
2253 Fig. 7.16. Notice that the efficiencies for the channel  $\tau^+\tau^-$  (blue line) are consistently  
2254 lower than the ones for the  $b\bar{b}$  channel (red line). This can be due to the fact that, for  
2255 each DM mass point, the neutrino spectrum coming from the  $b\bar{b}$  annihilation channel is  
2256 centered at lower energies when compared to the  $\tau^+\tau^-$  spectrum. This directly translates  
2257 into more low energy neutrinos undergoing QEL interactions, which give signals that  
2258 can be easily separated from the atmospheric background. This explanation also help us  
2259 understand why in both cases the signal acceptance drops when the DM mass increases.  
2260 In all cases, the background rejection took values between 99.8% to 99.9%. I will assume  
2261 a 99.8% background rejection value in all cases to keep our estimation conservative.

### 2262 7.6.3 Results

2263 In order to estimate the DM-nucleon cross section sensitivities in the present case I need  
2264 again to compute the expected number of background events. As I am now separating  
2265 events by interaction type Eq. (7.25) does not hold anymore, as in that case I integrated  
2266 over the total neutrino-argon cross section. In this instance, the expected background  
2267 events for DIS events is approximately given by:

$$N_B^{DIS} \simeq \eta_B^{DIS} \times (4.655 \times 10^3) \times \left( \frac{\text{exposure}}{400 \text{ kT yr}} \right), \quad (7.38)$$

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.17:** 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 [9] (solid lines) and the projected sensitivities for Pingu [10] (dashed lines) and Hyper-Kamiokande [11] (dash-dotted lines), as well as the direct detection limits from PICASSO [12] (solid green line) and PICO-60 C<sub>3</sub>F<sub>8</sub> [13] (dashed green line).

2268 whereas for QEL events we have:

$$N_B^{QEL} \simeq \eta_B^{QEL} \times (2.248 \times 10^4) \times \left( \frac{\text{exposure}}{400 \text{ kT yr}} \right). \quad (7.39)$$

2269 Now, using these together with Eqs. (7.26) and (7.27) one can obtain the 90% C.L.  
 2270 upper limit on the total annihilation rate at equilibrium for both kind of events. Then,  
 2271 applying the computed DM-nucleons capture rates I can translate these into limits on  
 2272 the DM-nucleon cross section by means of Eqs. (7.2), (7.5) and (7.6).

2273 Fig. 7.17 shows the obtained limits on the SD DM-nucleon cross section for DUNE,  
 2274 using the DIS (up triangles) and QEL (down triangles) events both for the  $\tau^+\tau^-$  (blue)  
 2275 and the  $b\bar{b}$  (red) samples, for an exposure of 400 kT yr. I also include the corresponding

## 7.7. Example: Leptophilic Dark Matter

2276 current limits from IceCube [9] (solid lines), as well as the projected sensitivities of Pingu  
2277 [10] (dashed lines) and Hyper-Kamiokande [11] (dash-dotted lines). For comparison, I  
2278 also show the reported direct detection limits from PICASSO [12] (solid green line) and  
2279 PICO-60 C<sub>3</sub>F<sub>8</sub> [13] (dashed green line).

2280 Notice that, for most of the mass range, the limits one can set by using the DIS  
2281 events are stronger than those of the QEL interactions, except for the low mass part  
2282 of both the  $\tau^+\tau^-$  and the  $b\bar{b}$  curves where the QEL events dominate. In general, the  
2283 expected sensitivity of DUNE for DM masses  $\lesssim 25$  GeV surpasses the stronger current  
2284 indirect limits. However, experiments like Hyper-Kamiokande are foreseen to have an  
2285 overall better sensitivity in this kind of searches, as they have a bigger active volume  
2286 and accept a broader energy range.

2287 A pending question is what happens when we add the RES and MEC charged-current  
2288 interaction contributions. In that case it would probably be more convenient to split  
2289 the samples by final state interaction topologies. Also, another necessary improvement  
2290 would be adding a full detector simulation and reconstructions. This will also require  
2291 considering the effect of poorly reconstructed events or final states containing neutral  
2292 particles such that they mimic the desired topology at the reconstruction level.

## 2293 7.7 Example: Leptophilic Dark Matter

2294 In general, the capture rate of DM particles by the Sun via interactions with electrons is  
2295 several orders of magnitude smaller than the capture via DM-nucleus scattering. Thus,  
2296 it would be sub-leading even when nucleon capture is loop suppressed. As I showed in  
2297 Fig. 7.2, the capture rate via scattering off electrons only surpasses the capture rates  
2298 via DM-nucleons interactions for DM masses  $\lesssim 100 - 500$  MeV.

2299 However, if one considers a model where DM-nucleon interactions are forbidden even  
2300 at loop level, then electron interactions will be the sole contributor to DM capture in  
2301 the Sun. One can describe such scenario where the DM particles couple to leptons but  
2302 not to the quark sector using effective operators.

## Chapter 7. DM searches with neutrinos from the Sun

2303 In general, assuming that the DM particle is a Dirac fermion, the dimension six  
 2304 operators describing the interaction between two DM particles and two leptons can be  
 2305 written as:

$$\mathcal{L}_{eff} = G \sum_i (\bar{\chi} \Gamma_\chi^i \chi) (\bar{\ell} \Gamma_\ell^i \ell), \quad (7.40)$$

2306 where  $G = 1/\Lambda^2$  is the effective coupling strength,  $\Lambda$  the cut-off of the effective field  
 2307 theory and  $\ell$  denotes any lepton. In principle, one should consider all the possible  
 2308 Lorentz structures  $\Gamma_f^i$  in order to have a complete set of effective operators.

2309 However, some combinations will induce interactions with nucleons at loop level. As  
 2310 we are specifically interested in interactions which forbid any communication with the  
 2311 quark sector, I will not consider those [97]. In addition, some of the effective operators  
 2312 give rise to velocity-suppressed scattering cross sections between DM particles and  
 2313 leptons. I will also neglect those, as the suppression goes with the square of the DM  
 2314 halo velocity which in units of the speed of light is  $\sim 10^{-6}$ .

2315 This way, the only Lorentz tensor structure that do not induce interactions with  
 2316 quarks at loop level and gives a contribution to the scattering cross section that is not  
 2317 velocity suppress is the axial-axial interaction. The effective Lagrangian is then given  
 2318 by:

$$\mathcal{L}_{eff} = \frac{c_A^\chi c_A^\ell}{\Lambda^2} (\bar{\chi} \gamma^\mu \gamma^5 \chi) (\bar{\ell} \gamma_\mu \gamma^5 \ell), \quad (7.41)$$

2319 where  $c_A^\chi$  and  $c_A^\ell$  are the couplings for the different species. As the DM coupling appears  
 2320 as a common factor for any lepton choice, I will redefine the corresponding coupling  $c_A^\ell$   
 2321 to absorb  $c_A^\chi$ . Also, for simplicity, I will assume that the couplings between the DM  
 2322 particles and the leptons are flavour independent, i.e. I have just two couplings,  $c_A^e$  for  
 2323 charged leptons and  $c_A^v$  for neutrinos.

2324 In the case of a scalar DM particle, the lowest order effective interaction with  
 2325 leptons happens through a dimension five operator, generating scalar and pseudoscalar  
 2326 interactions. However, the former induces interactions with quarks at two loop level  
 2327 whereas the latter gives a velocity suppressed scattering cross section.

2328 From the effective Lagrangian in Eq. (7.41) it can be shown that the axial-axial

## 7.7. Example: Leptophilic Dark Matter

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

$$\sigma_{\text{DM}-e}^{AA} = 3(c_A^e)^2 \frac{m_e^2}{\pi \Lambda^4}. \quad (7.42)$$

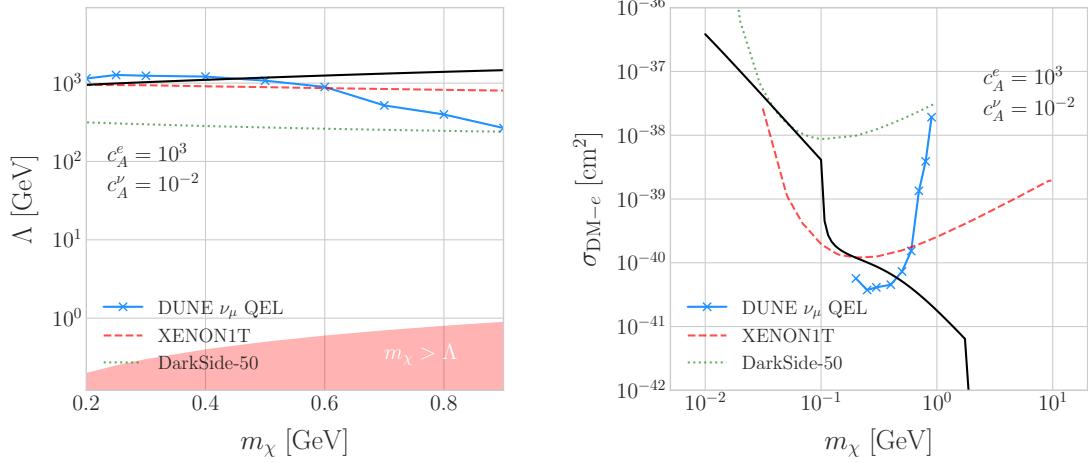
If the DM interacts exclusively with fermions, then the only annihilation channels that will give us a measurable neutrino flux coming out of the Sun are  $\tau^+\tau^-$  and  $\nu\bar{\nu}$ . The former channel, already explored previously in the more mainstream scenario of the DM capture via scattering off nucleons, is open only for  $m_{\text{DM}} > m_\tau \simeq 1776.86 \pm 0.12$  MeV [98], a mass region where the solar DM capture by electrons is at least one order of magnitude smaller than the capture via interactions with nucleons. On the contrary, the latter allows us to explore a region where the capture rate via scattering off electrons dominates over the rest.

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

In this case, as I have a specific realisation of the interaction between the DM and leptons, one can estimate the relic density of our DM for different values of the couplings and the effective field theory scale  $\Lambda$ . The first step to do so is compute the self-annihilation cross section. Because I consider cold relics, at the freeze-out time our DM particles were non-relativistic and so one can expand the annihilation cross section in terms of the relative velocity  $v$  between two annihilating DM particles as [99]:

$$\sigma_{\text{ann}}^{AA}|v| \approx \frac{1}{2\pi\Lambda^4} \sum_\ell \left(c_A^\ell\right)^2 m_\chi^2 \sqrt{1 - \frac{m_\ell^2}{m_\chi^2} \left[ \frac{m_\ell^2}{m_\chi^2} + \frac{1}{12} \left(2 - \frac{m_\ell^2}{m_\chi^2}\right) v^2 \right]}, \quad (7.43)$$

## Chapter 7. DM searches with neutrinos from the Sun



**Figure 7.18:** 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 [14] (dotted green line) and XENON1T [15] (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  $c_A^\nu = 10^{-2}$ .

2353 where the sum includes all the possible lepton final states with mass  $m_\ell$ .

2354 Solving the Boltzmann equation for the evolution of the DM density gives as a  
2355 solution a relic density of:

$$\Omega_\chi h^2 \approx \frac{(1.04 \times 10^9) x_F}{M_{Pl} \sqrt{g_*} (a + 3b/x_F)}, \quad (7.44)$$

2356 where  $x_F = m_\chi/T_F$  being  $T_F$  the freeze-out temperature,  $g_*$  the number of relativistic  
2357 degrees of freedom at freeze-out and  $a$  and  $b$  the terms in the annihilation cross section  
2358 expansion  $\sigma_{ann}|v| \approx a + bv^2 + \mathcal{O}(v^4)$ . Using the current best fit for the relic DM density  
2359  $\Omega_\chi h^2 = 0.1198 \pm 0.0012$  [100] one can use these relations to compute the required  
2360 effective theory scale  $\Lambda$  at which the correct density is achieved for any combinations of  
2361  $m_\chi$  and  $c_A^\ell$ .

2362 As discussed before, in the low DM mass region QEL interactions dominate. Moreover,  
2363 if I focus on direct annihilation to neutrinos, the energy of the muon neutrino flux is  
2364 known as it must be equal to the mass of the DM particle,  $E_\nu = m_\chi$ . That way, now  
2365 I do not need to use Eq. (7.33) in order to estimate the momentum transfer to the

## 7.7. Example: Leptophilic Dark Matter

2366 remnant nucleus, I can simply take:

$$\vec{p}_N = \hat{p}_\nu m_\chi - \vec{p}_\mu - \vec{p}_p. \quad (7.45)$$

2367 To estimate the signal efficiency and background rejection for this case I used again  
2368 the MLP classifier from `scikit-learn`, using the same specifications as before. The  
2369 only difference now is that I add also the reconstructed neutrino energy as one of the  
2370 features to train the classifier with, because the characteristic monoenergetic flux for  
2371 each  $m_\chi$  value will help to distinguish between signal and background events.

2372 In this case, for masses below  $\sim 500$  MeV I obtain a signal efficiency close to unity  
2373 while keeping a background rejection of 99.9%. For bigger values of the mass, the signal  
2374 efficiency drops significantly if I require to keep the background acceptance under 0.01%.  
2375 However, because this kind of search is dominated by the background, sacrificing the  
2376 signal acceptance to keep the background rejection to a minimum enhances the reach  
2377 of the analysis. This way, for DM masses of the order of  $m_\chi \sim 1$  GeV I end up with  
2378 efficiencies as low as 1%.

2379 Now, estimating the number of background events using Eq. (7.39) one can go on  
2380 and apply Eqs. (7.26) and (7.27) together with Eq. (7.42) to derive the sensitivity of  
2381 DUNE to this kind of model. Fig. 7.18 (left panel) shows the potential reach of DUNE  
2382 to constrain the EFT scale  $\Lambda$  this model containing only leptophilic DM axial-axial  
2383 interactions (blue line), for a choice of couplings  $c_A^e = 10^3$  and  $c_A^\nu = 10^{-2}$ . I also included  
2384 the current limits on the DM-electron scattering cross section from DarkSide-50 [14]  
2385 (dotted green line) and XENON1T [15] (dashed red line), reworked with Eq. (7.42) to  
2386 show their implications for the EFT scale. The values of  $\Lambda$  for which the correct DM relic  
2387 density value is achieved for each mass are also shown (black line). This tells us that,  
2388 for that specific choice of couplings, DUNE would be sensitive to DM configurations  
2389 allowed by the relic density constraint up to a mass of  $m_\chi \sim 400$  MeV.

2390 In Fig. 7.18 (right panel) I show the same upper limits but for the DM-electron  
2391 scattering cross section. From this view one can see that DUNE would be able to

## Chapter 7. DM searches with neutrinos from the Sun

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.

2398 Chapter 8

2399 Particle ID in GArSoft

2400 ND-GAr is a magnetised, high-pressure gaseous argon TPC (HPgTPC), surrounded by  
2401 an electromagnetic calorimeter (ECal) and a muon detector (commonly refer to as  $\mu$ ID).  
2402 A detailed discussion on the requirements, design, performance and physics of ND-GAr  
2403 can be found in the DUNE ND CDR [56] and the ND-GAr whitepaper (cite).

2404 In DUNE Phase II ND-GAr will fulfill the role of TMS, measuring the momentum  
2405 and sign of the charged particles exiting ND-LAr. Additionally, it will be able to measure  
2406 neutrino interactions inside the HPgTPC, achieving lower energy thresholds than those  
2407 of the ND and FD LArTPCs. By doing so ND-GAr will allow to constrain the relevant  
2408 systematic uncertainties for the LBL analysis even further.

2409 The goal of the present chapter is to review the requirements that the physics program  
2410 of DUNE impose on ND-GAr, present the current status of its design and describe the  
2411 GArSoft package, its simulation and reconstruction software.

2412 As decided during the DUNE Phase II workshop in June 2023 [reference], we want  
2413 to build ND-GAr physics case by showing:

- 2414 • That ND-GAr can constrain systematic uncertainties that ND-LAr might miss.
- 2415 • The impact on the neutrino oscillation results if such systematic uncertainties are  
2416 missed.
- 2417 • That ND-GAr is necessary to reach DUNE's main physics goals.

## Chapter 8. Particle ID in GArSoft

2418 This way, the design of ND-GAr will be physics driven.

2419 In order to study the effects of final state interactions (FSI) in CC interactions,  
2420 ND-GAr should be able to measure the spectrum of protons and charged pions at low  
2421 energies. ND-GAr also needs to be able to measure the pion multiplicity, specially for  
2422 energies above 100 MeV as at these energies the pions shower in the LAr, to inform the  
2423 pion mass correction in the ND and FD LArTPCs.

2424 In order to correctly identify electrons, muons, pions, kaons and protons ND-GAr  
2425 can use a combination of:  $dE/dx$  measurements in the HPgTPC,  $E_{ECAL}/p$  using the  
2426 ECAL total energy and the momentum obtained from magnetic spectroscopy in the  
2427 HPgTPC and penetration information through the ECAL and muon tagger.

### 2428 8.1 $dE/dx$ measurement in the TPC

2429 Among the parameters extracted from the track fitting, ionisation is particularly useful  
2430 for particle identification, as it is a function of the particle velocity. Although for the  
2431 case of relativistic particles this dependence is not very strong, measuring the track on  
2432 a large number of points may allow us to estimate the amount of ionisation accuratel.  
2433 This, paired with a measurement of the momentum, may allow us to identify the particle  
2434 type.

2435 The first calculation of the energy loss per unit length of relativistic particles using a  
2436 quantum-mechanical treatment is due to Bethe [101]. Using this approach, the mean  
2437 ionisation rate of a charged particle traveling through a material medium is (using  
2438 natural units  $G = \hbar = c = 1$ ):

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi Ne^4}{m_e \beta^2} z^2 \left( \log \frac{2m_e \beta^2 \gamma^2}{I} - \beta^2 \right), \quad (8.1)$$

2439 where  $N$  is the number density of electrons in the medium,  $e$  the elementary charge,  $m_e$   
2440 is the electron mass,  $z$  the charge of the particle in units of  $e$ ,  $\beta$  is the velocity of the  
2441 particle,  $\gamma = (1 - \beta^2)^{-1}$  and  $I$  denotes the effective ionisation potential averaged over  
2442 all electrons. This relation is known as the Bethe-Bloch formula.

## 8.1. $dE/dx$ measurement in the TPC

From Eq. (8.1) one can see that the ionisation loss does not depend explicitly on the mass of the charged particle, that for non-relativistic velocities it falls as  $\beta^{-2}$ , then goes through a minimum and increases as the logarithm of  $\gamma$ . This behaviour at high velocities is commonly known as the relativistic rise. The physical origin of this effect is partly due to the fact that the transverse electromagnetic field of the particle is proportional to  $\gamma$ , therefore as it increases so does the cross section.

It was later understood that the relativistic rise could not grow indefinitely with  $\gamma$ . A way to add this feature in the Bethe-Bloch formula is by introducing the so-called density effect term. It accounts for the polarisation effect of the atoms in the medium, which effectively shield the electromagnetic field of the charged particle halting any further increase of the energy loss [102]. Denoting the correction as  $\delta(\beta)$ , one can rewrite Eq. (8.1) as:

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi Ne^4}{m_e \beta^2} z^2 \left( \log \frac{2m_e \beta^2 \gamma^2}{I} - \beta^2 - \frac{\delta(\beta)}{2} \right). \quad (8.2)$$

In general, the form of  $\delta(\beta)$  depends on the medium and its state of aggregation, involving the usage of tabulated parameters and implicit relations [103].

Another standard method to compute the amount of ionisation a charged particle produces is the so-called photo-absorption ionisation (PAI) model proposed by Allison and Cobb [104]. Within their approach, the mean ionisation is evaluated using a semiclassical calculation in which one characterises the continuum material medium by means of a complex dielectric constant  $\epsilon(k, \omega)$ . However, in order to model the dielectric constant they rely on the quantum-mechanical picture of photon absorption and collision. Therefore, in the PAI model the computation of the ionisation loss involves a numerical integration of the measured photo-absorption cross-section for the relevant material.

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

## Chapter 8. Particle ID in GArSoft

2470 It happens that neither the Bethe-Bloch theory nor the PAI model from Allison and  
2471 Cobb offer a close mathematical form for the ionisation curve. This is the reason why a  
2472 full parametrisation of the ionisation curves can be useful. A parametrisation originally  
2473 proposed for the ALEPH TPC [105] and later used by the ALICE TPC [106] group that  
2474 manages to capture the features of the ionisation energy loss is:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \left( P_2 - \beta^{P_4} - \log \left[ P_3 + \frac{1}{(\beta\gamma)^{P_5}} \right] \right), \quad (8.3)$$

2475 where  $P_i$  are five free parameters. Hereafter, we will refer to Eq. (8.3) as the ALEPH  
2476  $dE/dx$  parametrisation.

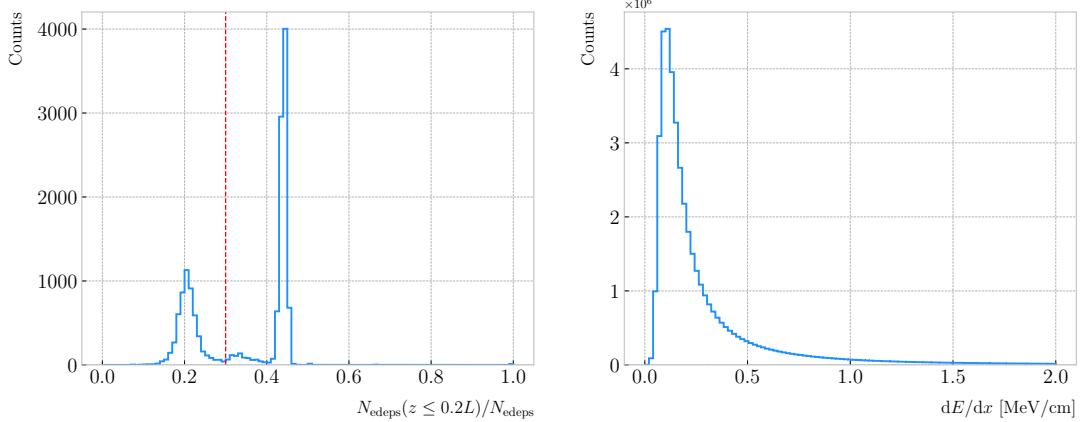
### 2477 8.1.1 Energy calibration

2478 In order to obtain the amount of energy loss by a charged particle due to ionisation  
2479 in our TPC we need to determine the conversion between the charge deposited in our  
2480 readout planes and the actual energy depositions. This procedure is known as energy  
2481 calibration.

2482 In a general, the first step of the calibration involves a non-uniformity correction,  
2483 to make sure that the detector response is uniform throughout the TPC. These are  
2484 typically divided into three categories, non-uniformities in the transverse  $YZ$  plane,  
2485 non-uniformities along the drift direction  $X$  and variations of the detector response  
2486 over time (would not apply to us as the detector is not built yet). These would correct  
2487 for effects such as electron diffusion and attenuation, space charge effects or channel  
2488 misconfiguration. However, because at the moment I am only interested in making sure  
2489 we recover a sensible result from our simulation, I will not apply uniformity corrections  
2490 to our charge deposits.

2491 Other effects, like electron-ion recombination or ADC saturation, lead to a non-linear  
2492 relation between the observed charge and the deposited energy in the detector, with the  
2493 observed readout charge saturating at high ionisation energies. In this case, because we  
2494 are dealing with gaseous argon and therefore recombination is not as important as in

## 8.1. $dE/dx$ measurement in the TPC



**Figure 8.1:** 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.

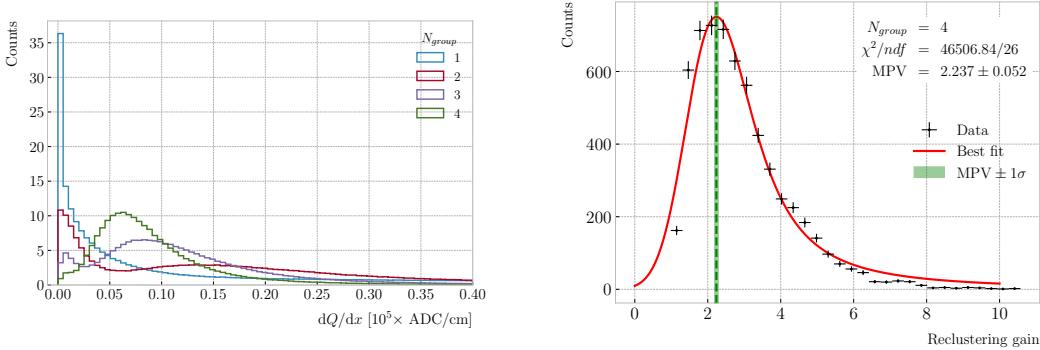
2495 liquid, we do not simulate recombination effects in the TPC. Even so, the simulation of  
 2496 the electronic response will still introduce charge saturation, and one needs to correct  
 2497 for it in order to obtain the exact amount of energy loss due to ionisation.

2498 By default, the track fitting algorithm in GArSoft provides a `TrackIonization`  
 2499 object associated to each reconstructed track. It contains two collections of charge  
 2500 deposits, one for each fitting direction, consisting on pairs of charge values ( $dQ$ , in ADC)  
 2501 and step sizes ( $dx$ , in cm).

2502 In order to estimate the ionisation loss in the ND-GAr TPC, I have used an MC  
 2503 sample consisting of single, isotropic protons propagating in the TPC. The starting points  
 2504 of the protons were sampled inside a  $50 \times 50 \times 25$  cm box centered at  $(100, -150, 1250)$ ,  
 2505 and their momenta are uniformly distributed in the range  $0.25 - 1.75$  GeV. I ran the  
 2506 simulated sample through GArSoft's default detector simulation and reconstruction, and  
 2507 then a custom analyser module that extracts the ionisation data together with other  
 2508 reconstructed track information from the Kalman fit.

2509 For studying the energy loss of the protons I select the reconstructed tracks that  
 2510 range out (i.e. slow down to rest) inside the TPC. A characteristic feature of the energy  
 2511 loss profile of any stopping ionising particle is the so-called Bragg peak, a pronounced

## Chapter 8. Particle ID in GArSoft



**Figure 8.2:** 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.

peak that occurs immediately before the particle comes to rest. From Eq. (8.1) 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 end of the track. In this case, I count the fraction of the Geant4 simulated energy deposits with a residual range value (the distance from a given energy deposit to the last deposit in the track trajectory) less than a 20% of the corresponding track length<sup>1</sup>. The distribution of this fraction of energy deposits for our proton sample is shown in Fig. 8.1 (left panel). We can clearly see two well separated peaks in this distribution, one centered at 0.2 and another, narrower, one centered at a higher value. The first one corresponds to non-stopping protons, as in that case the number of energy deposits towards the end of the track is uniformly distributed due to the absence of the Bragg peak. In that way, I apply a cut in this distribution, requiring that at least 30% of the simulated energy deposits sit in the last 20% of the tracks, to ensure that the Bragg peak is present.

Figure 8.1 (right panel) shows the distribution of the energy loss per unit length for the Geant4 simulated energy deposits of the selected stopping protons. We can see that

<sup>1</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.

## 8.1. $dE/dx$ measurement in the TPC

it follows the expected shape of a Landau distribution, which describes the fluctuations of the ionisation energy losses [107]. This distribution has a characteristic asymmetric PDF, with a long right tail that translates into a high probability for high-energy ionisation losses. The origin of these fluctuations is mainly the possibility of transferring a high enough energy to an electron, so it becomes a ionising particle itself.

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

Figure 8.2 (left panel) shows the distribution of the ionisation charge deposits per unit length for the track in the stopping proton sample (blue line). As one can notice, this distribution does not resemble the expected shape of the Landau PDF. This distribution peaks sharply at 0 and has a heavy tailed behaviour. Notice, however, how the distribution changes its shape as we group together  $N_{group}$  consecutive charge deposit pairs (red, purple and green lines). The distribution in the  $N_{group} = 4$  case already has a shape which resembles that of the Geant4-level ionisation per unit length, so I will proceed using this amount of reclustering for the reconstruction-level depositions.

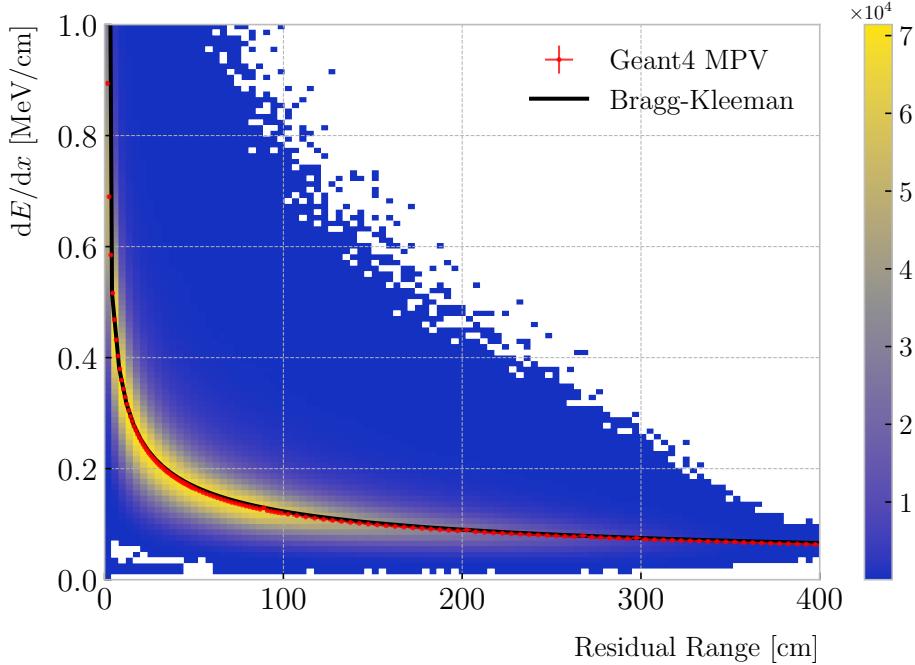
An extra factor I need to account for, when reclustering is applied, is how the overall  $dQ/dx$  per track changes. To do so, we can look at the ratio between the median  $dQ/dx$  after and before the reclustering. Figure 8.2 (right panel) shows the median enhancement in  $dQ/dx$  per track for the stopping proton sample in the case  $N_{group} = 4$ . Fitting a Landau distribution convolved with a Gaussian<sup>2</sup>, I estimate the most probable value of this ratio to be  $G_{group} = 2.24 \pm 0.05$ .

At this point, I am left with determining the conversion between the charge deposits per unit length  $dQ/dx$  and the energy deposits per unit length  $dE/dx$ . To this end, we need a way of comparing the two. I can use the residual range  $z$  to get a prediction of

---

<sup>2</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.

## Chapter 8. Particle ID in GArSoft



**Figure 8.3:** 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. (8.4).

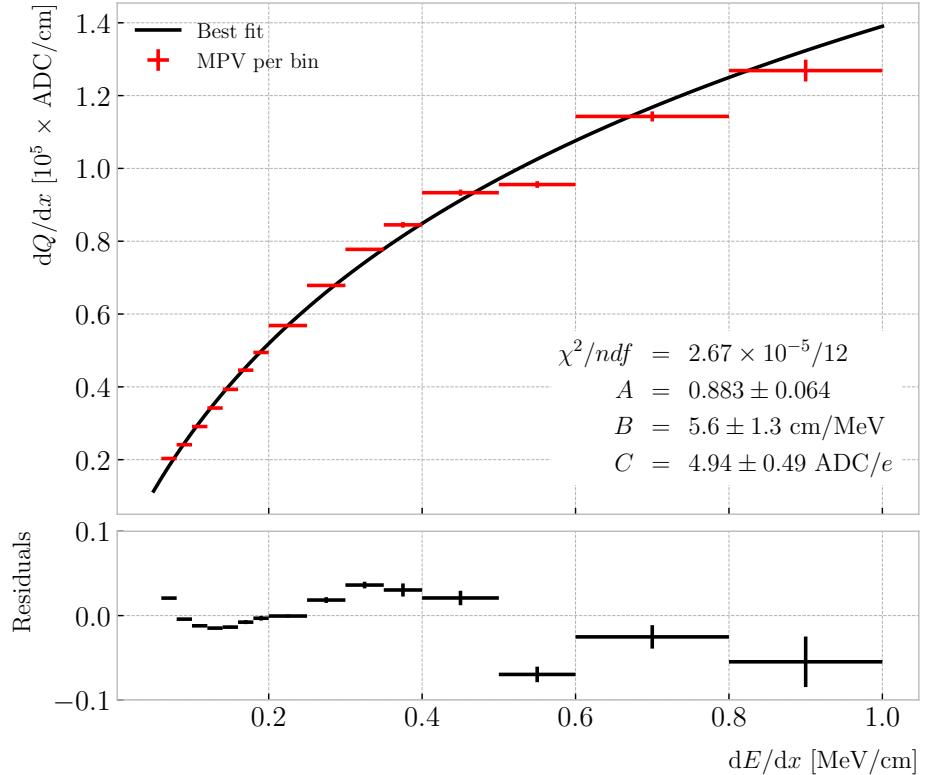
2556 the most probable  $dE/dx$  by using the following empirical parametrisation [108]:

$$\frac{dE}{dx}(z) = \frac{z^{\frac{1}{p}-1}}{p\Lambda^{\frac{1}{p}}}, \quad (8.4)$$

2557 which is quoted in the literature as the Bragg-Kleeman formula. In order to obtain the  
 2558  $p$  and  $\Lambda$  parameters I perform a fit using the energy losses and the residual ranges given  
 2559 by the Geant4 stage of our proton sample.

2560 Within our simulation, the residual range is sampled with a maximum size of  
 2561 5 mm. Therefore, to perform the fit to the Bragg-Kleeman formula, we can use a  
 2562 fine-grained residual range binning. For each of the residual range bins I extract the  
 2563  $dE/dx$  distribution and fit it to a LanGauss distribution, to obtain the value of the  
 2564 most probable  $dE/dx$  in the bin together with a statistical uncertainty. I then fit Eq.  
 2565 (8.4) to these most probable values and the centres of the residual range bins. This

## 8.1. $dE/dx$ measurement in the TPC



**Figure 8.4:** 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. (8.5).

procedure is depicted in Fig. 8.3, where I show the distribution of the energy loss per unit length versus the residual range, together with the most probable  $dE/dx$  values and their uncertainty in each bin (red points) and the curve with the best fit of the Bragg-Kleeman relation to those values (black line). The best fit is obtained for the parameter values  $p = 1.8192 \pm 0.0005$  and  $\Lambda = 0.3497 \pm 0.0008$  cm/MeV<sup>3</sup>.

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

In order to parametrise the charge saturation, we can use the following logarithmic

---

<sup>3</sup>These strange units for  $\Lambda$  come from dimensional analysis, just to keep the Bragg-Kleeman formula (8.4) consistent.

## Chapter 8. Particle ID in GArSoft

2575 function inspired by the modified box model for recombination:

$$\frac{dE}{dx} = \frac{e^{\frac{dQ}{dx}B\frac{W_{ion}}{G_{group}C}} - A}{B}, \quad (8.5)$$

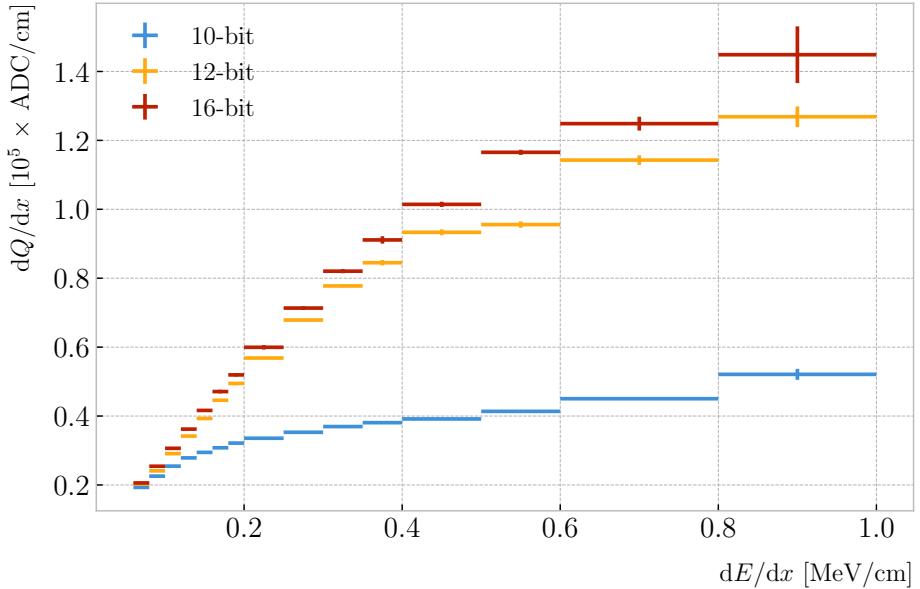
2576 where  $A$  and  $B$  are the calibration parameters we need to determine,  $W_{ion}$  is the average  
2577 energy to produce an electron-ion pair,  $G_{group}$  is the gain from the reclustering discussed  
2578 above and  $C$  is the calibration constant to convert number of electrons to ADC counts,  
2579 commonly refer to as gain (also to be obtained in the fit). In this case, I use a value  
2580 for the electron-ion production energy of  $W_{ion} = 26.4$  eV [109]. This value, used in our  
2581 simulation as well, was measured for gaseous argon in normal conditions, and therefore  
2582 should be checked in the future to describe correctly the high-pressure argon-CH<sub>4</sub> mixture  
2583 of ND-GAr.

2584 For the calibration fit I follow a procedure similar to the previous one for Eq. (8.4).  
2585 Binning the  $dE/dx$  range, I fit a LanGauss distribution to the corresponding  $dQ/dx$   
2586 distribution to obtain the most probable value. The resulting data points (red bars) are  
2587 shown in Fig. 8.4 (top panel), the horizontal error bars depict the width of the  $dE/dx$   
2588 bin whereas the vertical bars represent the error associated to the most probable value  
2589 estimation. A fit to the logarithmic function in Eq. (8.5) is also shown (black line).  
2590 For this I weighted the data points using the inverse of their relative error, obtaining  
2591 a reduced chi-square value of  $\chi^2/ndf = 2.22 \times 10^{-6}$ . The best fit parameters I found  
2592 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.  
2593 Figure 8.4 (bottom panel) shows the residuals between the data points and the fit.

2594 The value for the gain I obtained from the fit is in reasonable agreement with our  
2595 expectation. This value is set in GArSoft to 5 ADC/e by default.

2596 One interesting thing to check is what induces this non-linear relation between charge  
2597 and energy. The only effects that modify the amount of electrons reaching the readout  
2598 planes in the simulation are the transverse diffusion and the finite electron lifetime.  
2599 Once the electrons reach the readout chambers, the pad response functions are applied,  
2600 together with an electrons-to-ADC conversion and the ADC saturation limit.

## 8.1. $dE/dx$ measurement in the TPC



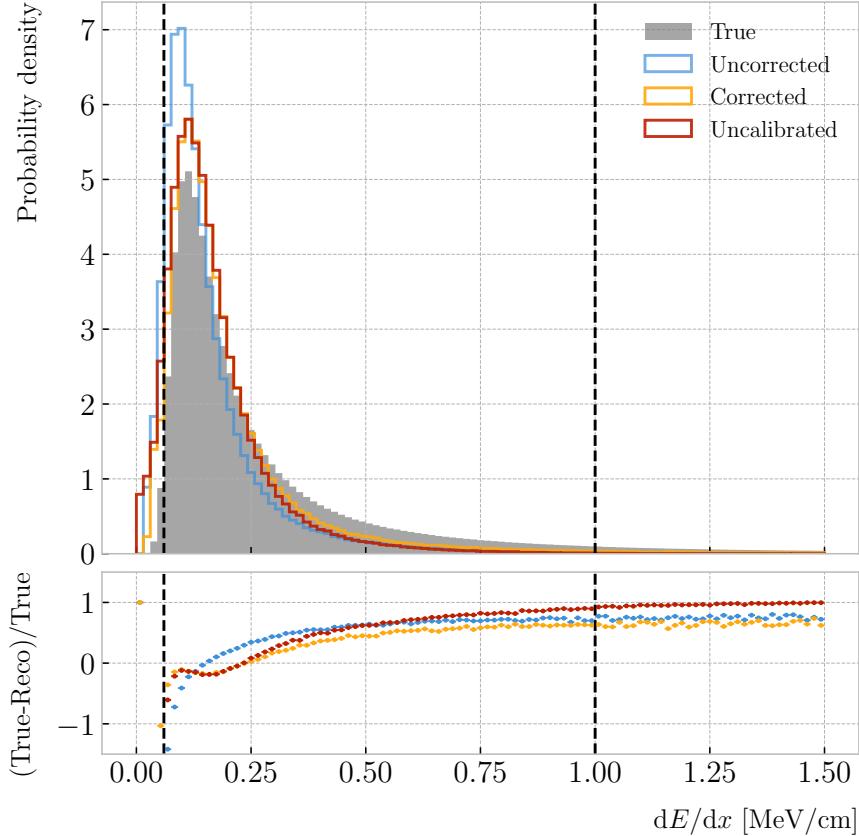
**Figure 8.5:** Fitted most probable  $dQ/dx$  values for each  $dE/dx$  bin for three different ADC bit limits, 10 (blue points), 12 (default, yellow points) and 16-bit (red points).

**Table 8.1:** Calibration parameters obtained from the fit of the ND-GAr simulated stopping proton sample to the calibration function from Eq. (8.5). 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	$1.44 \times 10^{-5}/12$	$0.949 \pm 0.024$	$3.53 \pm 0.58$	$4.52 \pm 0.29$

2601 By default, GArSot applies a 12-bit ADC limit, which can be changed in the  
 2602 simulation configuration. However, it can only be increased up to 16-bit, as we represent  
 2603 the ADC collection as a `std::vector<short>`. This way, I tried to change the saturation  
 2604 parameter to see how it affects the relation between reconstructed charge and energy.  
 2605 Figure 8.5 shows a comparison between the most probable  $dQ/dx$  for 10, 12 and 16-  
 2606 bit ADC limits. As expected, the lower the limit is the sooner the charge saturates.  
 2607 For higher ADC limits the relation between energy and charge remains linear up to  
 2608 higher  $dE/dx$  values, but even for the 16-bit limit the saturation is noticeable for values  
 2609  $\gtrsim 0.5$  MeV/cm.

## Chapter 8. Particle ID in GArSoft



**Figure 8.6:** 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.

Table 8.1 shows the results of fitting the samples with 10 and 16-bits ADC limits to the calibration function from Eq. (8.5), using the weights based on their relative error as described previously. One interesting feature to notice is how different the best fit points look for the 10-bit ADC saturation when compared to the other two, which are consistent with each other.

At this point we can compare the  $dE/dx$  distribution one gets from Geant4, i.e. the true energy loss distribution, and the distribution I found by applying the calibration function to our collection of reconstructed  $dQ/dx$  values. Figure 8.6 (top panel) shows the true (solid grey) and reconstructed (blue, labeled as uncorrected) distributions

## 8.1. $dE/dx$ measurement in the TPC

together. The dashed vertical lines indicate the region of validity of the calibration fit, i.e. the left and right edges of the first and last  $dE/dx$  bin respectively. Notice that these histograms are area-normalised, as the total number of true energy deposits is much higher than the number of reconstructed charge deposits. This is due to a combination of effects, like the finite spatial resolution of the detector, the hit clustering used in the track fitting and the reclustering we have applied here.

The two distributions are significantly different. That can be seen clearly when looking at the fractional residuals, shown in Fig. 8.6 (bottom panel). In particular, the position of the peak is off, which could bias the mean energy loss predictions. It seems like the difference between these may be due to an overall scaling factor. One possibility is to scale the most probable value of the reconstructed distribution to the most probable value predicted by Geant4. I do this by fitting both distributions using a LanGauss function, obtaining  $dE/dx_{MPV, true} = 0.1145 \pm 0.0005$  MeV/cm and  $dE/dx_{MPV, reco} = 0.0928 \pm 0.0005$  MeV/cm for the true and reconstructed most probable values respectively. These can be translated into a scaling factor  $S = 0.579 \pm 0.006$ .

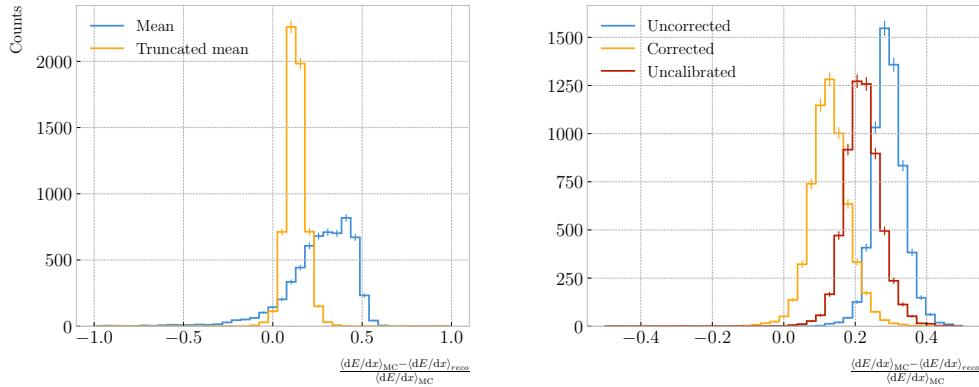
The result of applying the scaling correction can be seen in Fig. 8.6 (top panel). The corrected  $dE/dx$  distribution (yellow, labeled as corrected) peaks around the same value the true distribution does, as expected. Moreover, the high energy region is also slightly better described. For low ionisations, below the lower limit of the calibration fit, the differences between true and reconstructed are still significant. This low energy excess may be migration of some events from the peak region. The overall effect of the correction can be seen in the fractional residual plot in Fig. 8.6 (bottom panel).

One can also check what happens if instead of applying the logarithmic calibration we simply scale the  $dQ/dx$  distribution (post reclustering) to have the same most probable value as the true  $dE/dx$  distribution. In this case, following an analogous procedure to the one described earlier, I found the scaling factor  $S_{uncalibrated} = 0.414 \pm 0.002$  MeV/ADC<sup>4</sup>. The resulting distribution (red, labeled as uncalibrated) is also shown in in Fig. 8.6 (top panel). The behaviour of the new distribution is similar to the corrected case at low

---

<sup>4</sup>Notice that now the scaling factor is not dimensionless, as it acts more like a conversion factor here.

## Chapter 8. Particle ID in GArSoft



**Figure 8.7:** 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.

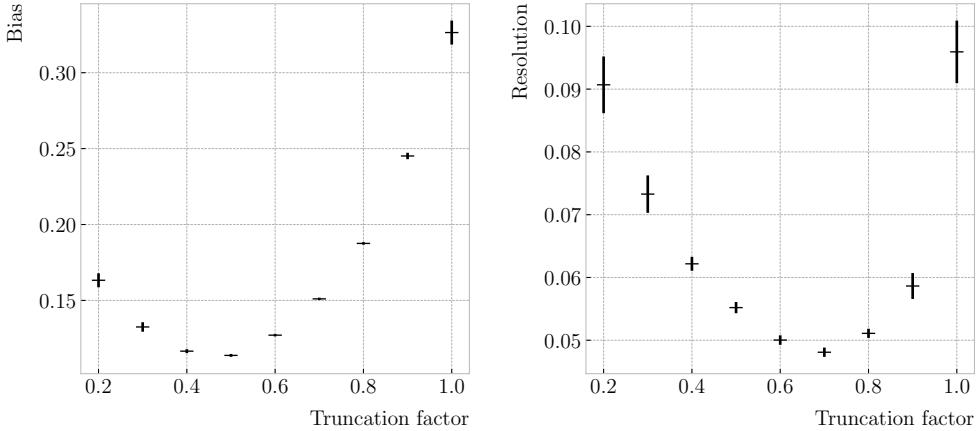
2647 energy losses, around the peak of the true distribution, but it is worse at describing the  
 2648 high energy tail. This is expected, it is in the high ionisation regime where saturation  
 2649 effects apply and therefore calibration is needed.

### 2650 8.1.2 Truncated $dE/dx$ mean

2651 Once we have a collection of  $dE/dx$  values for each reconstructed track, we can compute  
 2652 the corresponding most probable ionisation loss per unit length of the particle. This  
 2653 is the value predicted by the Bethe-Bloch or the PAI models, and together with a  
 2654 measurement of the momentum it allows for particle identification.

2655 However, estimating the most probable  $dE/dx$  value for each reconstructed track  
 2656 is not a trivial task. As mentioned before, the  $dE/dx$  distributions follow Landau-like  
 2657 distributions. Therefore, one should perform e.g. a LanGauss fit to correctly estimate  
 2658 the most probable values. Automating this kind of fits is often problematic, as they  
 2659 usually incur in convergence problems. Moreover, the reconstructed  $dE/dx$  distributions  
 2660 we obtain tend to have relatively small statistics, which may also produce poor fits. In  
 2661 practice, doing these unsupervised fits may degrade our performance, and a more robust  
 2662 method is preferred.

## 8.1. $dE/dx$ measurement in the TPC



**Figure 8.8:** 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.

2663 A possibility could be taking the mean of the reconstructed  $dE/dx$  distribution for  
 2664 each particle. The problem with this approach is that the high energy Landau tail,  
 2665 combined with our limited statistics, can induce large fluctuations in the computation  
 2666 of the mean. Imagine you have two protons with the same kinetic energy, but due to  
 2667 reconstruction problems in one case you did not get as many charge deposits reconstructed  
 2668 in its high ionisation loss region. If you do not remove the tails the computed  $dE/dx$   
 2669 means will be significantly different.

2670 In order to avoid those fluctuations, one can compute the mean of a truncated  $dE/dx$   
 2671 distribution instead. By keeping only a given fraction of the lowest energy deposits  
 2672 we obtain an estimate of the mean energy loss that is more resilient to reconstruction  
 2673 inefficiencies and statistical effects. Figure 8.7 (left panel) shows a comparison between  
 2674 the  $\langle dE/dx \rangle$  computed by taking the mean of the full distribution (blue line) and the  
 2675 60% lowest energy clusters (yellow line), for the stopping proton sample. The fractional  
 2676 residuals are computed for each proton, taking the corresponding means using their  
 2677 collections of true and reconstructed energy deposits. One can see that using the simple  
 2678 mean translates into a high bias and uncertainty in the  $\langle dE/dx \rangle$  estimation, whereas  
 2679 applying the truncation reduces both significantly.

2680 Additionally, I performed a comparison between the 60% truncated mean  $dE/dx$

## Chapter 8. Particle ID in GArSoft

2681 obtained using the different calibration methods discussed earlier, namely the uncorrected  
2682 (blue), corrected (yellow) and uncalibrated (red) distributions. The results are shown  
2683 in Fig. 8.7 (right panel). While the widths of these distributions are similar, the bias  
2684 obtained for the corrected sample, i.e. calibration function and correction factor applied,  
2685 is a factor of  $\sim 2$  lower than in the uncalibrated case and almost three times smaller  
2686 than for the uncorrected sample.

2687 The next step is to optimise the level of truncation we are going to apply to our  
2688 data. To do so, I used different truncation factors, i.e. the percentage of energy-ordered  
2689 reconstructed energy deposits we keep to compute the mean, on the corrected  $dE/dx$   
2690 sample of the stopping protons. Then, following the same procedure of computing the  
2691 fractional residuals as before, I fitted the resulting histograms using a double Gaussian  
2692 function. This is simply the sum of two Gaussian functions of the type:

$$g(x; \mu, \sigma, A) = A e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (8.6)$$

2693 I do not add the classical normalisation factor of the Gaussian,  $1/\sqrt{2\pi}\sigma$ , therefore  
2694 the amplitude  $A$  simply represents the maximum of the function. One of the two  
2695 Gaussian functions describes the core part of the distribution, while the other captures  
2696 the behaviour of the tails.

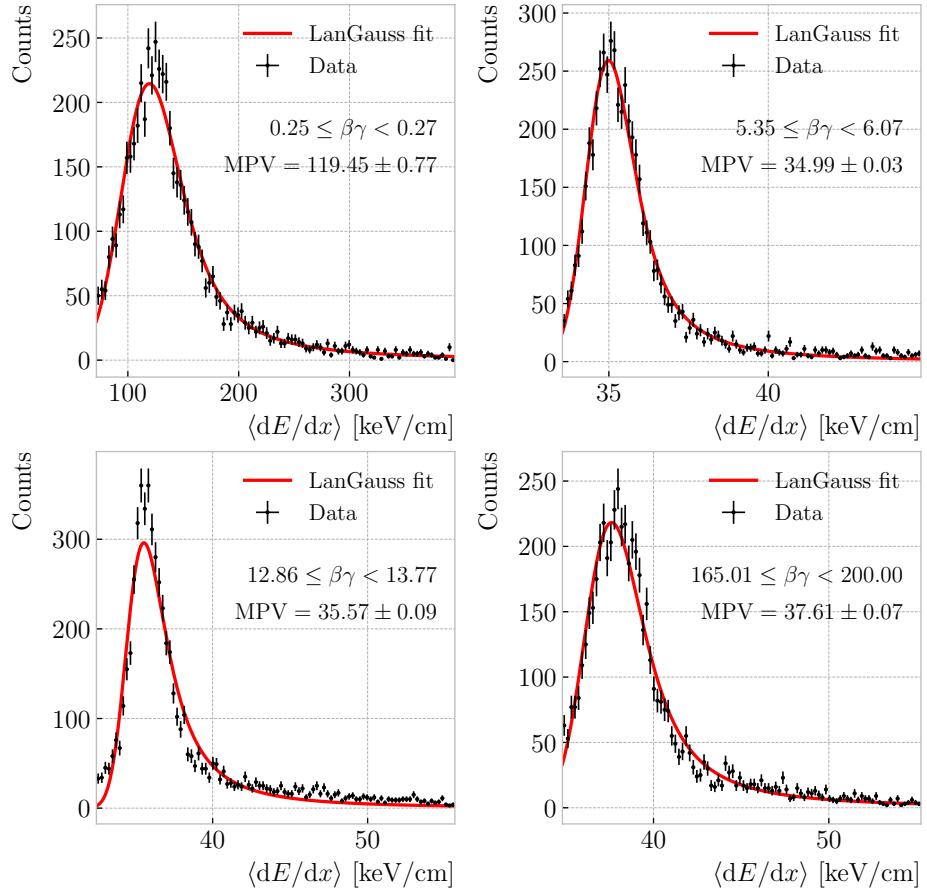
2697 For each truncation factor, I look at the bias and the resolution I obtain. I define  
2698 these as the weighted means of the corresponding parameters in the fits:

$$\bar{x} = \frac{A_{core} x_{core} + A_{tail} x_{tail}}{A_{core} + A_{tail}}, \quad (8.7)$$

2699 where  $A_{core}$  and  $A_{tail}$  are the amplitudes of the core and tail distributions respectively  
2700 and  $x$  is either the mean  $\mu$  or the width  $\sigma$  of said distributions.

2701 Figure 8.8 shows the bias (left panel) and the resolution (right panel) I obtained  
2702 for the stopping proton sample, using different values of the truncation. From these, it  
2703 can be seen that a truncation factor of 50% minimises the bias in the estimation, while

## 8.1. $dE/dx$ measurement in the TPC



**Figure 8.9:** Examples of the truncated mean  $dE/dx$  LanGauss fits for various  $\beta\gamma$  bins, from a simulated FHC neutrino sample.

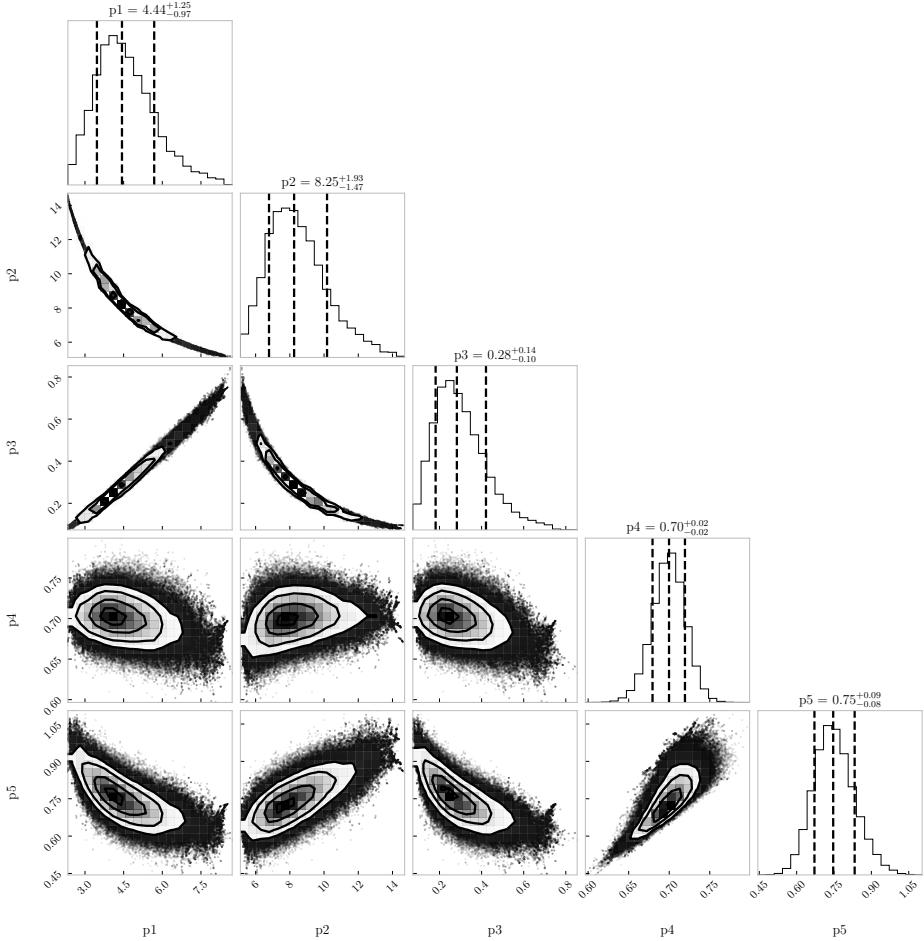
2704 70% gives the best resolution. That way, I settled on the intermediate value of 60%  
 2705 truncation, which yields a  $\langle dE/dx \rangle$  resolution of  $5.00 \pm 0.08$  % for stopping protons.

### 2706 8.1.3 Mean $dE/dx$ parametrisation

2707 Now that we have a way to estimate the mean energy loss of a particle in the HPgTPC,  
 2708 we can determine the value of the free parameters in the ALEPH formula, Eq. (8.3).  
 2709 For this, I used a sample of  $10^5$  reconstructed FHC neutrino events inside ND-GAr. In  
 2710 this case I cannot use the stopping proton sample, as we need to cover the full kinematic  
 2711 range of interest for the neutrino interactions in our detector.

2712 The original data does not contain an estimation of the velocity of the tracks, instead  
 2713 the tracks have a value for the reconstructed momentum and the associated PDG code

## Chapter 8. Particle ID in GArSoft



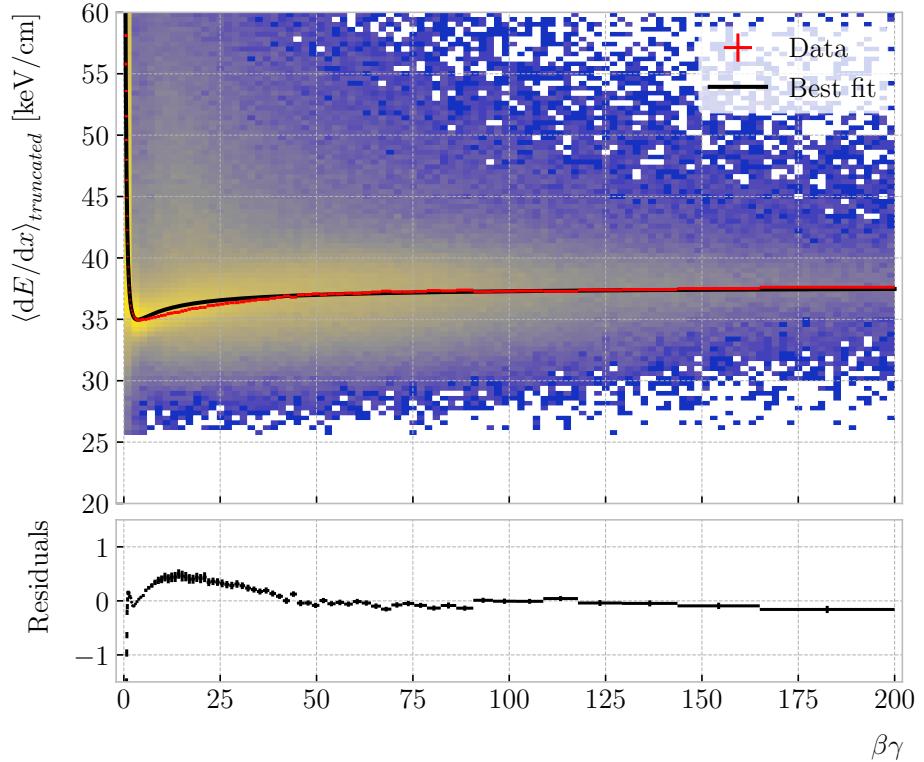
**Figure 8.10:** Resulting one and two dimensional projections of the posterior probability distributions of the ALEPH  $\langle dE/dx \rangle$  parameters obtained by fitting the 60% truncated mean  $dE/dx$  values from a FHC neutrino sample in ND-GAr. The vertical dashed lines in the 1D distributions represent the 16th, 50th and 84th percentiles.

2714 of the Geant4-level particle that created the track. Therefore, one can select some of the  
 2715 particles in the data, in this case I selected electrons, muons, pions and protons, and  
 2716 compute  $\beta$  and  $\gamma$  using the reconstructed momentum and their mass. In terms of  $\beta\gamma$   
 2717 the mean  $dE/dx$  does not depend on the particle species, so one can consider all the  
 2718 dataset as a whole. For this fit, I will express  $\beta$  in terms of the  $\beta\gamma$  product as:

$$\beta = \frac{\beta\gamma}{\sqrt{1 + (\beta\gamma)^2}}, \quad (8.8)$$

2719 which can be easily proven from the definition of  $\gamma$ .

### 8.1. $dE/dx$ measurement in the TPC

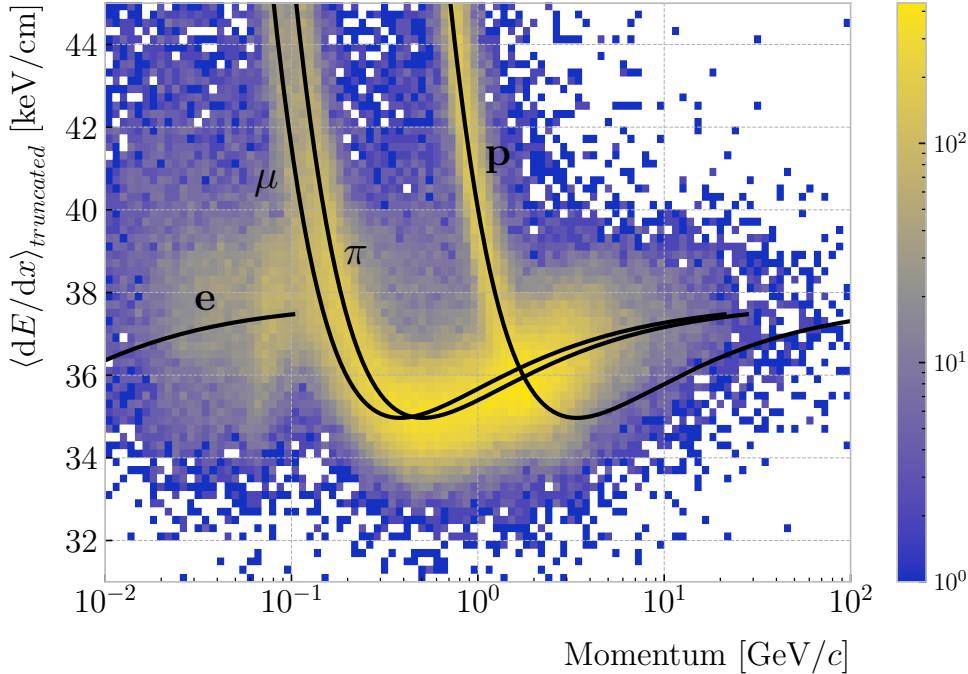


**Figure 8.11:** Truncated mean  $dE/dx$  obtained for the FHC neutrino sample as a function of the  $\beta\gamma$  product (upper panel). Also shown are the fitted most probable values for each  $\beta\gamma$  bin (red points) and the best fit obtained using the ALEPH parametrisation (black line). The residuals resulting from the fit are shown in the lower panel.

2720 Next, I bin the data in  $\beta\gamma$ . I chose a fine binning so as to capture the different  
 2721 features of the ionisation curve. Instead of fixing the bin width, I select them so each one  
 2722 has approximately the same statistics. Then, for each  $\beta\gamma$  slice, I compute the median  
 2723 and the interquartile range (IQR) of the  $\langle dE/dx \rangle$  distribution. Using these, I make a  
 2724 histogram in the range [median – IQR, median + 5 IQR], which I fit to a LanGauss  
 2725 function in order to extract the MPV. Using this range accounts for the asymmetric  
 2726 nature of the distributions, while also helps avoiding a second, lower maximum present  
 2727 at low  $\beta\gamma$ , probably a result of reconstruction failures.

2728 A few examples of these fits are shown in Fig. 8.9. The chosen values of  $\beta\gamma$  sit in  
 2729 very distinct points along the  $\langle dE/dx \rangle$  curve, going from the high ionisation region at  
 2730 low velocities (top left panel), to the minimum point (top right panel), the beginning of

## Chapter 8. Particle ID in GArSoft



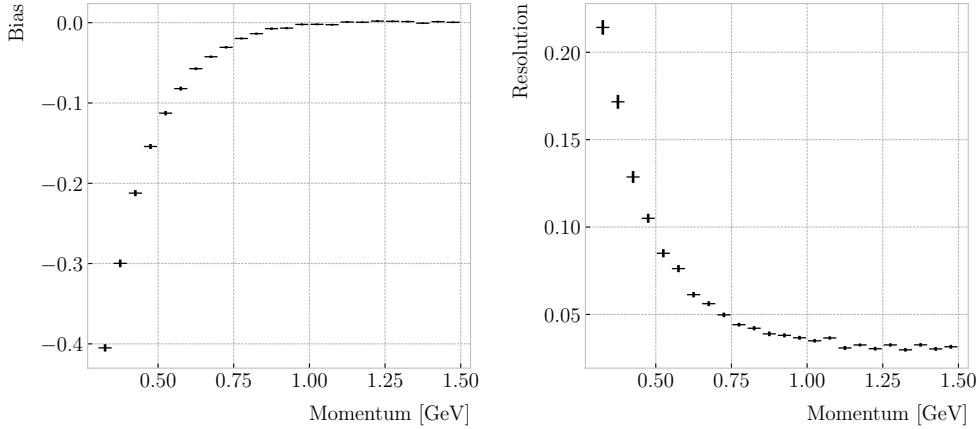
**Figure 8.12:** Distribution of the 60% truncated mean  $\langle dE/dx \rangle$  versus reconstructed momentum for the FHC neutrino sample. The black lines indicate the predictions of the ALEPH parametrisation for electrons, muons, charged pions and protons.

2731 the relativistic rise (bottom left panel), and the plateau produced by the density effect  
 2732 (bottom right panel).

2733 I used the resulting most probable  $\langle dE/dx \rangle$  values and the centres of the  $\beta\gamma$  bins as  
 2734 the points to fit to the ALEPH formula. For this particular fit I used the least-squares  
 2735 method to get a first estimation of the ALEPH parameters. Applying some uniform  
 2736 priors, I then used these values as the starting point of a 100000 steps MCMC. Figure 8.10  
 2737 shows the posterior probability distributions I obtain for each parameter. The reported  
 2738 best fit points are based on the 16th, 50th, and 84th percentiles in the marginalised  
 2739 distributions.

2740 The resulting fit (black line), compared to the data points (red points) and the  
 2741 underlying distribution is shown in Fig. 8.11 (top panel). The overall fit is good, with a  
 2742 reduced chi-squared of  $\chi^2/ndf = 1.02$ . However, there are some regions where the fit  
 2743 does not describe the data correctly, like the very low  $\beta\gamma$  regime, where the fit severely  
 2744 underestimates for energy losses  $\gtrsim 50$  keV/cm, and the start of the relativistic raise,

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.13:** Estimated values of the mean  $dE/dx$  bias (left panel) and resolution (right panel) obtained for the true protons in a FHC neutrino sample.

where we have a slight overestimation. This is a result of those points having a larger uncertainty when compared to the ones around the dip or the plateau areas. These differences can be better seen in the residual plot, Fig. 8.11 (bottom panel).

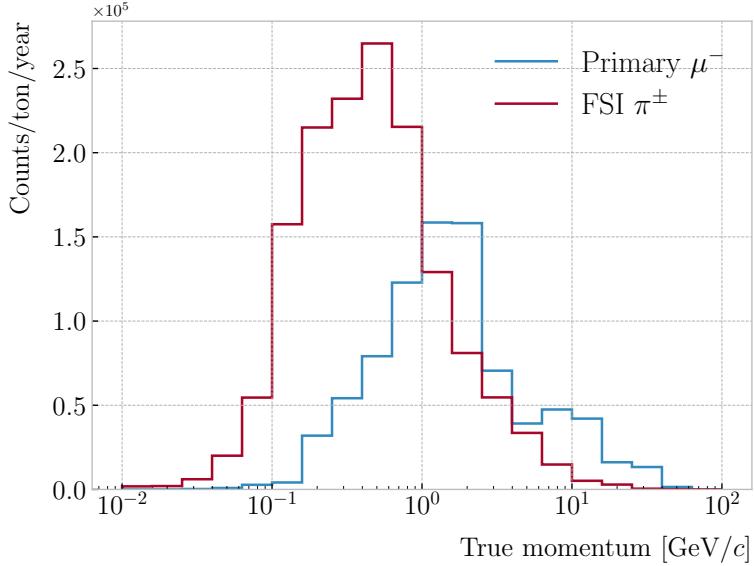
### 8.1.4 Particle identification

## 8.2 Muon and pion separation in the ECal and MuID

As it could be seen from Fig. 8.12, it is not possible to separate muons and charged pions in the HPgTPC using  $dE/dx$  for momenta  $\gtrsim 300 \text{ MeV}/c$ . In ND-GAr, approximately 70% of the interactions in FHC mode will be  $\nu_\mu$  CC (compared to the 47% of  $\bar{\nu}_\mu$  CC interactions when operating in RHC mode), while 24% are neutral currents. Out of these, around 53% and 47% of them will produce at least one charged pion in the final state, respectively. Figure 8.14 shows a comparison between the spectra of the primary muons and the charged pions for  $\nu_\mu$  CC interactions in ND-GAr producing one or more charged pions. From this, one can see that (i) the majority of muons and charged pions are not going to be distinguishable with a  $\langle dE/dx \rangle$  measurement, and that (ii) particle identification is necessary both to classify correctly the  $\nu_\mu$  CC events and identify the primary muon within them.

ND-GAr features two other subdetectors which can provide additional information

## Chapter 8. Particle ID in GArSoft



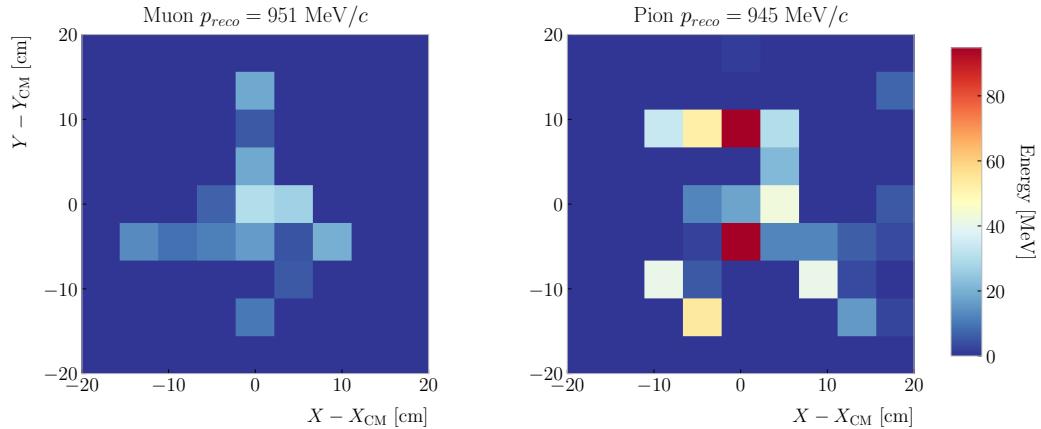
**Figure 8.14:** True momentum distribution for the primary muon in  $\nu_\mu$  CC  $N\pi^\pm$  interactions inside the fiducial volume of ND-GAr (blue line), compared to the post FSI charged pion spectrum (red line).

for this task, namely the ECal and MuID. The current ECal design, described in (ref section), consists of 42 layers, made of 5 mm of Pb, 7 mm of plastic scintillator and a 1 mm PCB board. The total thickness of this calorimeter is 1.66 nuclear interaction lengths or 1.39 pion interaction lengths. The MuID design is in a more conceptual stage, however it is envisioned to feature layers with 10 cm of Fe and 2 cm of plastic scintillator<sup>5</sup>. With its three layers, it will have a thickness of 1.87 or 1.53 nuclear or pion interaction lengths, respectively.

Because pion showers are dominated by inelastic nuclear interactions, the signatures of these particles in the calorimeter will look significantly different from those of muons. Although our ECal is not thick enough to fully contain the hadronic showers of the charged pions at their typical energies in FHC neutrino interactions, they can still be used to understand whether the original particle was more hadron-like or MIP-like. In Fig. 8.15 I show two examples of energy distributions created by a muon (left panel) and a charged pion (right panel) of similar momenta interacting in the ECal. These

<sup>5</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.

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.15:** 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.

2776 figures represent the transverse development of the interactions. For each of them, I  
 2777 computed the principal component and centre of mass of the interaction, projecting  
 2778 the position of the hits onto the plane perpendicular to that direction, and taking the  
 2779 distances relative to the centre. It can be seen that the muon follows an almost MIP-like  
 2780 behaviour, being the central bin in the histogram the one with the highest deposited  
 2781 energy. On the other hand, the pion not only deposits more energy overall, but also this  
 2782 energy is more spread-out among the different hits. It is this kind of information that  
 2783 would allow us to tell apart muons from pions.

2784 This way, I identify three main action points that need to be addressed if one wants  
 2785 to use these detectors to distinguish between muons and charged pions. These are:

- 2786 1. the way we make the associations between tracks in the HPgTPC to the activities  
 2787 (what in GArSoft we call clusters) in the ECal and the MuID,
- 2788 2. what variables or features one can extract from the calorimeters that encapsulate  
 2789 the information we are interested about,
- 2790 3. and how to carry out the classification problem.

## Chapter 8. Particle ID in GArSoft

### 2791 8.2.1 Track-ECal matching

2792 One of the main players in the muon and pion separation is the way we associate clusters  
2793 in the ECal to reconstructed tracks in the TPC. Missing some associations or making  
2794 wrong ones can bias the ECal quantities that we can use for classifying particles. The  
2795 current algorithm in GArSoft provides precise associations, i.e. most of the associations  
2796 that it produces are correct, but it appears to miss an important number of associations  
2797 (at least when using the default configuration).

2798 The current TPC track-ECal cluster association algorithm is divided in four parts.  
2799 It first checks whether the track end point fulfils certain conditions to be extrapolated.  
2800 There are two cut values in this step, one for the drift direction and other radial.

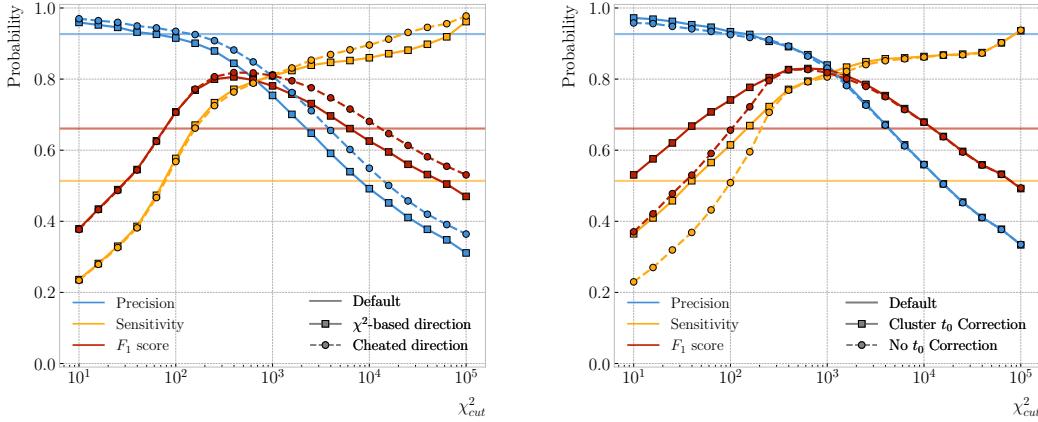
2801 If the point can be extrapolated, the code computes the coordinates of the centre  
2802 of curvature using the Kalman fit estimates at the track end ( $y, z, 1/R, \phi, \tan\lambda$ ). It  
2803 then compares the distance between this and the cluster in the  $(z, y)$  plane with  $R$ . This  
2804 introduces another cut in the perpendicular direction.

2805 The next step is different for clusters in the barrel or in one of the end caps. If it  
2806 is a barrel cluster the algorithm extrapolates the track up to the radial distance of the  
2807 cluster. There are three possible outcomes, the extrapolated helix can cut the cylinder  
2808 of radius  $r_{clus}$  two, one or zero times. I get the cut point that is closer to the cluster and  
2809 check that it is either in the barrel or the end caps. Computing the difference between  
2810 the  $x$  coordinates of the cluster and the extrapolated point, the module checks that this  
2811 is not greater than a certain cut. If the cluster is in an end cap, I propagate the track  
2812 up to the  $x$  position of the cluster. Then, the algorithm computes the angle in the  $(z, y)$   
2813 plane between the centre of curvature and the cluster,  $\alpha$ , and the centre of curvature  
2814 and the propagated point,  $\alpha'$ . A cut is applied to the quantity  $(\alpha - \alpha')R$ .

2815 If the cluster contains more than a certain number  $N$  of hits, I apply an extra cut to  
2816 the dot product of the direction of the track at the propagated  $x$  value and the cluster  
2817 direction.

2818 The code makes sure to only associate one end of the track (if any) to a cluster.

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.16:** 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).

2819 However, it can associate more than one track to the same cluster. This makes sense,  
2820 as different particles can contribute to the same cluster in the ECal, but it makes it  
2821 difficult to quantify the relative contributions of the tracks to a certain cluster.

2822 As a way of comparing the performance of this algorithm, a new, simpler association  
2823 module was written. The goal was to have a simple and robust algorithm, which depends  
2824 on as few parameters as possible and that can produce a one-to-one matching between  
2825 tracks and ECal clusters.

2826 For each reconstructed track, the new algorithms applies the same procedure to the  
2827 forward and the backward fits irrespective of their end point positions. It first gets the  
2828 Kalman fit parameters at the corresponding end point together with the  $X$  position,  $x_0$ ,  
2829 ( $y_0$ ,  $z_0$ ,  $1/R$ ,  $\phi_0$ ,  $\tan\lambda$ ).

2830 For each ECal cluster, I compute the radial distance to the centre of the TPC and  
2831 find the  $\phi$  value in the range  $[\phi_0, \phi_0 + \text{sign}(R)\phi_{max}]$  that makes the propagated helix  
2832 intersect with the circle defined with such radius. The  $(x, y, z)$  position of the helix for  
2833 the  $\phi$  value found (if any) is then computed. In case there are two intersections, I keep  
2834 the one that minimises the distance between  $(y, z)$  and  $(y_c, z_c)$ .

2835 I then calculate  $\chi^2$  value based on the Euclidean distance between the propagated

## Chapter 8. Particle ID in GArSoft

2836 point and the cluster:

$$\chi^2/ndf = \frac{\sum_{n=0}^2 (x^{(n)} - x_c^{(n)})^2}{3}. \quad (8.9)$$

2837 If there was no intersection I store a  $-1$  instead. In the end, for each reconstructed track  
2838 in the event one ends up with two collections of  $\chi^2$  values, one for each ECal cluster  
2839 and fit directions.

2840 The current code only supports having ECal clusters associated to one end of each  
2841 track. We have two options to decide what track end to keep. The first one tries to  
2842 cheat the selection, looking at the distance between the two track ends and the true  
2843 start position of the associated MC particle. The second one keeps the track end with  
2844 more  $\chi^2$  entries below the cut.

2845 This feature of only considering one track end limits the algorithm, making it not  
2846 suitable for reconstructing events with particles originating outside the TPC. However,  
2847 as for the moment the main concern of the group is the study of neutrino interactions  
2848 off the gaseous argon, this is an acceptable assumption.

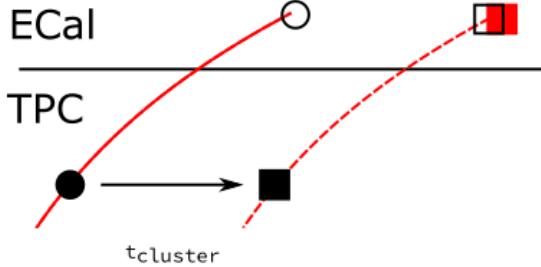
2849 In order to associate a cluster to a track, I take all clusters with a  $\chi^2$  value in the  
2850 range  $[0, \chi_{cut}^2]$ . If a cluster has been assigned to more than one track we leave it with  
2851 the one with the lowest  $\chi^2$ .

2852 This default behaviour of the algorithm can be modified to associate more than one  
2853 track to each cluster. Not only that, but the  $\chi^2$  values can be used to assign relative  
2854 weights to the different contributions.

2855 To evaluate the performance of the association method, I use a binary classification  
2856 approach. In this case, I check the leading MC Track IDs associated to the reconstructed  
2857 tracks and ECal clusters. I count an association as true positive (TP) if both Track  
2858 IDs coincide. An association is considered false positive (FP) when the Track IDs are  
2859 different. If a cluster has not been associated to any track but it shares the Track ID  
2860 with a reconstructed track it is counted as a false negative (FN).

2861 For the testing, I used a sample of 10000 FHC neutrino events inside the HPgTPC.  
2862 Figure 8.16 (left panel) shows the precision (blue line), sensitivity (yellow line), and  $F_1$

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.17:** 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,  $t_{cluster}$ .

2863 score (red line) I obtained for different values of  $\chi^2_{cut}$ . For comparison, the same metrics  
 2864 computed for the default algorithm with the current configuration are also shown (dashed  
 2865 lines). In the case of the new algorithm, I used both the  $\chi^2$ -based method to estimate  
 2866 the track direction described earlier (square markers) and the cheated direction from the  
 2867 Geant-level information (circle markers). For either of these we achieve similar values of  
 2868 the precision compared to the old code, while having a considerably higher sensitivity.  
 2869 It can be seen that cheating the direction of the tracks only makes a difference at high  
 2870  $\chi^2_{cut}$ , past the optimal value of the cut around the  $F_1$  score maximum. Therefore, I set  
 2871 the  $\chi^2$  method as the default.

2872 One of the possible weak points of this approach is that it relies on the position along  
 2873 the drift direction to make the decisions. Within the current ND-GAr design implemented  
 2874 in GArSoft, the timing information is provided by the ECal. That effectively means  
 2875 that prior to make the track-ECal associations the reconstructed  $x$  positions of the track  
 2876 trajectories differ from the simulated ones by an amount:

$$x_{reco}^{(n)} - x_{sim}^{(n)} = v_{drift} t_0, \quad (8.10)$$

2877 where  $v_{drift}$  is the mean drift velocity in our medium and the initial time is in the range  
 2878  $t_0 \in [0, t_{spill}]$  where  $t_{spill}$  is the spill length. For a 10  $\mu\text{s}$  spill this translates into a  
 2879 maximum 30 cm uncertainty on the drift direction position.

2880 The current default in GArSoft sets  $t_0 = 0$ , but the functionality to randomly sample

## Chapter 8. Particle ID in GArSoft

2881 this within the spill time is in place. Therefore, we need to understand what is the impact  
2882 of a non-zero  $t_0$  on the associations algorithm and foresee possible ways of minimising a  
2883 loss in performance.

2884 Figure 8.17 represents a possible option to tackle the association problem when  
2885 having events with a non-zero initial time  $t_0$ . The black and white circles represent the  
2886 original points, whereas the squares indicate the corrected positions. The end points of  
2887 the track and the propagated points up to the cluster radius are indicated using filled  
2888 and unfilled markers respectively. The red square represents the position of the cluster.

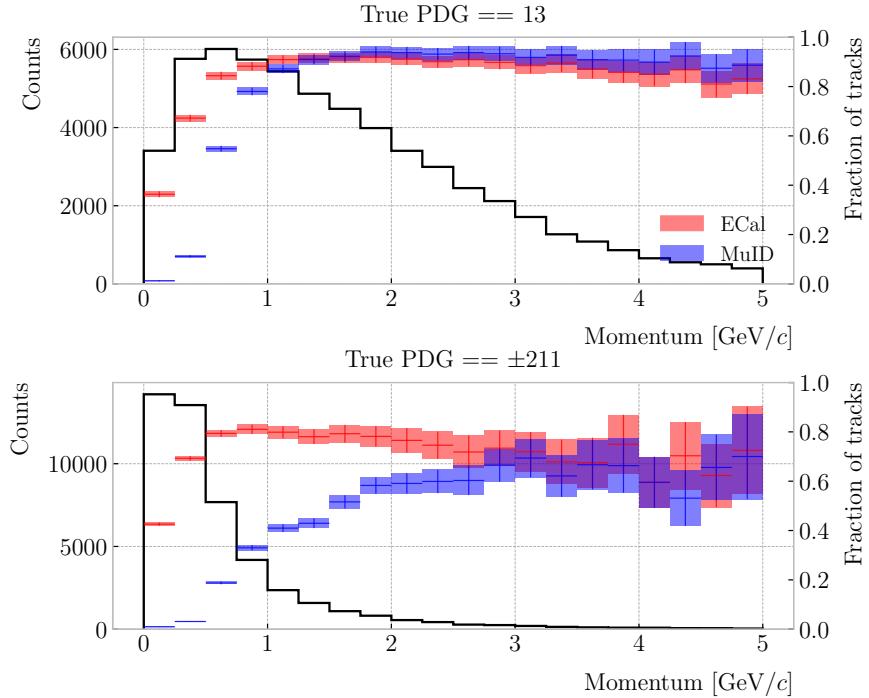
2889 Here I try to correct for the drift coordinate position using the time associated to the  
2890 cluster. Assuming that the drift time is much larger than the propagation time,  $t_{cluster}$   
2891 could be used as a good estimation of the  $t_0$ . An alternative can be using the earliest  
2892 time associated to a hit in said cluster. Doing this for each cluster before computing  
2893 the  $\chi^2$  value could be used as an alternative to knowing the specific value of the  $t_0$ , as  
2894 when the association is correct this will provide the right correction but its impact is  
2895 small enough to not change the position significantly in the case the cluster does not  
2896 correspond to a given track.

2897 I tested the effect of this correction again using a sample of 10000 FHC neutrino  
2898 events. Figure 8.16 (right panel) shows the precision (blue line), sensitivity (yellow line),  
2899 and  $F_1$  score (red line) for the case the cluster  $t_0$  correction is applied (square markers)  
2900 and for the no correction case (circle markers), as a function of  $\chi^2_{cut}$ . In this case, the  
2901 differences are particularly notorious at low values of the cut. It makes sense, as the  $t_0$   
2902 effect becomes subdominant when the distance we consider grows large. Overall, the  
2903 correction increases the sensitivity while keeping the precision almost unchanged. As a  
2904 result, I apply the  $t_0$  correction to the generated samples as the default.

### 2905 8.2.2 Classification strategy

2906 The problem of the muon and charged pion separation has to be viewed in the broader  
2907 context of the particle identification in our detector. Focusing on the beam neutrino  
2908 interactions, it is clear that we are going to have muons and pions spanning a broad

## 8.2. Muon and pion separation in the ECal and MuID

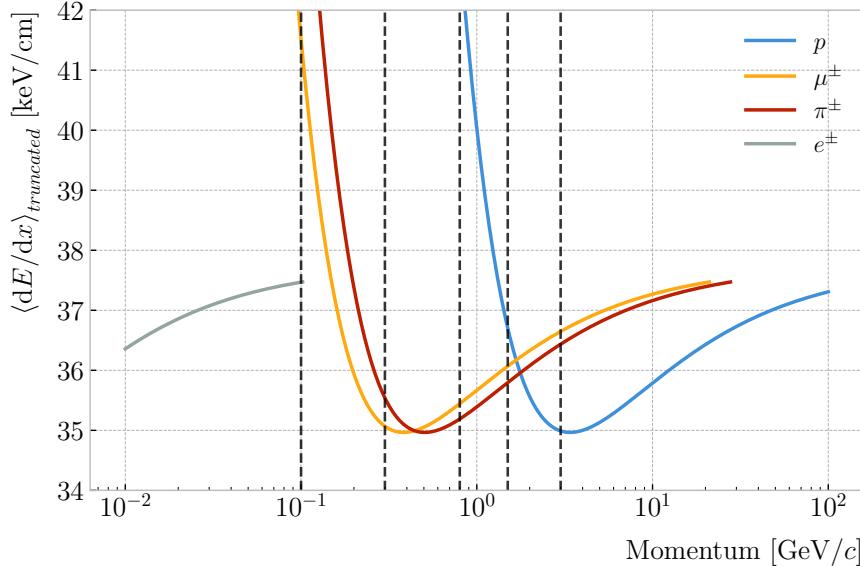


**Figure 8.18:** Momentum distribution for the reconstructed muons (top panel) and charged pions (bottom panel) in a FHC neutrino sample, together with the fraction of them reaching the ECal (red) and MuID (blue). Each entry corresponds to a reconstructed track, backtracked to a true muon or pion which has not produced any other reconstructed track.

momentum range. Not only that, but we will also have other particles with similar characteristics that will make the classification even more challenging. Therefore, we are presented with a task that will depend heavily on the kinematic range we are looking at each time, as both the available information and the possible impurities of other particle species vary.

For instance, distinguishing muons from pions could be difficult at low momenta, as a great number of them do not reach the ECal. Therefore, we could think of tailoring a version of the classification for that particular case, which could be complemented with a  $dE/dx$  measurement. Likewise, for momenta  $\gtrsim 1$  GeV muons and pions reach the calorimeters efficiently, but so do protons. Because of this, one can try to train another classifier for this energy range, and rely on other methods to remove as many of the protons as possible.

## Chapter 8. Particle ID in GArSoft



**Figure 8.19:** 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.

Figure 8.18 shows the momentum distribution of the reconstructed muons (top) and pions (bottom) in a FHC sample. It also contains the fraction of particles reaching the ECal (red) and MuID (blue), for the different momentum bins. In Fig. 8.19 I show the mean  $dE/dx$  of different particles as a function of the momentum, computed using the ALEPH parametrisation with the best fit parameters found in Subsec. 8.1.2.

Using these two figures as references, I decided to approach the classification by dividing the problem into six different momentum regions. A summary of these can be found in Tab. 8.2. The basic idea is to exploit all the information that is available in each region and . For the problem at hand, I prepared separated samples of isotropic single muons and pions, with momenta uniformly distributed along the corresponding momentum range. Each sample contains 50000 events of the corresponding particle species. I did not generate samples for the first region, as it is assumed that the separation can be achieved using  $dE/dx$  only. For the last region, I generated particles up to a momentum of 10  $\text{GeV}/c$ , as that is well above the typical energies of muons and pions from FHC neutrino interactions in ND-GAr.

## 8.2. Muon and pion separation in the ECal and MuID

**Table 8.2:** 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

2936     Additionally, I prepared another sample of 100000 FHC neutrino events. For each  
 2937    interaction, I select the reconstructed particles which were backtracked to true muons or  
 2938    charged pions. I use this dataset to perform validation checks, to see how the models  
 2939    trained with the single particle data generalise to a more realistic scenario.

2940     To tackle this classification problem, I make use of Boosted Decision Trees (BDT). A  
 2941    decision tree uses a flowchart-like structure to make decisions based on some input data.  
 2942    It starts from a root node, which represents the complete dataset, and then it splits  
 2943    this based on the variable or feature which gives the best separation between classes,  
 2944    creating two new nodes. The process repeats for each node until it reaches a certain  
 2945    limit, like a maximum number of splits or some tolerance criteria. The last set of nodes  
 2946    are often called leave nodes, and represent the final prediction of the classifier.

2947     Boosting refers to a family of methods to combine the predictions from multiple  
 2948    classifiers, following a sequential approach where each new model learns from the errors  
 2949    of the previous one. The process starts with a simple decision tree, which is used to  
 2950    make predictions on the training data. Then, the data points misclassified by the first  
 2951    model are assigned higher weights, and another decision tree is trained on the data with  
 2952    adjusted weights. The predictions of the two trees are then combined, and the cycle  
 2953    repeats for a predefined number of iterations. Gradient boosting uses the direction of  
 2954    the steepest error descent to guide the learning process and improve the accuracy with  
 2955    each iteration.

## Chapter 8. Particle ID in GArSoft

### 2956 8.2.3 Feature selection and importance

2957 Using the reconstructed tracks as a starting point, I compute a number of ECal and  
2958 MuID variables for each of them. As there can be more than one cluster associated to a  
2959 track, what I do is collect all associated clusters and compute these variables from the  
2960 complete collection of associated hits. For the MuID, because it only features three layers  
2961 and typically there will be less hits, I also allow single hits to be associated with tracks<sup>6</sup>.  
2962 I can roughly divide the variables in three types: energy-related, geometry-related and  
2963 statistical. In the following, I briefly describe the variables related exclusively to the  
2964 ECal:

2965 • Energy-related ECal

- 2966 – ECal total energy (ClusterTotalEnergy): sum of the energy of all the ECal  
2967 hits.
- 2968 – Mean ECal hit energy (HitMeanEnergy): mean of the hit energy distribution.
- 2969 – Standard deviation ECal hit energy (HitStdEnergy): standard deviation of  
2970 the hit energy distribution.
- 2971 – Maximum ECal hit energy (HitMaxEnergy): maximum of the hit energy  
2972 distribution.

2973 • Geometry-related ECal

- 2974 – Mean distance hit-to-cluster (DistHitClusterMean): mean of the distance  
2975 distribution between the hits and the corresponding cluster's main axis.
- 2976 – RMS distance hit-to-cluster (DistHitClusterRMS): root mean square of the  
2977 distance distribution between the hits and the corresponding cluster's main  
2978 axis.
- 2979 – Maximum distance hit-to-centre (DistHitCenterMax): maximum of the  
2980 distance distribution between the hits and the centre of the TPC.

---

<sup>6</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.

## 8.2. Muon and pion separation in the ECal and MuID

- 2981 – Time-of-Flight velocity (TOFVelocity): slope obtained when fitting a straight  
2982 line to the hit time versus hit distance to the centre (i.e.  $d = v \times t$ ).

2983 • **Energy and geometry ECal**

- 2984 – Radius 90% energy (Radius90E): distance in the hit-to-cluster distribution  
2985 for which 90% of the total energy is contained in the hits that are closer to  
2986 the axis (i.e. radius that contains 90% of the energy).

2987 • **Statistical ECal**

- 2988 – Number of hits (NHits): total number of hits associated to the track.  
2989 – Number of layers with hits (NLayers): not really a count of all layers with  
2990 hits but the difference between the last and the first layer with hits.

2991 Figure 8.20 shows the distributions of three different ECal variables, separating true  
2992 muons (blue) and charged pions (red), for the five momentum ranges considered. I chose  
2993 to show one feature from each category, namely the mean energy per hit (left column),  
2994 the mean distance between the hits and the centre of the cluster (middle column), and  
2995 the number of ECal layers with hits (right column). These give an idea of the separating  
2996 power of the different features, and how it changes considerably with the energy. In  
2997 the number of layers with hits distributions, the peak at 6 is due to the fact that the  
2998 first six ECal layers sit inside the pressure vessel<sup>7</sup>. Therefore, some of the particles get  
2999 stopped crossing it, never making it to the seventh layer.

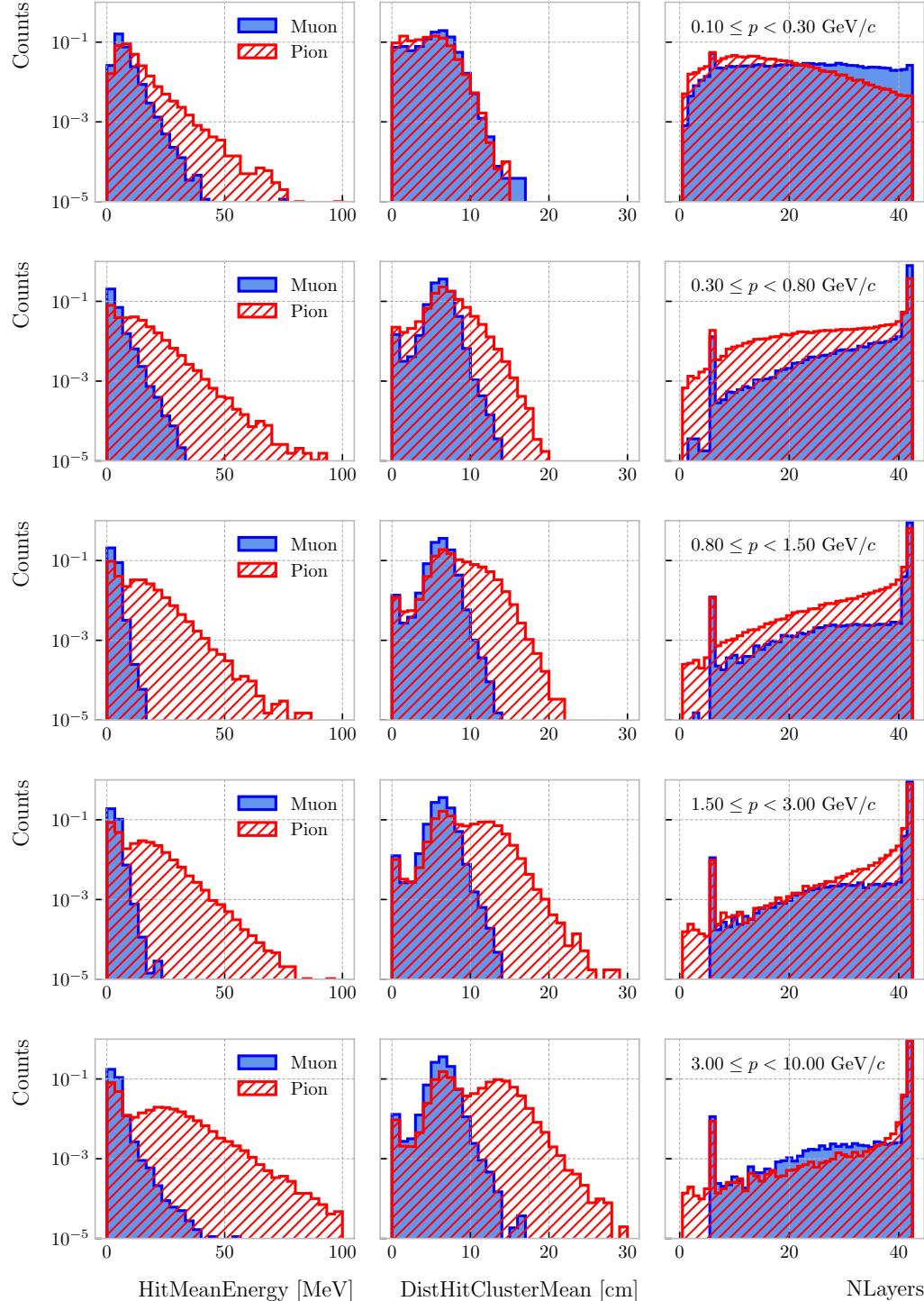
3000 In the case of the MuID, because at low momenta a significant fraction of the particles  
3001 do not make it past the ECal, I only consider the information coming from this detector  
3002 for momenta  $\geq 0.8$  GeV/ $c$ , i.e. for the last three momentum regions. The variables I  
3003 extract from it are the following:

3004 • **Energy-related MuID**

---

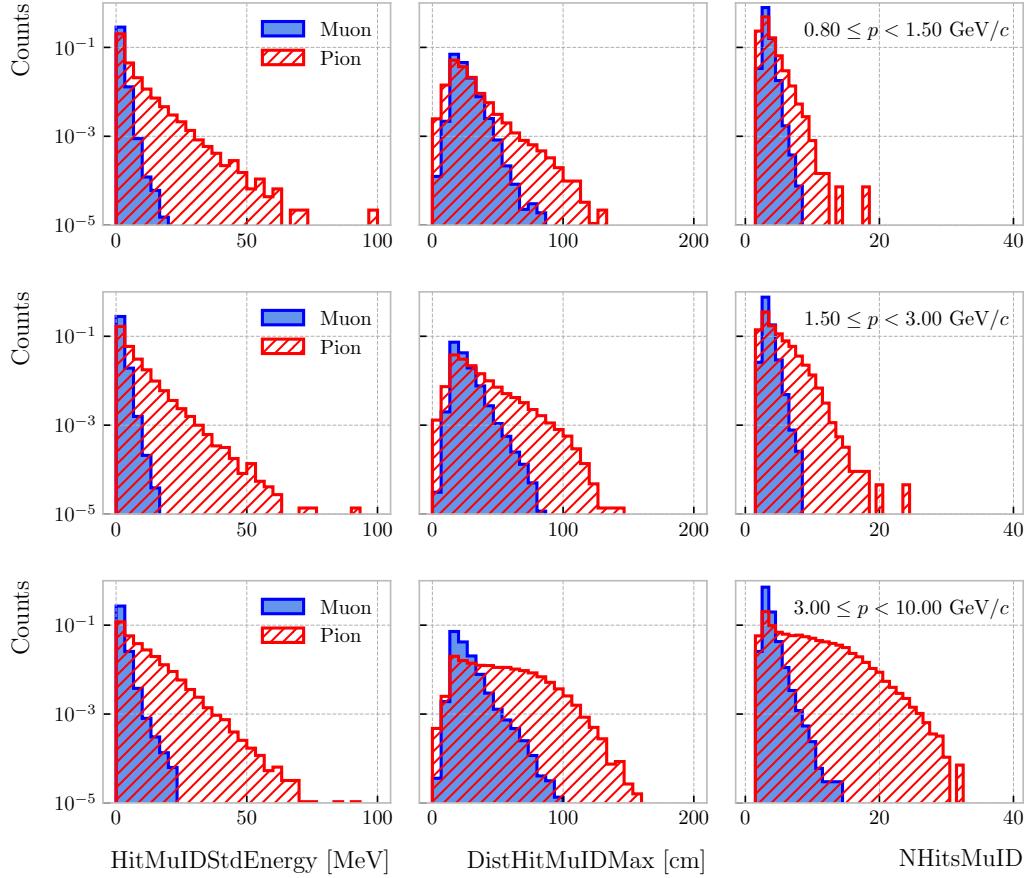
<sup>7</sup>Note to self: check this. I thought the ECal barrel had 8 layers of tiles, and that all of them were inside the pressure vessel.

## Chapter 8. Particle ID in GArSoft



**Figure 8.20:** 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.

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.21:** 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.

- 3005           – MuID total energy (ClusterMuIDTotalEnergy): sum of the energy of all the
- 3006            MuID hits.
- 3007           – Mean MuID hit energy (HitMuIDMeanEnergy): mean of the MuID hit energy
- 3008            distribution.
- 3009           – Standard deviation MuID hit energy (HitMuIDStdEnergy): standard deviation
- 3010            of the MuID hit energy distribution.
- 3011           – Maximum MuID hit energy (HitMuIDMaxEnergy): maximum of the MuID
- 3012            hit energy distribution.

- 3013           • **Geometry-related MuID**

## Chapter 8. Particle ID in GArSoft

- 3014        – Maximum distance MuID hit-to-hit (DistHitMuIDMax): maximum distance  
3015              between pairs of MuID hits (not sure this is a good variable, distribution  
3016              looks nuts).  
3017        – Maximum distance MuID hit-to-centre (DistHitCenterMuIDMax): maximum  
3018              of the distance distribution between the MuID hits and the centre of the  
3019              TPC.

3020        • **Statistical MuID**

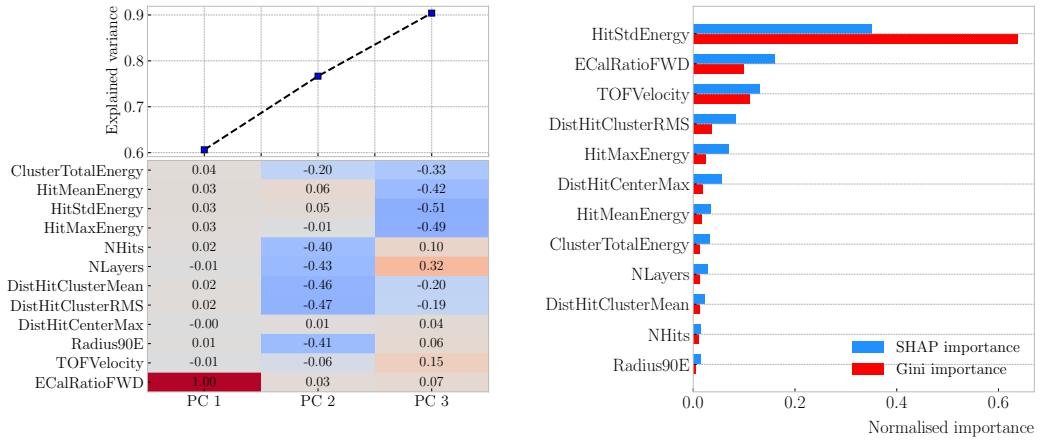
- 3021        – Number of hits (NHitsMuID): total number of MuID hits associated to the  
3022              track.  
3023        – Number of layers with hits (NLayersMuID): not really a count of all layers  
3024              with MuID hits but the difference between the last and the first layer with  
3025              MuIDhits.

3026        Figure 8.21 shows the distributions of three different MuID variables, separating true  
3027        muons (blue) and charged pions (red), for the three momentum ranges which use the  
3028        muon tagger information. In this case I decided to standard deviation of the MuID hit  
3029        energy distribution (left column), the maximum distance between the MuID hit pairs  
3030        (middle column), and the number of MuID hits (right column). These variables are used  
3031        together with the ECal features at high momenta, providing additional disambiguation  
3032        power.

3033        Once our features have been defined, one can do some exploratory analysis to  
3034        understand how well the variables describe the target class, and avoid the black-box  
3035        approach by what features are most relevant for the learning process. This way, I  
3036        performed a feature analysis for each of the momentum ranges I divided this classification  
3037        problem into. It follows three steps: first a principal component analysis (PCA), followed  
3038        by a feature importance study using Gini and Shapley values, and finally a feature  
3039        permutation importance analysis.

3040        The PCA is useful to understand the variance of the feature space. It is an  
3041        unsupervised machine learning technique that allows the user to perform a dimensionality

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.22:** Left panel: cumulative explained variance for the first three principal components (top panel) and contribution of the different features to the principal axes in feature space (bottom panel). Right panel: Shapley (blue) and Gini (red) feature importances for the different input features. Both figures correspond to the samples in the momentum range  $0.3 \leq p < 0.8$  GeV/c.

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:

$$\mathbf{C} = \mathbf{V}\mathbf{L}\mathbf{V}^\top, \quad (8.11)$$

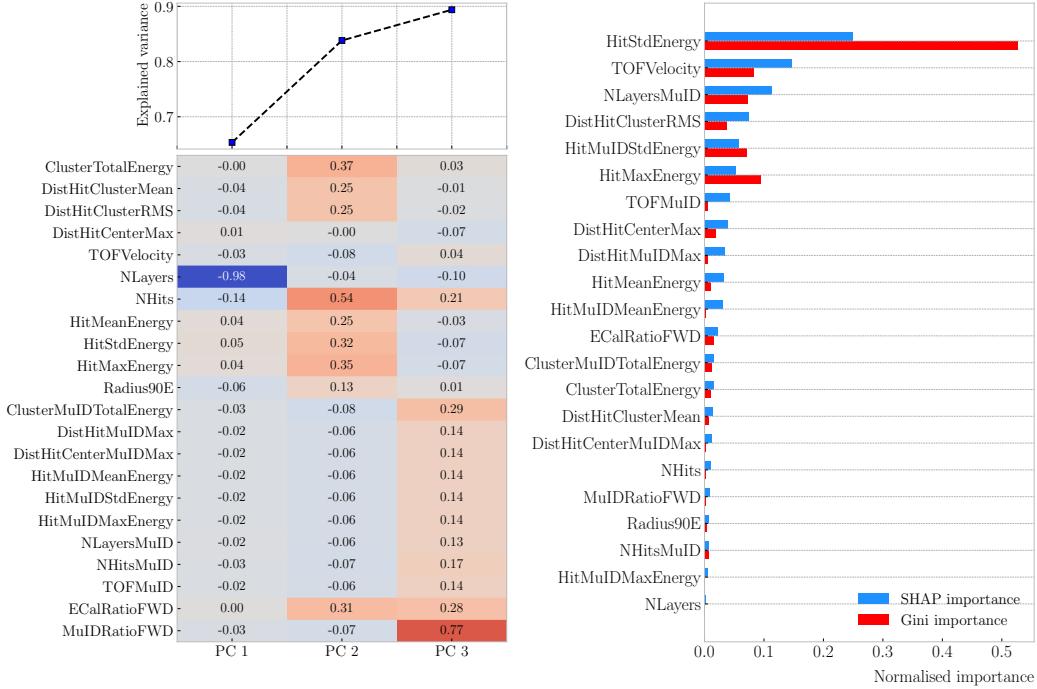
where  $\mathbf{V}$  is a matrix of eigenvectors and  $\mathbf{L}$  a diagonal matrix with eigenvalues  $\lambda_i$ . Then, performing SVD on  $\mathbf{X}$  gives us:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{W}^\top, \quad (8.12)$$

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 as right singular vectors. This way, we can write:

$$\mathbf{C} = \mathbf{W}\mathbf{S}\mathbf{U}^\top\mathbf{U}\mathbf{S}\mathbf{W}^\top/(n - 1) = \mathbf{W}\frac{\mathbf{S}^2}{n - 1}\mathbf{W}^\top. \quad (8.13)$$

## Chapter 8. Particle ID in GArSoft



**Figure 8.23:** 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  $0.8 \leq p < 1.5 \text{ GeV}/c$ .

meaning that the right singular vectors are also the eigenvectors of the covariance matrix.

The SVD can be computed numerically following an iterative approach.

This way, taking an input data vector  $X \in \mathbb{R}^n$ , the resulting feature vector  $Y \in \mathbb{R}^m$

is given by:

$$Y = \mathbf{C}_m^\top X. \quad (8.14)$$

The new features capture most of the variance of the original sample, while being lower dimensional, as  $m < n$ .

Before applying the PCA reduction one needs to centre and scale the input data. Centring is necessary when using SVD to obtain the eigenvectors of the covariance matrix, as only in that case we can do the identification with the right singular vectors from the input data. Scaling is needed when variables are on different scales, as some can then dominate the PCA procedure.

## 8.2. Muon and pion separation in the ECal and MuID

3064 I used the PCA module of `scikit-learn`, together with the `RobustScaler`, which  
 3065 centres the data and scales it based on the interquartile range. In Fig. 8.22 (left panel)  
 3066 and Fig. 8.23 (left panel) I show the results I obtained from the PCA for the momentum  
 3067 ranges  $0.3 \leq p < 0.8 \text{ GeV}/c$  and  $0.8 \leq p < 1.5 \text{ GeV}/c$ , respectively. Notice that in  
 3068 the second case the number of features increases considerably, as this is the first region  
 3069 which uses the MuID variables. I found that, in all the cases, adding a fourth PC does  
 3070 not add additional information. As it can be seen in the top panels of the figures, the  
 3071 cumulative explained variance is already over 80% with three PCs.

3072 The bottom panels show the contribution of the variables to the principal axes. For  
 3073 the two first momentum regions, I observe a tendency of the energy-related and the  
 3074 geometry-related ECal variables to be clustered together. For the other ranges, when  
 3075 I include the MuID variables, there seems to be a division between ECal and MuID  
 3076 variables. For these, it seems like the number of ECal layers with hits also plays an  
 3077 important role.

3078 The next step in the analysis is to quantify the importance of the features based on  
 3079 two additional metrics, namely the Gini and the Shapley values. The Gini importance,  
 3080 often called mean decrease impurity, is based on how much a feature contributes to the  
 3081 purity improvement at the splits in each decision tree. The purity is measured in terms  
 3082 of the Gini impurity index, defined as:

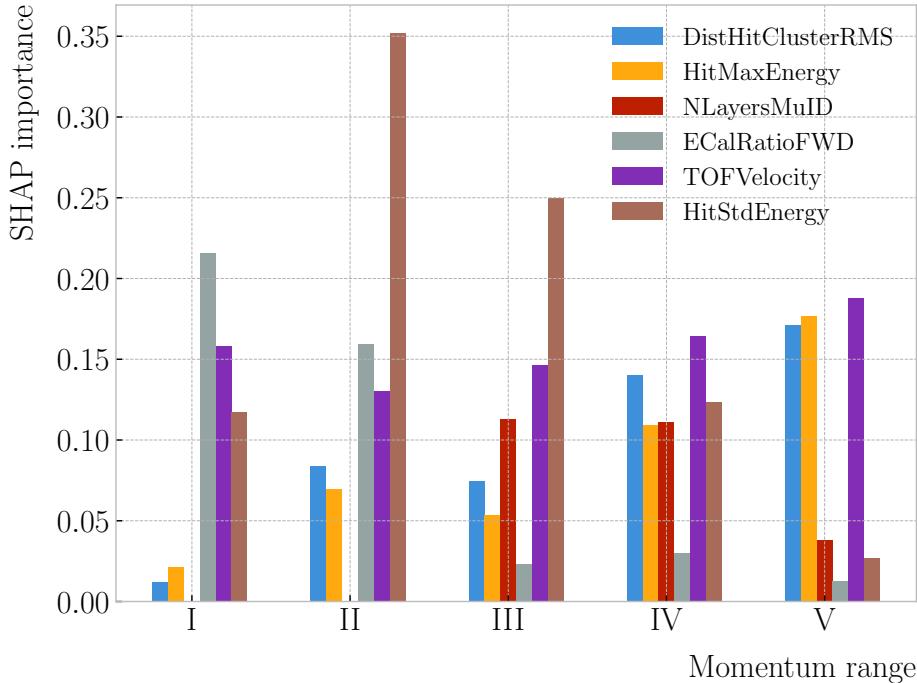
$$I_G = 1 - \sum_i f_i, \quad (8.15)$$

3083 where  $f_i$  is the fractional abundance of the  $i$ -th class. Then, for each split one can  
 3084 compute the weighted decrease in impurity as:

$$\Delta_G = \frac{N_t}{N} \left( I_G - \frac{N_t^R}{N_t} I_G^R - \frac{N_t^L}{N_t} I_G^L \right), \quad (8.16)$$

3085 where  $N$  represents the total number of samples,  $N_t$  the number of samples at the current  
 3086 node,  $N_t^R$  and  $N_t^L$  the number of samples in the right and left children respectively,

## Chapter 8. Particle ID in GArSoft



**Figure 8.24:** Evolution of the SHAP importance for the top six most important features across all five momentum ranges.

3087  $I_G$  is the Gini impurity at the current node, and  $I_G^R$  and  $I_G^L$  the Gini impurities of the  
 3088 resulting right and left children.

3089 For each decision tree, one will have a normalised vector with the accumulated  
 3090 decrease in Gini impurity for each feature. In the case of a BDT, the feature importances  
 3091 are simply the mean for all the estimators in the ensemble<sup>8</sup>.

3092 The concept of Shapley values originated in the context of game theory, and it  
 3093 measures the marginal contribution of a feature in enhancing the accuracy of a classifier.  
 3094 Take  $F$  to be the set of all features in a problem, and  $S \subseteq F$  the subset of features. To  
 3095 compute the Shapley value of the  $i$ -th feature, one has to train a model with that feature  
 3096 present,  $f_{S \cup \{i\}}$ , and another model trained without it,  $f_S$ . This has to be repeated for  
 3097 all possible combinations of subsets  $S \subset F \setminus \{i\}$ , and evaluating the models predictions

---

<sup>8</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.

## 8.2. Muon and pion separation in the ECal and MuID

3098 on the appropriate sets of data  $x_S$ . This way, the Shapley value results:

$$\varphi_i = \sum_{S \subset F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]. \quad (8.17)$$

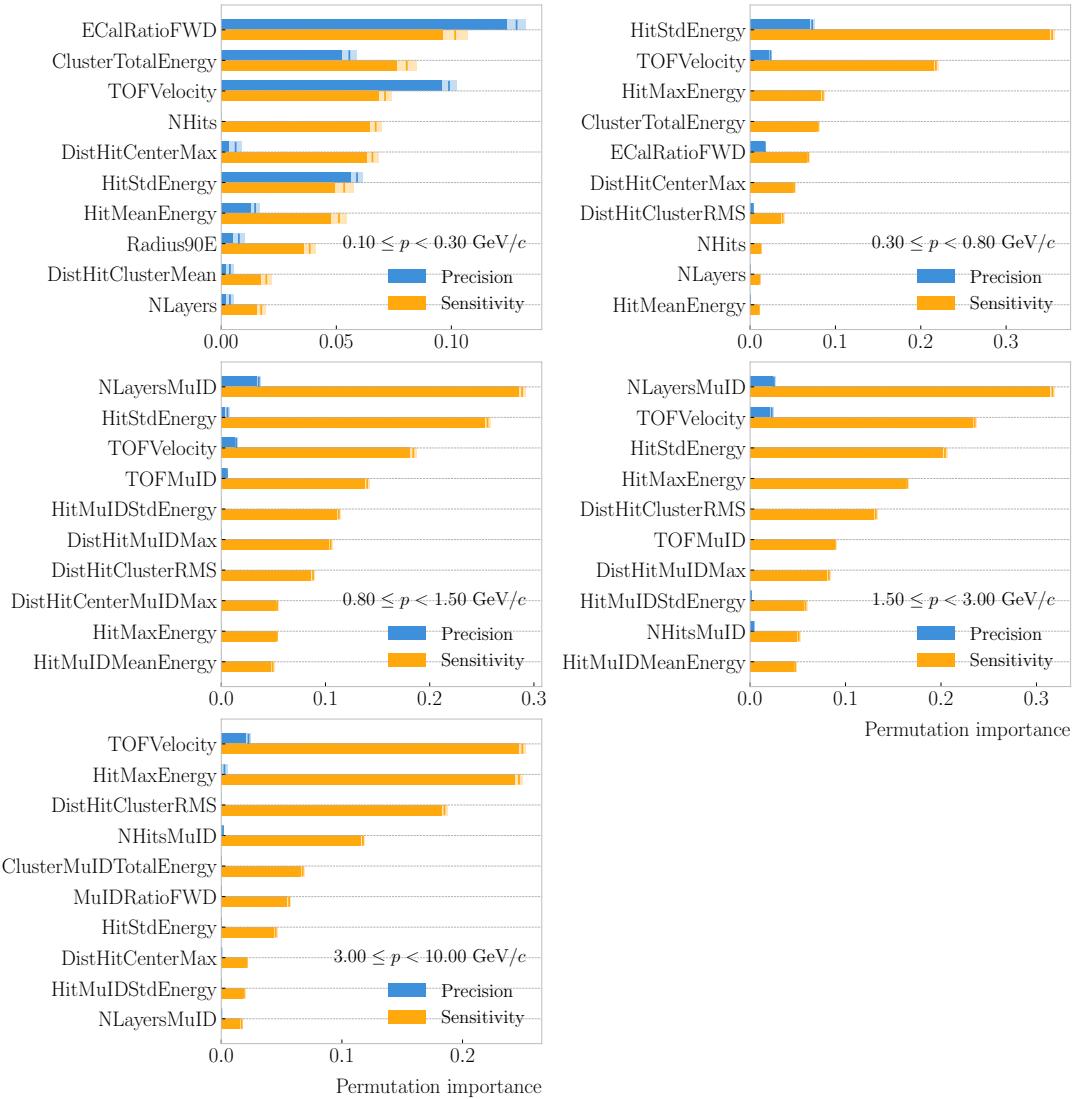
3099 I trained the `GradientBoostingClassifier` from `scikit-learn` with the default  
 3100 configuration in order to evaluate both the Gini and Shapley importances. The Gini  
 3101 scores are automatically computed by `scikit-learn`, using the training data. For the  
 3102 Shapley importance, I used the implementation from the `SHAP` package, computing  
 3103 it using the test sample. The results can be seen in Fig. 8.22 (right panel) and  
 3104 Fig. 8.23 (right panel), again for the momentum ranges  $0.3 \leq p < 0.8 \text{ GeV}/c$  and  
 3105  $0.8 \leq p < 1.5 \text{ GeV}/c$ . The length of the bars denote either the SHAP (blue) or the Gini  
 3106 (red) importance of the feature. One interesting thing to notice is that, when looking at  
 3107 the Gini importance, there is always one feature that dominates over the rest. This is  
 3108 not the case for the SHAP importance, where importances tend to be more balanced.

3109 Across all momentum ranges, I observe that the most important features are. For  
 3110 the five momentum ranges considered, only six variables sit in the top five at least once.  
 3111 Figure 8.24 shows the evolution of the SHAP importance of these six features. It is  
 3112 interesting to see that the time-of-flight variable keeps its importance almost unchanged  
 3113 for all momenta. Also, it looks like the ECal energy ratio gets less relevant the higher the  
 3114 momentum is, but the RMS of the hit-to-cluster distance distribution and the maximum  
 3115 ECal hit energy become more important in the last momentum ranges.

3116 The last step in the feature selection analysis is the feature permutation. This  
 3117 technique measures the contribution of each feature to the performance of a model by  
 3118 randomly shuffling its values and checking how some scores degrade. For the present  
 3119 case, I am interested in the precision or purity, and the sensitivity or efficiency, as these  
 3120 two are the most relevant metrics from a physics point of view. The `scikit-learn`  
 3121 module provides the user with a method to perform the permutation scans.

3122 The results of these are shown in Fig. 8.25. For the different momentum ranges  
 3123 I show the permutation importances for the ten most important features. For each

## Chapter 8. Particle ID in GArSoft



**Figure 8.25:** 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.

## 8.2. Muon and pion separation in the ECal and MuID

3124 of the variables I report the effect the permutations have on the precision (blue) and  
3125 sensitivity (yellow) of the models. The bars indicate the importance value, with the  
3126 lighter part representing one standard deviation around the mean (hinted as an additional  
3127 vertical line). Something to notice is that, in the first momentum region, the feature  
3128 permutations have an effect on both the precision and the sensitivity. However, for the  
3129 rest the precision is almost unaffected, while the sensitivity changes are considerably  
3130 larger.

3131 It is also interesting to see that most of the variables identified as important here  
3132 are the same I found when looking at the Shapley values. The behaviour of these across  
3133 the momentum ranges is also similar, with the same patterns of some features being  
3134 important at low momenta and then dropping in importance for the high momentum  
3135 ranges.

3136 Wit this, I conclude the study of the features. I have prepared the training and  
3137 testing datasets and understood what features are likely to have the largest impact on  
3138 the performance of the classifiers.

### 3139 8.2.4 Hyperparameter optimisation

3140 Any BDT requires the user to specify a number of parameters that will dictate its  
3141 behaviour. They can be divided into two categories: (i) tree-specific parameters, which  
3142 affect each individual tree in the model, and (ii) boosting parameters, which control the  
3143 boosting operation in the model. The value of these so-called hyperparameters affect the  
3144 performance and predictive power of the models. Therefore, one needs to carefully select  
3145 their optimal values in order to extract as much information as possible from the data.

3146 From all the parameters used to define a tree in the `scikit-learn` implementation  
3147 of the BDT classifier, I only consider a subset of them. This is due to the fact that some  
3148 are mutually exclusive, but also because I noticed that others have little effect on the  
3149 problem at hand. Therefore, the parameters I investigate are the following:

- 3150 • `min_samples_split`: defines the minimum number of samples required in a node  
3151 to be considered for splitting. Higher values prevent a model from learning relations

## Chapter 8. Particle ID in GArSoft

3152 which might be highly specific to the particular sample, but may lead to under-fitting  
3153 if the value is too low.

- 3154 • `min_samples_leaf`: defines the minimum samples required in a leaf node. For  
3155 imbalanced problems it should take a low value, as there will not be many cases  
3156 where the minority class dominates.
- 3157 • `max_depth`: maximum depth of a tree. Useful to prevent over-fitting, as higher  
3158 depth will allow a model to learn relations specific to the training sample.

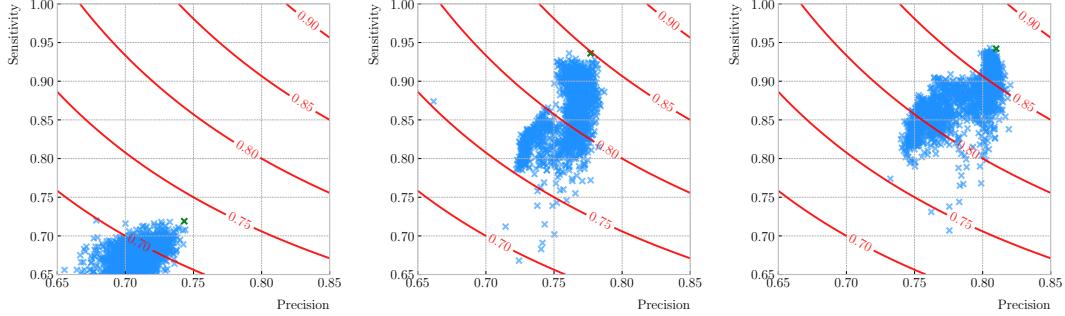
3159 In the case of the boosting parameters, the ones I look at are:

- 3160 • `learning_rate`: determines the impact of each tree on the final outcome. Low  
3161 values make the model robust to the specific characteristics of a tree, and thus  
3162 allow it to generalise well. However, that usually requires a large number of trees  
3163 to model the data properly.
- 3164 • `n_estimators`: number of sequential trees to be trained. In general, BDTs are  
3165 fairly robust at higher number of trees but it can still overfit at a point.
- 3166 • `subsample`: fraction of observations to be selected for each tree. Values slightly  
3167 less than 1 make the model robust by reducing the variance.

3168 In general, hyperparameters depend on each other. Thus, it is not possible to  
3169 optimise them independently. In the literature, we find two main strategies to explore  
3170 the hyperparameter space. We could use a grid search, in which one discretises a  
3171 portion of the space of hyperparameters and evaluates the model at each point. Another  
3172 approach is the randomised search, where a certain number of random configurations of  
3173 hyperparameters are explored.

3174 In this case, I used the random search to scan the hyperparameter space. Also,  
3175 because it is not guaranteed that a set of hyperparameters can be efficiently applied  
3176 across different datasets, I perform the optimisation for each of the momentum ranges  
3177 considered. Table 8.3 shows the list of hyperparameters considered, and the range within

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.26:** 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.

3178 which I let them vary. I decided to fix the number of estimators to 400 in all cases, as  
 3179 its value is correlated with that of the learning rate.

3180 I evaluate 10000 different hyperparameter configurations for each momentum range.  
 3181 For the hyperparameter tuning, I used subsamples containing 10% of the full datasets,  
 3182 keeping the original proportions between classes, in order to reduce the computational  
 3183 load. The performance of the models was assessed using a stratified 3-fold cross-validation  
 3184 with replacement. Cross-validation involves dividing the data in a number of subsets,  
 3185 training the model using some of them, and testing it with the rest. In our case, I  
 3186 divide the data in 3 equal-sized subsets, maintaining the class proportions of the original  
 3187 dataset. Then, for 3 consecutive iterations, I train the models using 2 of the subsets  
 3188 while I compute the precision and sensitivity scores with the other. This approach  
 3189 provides a more robust estimate of the performance on unseen data.

3190 Figure 8.26 shows the results in the precision versus sensitivity plane, for the  
 3191 momentum regions I, III and V (from left to right). The contours represent the curves  
 3192 of equal  $F_1$ -score, i.e. the harmonic mean of the precision and the sensitivity. In order  
 3193 to select the optimal configurations (indicated in the plots with a green cross), I chose  
 3194 the point with the highest  $F_1$ -score.

3195 The results for the different momentum ranges are summarised in Tab. 8.3. One  
 3196 can see some consistency in hyperparameter choices, with models generally preferring

## Chapter 8. Particle ID in GArSoft

**Table 8.3:** 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 8.4:** Performance metrics of the BDTs with optimal hyperparameters, for the different momentum ranges.

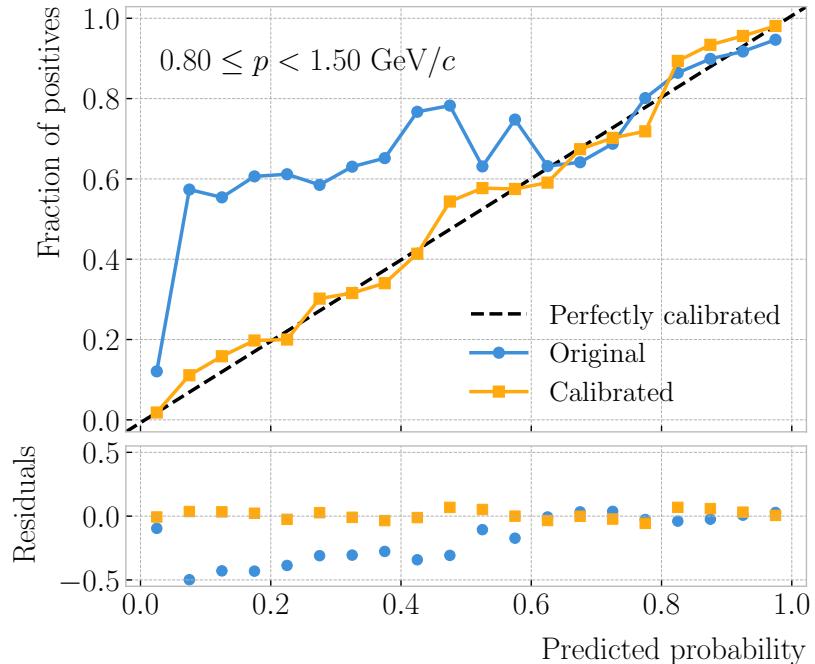
Metric	Value $\pm 1\sigma$				
	I	II	III	IV	V
Accuracy	$0.779 \pm 0.003$	$0.812 \pm 0.003$	$0.846 \pm 0.002$	$0.861 \pm 0.003$	$0.874 \pm 0.002$
Precision	$0.769 \pm 0.003$	$0.752 \pm 0.005$	$0.788 \pm 0.002$	$0.805 \pm 0.003$	$0.815 \pm 0.003$
Sensitivity	$0.745 \pm 0.009$	$0.921 \pm 0.006$	$0.965 \pm 0.002$	$0.967 \pm 0.002$	$0.976 \pm 0.001$
$F_1$ -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	$0.868 \pm 0.003$	$0.865 \pm 0.003$	$0.899 \pm 0.002$	$0.902 \pm 0.002$	$0.911 \pm 0.001$

3197 small values for the tree-specific parameters, small learning rate, and relatively large

3198 subsample sizes.

3199 Now that I have obtained the optimal values of the hyperparameters, I can train  
 3200 the different BDTs. In this case I use the complete datasets, keeping 20% of the data  
 3201 for testing. Table 8.4 shows the values of the different performance metrics obtained  
 3202 using the selected hyperparameters and 5-fold cross-validation. The last row indicates  
 3203 the value of the area under the receiver operating characteristic (ROC) curve. This  
 3204 represents the sensitivity of a model as a function of the false positive rate. I have  
 3205 included it here as it is a classic model metric used in the machine learning community.  
 3206 Overall, there is a clear trend of models performing better at higher momentum.

## 8.2. Muon and pion separation in the ECal and MuID



**Figure 8.27:** Reliability diagrams for the BDT classifier used in the momentum range  $0.3 \leq p < 0.8$  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).

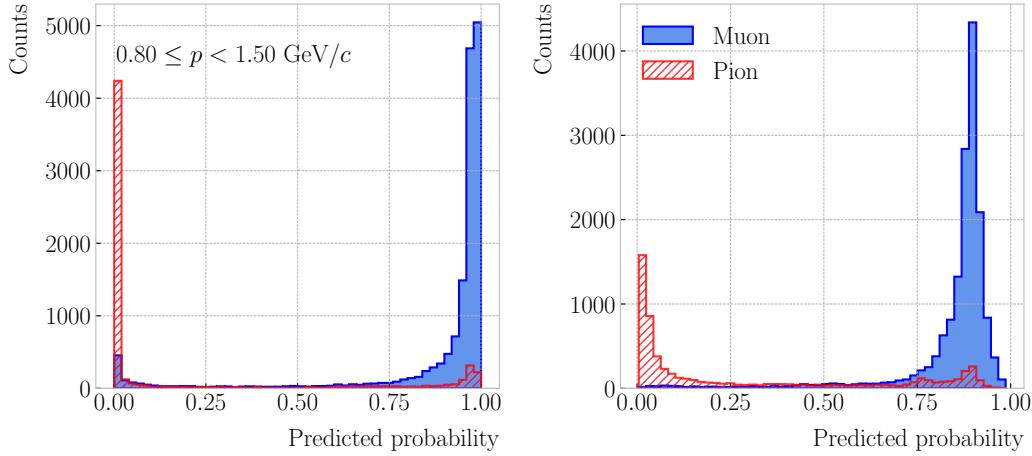
### 8.2.5 Probability calibration

So far, the trained BDTs are able to provide predictions of the class labels. Ideally, one would like the output of a classifier to give a confidence level about the prediction. However, it is not straightforward to interpret the outputs of our BDTs in terms of probabilities.

A way to visualise how well the predictions of a classifier are calibrated is using reliability diagrams [110]. They represent the probability of the positive label versus the probability predicted by the classifier. These can be obtained by binning the predicted probabilities, and then compute the conditional probability  $P(y_{true} = 1 | y_i \leq y_{pred} < y_{i+1})$  by checking the fraction of true positive instances in each bin. The reliability diagram of a perfectly calibrated classifier would be a diagonal line.

In this case, I try to correct the raw response of the classifiers by applying a sigmoid

## Chapter 8. Particle ID in GArSoft



**Figure 8.28:** 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  $0.3 \leq p < 0.8 \text{ GeV}/c$ .

3219 function:

$$\sigma(x; A, B) = \frac{1}{1 + e^{Ax+B}}, \quad (8.18)$$

3220 where the parameters  $A$  and  $B$  are real numbers determined using the method of least  
3221 squares.

3222 For each classifier, I perform a grid search to obtain the optimal values of  $A$  and  $B$ .  
3223 For any pair, I compute the predicted probabilities as  $y_{pred} = \sigma(y_{raw}; A, B)$ , where  $y_{raw}$   
3224 are the raw predictions of the classifier<sup>9</sup>. Then, I calculate the corresponding reliability  
3225 curve, and take the sum of the squared residuals between it and the response of the  
3226 perfectly calibrated classifier.

3227 Figure 8.27 shows the reliability diagrams for the original (blue) and calibrated  
3228 (yellow) probability predictions of the classifier for the III momentum range,  $0.3 \leq p <$   
3229  $0.8 \text{ GeV}/c$ . The original response of the classifier is given by  $y_{pred} = \sigma(y_{raw}; -2, 0)$ ,  
3230 which is the transformation applied by `scikit-learn` to produce the probability estimate.  
3231 Notice how the calibrated prediction matches the ideal response much better than the  
3232 original, across all the probability range.

3233 One can also compare the responses of the uncalibrated and calibrated classifiers

---

<sup>9</sup>In `scikit-learn` these correspond to the outputs of the `decision_function` method.

### 8.3. ECal time-of-flight

3234 broken down by true particle type, as shown in Fig. 8.28. It can be seen that the  
3235 distributions for both muons (blue) and charged pions (red) smoothen after calibration,  
3236 but still the separating power of the classifier remains unchanged.

3237 At this point, having the trained classifiers and the probability calibration parameters,  
3238 I am able to assess the performance of the classification strategy in a physics-relevant  
3239 case.

#### 3240 8.2.6 Performance

### 3241 8.3 ECal time-of-flight

3242 Looking at Fig. 8.19, it is clear that for momentum values in the range  $1.0 - 3.0 \text{ GeV}/c$   
3243 it is not possible to separate pions and protons using a  $\langle dE/dx \rangle$  measurement in the  
3244 HPgTPC. However, in the previous section I assumed that protons at those energies  
3245 could be identified by other means, and therefore were not an issue for the muon and  
3246 pion discrimination.

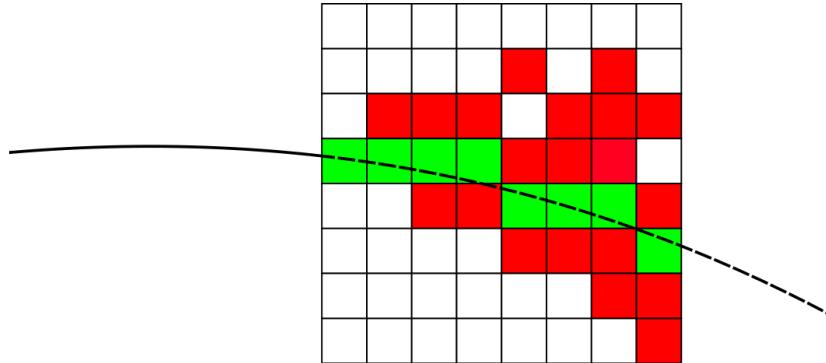
3247 Some detectors, like ALICE [111] or the ILD concept [112], complement the PID  
3248 capabilities of their gaseous trackers with time-of-flight measurements. The use of  
3249 fast timing silicon sensors, with hit time resolutions under 100 ps, would allow for the  
3250 identification of charged hadrons via a ToF measurement up to  $5.0 \text{ GeV}/c$ . In the case  
3251 of ND-GAr, one could think of using the inner layers of the ECal, the ones consisting of  
3252 high-granularity tiles, to obtain a ToF-based PID, with some inputs from the TPC.

3253 Measuring the momentum and the velocity of a charged particle allows for a  
3254 determination of the mass through the relativistic momentum formula:

$$m = \frac{p}{\beta} \sqrt{1 - \beta^2}. \quad (8.19)$$

3255 In our case, the momentum is measured in the TPC, using the curvature and the dip  
3256 angle of the helix inside the magnetic field. The velocity of the particle can be written

## Chapter 8. Particle ID in GArSoft



**Figure 8.29:** 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.

3257 as:

$$\beta = \frac{\ell_{track}}{c\tau}, \quad (8.20)$$

3258 where  $\ell_{track}$  is the length of the track, and  $\tau$  the arrival time to the ECal.

3259 In GArSoft, the track length is computed at the Kalman filter stage. It is simply the  
 3260 sum of the line segments along the track, either in the forward or backward fit. In this  
 3261 case, because we are only interested in the particles that make it to the ECal, I choose  
 3262 the fit direction based on the results of the track-cluster associations.

3263 Additionally, because the last 30 cm of the TPC radius are uninstrumented<sup>10</sup>, I need  
 3264 to correct for the length of the tracks. Using the track fit parameters to propagate the  
 3265 helix to its entry point in the ECal, one can write the total track length as:

$$\ell_{track} = \ell + \left| \frac{\phi_{EP} - \phi}{R^{-1}} \right| \sqrt{1 + \tan^2 \lambda}, \quad (8.21)$$

3266 where  $\phi_{EP}$  is the angle of rotation at the entry point to the calorimeter, and  $\ell$ ,  $\phi$ ,  $R$  and  
 3267  $\lambda$  are the track length, angle of rotation, radius of curvature and dip angle at the last  
 3268 point in the fit, respectively.

3269 To test the idea of performing a ToF measurement with the inner ECal, I generated  
 3270 two data samples. Each consists of 10000 single particle events, either charged pions or

---

<sup>10</sup>Note to self: check this number.

### 8.3. ECal time-of-flight

3271 protons. Their momenta are uniformly distributed in the range  $0.5 - 5.0 \text{ GeV}/c$ , and  
3272 their directions are isotropic. I process each sample using different values of the time  
3273 resolution, from  $\Delta\tau = 0$ , the perfect time resolution case for comparison, to the current  
3274 nominal value of  $\Delta\tau = 0.7 \text{ ns}$ , and the worse scenario of  $\Delta\tau = 1.0 \text{ ns}$ .

#### 3275 8.3.1 Arrival time estimations

3276 In the simulation, the limited time resolution of the ECal is taken into account by  
3277 applying a Gaussian smearing to the true hit times. Other effects, like the digitisation  
3278 of the signals, are not taken into account and fall beyond the scope of this study. After  
3279 the track-cluster, one ends up with a collection of ECal hits associated to each particle.  
3280 From these, the arrival time of the particle to the ECal can be extracted.

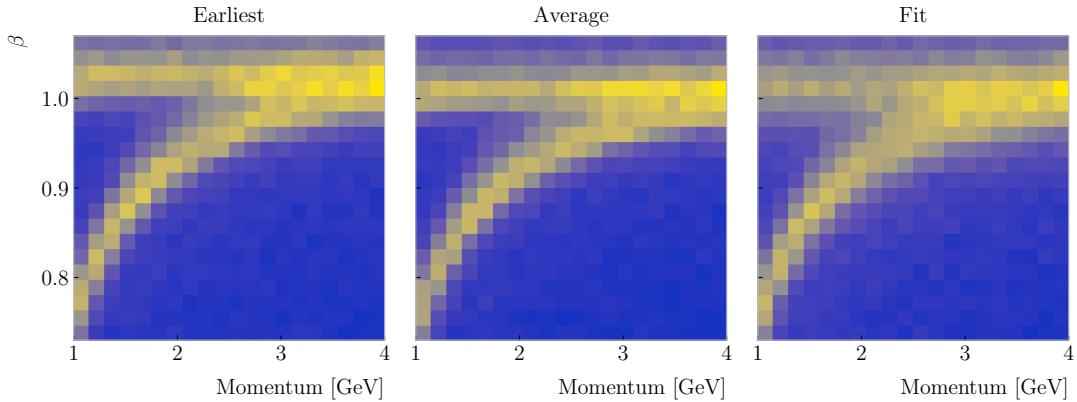
3281 The simplest possibilities are to either take the time of the earliest hit or the hit  
3282 closest to the entry point. Because these two coincide, in general, I focused only in  
3283 the earliest hit time. However, this needs to be corrected, to account for the distance  
3284 travelled from the entry point to the position of the hit:

$$\tau_{earliest} = \tau_{hit} - \frac{d_{EP-hit}}{c}, \quad (8.22)$$

3285 where  $\tau_{hit}$  is the time of the earliest hit, and  $d_{EP-hit}$  is the distance between that hit  
3286 and the entry point of the particle to the ECal. This is computed as the arc length  
3287 between the entry point and the point of the extrapolated helix up to the layer of the  
3288 hit. This way of correcting the time assumes  $c$  for the propagation of the particle, which  
3289 may lead to biased estimates.

3290 I also tried to estimate the arrival times using information from the rest of the hits.  
3291 In order to do this, as a simplifying assumption, I approximate the hadronic shower  
3292 considering only its MIP component. For each layer, I keep only the hit in the tile closest  
3293 to the point of the extrapolated track up to that layer. Figure 8.29 shows an example of  
3294 how this hit selection works. The dashed line represents the extrapolated track, while  
3295 the coloured squares are the tiles containing hits. Green indicates the tiles closer to the

## Chapter 8. Particle ID in GArSoft



**Figure 8.30:** 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.

3296 track in each layer (in the sketch they correspond to the grid columns).

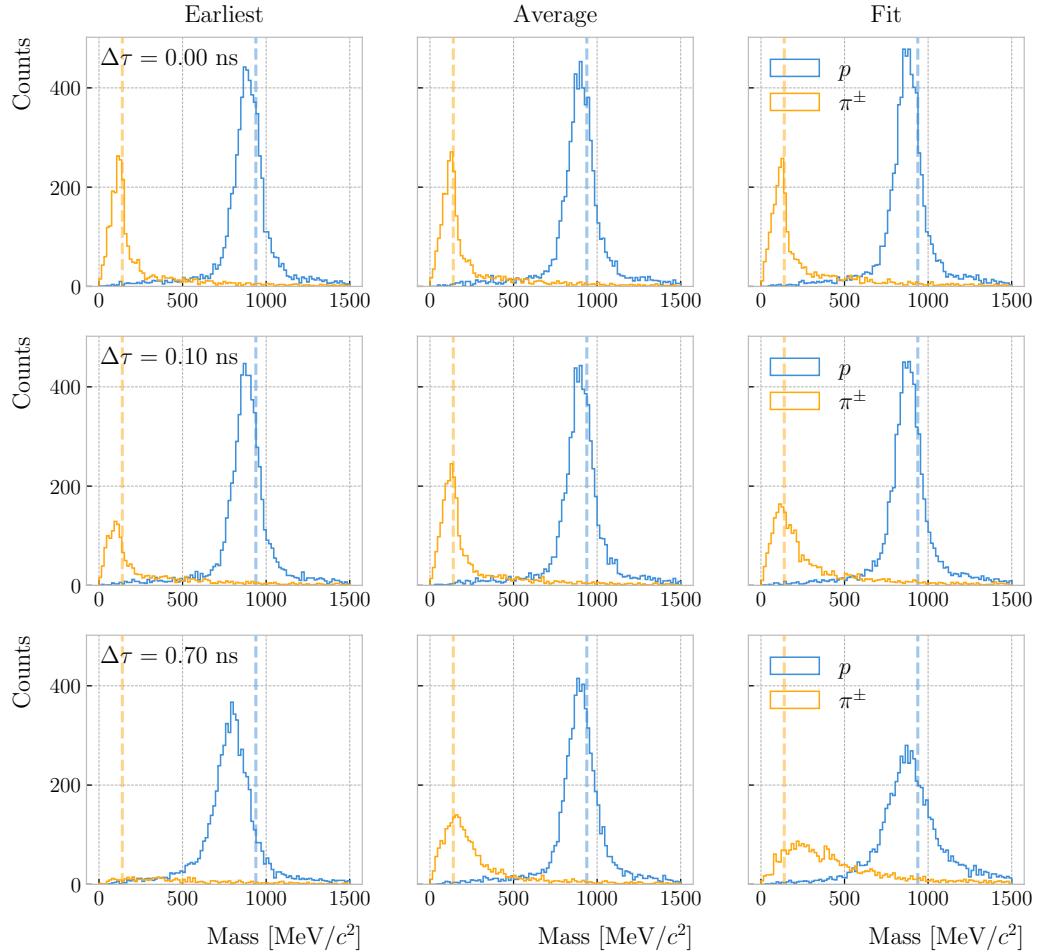
3297 Now, I can use these collections of hits to estimate the arrival times. A possibility  
 3298 is to take the average of the times of the selected hits, denoted  $\tau_{average}$ . For that to  
 3299 work, one needs to correct these times, in a similar way as in Eq. (8.22), before taking  
 3300 the average. However, as before, this correction assumes that the particle travels at the  
 3301 speed of light inside the ECal. Another option is to perform a linear fit to the hit times  
 3302 and the distances to the entry point. In that case, the arrival time would be the fitted  
 3303 value of the intercept,  $\tau_{fit}$ . This method would not assume a speed of light propagation.

3304 Figure 8.30 shows the velocity estimations as a function of the particle momentum,  
 3305 for the earliest hit time (left panel), average hit time (middle panel), and fitted hit time  
 3306 (right panel). The two bands correspond to the  $\pi^\pm$  and the  $p$  particles.  $\Delta\tau = 0.1$  ns.  
 3307 Notice how, for the earliest hit time method, the velocities are significantly biased  
 3308 towards larger values. For the multi-hit methods, the  $\tau_{fit}$  estimate appears to produce a  
 3309 larger variance than when using the  $\tau_{average}$  method.

### 3310 8.3.2 Proton and pion separation

3311 Once we have the velocities of the particles, one can estimate their masses through  
 3312 Eq. (8.19). The resulting mass spectra are shown in Fig. 8.31. I computed the masses

### 8.3. ECal time-of-flight

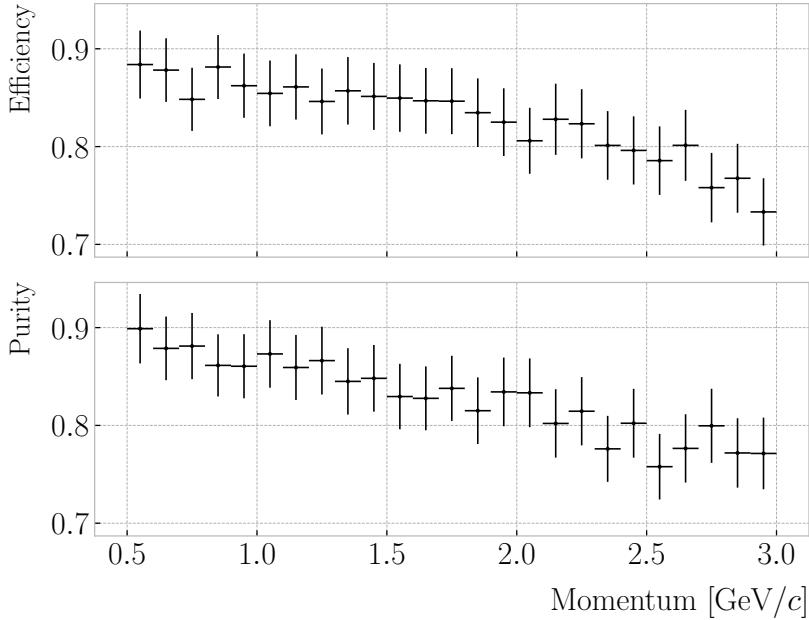


**Figure 8.31:** Mass spectra for  $p$  (blue) and  $\pi^\pm$  (yellow) particles, using different ECal time resolution values (from top to bottom, in ascending order), and arrival time estimates. From left to right: earliest hit time, average hit time, and fitted hit time. The dashed lines indicate the true masses of the particles.

for the three arrival time estimates discussed above, and three different values of the time resolution:  $\Delta\tau = 0.00$  (perfect time resolution),  $\Delta\tau = 0.10$  ns, and  $\Delta\tau = 0.70$  ns. Although in all cases we have the same number of events, it appears as if the entries in the histograms decrease as the time resolution increases. Sometimes, the particles get unphysical values of  $\beta > 1$ , and in turn they do not contribute to the mass spectra. This is more likely to happen for higher values of  $\Delta\tau$ .

As noted before, the average hit time method produces the most robust estimates when increasing  $\Delta\tau$ . Intuitively this makes sense, as by taking the mean one averages

## Chapter 8. Particle ID in GArSoft



**Figure 8.32:** Efficiency (top panel) and purity (bottom panel) for the proton selection as a function of the momentum, for  $\Delta\tau = 0.10$  ns.

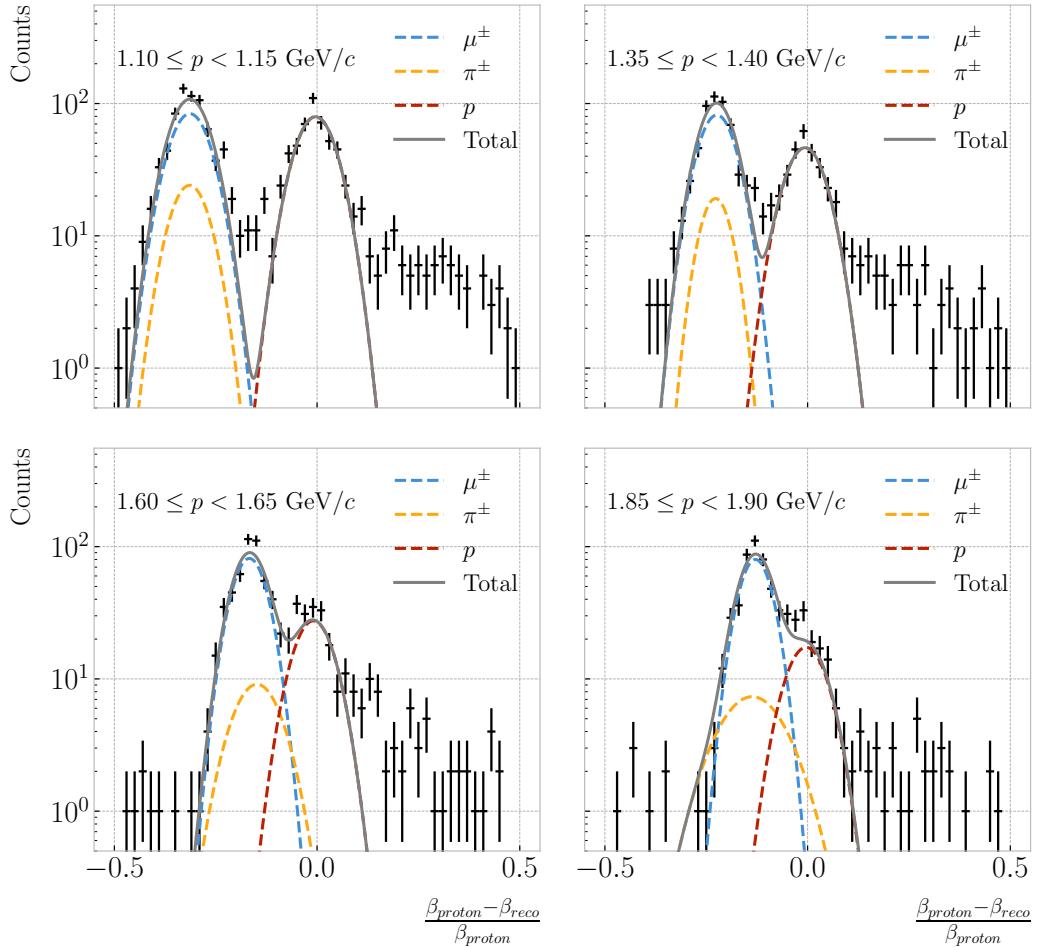
3321 out the effect of the Gaussian smearing. Going forward, I will use this arrival time  
 3322 estimator, as it appears to be the best performing one.

3323 It is possible to use the velocity estimations to select a sample of protons. In this  
 3324 case, I do so by dividing the relevant momentum range in bins of  $0.1$   $\text{GeV}/c$ . For each  
 3325 momentum bin, I compute the expected velocity for the protons via the inverse of Eq.  
 3326 (8.19), and then take the fractional residuals of the measured velocities. Using that  
 3327 distribution, I choose the cut that maximises the  $F_1$ -score of the proton selection.

3328 The results can be seen in Fig. 8.32, for the case  $\Delta\tau = 0.10$  ns. As expected from  
 3329 Fig. 8.30, the performance of the selection degrades rapidly with increasing momentum.  
 3330 However, the purity is still around 75% at  $3.0$   $\text{GeV}/c$ . This is likely to be sufficient, as  
 3331 we do not expect protons or charged pions with higher energies from the beam neutrino  
 3332 interactions.

3333 Figure 8.33

## 8.4. Charged pion decay in flight



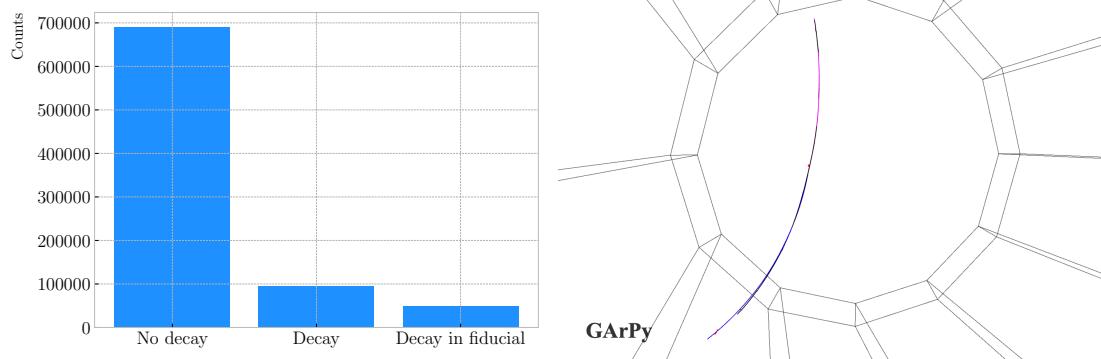
**Figure 8.33:** Distributions of the velocities measured by ToF with the inner ECal, for different momentum bins, in a FHC neutrino interaction sample. The Gaussian fits are performed around the maxima for each particle species.

## 3334 8.4 Charged pion decay in flight

3335 As discussed previously, in GArSoft the TPC tracks are formed after a pattern recognition  
 3336 algorithm and a Kalman filter are applied to the TPC clusters. These two steps can  
 3337 find discontinuities in the track candidates (e.g. due to a particle decay) when these  
 3338 so-called breakpoints are large enough. However, for some, more subtle, cases they may  
 3339 miss them and form a single reconstructed track. It has been noted in the literature  
 3340 that Kalman filters offer, as a by-product, additional information to form test statistics  
 3341 to identify these breakpoints [113, 114].

3342 Considering the mean life of the charged pion,  $\tau = (2.6033 \pm 0.0005) \times 10^{-8} \text{ s}$ , one

## Chapter 8. Particle ID in GArSoft



**Figure 8.34:** Left panel: number of non-decaying, decaying and decaying in the fiducial volume pions for a MC sample of 100000,  $p = 500 \text{ MeV}/c$  isotropic positively charged pions inside the TPC. Right panel: event display for a positive pion decaying inside the fiducial volume, with a single reconstructed track for the pion and muon system.

3343 can estimate that about 12% of the pions with momentum  $p \sim \mathcal{O}(500 \text{ MeV}/c)$  (roughly  
 3344 the peak of the pion momentum distribution in  $\nu_\mu$  CC interactions off argon) decay  
 3345 inside the TPC. Figure 8.34 (left panel) shows the amount of charged pions decaying in  
 3346 the full TPC and fiducial volumes from an isotropic, monoenergetic sample of 100000  
 3347 negatively charged pions with  $p = 500 \text{ MeV}/c$ . We see that about 10% of those decayed,  
 3348 with more than half of them decaying inside the TPC fiducial volume.

3349 Figure 8.34 (right panel) shows an example event display of a charged pion (magenta  
 3350 line) decays in flight inside the TPC, but because the angle of the muon (blue line) is  
 3351 small both were reconstructed as one single track (black line). In this case, the composite  
 3352 track reaches the ECal, where it undergoes a muon-like interaction, thus being classified  
 3353 as a muon.

3354 A way to understand what decaying pion tracks were totally or partially reconstructed  
 3355 together with the daughter muon is looking at the relative energy contributions to the  
 3356 reconstructed track. In order to select a sample of such events, I require that a minimum  
 3357 50% of the total energy comes from the pion and at least 20% from the muon.

### 3358 8.4.1 Track breakpoints

3359 To identify potential decays we can use the information we obtain from the Kalman  
 3360 filter at each step of the fitted track. The simplest test we can think about is computing

## 8.4. Charged pion decay in flight

3361 the  $\chi^2$  of the mismatch between all the parameters in the forward and the backward fits:

$$\chi_k^{2 (FB)} = (\hat{x}_k^B - \hat{x}_k^F)^T [V^{(\hat{x}_k, B)} + V^{(\hat{x}_k, F)}]^{-1} (\hat{x}_k^B - \hat{x}_k^F), \quad (8.23)$$

3362 where  $\hat{x}_k^F$ ,  $\hat{x}_k^B$  are the Kalman filter state vector estimates at step  $k$  in the forward and  
3363 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.  
3364 Using the values of the  $\chi^2$  at measurement  $k$  for the forward and backward fits we can  
3365 compute another  $\chi^2$  value that characterises the overall track fit:

$$\chi_{track}^2 = \chi_k^{2 (F)} + \chi_k^{2 (B)} + \chi_k^{2 (FB)}, \quad (8.24)$$

3366 which remains approximately constant for all  $k$ .

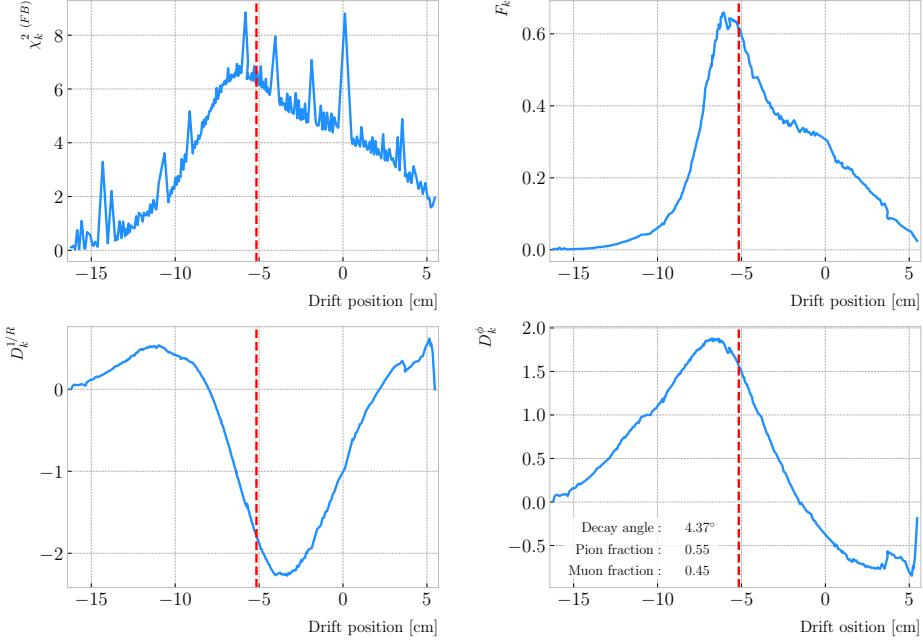
3367 An alternative approach proposed in the context of the NOMAD experiment was  
3368 using a fit with a more elaborate breakpoint hypothesis, so we can perform a comparison  
3369 of the  $\chi^2$  with and without breakpoints. This can be achieved by using some alternative  
3370 parametrisation with extra parameters, which allows some of the track parameters to  
3371 be discontinuous at certain points. A decay changes the momentum magnitude and  
3372 direction, so we can use the new state vector:

$$\alpha = (y, z, 1/R_F, 1/R_B, \phi_F, \phi_B, \tan\lambda_F, \tan\lambda_B)^T. \quad (8.25)$$

3373 As we already have the estimates from the standard Kalman filter and their  
3374 covariance matrices at each point, we do not need to repeat the Kalman fit for the new  
3375 parametrisation. Instead, I can compute the values of  $\alpha$  at each point  $k$  that minimise  
3376 the  $\chi^2$  resulting from comparing them to  $\{\hat{x}_k^B, \hat{x}_k^F\}$ . Introducing the two  $5 \times 8$  matrices:

$$H^F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad H^B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (8.26)$$

## Chapter 8. Particle ID in GArSoft



**Figure 8.35:** 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.

3377 we can write this as:

$$\begin{aligned}\chi_k^{2(FB)}(\alpha) &= (\hat{x}_k^F - H^F \alpha)^T \left[ V^{(\hat{x}_k, F)} \right]^{-1} (\hat{x}_k^F - H^F \alpha) \\ &\quad + (\hat{x}_k^B - H^B \alpha)^T \left[ V^{(\hat{x}_k, B)} \right]^{-1} (\hat{x}_k^B - H^B \alpha).\end{aligned}\tag{8.27}$$

3378 The minimum of  $\chi_k^{2(FB)}(\alpha)$  is found when the measured new state vector takes the

3379 value:

$$\hat{\alpha}_k = V^{(\hat{\alpha}_k)} H^T (V^{(\hat{x}_k)})^{-1} \hat{X},\tag{8.28}$$

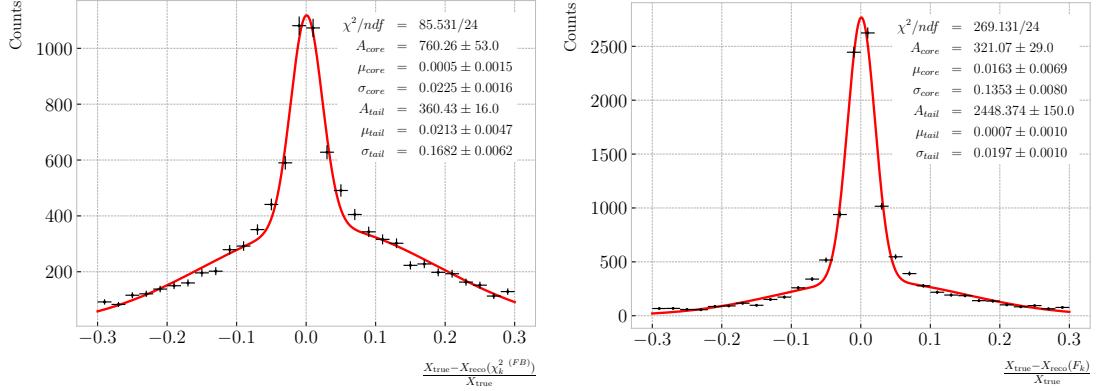
3380 where  $\hat{X} = \{\hat{x}_k^B, \hat{x}_k^F\}$ ,  $V^{(\hat{x}_k)}$  is the block diagonal matrix formed by  $V^{(\hat{x}_k, F)}$  and  $V^{(\hat{x}_k, B)}$

3381 and  $V^{(\hat{\alpha}_k)}$  is the covariance matrix of  $\hat{\alpha}_k$ , given by:

$$V^{(\hat{\alpha}_k)} = \left( H^T (V^{(\hat{x}_k)})^{-1} H \right)^{-1}.\tag{8.29}$$

3382 From these new fit estimates we can compute the  $F$  statistic, which tells us whether

## 8.4. Charged pion decay in flight



**Figure 8.36:** 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).

3383 the model with breakpoint provides a statistically significant better fit:

$$F_k = \left( \frac{\chi_{\text{track},k}^2 - \chi_{\text{full},k}^2}{8 - 5} \right) / \left( \frac{\chi_{\text{full},k}^2}{N - 8} \right). \quad (8.30)$$

3384 One can also compute the signed difference of the duplicated variables divided by  
 3385 their standard deviation at each point. These represent how significant the discontinuity  
 3386 in each variable is. For any variable  $\eta$  we can write it as:

$$D_k^\eta = \frac{\hat{\eta}_k^B - \hat{\eta}_k^F}{\sqrt{\text{Var}[\hat{\eta}_k^F] + \text{Var}[\hat{\eta}_k^B] - 2\text{Cov}[\hat{\eta}_k^F, \hat{\eta}_k^B]}}. \quad (8.31)$$

3387 In our case, the relevant ones to look at are  $D_k^{1/R}$  and  $D_k^\phi$ .

3388 Figure 8.35 shows the values of  $\chi_k^2(FB)$ ,  $F_k$ ,  $D_k^{1/R}$  and  $D_k^\phi$  as functions of the position  
 3389 along the drift direction, for an example reconstructed track with 55.5% of the energy  
 3390 coming from the charged pion and 45.5% from the daughter muon. The true position of  
 3391 the decay is indicated (dashed red lines). Notice how  $\chi_k^2(FB)$  and  $F_k$ ,  $D_k^{1/R}$  reach their  
 3392 maxima near the decay point. In the former case this indicates a large forward-backward  
 3393 difference in the track fit. In the later it represents that the extended state vector  
 3394 improves the fit particularly around that point.

## Chapter 8. Particle ID in GArSoft

3395 I can estimate the decay position finding resolution by computing the difference  
 3396 between the  $X$  position of the maxima of  $\chi_k^2(FB)$  and  $F_k$  and the  $X$  position of the  
 3397 true decay. Figure 8.36 represent the the fractional residual distributions for both  
 3398 cases, from the sample of tracks containing pion decays. Fitting a double Gaussian to  
 3399 the distributions (red lines) I find a resolution of  $(3.31 \pm 0.15)\%$  and  $(6.94 \pm 0.31)\%$   
 3400 respectively.

3401 In principle, the  $F$ -statistic should follow a Fisher distribution with  $(8 - 5)$  and  
 3402  $(N - 8)$  degrees of freedom under the null hypothesis. In most of our cases  $N \sim \mathcal{O}(100)$ ,  
 3403 so the probability density functions will look very similar. In this case, it is safe to take  
 3404 the limit  $N \rightarrow \infty$  in the Fisher PDF:

$$\begin{aligned}\tilde{f}(x; a - b) &= \lim_{N \rightarrow \infty} f(x; a - b, N - a) \\ &= \frac{2^{-\frac{a-b}{2}}}{\Gamma(\frac{a-b}{2})} (a - b)^{\frac{a-b}{2}} x^{\frac{a-b}{2}-1} e^{-\frac{a-b}{2}x}.\end{aligned}\tag{8.32}$$

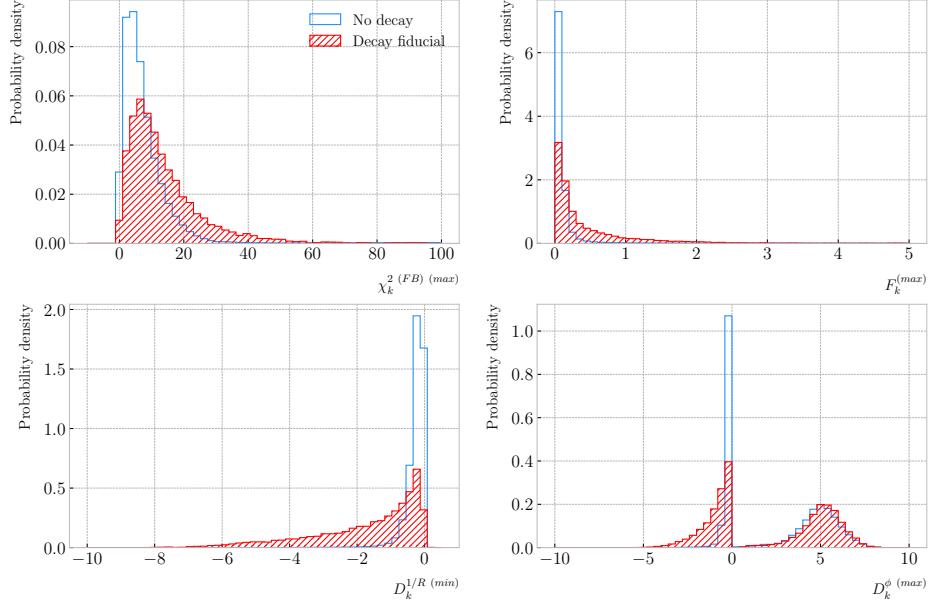
3405 In our case  $a - b = 8 - 5 = 3$ , so we would obtain a p-value of 0.05 at  $x = 2.60$ .

3406 Figure 8.37 contains the distributions of the maxima of  $\chi_k^2(FB)$ ,  $F_k$  and  $D_k^\phi$  and the  
 3407 minima of  $D_k^{1/R}$  for a sample of non-decaying pion tracks (blue) and another sample of  
 3408 reconstructed tracks containing part of the pion and the daughter muon from a decay  
 3409 inside the fiducial volume (red). Notice that, even though the values of  $F_k^{(max)}$  for the  
 3410 decay sample are typically larger than for the non-decaying one, just a small fraction of  
 3411 the events go beyond the aforementioned value of  $F = 2.60$ . Therefore, from a practical  
 3412 point of view, it is not the most efficient variable to use for selecting the decay events.

3413 However, looking at the  $D_k^{1/R \ (min)}$  distribution we can see there is a big difference  
 3414 between non-decaying and decaying events in this variable. One can use a combination  
 3415 of these four variables to distinguish between the pion decay events (signal) and the  
 3416 non-decaying pions (background).

3417 An approach to this classification could be using a boosted decision tree (BDT). One  
 3418 of the advantages of BDTs is that they are easy to interpret and identify the relative  
 3419 importance of the different input variables. Training a BDT with 400 estimators and a

## 8.4. Charged pion decay in flight



**Figure 8.37:** 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).

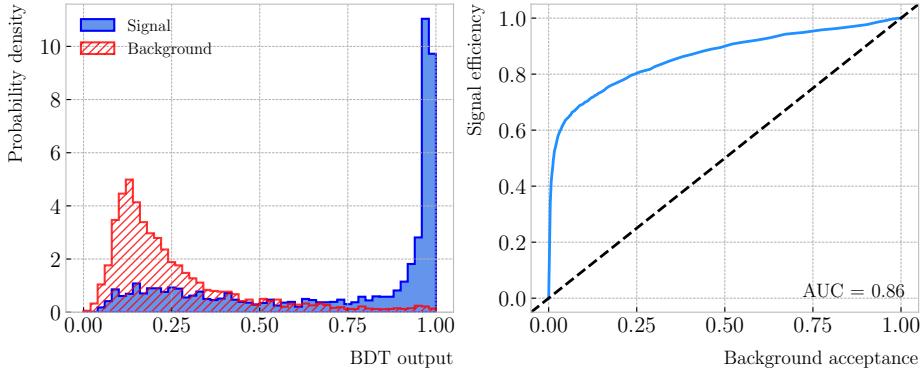
maximum depth of 4 I can obtain an efficient classification without overtraining. Figure 8.38 (left panel) shows the distribution of probabilities predicted by the BDT for a test sample. The signal efficiency as a function of background acceptance, the so-called ROC curve, is shown in Fig. 8.38 (right panel). With a relative importance of 0.83, the most important variable turned out to be  $D_k^{1/R} \text{ (min)}$ .

One thing we can check is how the resolution to the decay and the signal efficiency in the classification changes with the true decay angle. Using an equal-frequency binning for the decay angles, we can repeat the previous steps for each bin.

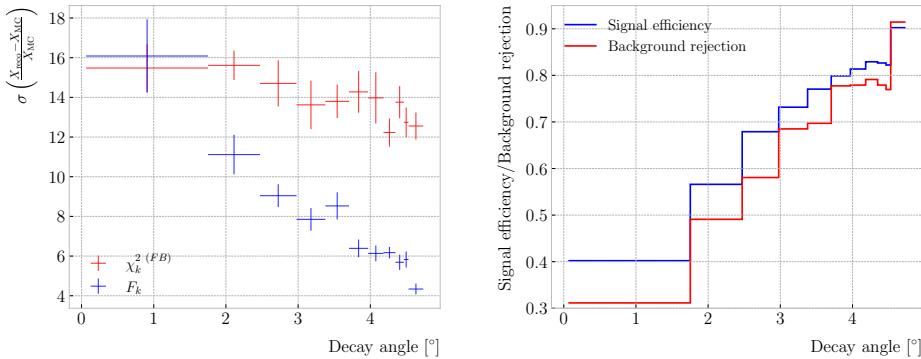
Figure 8.39 (left panel) shows the dependence on the decay angle of the decay finding resolution. We can see that for the  $\chi_k^2(FB)$  maximum location method the resolution consistently lies between 12 to 16%. However, the  $F_k^{(max)}$  approach gives a significantly better resolution for high 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

## Chapter 8. Particle ID in GArSoft



**Figure 8.38:** 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 8.39:** 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.

<sup>3434</sup> signal efficiency in each bin for a fixed value of the background rejection, in this case  
<sup>3435</sup> 90%. Similarly, for the background rejection estimation I use a fixed signal efficiency  
<sup>3436</sup> value of 90%. Figure 8.39 (right panel) represents the change in signal efficiency (blue)  
<sup>3437</sup> and background rejection (red) with the value of the true decay angles.

## 8.5. Neutral particle identification

### 3438 8.5 Neutral particle identification

#### 3439 8.5.1 ECal clustering

3440 Another important reconstruction item is the clustering algorithm of ECal hits in  
3441 GArSoft. The default module features a NN algorithm that treats all hits in the same  
3442 way, independently of the layer each hit comes from. However, the current ECal design  
3443 of ND-GAr has two very different types of scintillator layers. The inner layers are made  
3444 out of tiles, which provide excellent angular and timing resolutions. On the other hand,  
3445 the outer layers are cross scintillator strips. That way, an algorithm that treats hits  
3446 from both kinds of layers differently may be able to improve the current performance.

3447 Inspired by the reconstruction of T2K's ND280 downstream ECal [115], the idea  
3448 was to put together a clustering module that first builds clusters for the different ECal  
3449 views (tiles, strips segmented in the  $X$  direction and strips segmented in  $Y$  direction),  
3450 and then tries to match them together to form the final clusters.

3451 Working on a module-by-module basis, the algorithm first separates the hits depending  
3452 on the layer type they come from. Then, it performs a NN clustering for the 3 sets of  
3453 hits separately. For the tile hits it clusters together all the hits which are in nearest-  
3454 neighbouring tiles and nearest-neighbouring layers, for strip hits it looks at nearest-  
3455 neighbouring strips and next-to-nearest-neighbouring layers (as the layers with strips  
3456 along the two directions are alternated). For strip clusters an additional cut in the  
3457 direction along the strip length is needed.

3458 After this first clustering I then apply a recursive re-clustering for each collection  
3459 of strip clusters based on a PCA method. In each case, we loop over the clusters with  
3460  $N_{hits} \geq 2$ , computing the centre of mass and three principal components. Propagating  
3461 these axes up to the layers of the rest of the clusters, we check if the propagated point  
3462 and the centre of mass of the second cluster are within next-to-nearest-neighbouring  
3463 strips. An additional cut in the direction along the strip length is also needed. Moreover,  
3464 I require that the two closest hits across the two clusters are at most in next-to-nearest-  
3465 neighbouring strips. I merge the clusters if these three conditions are satisfied. The

## Chapter 8. Particle ID in GArSoft

3466 re-clustering is repeated until no more cluster pairs pass the cuts.

3467 The clusters in each strip view are combined if their centres of mass are close enough  
3468 and they point in the same direction. An alternative approach for the strip cluster  
3469 merging could be to compute the overlap between the ellipsoids defined by the principal  
3470 axes of the clusters, and then merge the pair if the overlap exceeds some threshold.  
3471 Further study is needed to understand if this change would have an impact in the overall  
3472 clustering performance.

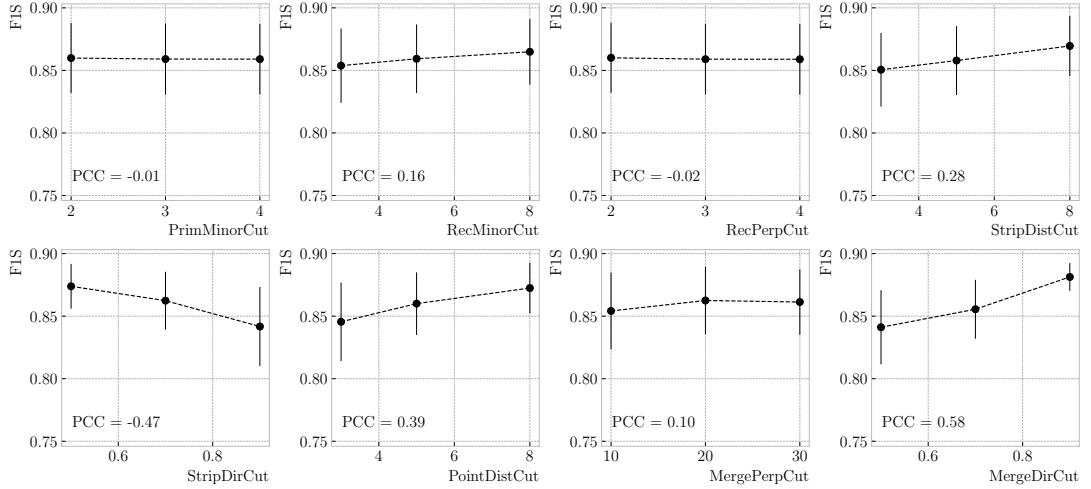
3473 To merge the tile clusters to the combined strip clusters I propagate the principal  
3474 axis of the strip cluster towards the inner layers, up to the centre of mass layer of the  
3475 tile cluster. I merge the clusters if the distance between the propagated point and the  
3476 centre of mass is below a certain cut.

3477 The last step is to check if clusters in neighbouring modules should be merged  
3478 together, both across two barrel modules, across end cap modules and between barrel  
3479 end cap modules. I check the distance between the two closest hits in the pair of clusters  
3480 and merge them if it passes this and an additional direction cut.

3481 Figure ?? presents an example of the clustering steps relevant for strip layer hits, from  
3482 the input hits (top left panel) to the NN clustering (top right panel) and re-clustering  
3483 (bottom left panel) for each strip view and the final merging strip clusters (bottom  
3484 right panel). It shows the hits from a single ECal barrel module in a  $\nu_\mu$  CC interaction  
3485 event with a neutral pion and a proton in the final state. The two clusters on the left  
3486 correspond to the photon pair from the  $\pi^0$  decay and the one on the upper right corner  
3487 is associated to the proton.

3488 This algorithm has a total number of eight free parameters that need to be optimised.  
3489 I used a sample of 1000  $\nu_\mu$  CC interactions in order to obtain the optimal configuration of  
3490 clustering parameters. This sample was generated up to the default ECal hit clustering  
3491 level, so then I could run the new clustering algorithm each time with a different  
3492 configuration of parameters. As the number of parameters is relatively large, I only  
3493 performed a coarse-grained scan of the parameter space. Sampling each of the eight  
3494 parameters at three different points each I obtain 6561 different configurations. These

## 8.5. Neutral particle identification



**Figure 8.40:** 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 8.5:** 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

parameters, together with the used values, are summarised in Tab. 8.5.

In order to measure the performance of the clustering, I use a binary classification approach. For each formed cluster, I identify the Geant4 Track ID of the matching MC particle and the energy fraction of each hit. Then, I assign to each cluster the Track ID with the highest total energy fraction. For each of the different Track IDs associated to the clusters, I select the cluster with the highest energy (only from the hits with the same Track ID). I identify such a cluster as the main cluster for that Track ID. I count

## Chapter 8. Particle ID in GArSoft

as true positives (TPs) the hits with the correct Track ID in each main cluster. False positives (FPs) are the hits with the incorrect Track ID for the cluster they are in, not only main clusters. The false negatives (FNs) are the hits with the correct Track ID in clusters other than the main.

Figure 8.40 shows the computed  $F_1$ -score values for the different cuts. In each case, the central value represents the mean of the  $F_1$ -score distribution for the specified value of the corresponding variable and the vertical error bar represents one standard deviation around the mean. Also shown are the Pearson correlation coefficients of these central values. We can see that five of the variables have a sizeable effect on the  $F_1$ -score, with an absolute difference between the last and first values as big as 4%.

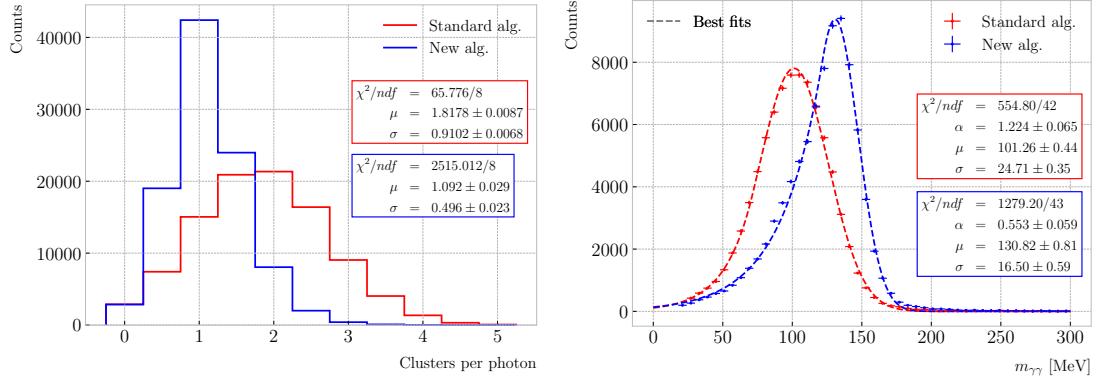
The working configuration is obtained as follows. I first select all configurations with purity  $\geq 90\%$ . Among those, I choose the combinations that yield the maximum  $F_1$ -score. If more than one configuration remains I select the one with the highest sensitivity. Doing so, I end up with a parameter configuration with an efficiency of 88% and a 90% purity. Compared with the default algorithm, which gives an efficiency of 76% and a purity of 91% for the same sample, I have managed to improve the efficiency by a factor of 1.16.

### 8.5.2 $\pi^0$ reconstruction

One of the potential applications of the new ECal hit clustering is the reconstruction of neutral particles, in particular pions. Neutral pions decay promptly after being produced, through the  $\pi^0 \rightarrow \gamma\gamma$  channel ( $98.823 \pm 0.034\%$ ) of the time. The photon pair does not leave any traces in the HPgTPC (unless one or both of them converts into an electron-positron pair), but each of them will produce an electromagnetic shower in the ECal.

To test the potential impact of the new algorithm in  $\pi^0$  reconstruction, I generated a MC sample of single, isotropic neutral pions inside the HPgTPC. All pions were generated with a momentum  $p = 500$  MeV and their initial positions were uniformly sampled inside a  $2 \times 2 \times 2$  m box aligned with the centre of the TPC. I ran both the

## 8.5. Neutral particle identification



**Figure 8.41:** 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.

3530 default and the new clustering algorithms, using for the latter the optimised configuration  
3531 discussed above.

3532 The first thing to notice is that the number of clusters produced per photon has  
3533 decreased. Figure 8.41 (left panel) shows these distributions for the default (red) and  
3534 new (blue) algorithms. Using a simple Gaussian fit, we see that the mean number of  
3535 ECal clusters per photon went from  $1.82 \pm 0.01$  to  $1.09 \pm 0.03$ . This effectively means that  
3536 with the new algorithm the ECal activity of one true particle is typically reconstructed  
3537 as a single object. From the reconstruction point of view this can be an advantage. As  
3538 now most of the photon energy ends up in a single ECal cluster, I can simply use cluster  
3539 pairs to identify the  $\pi^0$  decay.

3540 In general, one calculates the invariant mass of the photon pair as:

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos \theta)}, \quad (8.33)$$

3541 where  $E_i$  are the energies of the photons and  $\theta$  the opening angle between them. In this  
3542 case I can use the energies deposited in the ECal and their incident directions. This  
3543 quantity is computed for all possible pairs of clusters, using their position together with  
3544 the true decay point. In a more realistic scenario, e.g.  $\nu_\mu$  CC interaction, one could use  
3545 the position of the reconstructed primary vertex instead. I also tried to use the principal

## Chapter 8. Particle ID in GArSoft

3546 direction of the clusters, but that approach gave considerably worse results. For each  
3547 event I only keep the pair with an invariant mass closer to the true  $\pi^0$  mass value.

3548 Figure 8.41 (right panel) shows the invariant mass distributions for the photon pairs  
3549 we get using the default (red) and the new (blue) ECal clustering algorithms. For the fit  
3550 I used a modified version of the Crystal Ball function [116], obtained by taking the limit  
3551 where the parameter controlling the power-law tail goes to infinity:

$$f(x; N, \mu, \sigma, \alpha) = N \cdot \begin{cases} e^{\frac{\alpha(2x-2\mu+\alpha\sigma)}{2\sigma}}; & x \leq \mu - \alpha\sigma, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}; & x > \mu - \alpha\sigma. \end{cases} \quad (8.34)$$

3552 Comparing the fitted mean and standard deviation values for the Gaussian cores, we  
3553 see that the distribution for the new algorithm is a 67% narrower and also peaks much  
3554 closer to the true  $m_{\pi^0}$  value, going from  $101.3 \pm 0.4$  MeV to  $130.8 \pm 0.6$  MeV.

3555 Chapter 9

3556 Event selection in ND-GAr

3557 9.1 CAFs and CAFAna

3558 9.2 Event selection

3559 9.2.1  $\nu_\mu$  CC selection

3560 9.2.2 Charged pion multiplicity



<sub>3561</sub> Chapter 10

<sub>3562</sub> Conclusions



<sup>3563</sup> Appendix A

<sup>3564</sup> An appendix



# 3565 Bibliography

- 3566 [1] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*  
3567 *Detector Technical Design Report, Volume I Introduction to DUNE, JINST* **15**  
3568 (2020) T08008 [2002.02967]. 15, 49, 50, 51, 54, 55, 56, 62, 63, 64, 71, 72
- 3569 [2] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far*  
3570 *Detector Technical Design Report, Volume IV: Far Detector Single-phase*  
3571 *Technology, JINST* **15** (2020) T08010 [2002.03010]. 15, 59
- 3572 [3] J.N. Bahcall, A.M. Serenelli and S. Basu, *New solar opacities, abundances,*  
3573 *helioseismology, and neutrino fluxes, Astrophys. J. Lett.* **621** (2005) L85  
3574 [[astro-ph/0412440](#)]. 21, 119, 121
- 3575 [4] R. Garani and S. Palomares-Ruiz, *Dark matter in the Sun: scattering off electrons*  
3576 *vs nucleons, JCAP* **05** (2017) 007 [[1702.02768](#)]. 21, 117, 118, 120, 121, 122, 145
- 3577 [5] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara and S. Midorikawa,  
3578 *Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model,*  
3579 *Phys. Rev. D* **92** (2015) 023004 [[1502.03916](#)]. 22, 125, 126
- 3580 [6] M. Colom i Bernadich and C. Pérez de los Heros, *Limits on Kaluza-Klein dark*  
3581 *matter annihilation in the Sun from recent IceCube results, Eur. Phys. J. C,* **80** 2  
3582 (2020) 129 **80** (2019) [[1912.04585](#)]. 22, 130
- 3583 [7] ANTARES collaboration, *Search for Dark Matter in the Sun with the ANTARES*  
3584 *Neutrino Telescope in the CMSSM and mUED frameworks, Nucl. Instrum. Meth.*  
3585 *A* **725** (2013) 76 [[1204.5290](#)]. 22, 130

## BIBLIOGRAPHY

- 3586 [8] N. Deutschmann, T. Flacke and J.S. Kim, *Current LHC Constraints on Minimal*  
3587 *Universal Extra Dimensions*, *Phys. Lett. B* **771** (2017) 515 [[1702.00410](#)]. 22, 130,  
3588 131
- 3589 [9] ICECUBE collaboration, *Search for GeV-scale dark matter annihilation in the Sun*  
3590 *with IceCube DeepCore*, *Phys. Rev. D* **105** (2022) 062004 [[2111.09970](#)]. 24, 142,  
3591 143
- 3592 [10] C.-S. Chen, F.-F. Lee, G.-L. Lin and Y.-H. Lin, *Probing Dark Matter*  
3593 *Self-Interaction in the Sun with IceCube-PINGU*, *JCAP* **10** (2014) 049  
3594 [[1408.5471](#)]. 24, 142, 143
- 3595 [11] N.F. Bell, M.J. Dolan and S. Robles, *Searching for dark matter in the Sun using*  
3596 *Hyper-Kamiokande*, *JCAP* **11** (2021) 004 [[2107.04216](#)]. 24, 142, 143
- 3597 [12] E. Behnke et al., *Final Results of the PICASSO Dark Matter Search Experiment*,  
3598 *Astropart. Phys.* **90** (2017) 85 [[1611.01499](#)]. 24, 142, 143
- 3599 [13] PICO collaboration, *Dark Matter Search Results from the Complete Exposure of*  
3600 *the PICO-60 C<sub>3</sub>F<sub>8</sub> Bubble Chamber*, *Phys. Rev. D* **100** (2019) 022001  
3601 [[1902.04031](#)]. 24, 142, 143
- 3602 [14] DARKSIDE collaboration, *Constraints on Sub-GeV Dark-Matter–Electron*  
3603 *Scattering from the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018) 111303  
3604 [[1802.06998](#)]. 24, 146, 147
- 3605 [15] XENON collaboration, *Light Dark Matter Search with Ionization Signals in*  
3606 *XENON1T*, *Phys. Rev. Lett.* **123** (2019) 251801 [[1907.11485](#)]. 24, 146, 147
- 3607 [16] P.F. de Salas, D.V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena,  
3608 C.A. Ternes et al., *2020 global reassessment of the neutrino oscillation picture*,  
3609 *JHEP* **02** (2021) 071 [[2006.11237](#)]. 31, 46, 51
- 3610 [17] W. Pauli, *Dear radioactive ladies and gentlemen*, *Phys. Today* **31N9** (1978) 27. 37

## BIBLIOGRAPHY

- 3611 [18] F. Reines and C.L. Cowan, *Detection of the free neutrino*, *Phys. Rev.* **92** (1953)  
3612 830. 37
- 3613 [19] S.L. Glashow, *Partial-symmetries of weak interactions*, . 38
- 3614 [20] S. Weinberg, *A model of leptons*, . 38
- 3615 [21] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968)  
3616 367. 38
- 3617 [22] A. Pich, *The Standard Model of Electroweak Interactions*, in *2010 European  
School of High Energy Physics*, pp. 1–50, 1, 2012 [[1201.0537](#)]. 38
- 3619 [23] ALEPH, DELPHI, L3, OPAL, SLD, LEP ELECTROWEAK WORKING GROUP,  
3620 SLD ELECTROWEAK GROUP, SLD HEAVY FLAVOUR GROUP collaboration,  
3621 *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006)  
3622 257 [[hep-ex/0509008](#)]. 41
- 3623 [24] SUPER-KAMIOKANDE collaboration, *Evidence for oscillation of atmospheric  
neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [[hep-ex/9807003](#)]. 42
- 3625 [25] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, *Zh.  
Eksp. Teor. Fiz.* **34** (1957) 247. 42
- 3627 [26] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary  
particles*, *Prog. Theor. Phys.* **28** (1962) 870. 42
- 3629 [27] L. Wolfenstein, *Neutrino Oscillations in Matter*, *Phys. Rev. D* **17** (1978) 2369. 44
- 3630 [28] B.T. Cleveland, T. Daily, R. Davis, Jr., J.R. Distel, K. Lande, C.K. Lee et al.,  
3631 *Measurement of the solar electron neutrino flux with the Homestake chlorine  
detector*, *Astrophys. J.* **496** (1998) 505. 45
- 3633 [29] F. Kaether, W. Hampel, G. Heusser, J. Kiko and T. Kirsten, *Reanalysis of the  
GALLEX solar neutrino flux and source experiments*, *Phys. Lett. B* **685** (2010) 47  
3634 [[1001.2731](#)]. 45

## BIBLIOGRAPHY

- 3636 [30] SAGE collaboration, *Measurement of the solar neutrino capture rate with gallium*  
3637 *metal. III: Results for the 2002–2007 data-taking period*, *Phys. Rev. C* **80** (2009)  
3638 015807 [0901.2200]. 45
- 3639 [31] G. Bellini et al., *Precision measurement of the  $^7\text{Be}$  solar neutrino interaction rate*  
3640 *in Borexino*, *Phys. Rev. Lett.* **107** (2011) 141302 [1104.1816]. 45
- 3641 [32] SUPER-KAMIOKANDE collaboration, *Solar neutrino measurements in*  
3642 *super-Kamiokande-I*, *Phys. Rev. D* **73** (2006) 112001 [[hep-ex/0508053](#)]. 45
- 3643 [33] SNO collaboration, *Combined Analysis of all Three Phases of Solar Neutrino*  
3644 *Data from the Sudbury Neutrino Observatory*, *Phys. Rev. C* **88** (2013) 025501  
3645 [1109.0763]. 45
- 3646 [34] SUPER-KAMIOKANDE collaboration, *Atmospheric neutrino oscillation analysis*  
3647 *with external constraints in Super-Kamiokande I-IV*, *Phys. Rev. D* **97** (2018)  
3648 072001 [1710.09126]. 45
- 3649 [35] ICECUBE collaboration, *Measurement of Atmospheric Neutrino Oscillations at*  
3650 *6–56 GeV with IceCube DeepCore*, *Phys. Rev. Lett.* **120** (2018) 071801  
3651 [1707.07081]. 45
- 3652 [36] KAMLAND collaboration, *Reactor On-Off Antineutrino Measurement with*  
3653 *KamLAND*, *Phys. Rev. D* **88** (2013) 033001 [1303.4667]. 45
- 3654 [37] RENO collaboration, *Measurement of Reactor Antineutrino Oscillation Amplitude*  
3655 *and Frequency at RENO*, *Phys. Rev. Lett.* **121** (2018) 201801 [1806.00248]. 45
- 3656 [38] DAYA BAY collaboration, *Measurement of the Electron Antineutrino Oscillation*  
3657 *with 1958 Days of Operation at Daya Bay*, *Phys. Rev. Lett.* **121** (2018) 241805  
3658 [1809.02261]. 45
- 3659 [39] K.J. Kelly, P.A.N. Machado, S.J. Parke, Y.F. Perez-Gonzalez and R.Z. Funchal,  
3660 *Neutrino mass ordering in light of recent data*, *Phys. Rev. D* **103** (2021) 013004.  
3661 46

## BIBLIOGRAPHY

- 3662 [40] P. Dunne, “Latest Neutrino Oscillation Results from T2K.” Jul, 2020. 46
- 3663 [41] MINOS collaboration, *Combined analysis of  $\nu_\mu$  disappearance and  $\nu_\mu \rightarrow \nu_e$  appearance in MINOS using accelerator and atmospheric neutrinos*, *Phys. Rev. Lett.* **112** (2014) 191801 [[1403.0867](#)]. 46
- 3664 [42] OPERA collaboration, *Final Results of the OPERA Experiment on  $\nu_\tau$  Appearance in the CNGS Neutrino Beam*, *Phys. Rev. Lett.* **120** (2018) 211801
- 3665 [1804.04912]. 46
- 3666 [43] K2K collaboration, *Measurement of Neutrino Oscillation by the K2K Experiment*, *Phys. Rev. D* **74** (2006) 072003 [[hep-ex/0606032](#)]. 46
- 3667 [44] DUNE collaboration, *Long-baseline neutrino oscillation physics potential of the DUNE experiment*, *Eur. Phys. J. C* **80** (2020) 978 [[2006.16043](#)]. 46
- 3668 [45] HYPER-KAMIOKANDE collaboration, *Physics potential of Hyper-Kamiokande for neutrino oscillation measurements*, *PoS NuFact2019* (2019) 040. 46
- 3669 [46] SUPERNEMO collaboration, *Probing New Physics Models of Neutrinoless Double Beta Decay with SuperNEMO*, *Eur. Phys. J. C* **70** (2010) 927 [[1005.1241](#)]. 47
- 3670 [47] SNO+ collaboration, *Current Status and Future Prospects of the SNO+ Experiment*, *Adv. High Energy Phys.* **2016** (2016) 6194250 [[1508.05759](#)]. 47
- 3671 [48] NEXT collaboration, *Sensitivity of a tonne-scale NEXT detector for neutrinoless double beta decay searches*, *JHEP* **2021** (2021) 164 [[2005.06467](#)]. 47
- 3672 [49] DUNE collaboration, *Snowmass Neutrino Frontier: DUNE Physics Summary*, [2203.06100](#). 51, 52
- 3673 [50] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics*, [2002.03005](#). 51, 53, 67, 69, 125, 134, 138

## BIBLIOGRAPHY

- 3686 [51] SUPER-KAMIOKANDE collaboration, *Search for Proton Decay via  $p \rightarrow e^+ \pi_0$  and*
- 3687  $p \rightarrow \mu^+ \pi_0$  in a Large Water Cherenkov Detector, *Phys. Rev. Lett.* **102** (2009)
- 3688 141801 [0903.0676]. 52
- 3689 [52] S. Raby, *Grand Unified Theories*, in *2nd World Summit: Physics Beyond the*
- 3690 *Standard Model*, 8, 2006 [[hep-ph/0608183](#)]. 52
- 3691 [53] KAMIOKANDE-II collaboration, *Observation of a Neutrino Burst from the*
- 3692 *Supernova SN 1987a*, *Phys. Rev. Lett.* **58** (1987) 1490. 53
- 3693 [54] R.M. Bionta et al., *Observation of a Neutrino Burst in Coincidence with*
- 3694 *Supernova SN 1987a in the Large Magellanic Cloud*, *Phys. Rev. Lett.* **58** (1987)
- 3695 1494. 53
- 3696 [55] DUNE collaboration, *The DUNE Far Detector Vertical Drift Technology,*
- 3697 *Technical Design Report*, 2312.03130. 57, 58
- 3698 [56] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE) Near*
- 3699 *Detector Conceptual Design Report, Instruments* **5** (2021) 31 [2103.13910]. 61,
- 3700 62, 65, 69, 149
- 3701 [57] J. Strait, E. McCluskey, T. Lundin, J. Willhite, T. Hamernik, V. Papadimitriou
- 3702 et al., *Long-baseline neutrino facility (lbnf) and deep underground neutrino*
- 3703 *experiment (dune) conceptual design report volume 3: Long-baseline neutrino*
- 3704 *facility for dune june 24, 2015*, 1601.05823. 66
- 3705 [58] C. Green, J. Kowalkowski, M. Paterno, M. Fischler, L. Garren and Q. Lu, *The*
- 3706 *Art Framework*, *J. Phys. Conf. Ser.* **396** (2012) 022020. 73
- 3707 [59] DUNE DAQ, “dtp-firmware.”
- 3708 <https://gitlab.cern.ch/dune-daq/readout/dtp-firmware>, 2020. 81
- 3709 [60] DUNE DAQ, “dtp-simulation.”
- 3710 <https://gitlab.cern.ch/dune-daq/readout/dtp-simulation>, 2020. 84

## BIBLIOGRAPHY

- 3711 [61] DUNE DAQ, “dtpemulator.”  
3712 [https://github.com/DUNE-DAQ/dtpemulator/tree/fmlopez/filter\\_ana](https://github.com/DUNE-DAQ/dtpemulator/tree/fmlopez/filter_ana),  
3713 2022. 84
- 3714 [62] J. McClellan and T. Parks, *A personal history of the Parks-McClellan algorithm*,  
3715 *IEEE Signal Processing Magazine* **22** (2005) 82. 85
- 3716 [63] G. Turin, *An introduction to matched filters*, *IEEE Transactions on Information*  
3717 *Theory* **6** (1960) 311. 88
- 3718 [64] J.W. Goodman, *Statistical Optics*, Wiley (1985). 90
- 3719 [65] B. Dwork, *Detection of a pulse superimposed on fluctuation noise*, *Proceedings of*  
3720 *the IRE* **38** (1950) 771. 91
- 3721 [66] L. Wainstein and V. Zubakov, *Extraction of Signals from Noise*, Prentice-Hall  
3722 (1962). 91
- 3723 [67] E.D. Church, *Larsoft: A software package for liquid argon time projection drift*  
3724 *chambers*, **1311.6774**. 94, 95
- 3725 [68] S.V. Stehman, *Selecting and interpreting measures of thematic classification*  
3726 *accuracy*, *Remote Sensing of Environment* **62** (1997) 77. 108
- 3727 [69] A.A. Taha and A. Hanbury, *Metrics for evaluating 3d medical image*  
3728 *segmentation: analysis, selection, and tool*, *BMC Medical Imaging* **15** (2015) . 109
- 3729 [70] J. Silk, K.A. Olive and M. Srednicki, *The Photino, the Sun and High-Energy*  
3730 *Neutrinos*, *Phys. Rev. Lett.* **55** (1985) 257. 115
- 3731 [71] M. Srednicki, K.A. Olive and J. Silk, *High-Energy Neutrinos from the Sun and*  
3732 *Cold Dark Matter*, *Nucl. Phys. B* **279** (1987) 804. 115
- 3733 [72] J.S. Hagelin, K.W. Ng and K.A. Olive, *A High-energy Neutrino Signature From*  
3734 *Supersymmetric Relics*, *Phys. Lett. B* **180** (1986) 375. 115

## BIBLIOGRAPHY

- 3735 [73] T.K. Gaisser, G. Steigman and S. Tilav, *Limits on Cold Dark Matter Candidates*  
3736 *from Deep Underground Detectors*, *Phys. Rev. D* **34** (1986) 2206. 115
- 3737 [74] N. Bernal, J. Martín-Albo and S. Palomares-Ruiz, *A novel way of constraining*  
3738 *WIMPs annihilations in the Sun: MeV neutrinos*, *JCAP* **08** (2013) 011  
3739 [1208.0834]. 115, 116, 117, 123
- 3740 [75] C. Rott, J. Siegal-Gaskins and J.F. Beacom, *New Sensitivity to Solar WIMP*  
3741 *Annihilation using Low-Energy Neutrinos*, *Phys. Rev. D* **88** (2013) 055005  
3742 [1208.0827]. 115, 122
- 3743 [76] C. Rott, S. In, J. Kumar and D. Yavlali, *Dark Matter Searches for Monoenergetic*  
3744 *Neutrinos Arising from Stopped Meson Decay in the Sun*, *JCAP* **11** (2015) 039  
3745 [1510.00170]. 115
- 3746 [77] DUNE collaboration, *Searching for solar KDAR with DUNE*, *JCAP* **10** (2021)  
3747 065 [2107.09109]. 115
- 3748 [78] G. Busoni, A. De Simone and W.-C. Huang, *On the Minimum Dark Matter Mass*  
3749 *Testable by Neutrinos from the Sun*, *JCAP* **07** (2013) 010 [1305.1817]. 116
- 3750 [79] V.A. Bednyakov and F. Simkovic, *Nuclear spin structure in dark matter search:*  
3751 *The Zero momentum transfer limit*, *Phys. Part. Nucl.* **36** (2005) 131  
3752 [hep-ph/0406218]. 117
- 3753 [80] A. Gould, *WIMP Distribution in and Evaporation From the Sun*, *Astrophys. J.*  
3754 **321** (1987) 560. 118
- 3755 [81] J.N. Bahcall, M.H. Pinsonneault and S. Basu, *Solar models: Current epoch and*  
3756 *time dependences, neutrinos, and helioseismological properties*, *Astrophys. J.* **555**  
3757 (2001) 990 [astro-ph/0010346]. 120
- 3758 [82] T. Golan, J.T. Sobczyk and J. Zmuda, *NuWro: the Wroclaw Monte Carlo*  
3759 *Generator of Neutrino Interactions*, *Nucl. Phys. B Proc. Suppl.* **229-232** (2012)  
3760 499. 124

## BIBLIOGRAPHY

- 3761 [83] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for  
3762 likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [[1007.1727](#)].  
3763 125
- 3764 [84] T. Kaluza, *Zum Unitätsproblem der Physik, Sitzungsber. Preuss. Akad. Wiss.  
3765 Berlin (Math. Phys.)* **1921** (1921) 966 [[1803.08616](#)]. 127
- 3766 [85] O. Klein, *Quantum Theory and Five-Dimensional Theory of Relativity. (In  
3767 German and English)*, *Z. Phys.* **37** (1926) 895. 127
- 3768 [86] T. Appelquist, H.-C. Cheng and B.A. Dobrescu, *Bounds on universal extra  
3769 dimensions*, *Phys. Rev. D* **64** (2001) 035002 [[hep-ph/0012100](#)]. 127
- 3770 [87] N. Arkani-Hamed, S. Dimopoulos and G.R. Dvali, *The Hierarchy problem and new  
3771 dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263 [[hep-ph/9803315](#)]. 127
- 3772 [88] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra  
3773 dimension*, *Phys. Rev. Lett.* **83** (1999) 3370 [[hep-ph/9905221](#)]. 127
- 3774 [89] G. Servant and T.M.P. Tait, *Is the lightest Kaluza-Klein particle a viable dark  
3775 matter candidate?*, *Nucl. Phys. B* **650** (2003) 391 [[hep-ph/0206071](#)]. 127
- 3776 [90] H.-C. Cheng, K.T. Matchev and M. Schmaltz, *Radiative corrections to  
3777 Kaluza-Klein masses*, *Phys. Rev. D* **66** (2002) 036005 [[hep-ph/0204342](#)]. 129
- 3778 [91] M. Blennow, J. Edsjö and T. Ohlsson, *Neutrinos from WIMP annihilations using  
3779 a full three-flavor Monte Carlo*, *JCAP* **01** (2008) 021 [[0709.3898](#)]. 129, 132
- 3780 [92] J. Edsjö, J. Elevant and C. Niblaeus, *WimpSim Neutrino Monte Carlo*. 129, 132
- 3781 [93] U. Haisch and A. Weiler, *Bound on minimal universal extra dimensions from  
3782 anti-B → X(s)gamma*, *Phys. Rev. D* **76** (2007) 034014 [[hep-ph/0703064](#)]. 131
- 3783 [94] A. Freitas and U. Haisch, *Anti-B → X(s) gamma in two universal extra  
3784 dimensions*, *Phys. Rev. D* **77** (2008) 093008 [[0801.4346](#)]. 131

## BIBLIOGRAPHY

- 3785 [95] C. Rott, D. Jeong, J. Kumar and D. Yaylali, *Neutrino Topology Reconstruction at*  
3786 *DUNE and Applications to Searches for Dark Matter Annihilation in the Sun,*  
3787 *JCAP* **07** (2019) 006 [[1903.04175](#)]. 134
- 3788 [96] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al.,  
3789 *Scikit-learn: Machine learning in Python, Journal of Machine Learning Research*  
3790 **12** (2011) 2825. 140
- 3791 [97] J. Kopp, V. Niro, T. Schwetz and J. Zupan, *DAMA/LIBRA and leptонically*  
3792 *interacting Dark Matter, Phys. Rev. D* **80** (2009) 083502 [[0907.3159](#)]. 144
- 3793 [98] PARTICLE DATA GROUP collaboration, *Review of Particle Physics, PTEP* **2020**  
3794 (2020) 083C01. 145
- 3795 [99] M. Beltran, D. Hooper, E.W. Kolb and Z.C. Krusberg, *Deducing the nature of*  
3796 *dark matter from direct and indirect detection experiments in the absence of*  
3797 *collider signatures of new physics, Phys. Rev. D* **80** (2009) 043509 [[0808.3384](#)].  
3798 145
- 3799 [100] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters, Astron.*  
3800 *Astrophys.* **641** (2020) A6 [[1807.06209](#)]. 146
- 3801 [101] H. Bethe, *Zur theorie des durchgangs schneller korpuskularstrahlen durch materie,*  
3802 *Annalen der Physik* **397** (1930) 325. 150
- 3803 [102] E. Fermi, *The ionization loss of energy in gases and in condensed materials,*  
3804 *Physical Review* **57** (1940) 485. 151
- 3805 [103] R. Sternheimer, M. Berger and S. Seltzer, *Density effect for the ionization loss of*  
3806 *charged particles in various substances, Atomic Data and Nuclear Data Tables* **30**  
3807 (1984) 261. 151
- 3808 [104] W.W.M. Allison and J.H. Cobb, *Relativistic Charged Particle Identification by*  
3809 *Energy Loss, Ann. Rev. Nucl. Part. Sci.* **30** (1980) 253. 151

## BIBLIOGRAPHY

- 3810 [105] W. Blum, L. Rolandi and W. Riegler, *Particle detection with drift chambers*,  
3811 Particle Acceleration and Detection (2008), 10.1007/978-3-540-76684-1. 152
- 3812 [106] ALICE TPC collaboration, *Particle identification of the ALICE TPC via dE/dx*,  
3813 *Nucl. Instrum. Meth. A* **706** (2013) 55. 152
- 3814 [107] L. Landau, *On the energy loss of fast particles by ionization*, *J. Phys. (USSR)* **8**  
3815 (1944) 201. 155
- 3816 [108] W. Ulmer and E. Matsinos, *Theoretical methods for the calculation of Bragg  
3817 curves and 3D distributions of proton beams*, *The European Physical Journal  
3818 Special Topics* **190** (2010) 1. 156
- 3819 [109] E. Aprile, A.E. Bolotnikov, A.L. Bolozdynya and T. Doke, *Noble Gas Detectors*,  
3820 Wiley (Oct., 2008), 10.1002/9783527610020. 158
- 3821 [110] D.S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press. 195
- 3822 [111] ALICE collaboration, *Production of pions, kaons and protons in pp collisions at  
3823  $\sqrt{s} = 900 \text{ gev with alice at the lhc}$* , **1101.4110**. 197
- 3824 [112] U. Einhaus, *Charged hadron identification with de/dx and time-of-flight at future  
3825 higgs factories*, **2110.15115**. 197
- 3826 [113] R. Frühwirth, *Application of filter methods to the reconstruction of tracks and  
3827 vertices in events of experimental high energy physics*, Ph.D. thesis, Technischen  
3828 Universität Wien, 1988. 203
- 3829 [114] P. Astier, A. Cardini, R.D. Cousins, A. Letessier-Selvon, B.A. Popov and  
3830 T. Vinogradova, *Kalman filter track fits and track breakpoint analysis*, *Nuclear  
3831 Instruments and Methods in Physics Research Section A: Accelerators,  
3832 Spectrometers, Detectors and Associated Equipment* **450** (2000) 138. 203
- 3833 [115] T2K UK collaboration, *The Electromagnetic Calorimeter for the T2K Near  
3834 Detector ND280*, *JINST* **8** (2013) P10019 [1308.3445]. 211

## BIBLIOGRAPHY

- 3835 [116] J.E. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the  $J/\psi$  and  $\psi'$ ,*  
3836 Ph.D. thesis, Stanford University, 1982. 216