

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/154435>

**Copyright and reuse:**

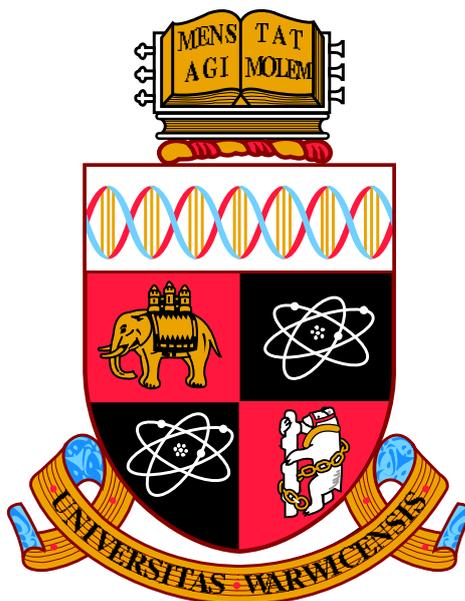
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)



**Electromagnetic Shower  
Characterisation and a  $p \rightarrow K^+ \bar{\nu}$   
sensitivity measurement at DUNE**

by

**James Pillow**

**Thesis**

Submitted to the University of Warwick

for the degree of

**Doctor of Philosophy**

**Department of Physics**

2021

# Contents

<b>List of Tables</b>	<b>1</b>
<b>List of Figures</b>	<b>2</b>
<b>Acknowledgments</b>	<b>10</b>
<b>Declarations</b>	<b>11</b>
<b>Abstract</b>	<b>12</b>
<b>Acronyms</b>	<b>13</b>
<b>Chapter 1 Introduction</b>	<b>16</b>
<b>Chapter 2 Neutrino Theory</b>	<b>18</b>
2.1 Neutrino History . . . . .	18
2.2 Neutrino Interactions . . . . .	23
2.3 Oscillations . . . . .	24
2.3.1 Oscillation Formalism . . . . .	25
2.4 CP Violation . . . . .	27
2.5 MSW Matter Effect . . . . .	28
2.6 Mass Ordering . . . . .	29
2.6.1 Majorana Mass . . . . .	30
<b>Chapter 3 The Deep Underground Neutrino Experiment</b>	<b>31</b>
3.1 Physics Goals of DUNE . . . . .	31
3.2 Far Detector . . . . .	33
3.2.1 Liquid Argon Time Projection Chambers . . . . .	34
3.2.2 Single-Phase . . . . .	36
3.2.2.1 Anode Plane Assemblies . . . . .	37
3.2.2.2 Cathode Plane Assemblies . . . . .	40

3.2.3	Dual-Phase . . . . .	40
3.3	The LBNF Beamline and Target . . . . .	41
3.4	Near Detector . . . . .	44
3.5	ProtoDUNE . . . . .	46
3.5.1	H4-VLE Beamline . . . . .	47
3.5.2	ProtoDUNE-SP . . . . .	47
3.5.2.1	Online Monitor . . . . .	50
<b>Chapter 4 Particle Reconstruction, Identification, and Tagging</b>		<b>52</b>
4.1	Monte Carlo Simulations . . . . .	52
4.2	Data Reconstruction . . . . .	53
4.3	The <i>art</i> Framework . . . . .	54
4.4	The LArSoft Framework . . . . .	54
4.5	The Pandora SDK . . . . .	56
4.5.1	Pandora Reconstruction Chains . . . . .	57
4.5.1.1	Pandora Cosmic . . . . .	57
4.5.1.2	Pandora Test Beam . . . . .	57
4.5.1.3	Particle Stitching . . . . .	57
4.5.1.4	Pandora Consolidated reconstruction . . . . .	59
4.5.2	Track/Shower Separation . . . . .	60
4.5.3	Performance . . . . .	61
<b>Chapter 5 Characterisation of EM Showers</b>		<b>64</b>
5.1	Datasets . . . . .	65
5.2	Space Charge Effect . . . . .	67
5.2.1	Event Selection . . . . .	68
5.3	Shower Principal Components Analysis . . . . .	70
5.3.1	Pandora's PCA . . . . .	71
5.3.2	Recursive Weighted PCA . . . . .	75
5.4	Shower Direction . . . . .	76
5.5	Fractional Explained Variance . . . . .	79
5.6	Shower Length . . . . .	81
5.7	Shower Energy Estimation . . . . .	84
5.7.1	Energy Estimation Corrections . . . . .	87
5.7.1.1	Deposition Correction . . . . .	87
5.7.1.2	Missed Hits Correction . . . . .	88
5.7.1.3	No Hits Correction . . . . .	89
5.7.1.4	Contamination Correction . . . . .	90
5.7.1.5	Hit Energy Correction . . . . .	91

5.7.1.6	Total Correction . . . . .	92
5.8	Shower $dE/dx$ . . . . .	96
5.9	Shower Longitudinal Profile . . . . .	97
5.9.1	Cascade Start . . . . .	99
5.10	Shower Transverse Profile . . . . .	102
5.11	Summary . . . . .	104
<b>Chapter 6</b>	<b>Proton Decay</b>	<b>105</b>
6.1	Event Reconstruction and Classification . . . . .	109
<b>Chapter 7</b>	<b>Summary</b>	<b>116</b>

# List of Tables

3.1	Single-phase LArTPC Specifications [46]. . . . .	37
3.2	Bias voltages for the wire planes of an APA [46]. . . . .	38
3.3	Number of beam triggers recorded at different energies provided by the test beam [54]. . . . .	48
4.1	Important LArSoft data products, and their descriptions. . . . .	55
5.1	Beam data run numbers. . . . .	66
5.2	ProtoDUNE-SP beam simulation MC datasets. . . . .	66
5.3	Number of 2D hit thresholds for PDSP-MC and PDSP-BD datasets. . . . .	69
5.4	Fit parameters for Fig. 5.28 . . . . .	96
6.1	The six most probable decay modes for a $K^+$ [34]. . . . .	106
6.2	Expected rate of atmospheric neutrino interactions in $^{40}\text{Ar}$ for a 10 kt · year exposure (not including oscillations) [28]. . . . .	108
6.3	Features used in the BDT for separating nucleon decay signal, $p \rightarrow K^+\bar{\nu}$ , from atmospheric neutrino background. . . . .	111

# List of Figures

2.1	The Standard Solar Model, showing the flux of neutrinos produced by various processes in the Sun, and their energies [16]. . . . .	21
2.2	Top left: NC scattering of a neutrino off of a nucleon (N). Top right: Quasi-elastic CC scattering of a neutrino with a nucleon, producing a lepton and change neutrons to protons and vice-versa. Bottom left: An example of one possible CC resonance scattering where a neutrino scattering off a proton produces a lepton, and a $\Delta^{++}$ resonance which decays to a proton and a charged pion. Bottom right: A deep inelastic scattering interaction. A high energy neutrino interacting with a proton produces a lepton and a jet of hadronic particles. . . . .	23
2.3	The effect that different values of $\delta_{CP}$ has on the oscillation probability of $\nu_\mu \rightarrow \nu_e$ (left) and $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ (right) with a baseline of 1300 km, and normal mass ordering. Figure taken from [29] . . . . .	28
2.4	The current latest results from the T2K Collaboration. CP conservation, $\delta_{CP} = 0, \pi$ , is ruled out to the 95% confidence level, but are not excluded by the 99.73% confidence level. Figure <b>a</b> shows the 68.7% confidence level for $\delta_{CP}$ vs $\sin^2 \theta_{13}$ as measured by T2K, with the star representing the best fit point for T2K + reactors for the normal mass ordering (Sec.2.6). Figure <b>b</b> shows the 68.27% and 99.73% confidence intervals for $\delta_{CP}$ vs $\sin^2 \theta_{23}$ as measured by T2K. The colour scheme represents the value of negative two times the logarithm of the likelihood for each parameter value. Figure <b>c</b> shows the 68.27% (box) and 99.73% (whiskers) confidence intervals for the normal mass ordering, and the 68.27% confidence interval of the inverted mass ordering, both as measured by T2k [30]. . . . .	29
2.5	The two possible neutrino mass orderings - normal or inverted. Each mass eigenstate is also split into its fractional flavour content, with the $\delta_{CP}$ varied from 0 to $\pi$ (bottom of mass bar to top of mass bar). Figure taken from [28] . . . . .	30

3.1	An overview of the planned LBNF and DUNE projects, and the two locations for the project - Fermilab, the near site, and SURF, the far site [29]. . . . .	32
3.2	The caverns at SURF. The large red boxes illustrate the cryostats of the first two far detector modules. The central cavern, housing the DAQ system and cryogenics, is located between the two cryostat caverns. The large vertical shaft is the access elevator [29]. . . . .	33
3.3	A graphic showing the size of the cryostat housing for a LArTPC FD module [38]. . . . .	34
3.4	A cartoon illustrating the principals of how a single-phase LArTPC operates. Electrons are liberated from their parent nuclei as charge particles traverse the LAr volume. These ionisation electrons are then drifted, under a large electric field, towards a set of instrumented wire planes where the charge information is read out [46]. . . . .	35
3.5	A schematic of the layout of a SP LArTPC. The anode (A) will be 2 APAs high, and the cathode (C) 6 CPA units high. The field cage can also be seen covering the top, bottom and end of the LArTPC [46].	36
3.6	A face on schematic of an APA. The blue boxes on the right hand edge are the readout electronics. The green and magenta lines represent the two induction planes, and the blue lines represent the collection and shielding wires [46]. . . . .	38
3.7	A side on schematic of the top of an APA frame, showing the order of the G, U, V, and X wire planes [46]. . . . .	39
3.8	The operational principle of the dual-phase design. The ionisation electrons are drifted vertically to be extracted and multiple in a gaseous phase to increase signal to noise ratio [29]. . . . .	41
3.9	A side view of the beamline, target, and ND complexes. The proton beam comes from the right, provided by the Fermilab Main Injector, before being taken up through an embankment and directed towards the target. After striking the target, decay products from proton interactions are directed down a decay pipe, where they will decay into neutrinos before continuing on towards the ND and FD facilities [29]. . . . .	42

3.10	A cartoon showing the layout of the target hall and decay pipe at the near site. The proton beam comes in from the left before hitting the graphite target, creating a secondary set of charged particles which are directed towards the decay pipe by the focusing horns located immediately downstream of the target. The neutrinos produced in the decay pipe then travel on to the far site [50]. . . . .	43
3.11	The layout of the DUNE ND complex. ArgonCube (pink box) can be seen to be the most upstream detector, SAND (red box) the most downstream, and the MPD (green box) sandwiched between the two. In the top down view you can also see the area where the DUNE-PRISM mechanism can move ArgonCube and the MPD off-axis [29].	45
3.12	The EHN1 building extension at CERN's Preveessin site - home to the two ProtoDUNEs. ProtoDUNE-SP can be seen in the foreground, while ProtoDUNE-DP is in the background. Located between the two detectors, surrounded by the large concrete blocks, is the end of the H4 beamline pointing towards ProtoDUNE-SP. . . . .	46
3.13	Diagram showing a top down view of the detector with the upstream, midstream, downstream, beam left, and beam right areas labelled. The $x-z$ plane origin is located on the upstream edge of the CPA wall, with $x$ increasing towards the beam left side of the detector. The beam is shown as entering the detector with a slight offset towards the beam right side, and with an angle of $\sim 13^\circ$ to the central CPA wall.	49
3.14	Diagram showing the upstream face of the detector, labelling the TPC and cryostat faces of the APAs. The entry point of the beam is also shown slightly offset to the beam right side, and about three-quarters of the way up the upstream detector face. The $x-y$ plane origin is located at the bottom of the CPA wall, with $y$ increasing vertically, and $x$ increasing towards the beam left side. . . . .	49
3.15	An example event display as seen on the online monitor's webpage. This particular display shows the area of the TPC where particles from the test beam enter the detector. Collaborators find such a display useful as they can easily see the quality of data being recorded in the most important area of the detector. The colour scheme indicates collected ionisation charge. . . . .	51
3.16	An example of what can happen to the event display if the pedestal values for the online monitoring system are not updated regularly. Each vertical line of colour represents a single wire. The colour scheme indicates collected ionisation charge. . . . .	51

4.1	An example of the input hits from the collection planes of the APAs that get passed to Pandora. The hits in blue, red, and black represent hits caused by the test beam particle, test beam halo particles, cosmic-ray particles respectively [64]. . . . .	56
4.2	An example of a reconstructed test beam particle. The colours of particles represent the hierarchy, with red being the initial incident test beam particle, blue the daughters of the test beam particle, and green the granddaughters [64]. . . . .	58
4.3	An example of how the stitching works for CPA crossing cosmic-rays. The red and blue tracks are the initial reconstructed particles, the green arrows show the direction of the drift time offset, and the black track represents the stitched together cosmic-ray [64]. . . . .	58
4.4	The Pandora consolidated reconstruction. The consolidated reconstruction first removes clear cosmic-ray particles by reconstructing all input hits under the Pandora Cosmic hypothesis. The non clear cosmic-ray hits are separated into 3D slices before being reconstructed under both the Pandora Test Beam and Pandora Cosmic hypotheses. The best outcome of which is decided by a boosted decision tree (BDT) [64]. . . . .	59
4.5	A comparison of Pandora’s original cut based approach (red) with the newer BDT approach (blue) to track separation. The BDT approach clearly has a much better accuracy at lower numbers of hits. . . . .	61
4.6	A breakdown of Pandora’s reconstruction efficiencies, for all beam particles, with increasing simulation complexity. Removing cosmic-ray muons and beam halo particles shows the scale of effect they have on decreasing the reconstruction efficiency. Across all momenta, the cosmic ray-muons decrease efficiency by a constant amount, while the beam halo particles have a much more pronounced effect at the higher momentum [64]. . . . .	62
4.7	Pandora’s reconstruction efficiency against true particle momenta for real (red) and simulated (black) data. The simulated data’s efficiency differs from the blue line in Fig. 4.6 because of the different definition used for efficiency as described in the final paragraph of Sec. 4.5.3 [64]. . . . .	63
5.1	An example diagram of an electromagnetic shower induced by an incoming electron. After traversing one radiation length ( $X_0$ ) of a medium, indicated here by the blue hatched region, the electron radiates a bremsstrahlung photon. This process repeats until the EM shower dies out due to energy losses through ionisation. . . . .	65

5.2	The uncorrected and corrected positions for 3D space points in the PDSP-MC dataset. The impact of SCE is most notable in the z-positions due to the squeezing of the drift charge. . . . .	67
5.3	Number of reconstructed 2D hits <b>(a)</b> and length <b>(b)</b> for 1 GeV/c (red) and 6 GeV/c (azure) single particle gun MC events. . . . .	68
5.4	Length distribution before <b>(a)</b> and after <b>(b)</b> applying a 2D hit threshold to events in the PDSP-MC and PDSP-BD datasets. The fits in <b>(b)</b> are computed for the PDSP-MC distributions, and use the skewed Gaussian function of Eqn. 5.2. . . . .	70
5.5	An example of the resultant eigenvectors, orange and maroon arrows, for a distribution of points, with the length of the arrow being three times the square root of the eigenvector's corresponding eigenvalue. . . . .	71
5.6	An example of a reconstructed 3D shower with its PCA primary (red), secondary (green), and tertiary (blue) axes overlaid. The secondary and tertiary axes do not appear orthogonal to the primary axis because of the different scale plot axes and the perspective of the image. . . . .	73
5.7	A shower that has had its 3D space points projected into the PCA axes coordinate system. The top plot shows the secondary axis projections against the primary projections, and the bottom plot shows the tertiary axis projections against the primary projections. . . . .	74
5.8	An example of a reconstructed 3D shower with its unweighted PCA and recursive weighted PCA axes overlaid. The unweighted PCA axes are the dotted lines, and the recursive weighted PCA axes are the solid lines. . . . .	76
5.9	A shower that has had its 3D space points projected into the PCA axes coordinate system. . . . .	77
5.10	Plots comparing the angle between the direction of the Monte Carlo true particle and the primary axis from the unweighted PCA (blue) and the recursive weighted PCA (orange), for electrons from the single particle gun MC dataset. . . . .	78
5.11	Comparison of the angle between true and reconstructed direction using the recursive weighted PCA on PDSP-MC and PDSP-BD. The truth value for PDSP-BD is taken from the final beam monitors. . . . .	79
5.12	Box plots for the fractional explained variance of the three axes from the unweighted PCAs at each momentum interval for the PDSP-MC (orange) and PDSP-BD (blue) datasets. The decrease in the values for the primary axis (blue boxes) shows that EM showers of higher momentum are more spread out laterally relative to their length. . . . .	80

5.13	Box plots for the fractional explained variance of the three axes from the recursive weighted PCA at each momentum interval for the PDSP-MC (orange) and PDSP-BD (blue) datasets. . . . .	81
5.14	A 1 GeV/c single particle gun MC shower that has had its 3D space points projected into the PCA axes coordinate system. Overlaid are the two different definitions of length for the shower, with the eigenvalue length being about 50 cm longer than the projection length.	82
5.15	Comparison of eigenvalue length and max projection length for single particle gun MC electrons. . . . .	83
5.16	Comparison of the projection lengths for PDSP-MC (orange) and PDSP-BD (blue). . . . .	83
5.17	Recombination factors for simulated 1 GeV/c electrons. . . . .	84
5.18	True momentum for PDSP-MC (histograms), and beam momentum for PDSP-BD (points). The measured beam momenta for data appears to be more accurately reconstructed than anticipated by the MC, and this has been corrected for in the next MC production. . . . .	85
5.19	Energy estimations using Eqn. 5.10 for electrons generated at different momenta. It can be seen that there is a systematic underestimation across all momenta. . . . .	86
5.20	The ratio of the deposition correction to the mean expected energy for different momenta. . . . .	88
5.21	The ratio of the missed hits correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2. . . . .	89
5.22	The ratio of the no hits correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2. . . . .	90
5.23	The ratio of the no hits correction to the mean expected energy for different momenta. . . . .	91
5.24	The ratio of the hit energy correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2. . . . .	92
5.25	The ratio of the total correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2. . . . .	93
5.26	The corrected energy estimation for the PDSP-MC dataset, using an event-by-event correction. . . . .	94
5.27	Residuals for corrected energy estimations of the PDSP-BD dataset. The correction factor is calculated as the peak values from Fig. 5.25. The fits are Gaussians as per Eqn. 5.1. . . . .	95

5.28	Energy resolution for the PDSP-BD dataset. The left plot shows good agreement between the corrected estimated energy and the true energy. The right plots shows that, as expected, the resolution improves as the expected energy increases. . . . .	96
5.29	Comparison of measured $dE/dx$ for the first 5 cm of electron showers in the PDSP-MC (histograms) and the PDSP-BD (points) datasets. Each distribution is area normalised. . . . .	97
5.30	Energy deposition per radiation length for PDSP-BD dataset events.	98
5.31	Averages of the binned cross-sectional areas for the PDSP-BD dataset (dashed lines), and the PDSP-MC dataset (shaded regions). As seen in Fig. 5.12, the PDSP-BD EM showers appear to be wider than the EM showers in the MC. . . . .	99
5.32	An example plot of the binning method used to identify the start of an EM shower's cascading section. The 3D space points have been rotated onto the PCA axes, and placed into 1 cm bins (red lines). The blue shaded areas then represent the standard deviations for the 3D Space points in that bin, along each axis. The bottom plot shows the product of the standard deviations for each bin. The green dotted line shows the bin that went above the threshold selection for the start of the EM cascade. . . . .	101
5.33	The start positions, in radiation lengths, of the cascading part of the EM showers from the PDSP-BD dataset. . . . .	102
5.34	Transverse profiles for EM showers in the PDSP-MC and PDSP-BD datasets. . . . .	103
6.1	Summary of the 90% confidence level lower limits from Super-K for $p \rightarrow K^+\bar{\nu}$ and $p \rightarrow e^+\pi^0$ , and predicted lifetime from various GUT models. Adapted from [28]. . . . .	106
6.2	DUNE (cyan) and Hyper-Kamiokande's (red) expected sensitivity to $p \rightarrow e^+\pi^0$ (left) and $p \rightarrow K^+\bar{\nu}$ (right) [83]. . . . .	107
6.3	Images for a $p \rightarrow K^+\bar{\nu}$ signal event (top), and a CC $\nu_\mu$ background event (bottom). . . . .	108
6.4	True kaon kinetic energy for $p \rightarrow K^+\bar{\nu}$ . . . . .	109
6.5	Efficiency for reconstructed kaons using Pandora (top). . . . .	110
6.6	Two example BDT features. . . . .	111
6.7	Correlation matrix for the features of the BDT. The colour scheme denotes the level of correlation between two variables. . . . .	112
6.8	BDT response for $p \rightarrow K^+\bar{\nu}$ for background (blue hatch) and signal (red), with linear (a) and log (b) axes. . . . .	113

6.9	Background rejection (left) and signal efficiency (right) for Poisson fluctuated BDT responses to the non-training sample of events. . . .	114
-----	---	-----

# Acknowledgments

I'd like to thank Professor Gary Barker, Dr John Marshall, and Dr Nick Grant for all the guidance, encouragement, and general help they have provided with all aspects of my work.

To all the members of the EPP group at Warwick, and the members of the DUNE Collaboration I've met along the way, I thank you for providing many needed laughs and memorable moments throughout my PhD. I'd like to make a special mention of Sammy Valder and Ed Millard. You have both provided an immeasurable amount of support during the PhD, without out which I am not sure I'd have made it all the way.

For my father and sister, and every ounce of support they have provided, and their unwavering certainty that I'd make it through.

Mark. You have dealt with the cycle of me enjoying my PhD, to wanting nothing more than to quit so many times over its course. Every time, you have been there providing the love and support I have so sorely needed. Words cannot describe how thankful I am for, and how much I appreciate, everything you have put up with, and done for me, over the past few years.

I would also like to thank Frank Herbert for writing the 1965 science fiction book, Dune. Not only has this proved to be a fantastic sleeping aide during various points of my PhD, it has also provided some fantastic, and surprisingly relevant, quotes for the chapters of my thesis.

# Declarations

The material presented in this document has not been published or submitted for examination at another institute. It represents the author's own work unless stated otherwise, complete with references.

Chapter 2 provides a background and overview of the current state of neutrino physics, and has been composed from scientific journals, textbooks, non-fiction general interest books, and other publicly available media formats. Chapter 3 provides an overview of the technical nature of the DUNE detectors and their associated technologies, and is derived from information available in the collaboration's technical design reports.

Chapter 4 describes the reconstruction methods of Pandora, and the LArSoft and *art* frameworks. All are produced as collaborative works for a number of different particle experiments, and extensively used by the DUNE collaboration. While the author has not directly contributed to the writing of software for the Pandora group, the group is located at Warwick and the author has worked very closely with, and often given feedback to, the Pandora group.

The work in chapter 5 uses the software described in chapter 4, but all event selection and analysis software and techniques have been developed by the author. The author also contributed to the collection of the data in the PDSP-BD dataset while based at CERN, and helped develop monitoring software for data taking. Chapter 6 then expands on the selection and analysis software written for chapter 5.

# Abstract

DUNE is a future on-axis long-baseline neutrino oscillation experiment under construction in South Dakota, USA. Preceding it are two prototype experiments currently located at CERN - ProtoDUNE-SP and ProtoDUNE-DP. During the final months of the LHC 2018 running period, a mixed particle beam was provided to the ProtoDUNE-SP detector, where over 4 million beam trigger events were recorded. This data will be used to benchmark many aspects of the detector technology, and the results from it are vital to the success of DUNE's full scale detector.

Analysis of electron induced electromagnetic showers using ProtoDUNE-SP's 2018 beam data and 'Production 2' Monte Carlo has been performed. A number of modifications to existing reconstructed variables have been suggested to better meet the requirements of their intended use cases. Energy estimations have been shown to provide resolutions within the expected range of a calorimeter, and  $dE/dx$  measurements align with expected values also.

The Pandora reconstruction software has been applied to the nucleon decay search through the  $p \rightarrow K^+ \bar{\nu}$  channel, and a BDT has been implemented with a sensitivity of  $8.0 \times 10^{33}$  years.

# Acronyms

**ADC** Analogue to digital converter.

**APA** Anode plane assembly.

**Ar** Argon.

**B** Boron.

**BDT** Boosted decision tree.

**Be** Beryllium.

**BSM** Beyond standard model.

**CC** Charged current.

**Cd** Cadmium.

**CKM** Cabibbo, Kobayashi, Maskawa.

**Cl** Chlorine.

**CP** Charge-Parity.

**CPA** Cathode plane assembly.

**DAQ** Data acquisition.

**DoE** Department of energy.

**DP** Dual-phase.

**DUNE** Deep underground neutrino experiment.

**ECAL** Electromagnetic calorimeter.

**EM** Electromagnetic.

**FD** Far detector.

**FHiCL** Fermilab hierarchical configuration language.

**FSI** Final state interactions.

**GeV** Gigaelectronvolt.

**GUT** Grand unified theory.

**He** Helium.

**HPgTPC** High-pressure gaseous argon time projection chamber.

**kV** Kilovolts.

**LAr** Liquid argon.

**LArSoft** Liquid argon software.

**LArTPC** Liquid argon time projection chamber.

**LBNF** Long baseline neutrino facility.

**MC** Monte carlo.

**MPa** Megapascal.

**MPD** Multi-purpose detector.

**MSW** Mikheyev, Smirnov, Wolfenstein.

**MW** Megawatts.

**NC** Neutral current.

**ND** Near detector.

**PCA** Principal components analysis.

**PCB** Printed circuit board.

**PDSP-MC** ProtoDUNE-SP monte carlo dataset.

**PDSP-BD** ProtoDUNE-SP beam dataset.

**PFParticle** Particle flow particle.

**PMA** Projection matching algorithm.

**PMNS** Pontecorvo, Maki, Nakagawa, Sakata.

**ppm** Parts per million.

**ppt** Parts per trillion.

**PRISM** Precision reaction-independent spectrum measurement.

**QED** Quantum electrodynamic.

**SAND** System for on-axis neutrino detection.

**SCE** Space charge effect.

**SDK** Software development kit.

**SM** Standard model.

**SNO** Sudbury neutrino observatory.

**SP** Single-phase.

**SPS** Super proton synchrotron.

**SSM** Standard solar model.

**SURF** Sanford underground research facility.

**T2K** Tokai to Kamioka.

**TDR** Technical design report.

**TPC** Time projection chamber.

**VUV** Very ultraviolet.

# 1

## Introduction

---

*“Behold, as a wild ass in the desert, go I forth to my work.”*

---

**Frank Herbert - *Dune* - 1965**

**H**igh energy particle physics sits at the very forefront of modern day physics, consisting of experiments worth multiple billions of pounds, driven by complex collaborations of scientists numbering in the thousands. Chief among which are the neutrino detectors – such as the current T2K, and future DUNE and Hyper-K – poised to make the next<sup>1</sup> large discovery of physics beyond the Standard Model.

From the emergence of particle accelerators in the second quarter of the 20<sup>th</sup> century, the observance of the composite nature of the proton in the 1960s [1, 2], and the discovery of the Higgs Boson at CERN in 2012 [3, 4], particle physics has been moving forward at a phenomenal pace. The crowning achievement of the field however, was the formalisation of the Standard Model, and its predictive capabilities, in the 1970s. Despite its power, and the experimental confirmation of its predictions, the Standard Model has many shortcomings. The model provides no quantised description of General Relativity, nor a particle that explains the observations of dark matter and dark energy in the universe. The Standard Model also does not predict the existence of neutrino mass, and subsequently their oscillations; an observed phenomenon [5].

---

<sup>1</sup>In the opinion of the author.

The upcoming DUNE collaboration's goals include furthering the understanding of the beyond the Standard Model physics observed by previous, and currently running, neutrino experiments around the world. To do this it will deploy a 40 kt state of the art liquid argon time projection chamber - a technology with the ability to image particle interactions in phenomenal detail. This leap forward in detector abilities will pave the way for novel methods of data analysis, such as utilising the full potential of modern deep learning methodologies.

Neutrino theory, DUNE's goals and detector design, and data reconstruction techniques are discussed in Chapters 2-4.

Chapter 5 of this thesis investigates the reconstruction ability of electron induced electromagnetic showers. Using data from the ProtoDUNE Single-Phase run I dataset, as well as DUNE collaboration 'Production 2' Monte Carlo, a number of key variables are assessed:

- Principal components analysis.
- Energy estimation.
- Measurement of the  $dE/dx$  at the start of a shower.
- Longitudinal and transverse profiles.
- A method for identifying the start position of a shower's cascade.

In Chapter 6, a sensitivity study on the lower lifetime limit per branching fraction for the  $p \rightarrow K^+ \bar{\nu}$  is done using the Pandora reconstruction software, and a BDT for signal-background separation.



## Neutrino Theory

---

---

*“Science is made up of so many things that appear obvious after they are explained.”*

---

**Frank Herbert - Dune - 1965**

### 2.1 Neutrino History

**D**uring the first two decades of the 20<sup>th</sup> century, a number of experiments were being conducted by numerous scientists on the recently discovered phenomenon of radioactivity. Within the first few years of the century, it had been determined that there were three distinct types of radiation; alpha, beta, and gamma. Results from the experiments showed that alpha and gamma radiation both conserved energy, as observed by having a narrow energy distribution. However, in 1914, James Chadwick showed that the energy spectrum of the electron emitted in the beta decay process was continuous [6]; a result not expected based on the nuclear models of the time<sup>1</sup>, nor if energy conservation was to be universal.

Thus, for many years scientists were left rather nonplussed by this observation, and it was not until 1930 when Wolfgang Pauli suggested “einen verzweifelten Ausweg”: a desperate solution [7]. Pauli suggested that there must be a third constituent to the nucleus in the form of a small neutral particle. Which he named the neutron - later

---

<sup>1</sup>That being atomic nuclei consisted of only electrons and protons, and beta decay the emission of a sole electron.

renamed the neutrino following Chadwick’s discovery of the large<sup>2</sup> nucleon, which Chadwick also named the neutron. In 1934 Fermi formalised a neutrino-inclusive beta decay theory, describing the reactions [8–10]:

$$n \rightarrow p + e^{-} + \bar{\nu}_e \quad (2.1)$$

$$p \rightarrow n + e^{+} + \nu_e \quad (2.2)$$

where  $n$  and  $p$  are neutrons and protons,  $e^{-/+}$  an electron/positron, and  $\nu_e/\bar{\nu}_e$  the neutrino/antineutrino. Note that Eqn. 2.2 can only occur inside a nucleus.

However, it was not until 1956 [8, 11–13] that the existence of Pauli’s proposed neutrino was confirmed experimentally by the Cowan-Reines neutrino experiment. The experiment was initially designed to use the inverse beta decay process,

$$\bar{\nu}_e + p \rightarrow n + e^{+} \quad (2.3)$$

to detect antineutrinos emitted from a nearby nuclear explosion. The detector would be filled with a liquid scintillator that would produce flashes of light as the positron emitted by the process of eqn. 2.3 annihilates with a nearby electron. These light flashes could be detected by photomultiplier tubes on the detector’s edge, and their intensity used to calculate the energy of the initial two gamma photons. Despite receiving approval for their rather radical experiment, Cowan and Reines were persuaded to attempt the detection of antineutrinos emitted from the reactions inside a nuclear reactor instead. The change in direction came following a realisation that it could be possible to also detect the neutron emitted in eqn. 2.3. By doping the target material with a cadmium salt, it would be possible to quickly recapture the neutron released by the process in eqn. 2.3. As the neutron travels through the target material, it slows down and eventually is absorbed by a cadmium atom in the process:



As the excited cadmium atom relaxes, it emits a secondary burst of photons. This delayed secondary burst of photons, in coincidence with the prompt photons from the positron annihilating, can be used to verify the observation of an inverse beta decay process, and as such show the existence of the antineutrino.

Amidst the tribulations surrounding the missing energy in beta-decay, other

---

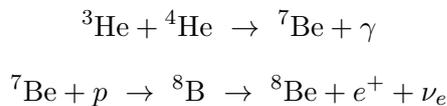
<sup>2</sup>Relative to the size suggested by Pauli.

scientists had their focus set on the problem surrounding the age of the Sun. What physical process was allowing the Sun to burn so ferociously? Anaxagoras had first theorised the Sun was a ‘mass of fiery metal’ [14] in the 5<sup>th</sup> century BCE, an idea that went largely unquestioned until the industrial revolution. During this time, as the understanding of thermodynamics increased, it became apparent that no physical process could account for a large ball of metal remaining so hot for the known history of humanity. As such, other explanations for the process that fuelled the Sun were floated about - such as the idea that the Sun was collapsing inwards, an idea from Lord Kelvin - but none provided processes that could last the timescales suggested by geologists, or required by Darwin’s theory of evolution. It was not until Rutherford suggested that radioactivity could be the source of the Sun’s energy, did the field pick up pace. This suggestion eventually led to Hans Bethe’s theorisation of the carbon-nitrogen-oxygen cycle, and upon further investigation, the proton-proton chain. During both of these processes neutrinos are emitted - in vast quantities.

Ray Davis led the first large scale experiment - a detector consisting of 400 000 l of tetrachloroethylene, C<sub>2</sub>Cl<sub>4</sub><sup>3</sup>, 1 480 m underground at the Homestake mine in Lead, South Dakota - to measure this neutrino flux from the Sun. John Bahcall noted in his 1964 Phys. Rev. Letter [15], that the 5.1 MeV excited state of argon is superallowed as it is the analogue for the ground state of chlorine. Thus, the experiment would rely on the capture of neutrinos by chlorine atoms in the process



The neutrino flux measured by the experiment would be that of boron-8, shown by the curve labelled <sup>8</sup>B in Fig. 2.1. The boron-8 neutrino flux is resultant from the process



where the initial <sup>3</sup>He and <sup>4</sup>He are products of the proton-proton chain. However, during the 20 years that the experiment was running, the number of detected neutrinos was falling short of the amount predicted by the Standard Solar Model (SSM), leading to the anomaly known as the Solar Neutrino Problem.

Before Davis had started the Homestake experiment, another experiment at the Brookhaven National Laboratory had discovered the existence of the muon neutrino

---

<sup>3</sup>Also a common cleaning fluid.

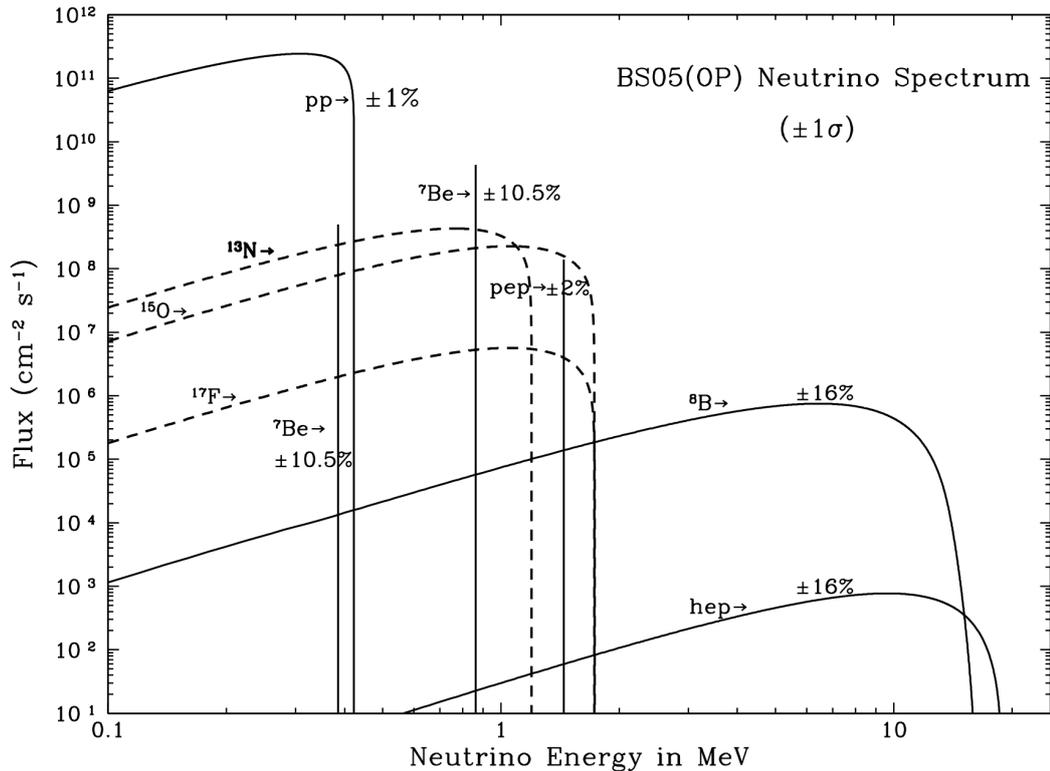


Figure 2.1: The Standard Solar Model, showing the flux of neutrinos produced by various processes in the Sun, and their energies [16].

using a high energy neutrino beam produced by the decay [17]

$$\pi^\pm \rightarrow \mu^\pm + \nu/\bar{\nu} \quad (2.6)$$

while the flavour of the neutrino is not directly measured, the experiment found that the neutrinos produced by the decay in Eqn. 2.6 always produced muons when interacting in the detector. This led to the assertion that the neutrinos produced in the beta decay must be different to the neutrinos produced by pion decay. The discovery of the distinct electron and muon neutrinos led Pontecorvo, Maki, Nakagawa, and Sakata to postulate the idea of neutrino mixing and oscillations in the vacuum [18–20].

Over the 30 years following the start of the Homestake experiment, a number of other experiments also started taking data on neutrino fluxes, and also found deficiencies in the number of neutrinos detected. The Kamiokande and Super-Kamiokande experiments, both water Cherenkov detectors, observed deficits in the number of solar neutrinos consistent with the results from the Homestake experiment. Both Kamiokande and Super-Kamiokande detected solar neutrinos through the

elastic scattering process:

$$\nu + e^- \rightarrow \nu + e^- \quad (2.7)$$

where  $\nu$  is left flavourless as all neutrino flavours can scatter in this way [21, 22]. However, as electron neutrinos have a significantly larger cross-section as they are able to undergo the process described by Eqn. 2.7 through both neutral and charged current interactions. The neutrino flux measured by Kamiokande and Super-Kamiokande is therefore dominated by the electron neutrino.

Like the Homestake experiment, Kamiokande and Super-Kamiokande were only able to measure the flux of the  $^8\text{B}$  neutrinos. It was not until the 1990's that the proton-proton chain neutrino flux was able to be investigated using two gallium based experiments - GALLEX/GNO and SAGE. Both experiments made use of the process:

$${}^{71}\text{Ga} + \nu_e \rightarrow {}^{71}\text{Ge} + e^- \quad (2.8)$$

The experiments produced compatible results, again showing a deficit in the number of expected neutrinos [23–26].

So far all experimental evidence was pointing towards two possibilities: i) The SSM was incorrect, ii) neutrinos had properties that were not known about. To solve this conundrum, there needed to be an experiment that could measure all neutrino fluxes, independent of flavour and solar model. To do this the Sudbury Neutrino Observatory (SNO) was made. SNO, like Super-Kamiokande, would be a water Cherenkov detector, but would use heavy water ( $\text{D}_2\text{O}$ ) as opposed to normal water. The advantage of using  $\text{D}_2\text{O}$  is the sensitivity to two extra processes alongside Eqn. 2.7:

$$\nu_e + d \rightarrow e^- + p + p \quad (2.9)$$

$$\nu + d \rightarrow \nu + p + n \quad (2.10)$$

Eqns. 2.9 & 2.10 describe the charged and neutral weak current processes respectively. The charged current data, sensitive to only electron neutrinos, once again showed a deficit in the expected value of the flux from the Sun. However, the neutral current data showed a flux consistent with the SSM [5] - the first direct evidence that the electron neutrinos could be oscillating into other flavours on their journey from the Sun to the Earth.

## 2.2 Neutrino Interactions

It is only possible for neutrinos to interact through the weak force, as they contain neither electric charge nor colour charge. As such, it is possible for neutrino interactions to proceed through either a charged current (CC) interaction, mediated through the W boson, or a neutral current (NC) interaction, mediated through the Z boson.

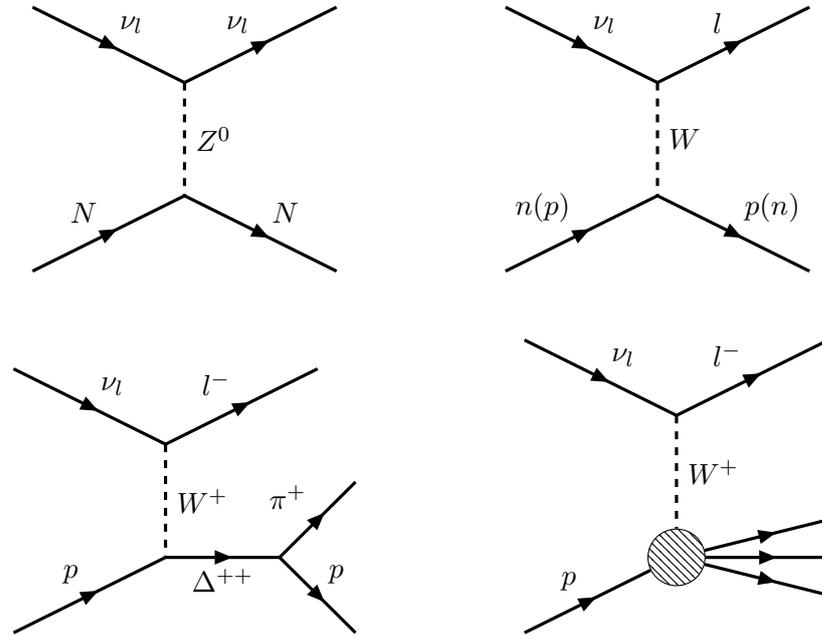


Figure 2.2: Top left: NC scattering of a neutrino off of a nucleon (N). Top right: Quasi-elastic CC scattering of a neutrino with a nucleon, producing a lepton and change neutrons to protons and vice-versa. Bottom left: An example of one possible CC resonance scattering where a neutrino scattering off a proton produces a lepton, and a  $\Delta^{++}$  resonance which decays to a proton and a charged pion. Bottom right: A deep inelastic scattering interaction. A high energy neutrino interacting with a proton produces a lepton and a jet of hadronic particles.

At the lowest energy scale, all neutrino flavours are able to elastically scatter off of nucleons through the NC interaction process:  $\nu + N \rightarrow \nu + N$ . If the neutrino has enough energy to create its corresponding charged lepton's mass, then it is able to scatter 'quasi-elastically' off a nucleon in the CC processes:  $\nu + n \rightarrow l^- + p$  and  $\bar{\nu} + p \rightarrow l^+ + n$ . Because electrons exist in normal matter, it is also possible for electron neutrinos to scatter off electrons through the process:  $\nu_e + e^- \rightarrow e^- + \nu_e$ .

With higher energies, it is possible that a neutrino interacting with a nucleon causes

the nucleon to go into a baryonic resonance, which then decays back into the nucleon plus another charged particle. At the highest energies, it is possible for the neutrino to interact directly with the quarks inside the nucleons, producing a jet of hadronic particles.

## 2.3 Oscillations

The experimental observation of neutrino oscillations gives rise to the need for neutrinos to have non-zero masses. As briefly mentioned in Chapter 1, this is one of the shortcomings of the Standard Model (SM). The SM does not explicitly require that neutrinos have mass, however, for neutrino oscillations to take place, the mass and flavour eigenstates of the neutrinos must be distinct. It is then possible to express the traditional flavour eigenstates - electron ( $\nu_e$ ), muon ( $\nu_\mu$ ), tau ( $\nu_\tau$ ) - as a superposition of the mass eigenstates

$$|\nu_\alpha\rangle = \sum_i^N U_{\alpha i} |\nu_i\rangle \quad (2.11)$$

where  $|\nu_\alpha\rangle$  are the flavour eigenstates,  $|\nu_i\rangle$  the mass eigenstates, and  $U_{\alpha i}$  a unitary mixing matrix<sup>4</sup>. Under current experimental results  $N = 3$ , but this is not necessarily limited by theory. The mixing matrix,  $U_{\alpha i}$ , is analogous to the CKM<sup>5</sup> mixing matrix of the quark sector, and is known as the PMNS<sup>6</sup> mixing matrix ( $U_{\text{PMNS}}$ ). An  $n \times n$  unitary matrix consists of  $n^2$  parameters, but by fixing the relative phases between the six neutrino states, it can be parametrised using three weak mixing angles,  $\theta_{12}$ ,  $\theta_{23}$ ,  $\theta_{13}$ , and one CP-violating phase,  $\delta_{\text{CP}}$ .

The  $U_{\text{PMNS}}$  matrix can be written as [27]:

$$U_{\text{PMNS}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{\text{CP}}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{\text{CP}}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.12)$$

where  $c_{ij} = \cos\theta_{ij}$  and  $s_{ij} = \sin\theta_{ij}$ . The left most matrix, containing only  $\theta_{23}$ , describes the oscillations of  $\nu_\mu \rightarrow \nu_\tau$  - the atmospheric oscillations. The central matrix, containing  $\theta_{13}$  and  $\delta_{\text{CP}}$ , describes  $\nu_\mu \rightarrow \nu_e$  oscillations - the reactor oscillations. The final matrix governs the oscillations  $\nu_e \rightarrow \nu_\mu$  and  $\nu_e \rightarrow \nu_\tau$  - the solar oscillations.

---

<sup>4</sup> $U^\dagger U = I$ . Where  $I$  is the identity matrix.

<sup>5</sup>Cabibbo-Kobayashi-Maskawa

<sup>6</sup>Pontecorvo-Maki-Nakagawa-Sakata

### 2.3.1 Oscillation Formalism

The following formalism primarily draws from *Neutrino Physics Second Edition* by K. Zuber [27].

Assuming a plane wave solution to the time dependant Schrödinger equation, the mass eigenstates  $|\nu_i\rangle$  have a time dependence

$$|\nu_i(\vec{x}, t)\rangle = e^{-i\phi_i} |\nu_i(\vec{0}, 0)\rangle \quad (2.13)$$

where  $\phi_i = E_i t - \vec{p} \cdot \vec{x}$ . Given that at a some time  $t$  and position  $\vec{x}$  the neutrino flavour state  $|\nu_\beta(\vec{x}, t)\rangle$  will be measured, it can be written as

$$|\nu_\beta(\vec{x}, t)\rangle = \sum_i U_{\beta i} |\nu_i(\vec{x}, t)\rangle \quad (2.14)$$

and combined with eqn. 2.13 to obtain

$$|\nu_\beta(\vec{x}, t)\rangle = \sum_i U_{\beta i} e^{-i\phi_i} |\nu_i(\vec{0}, 0)\rangle \quad (2.15)$$

Assuming that the initial neutrino state is a pure  $|\nu_\alpha\rangle$  state, then the transition amplitude  $\nu_\alpha \rightarrow \nu_\beta$  is given by

$$A(\alpha \rightarrow \beta)(t) = \langle \nu_\beta(\vec{x}, t) | \nu_\alpha(\vec{0}, 0) \rangle = \sum_i \sum_j U_{\beta i}^* e^{i\phi_i} U_{\alpha j} \langle \nu_i(\vec{0}, 0) | \nu_j(\vec{0}, 0) \rangle \quad (2.16)$$

using

$$\langle \nu_\beta(\vec{x}, t) | = \sum_i U_{\beta i}^* e^{i\phi_i} \langle \nu_i(\vec{0}, 0) | \quad (2.17)$$

and eqn. 2.11 where  $i \rightarrow j$ . Because the mass eigenstates are orthogonal, the relations

$$\langle \nu_i(\vec{0}, 0) | \nu_j(\vec{0}, 0) \rangle = 1 \text{ for } i = j \text{ and } \langle \nu_i(\vec{0}, 0) | \nu_j(\vec{0}, 0) \rangle = 0 \text{ for } i \neq j$$

can be used to reduce eqn. 2.16 to

$$A(\alpha \rightarrow \beta)(t) = \sum_i U_{\beta i}^* e^{i\phi_i} U_{\alpha i} \quad (2.18)$$

Now under the assumption that the neutrino travels in the  $x$ -direction,  $\phi_i$  becomes

$$\phi_i = E_i t - \vec{p} \cdot \vec{x} = E_i t - p_i x \quad (2.19)$$

and assuming that the momenta of each mass state is the same

$$E_i = \sqrt{m_i^2 + p_i^2} \simeq p_i + \frac{m_i^2}{2p_i} \quad (2.20)$$

substituting eqn. 2.20 into eqn. 2.19

$$\phi_i = p_i t + \frac{m_i^2 t}{2p_i} - p_i x \quad (2.21)$$

and then using the relativistic assumptions,  $L = x = t$  and  $E \approx p$ , eqn. 2.21 can be simplified to

$$\phi_i = EL + \frac{m_i^2 L}{2E} - EL = \frac{m_i^2 L}{2E} \quad (2.22)$$

This can then be substituted back into the transition amplitude, eqn. 2.18

$$A(\alpha \rightarrow \beta)(t) = \sum_i U_{\beta i}^* e^{i \frac{m_i^2 L}{2E}} U_{\alpha i} \quad (2.23)$$

It is then possible to obtain the transition probability  $P(\alpha \rightarrow \beta)(t)$  by squaring the transition amplitude to obtain:

$$P(\alpha \rightarrow \beta)(t) = A(\alpha \rightarrow \beta)(t) = \sum_i \sum_j U_{\beta i}^* U_{\beta j} e^{i \frac{(m_i^2 - m_j^2)L}{2E}} U_{\alpha i} U_{\alpha j}^* \quad (2.24)$$

which when expanded, gives

$$\begin{aligned} P(\alpha \rightarrow \beta)(t) &= \sum_i U_{\alpha i} U_{\beta i}^* \sum_j U_{\alpha j}^* U_{\beta j} \\ &+ 2 \operatorname{Re} \sum_{i>j} U_{\beta i}^* U_{\beta j} U_{\alpha i} U_{\alpha j}^* \left[ e^{i \frac{(m_i^2 - m_j^2)L}{2E}} - 1 \right] \end{aligned} \quad (2.25)$$

Because the masses, lengths, and energy are real observables, we can separate the exponential into real and complex parts

$$\begin{aligned} \operatorname{Re} \left( e^{i \frac{\Delta m_{ij}^2 L}{2E}} - 1 \right) &= -2 \sin^2 \left[ \frac{\Delta m_{ij}^2 L}{4E} \right] \\ \operatorname{Im} \left( e^{i \frac{\Delta m_{ij}^2 L}{2E}} - 1 \right) &= \sin \left[ \frac{\Delta m_{ij}^2 L}{4E} \right] \end{aligned} \quad (2.26)$$

where  $\Delta m_{ij}^2 = m_i^2 - m_j^2$  has been used for simplification. Because  $U$  is a unitary matrix, the first term in Eqn. 2.25 becomes  $\delta_{\alpha\beta}$ . Finally, by inserting Eqns. 2.26 into

Eqn. 2.25, we can obtain our final oscillation probability of:

$$\begin{aligned}
P(\alpha \rightarrow \beta)(t) &= \delta_{\alpha\beta} \\
&- 4 \operatorname{Re} \sum_{i>j} U_{\beta i}^* U_{\beta j} U_{\alpha i} U_{\alpha j}^* \sin^2 \left[ \frac{\Delta m_{ij}^2 L}{4E} \right] \\
&+ 2 \operatorname{Im} \sum_{i>j} U_{\beta i}^* U_{\beta j} U_{\alpha i} U_{\alpha j}^* \sin \left[ \frac{\Delta m_{ij}^2 L}{4E} \right]
\end{aligned} \tag{2.27}$$

## 2.4 CP Violation

The simplest models of the Big Bang predict that the universe was created with equal amounts of matter and anti-matter. However, the universe appears to be dominated by normal matter. Thus following the simplest assumption of the Big Bang models, some asymmetry in physical processes must exist to yield such a state. This asymmetry lies in the violation of the charge-parity (CP) operator. The charge (C) operator is the conjugation of a particle for its anti-particle (and vice-versa), while the parity (P) operator flips the signs of all spatial coordinates for a particle. The CP operator is the combination of the two.

If Eqn. 2.27 is applied to the three-flavour paradigm, and all parts of the  $U_{\text{PMNS}}$  mixing matrix are considered, then CP violating behaviour can be parameterised in terms of a CP violating phase  $\delta_{\text{CP}}$  - which can be seen in the oscillation probability between two neutrino flavour states. For example, if we take the first order approximation for the  $\nu_\mu \rightarrow \nu_e$  oscillation [28],

$$\begin{aligned}
P(\nu_\mu \rightarrow \nu_e) &\simeq \sin^2(\theta_{23}) \sin^2(2\theta_{13}) \frac{\sin^2(\Delta_{31} - aL)}{(\Delta_{31} - aL)^2} \Delta_{31}^2 \\
&+ \left[ \sin(2\theta_{23}) \sin(2\theta_{13}) \sin(2\theta_{12}) \right. \\
&\quad \left. \frac{\sin(\Delta_{31} - aL)}{(\Delta_{31} - aL)} \Delta_{31} \frac{\sin(aL)}{(-aL)} \Delta_{21} \cos(\Delta_{31} + \delta_{\text{CP}}) \right] \\
&+ \cos^2(\theta_{23}) \sin^2(2\theta_{12}) \frac{\sin^2(aL)}{(aL)^2} \Delta_{21}^2,
\end{aligned} \tag{2.28}$$

$$\text{with } \Delta_{ij} = \frac{\Delta m_{ij}^2 L}{4E},$$

it can be seen that  $\delta_{\text{CP}}$  (highlighted in red) plays a role in the second term. When considering the conjugate oscillation probability ( $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ), the sign of  $\delta_{\text{CP}}$  changes.

The effects that different values of  $\delta_{\text{CP}}$  have on the oscillation probability, likely to be observed at the DUNE far detector, are shown in Fig. 2.3

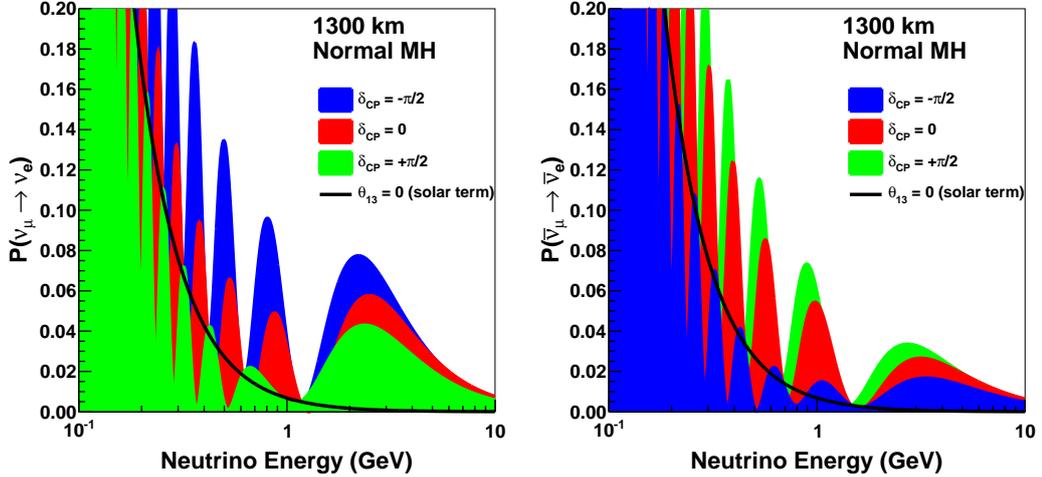


Figure 2.3: The effect that different values of  $\delta_{\text{CP}}$  has on the oscillation probability of  $\nu_{\mu} \rightarrow \nu_e$  (left) and  $\bar{\nu}_{\mu} \rightarrow \bar{\nu}_e$  (right) with a baseline of 1300 km, and normal mass ordering. Figure taken from [29]

The current best results from the T2K collaboration, seen in Fig. 2.4, suggests that within a 95% confidence level the values of  $\delta_{\text{CP}} = 0$  and  $\delta_{\text{CP}} = \pi$  are ruled out, thus hinting towards CP being violated in the lepton sector [30]. However, the 99.73% confidence level does not exclude the CP conserving points, and so while this gives strong hints as to level of CP violation in the lepton sector, definitive results at the five sigma level require a next generation oscillation experiments, such as DUNE.

## 2.5 MSW Matter Effect

In Eqn. 2.28 there is present another factor that affects the oscillation probability,  $a = G_F N_e / \sqrt{2}$ , where  $G_F$  is the Fermi constant and  $N_e$  the number density of electrons in the Earth. This is known as the matter effect, or the MSW effect - named after Mikheyev, Smirnov, and Wolfenstein for their work on the subject [31, 32]. The cause of this parameter is simply the fact that electrons are the only leptons to exist within normal matter, thus enabling electron neutrinos to undergo a CC scattering process as well as the NC scattering that all neutrinos can undergo. This then introduces another effect to the neutrino-antineutrino asymmetry as the sign of  $a$  is reversed for neutrinos and antineutrinos.

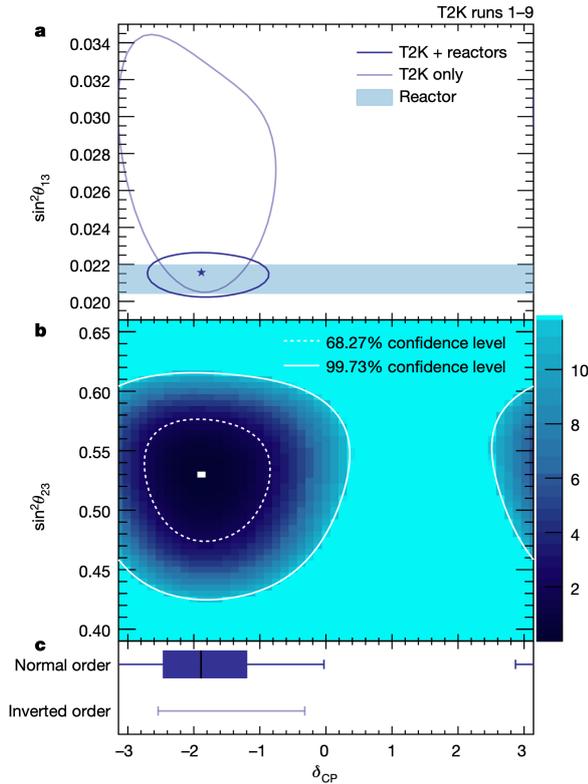


Figure 2.4: The current latest results from the T2K Collaboration. CP conservation,  $\delta_{CP} = 0, \pi$ , is ruled out to the 95% confidence level, but are not excluded by the 99.73% confidence level. Figure **a** shows the 68.7% confidence level for  $\delta_{CP}$  vs  $\sin^2 \theta_{13}$  as measured by T2K, with the star representing the best fit point for T2K + reactors for the normal mass ordering (Sec. 2.6). Figure **b** shows the 68.27% and 99.73% confidence intervals for  $\delta_{CP}$  vs  $\sin^2 \theta_{23}$  as measured by T2K. The colour scheme represents the value of negative two times the logarithm of the likelihood for each parameter value. Figure **c** shows the 68.27% (box) and 99.73% (whiskers) confidence intervals for the normal mass ordering, and the 68.27% confidence interval of the inverted mass ordering, both as measured by T2k [30].

## 2.6 Mass Ordering

Due to the neutrino oscillation measurements being sensitive only to the square of the neutrino mass, as seen in Eqn. 2.27, it is not possible to make a direct mass measurement with oscillation experiments. It is, however, possible to make measurements of the mass-squared differences:  $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$ .

The smallest mass-squared difference,  $\Delta m_{21}^2$ , is defined as being positive-definite such that  $m_2^2 > m_1^2$ , and has magnitude of order  $10^{-4} \text{ eV}^2$ . The two remaining mass-squared differences will be an order of magnitude larger, and so can be treated

as being similar:  $|\Delta m_{31}^2| \sim |\Delta m_{32}^2|$ . Measuring the sign of these two remaining differences is the crux of the mass ordering problem in physics. If the third mass eigenstate is more massive than the two other mass eigenstates then the mass ordering is said to be ‘normal’, otherwise it is said to be ‘inverted’.

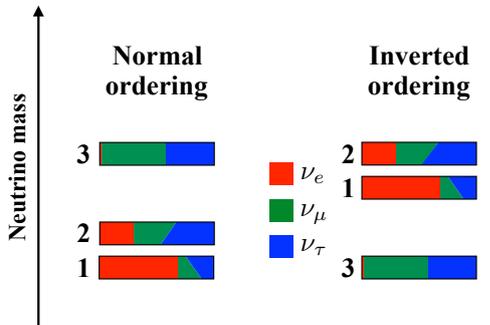


Figure 2.5: The two possible neutrino mass orderings - normal or inverted. Each mass eigenstate is also split into its fractional flavour content, with the  $\delta_{\text{CP}}$  varied from 0 to  $\pi$  (bottom of mass bar to top of mass bar). Figure taken from [28]

### 2.6.1 Majorana Mass

It is possible to generate the mass of particles in the SM through spontaneous symmetry breaking, however this process relies on the existence of right- and left-handed particles - neutrinos, however, only exist as left-handed particles. It is possible however to generate a mass term for neutrinos using a theory proposed by Ettore Majorana in 1937 [33].

In the case that neutrinos are Majorana particles, then it is necessary to extend the neutrino unitary mixing matrix, Eqn. 2.12, to include two Majorana phases,  $\eta_1$  and  $\eta_2$ .

$$U = U_{\text{PMNS}} \begin{bmatrix} e^{i\eta_1} & 0 & 0 \\ 0 & e^{i\eta_2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.29)$$

Because the oscillation probability depends on the square modulus of the mixing matrix elements, these complex phases are cancelled out and do not affect the neutrino oscillations. This is evidenced by a lack of neutrino-antineutrino oscillation observations [34]. As such, it is necessary to have a different type of experiment that involves the mixing of antineutrinos and neutrinos to measure the existence of these Majorana phases. A number of neutrinoless double-beta decay experiments - such as SuperNEMO [35], GERDA [36], and EXO [37] - are investigating this.

# 3

## The Deep Underground Neutrino Experiment

---

*“What senses do we lack that we cannot see and cannot hear another world all around us?”*

**Frank Herbert - *Dune* - 1965**

**T**he Deep Underground Neutrino Experiment (DUNE), is a next generation long-baseline neutrino oscillation experiment using liquid argon time projection chamber (LArTPC) technology, in conjunction with the world’s most intense neutrino beam. DUNE will be based at two locations: the U.S. Department of Energy’s Fermi National Accelerator Laboratory (Fermilab), just outside Chicago, Illinois; and the Sanford Underground Research Facility (SURF), in Lead, South Dakota. An overview of the locations of the near site (Fermilab), the far site (SURF), and the baseline connecting them is shown in Fig. 3.1.

The infrastructure for the near and far detectors, as well as the beamline, will be provided by the Long-Baseline Neutrino Facility (LBNF), a U.S. Department of Energy (DoE) project. The detectors themselves will be jointly funded and run by the U.S. DoE and multiple international collaborators.

### 3.1 Physics Goals of DUNE

Neutrino physics has taken a great many leaps forward in uncovering some fundamental questions in particle physics in recent years. The observed neutrino oscillations

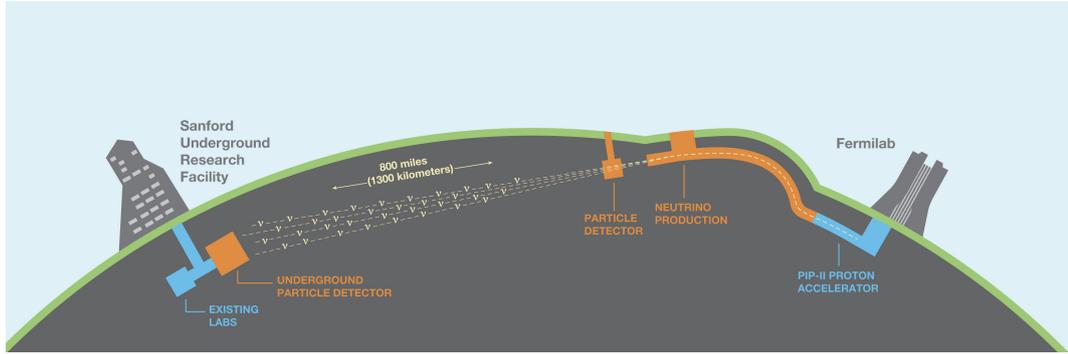


Figure 3.1: An overview of the planned LBNF and DUNE projects, and the two locations for the project - Fermilab, the near site, and SURF, the far site [29].

clearly show that we do not hold all the answers in the Standard Model, and that there is physics beyond it. DUNE will attempt to answer questions further, and also hope to answer some currently open questions. DUNE will do this by having a set of primary and secondary physics goals.

DUNE's primary physics goals are [29]:

- Precision measurements of neutrino oscillation parameters:
  - Determination of the neutrino mass ordering - sign of  $\Delta m_{31}^2$ .
  - Measurements of the charge-parity violating phase  $\delta_{CP}$ .
  - Measurements of the  $\theta_{23}$  mixing angle.
- Search for the evidence of proton decay, and other processes that violate baryon number.
- Measurement of  $\nu_e$  flux from an intra-galactic core-collapse supernova.

DUNE's secondary physics goals are:

- Search for sterile neutrinos in Beyond Standard Model (BSM) physics.
- Use of atmospheric neutrinos in measurements of neutrino oscillation phenomena.
- Tau neutrino appearance measurements.
- Neutrino interaction physics using the near detector - such as neutrino interaction cross sections, and various nuclear effects.
- Dark matter searches.

## 3.2 Far Detector

The far detector (FD) modules will be located at a far site at the SURF facility in South Dakota. The location will provide a baseline of 1 300 km, allowing access to the first and second oscillation maxima across a range of energies (see Fig. 2.3). The excavation of the caverns, located 1.5 km underground, will be provided by the LBNF project. The caverns, as seen in Fig. 3.2, will consist of two larger caverns for housing the cryostats - seen in Fig. 3.3 - of the FD modules, a smaller central cavern for the data acquisition (DAQ) system and the cryogenics, as well as various connecting tunnels and a refitted elevator shaft. The two larger caverns will provide enough space to each house two cryostats. The project's first phase will see two far detector modules installed by 2028, with the remaining two modules being installed over the following two years - or as quickly as funding allows.

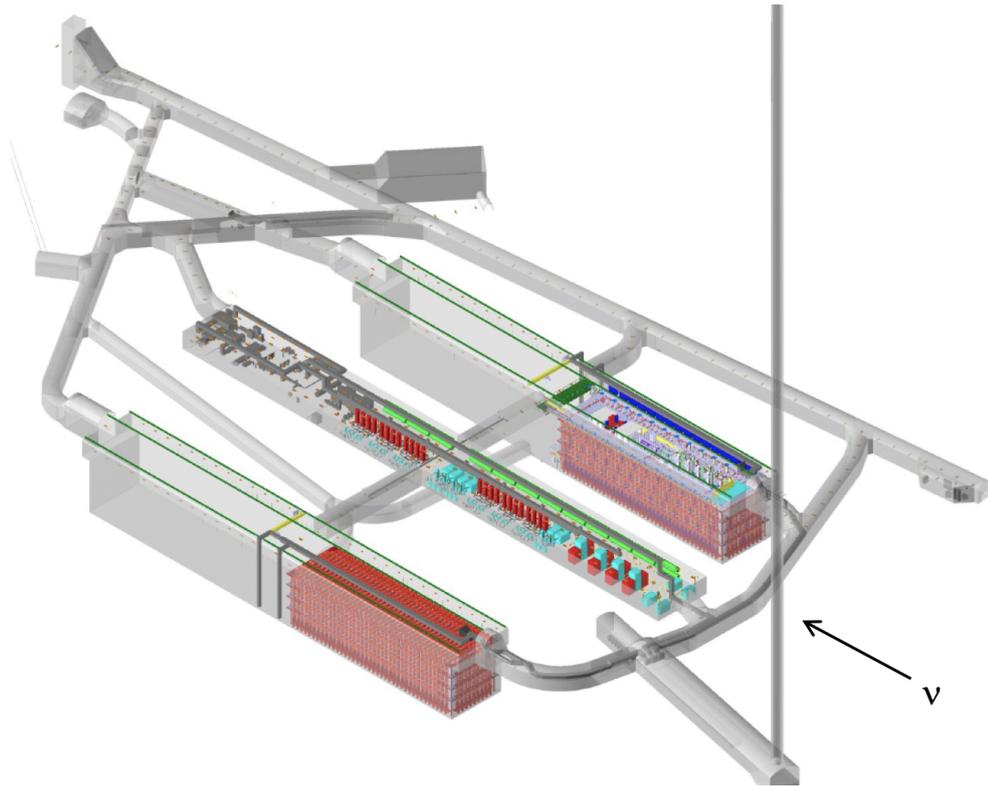


Figure 3.2: The caverns at SURF. The large red boxes illustrate the cryostats of the first two far detector modules. The central cavern, housing the DAQ system and cryogenics, is located between the two cryostat caverns. The large vertical shaft is the access elevator [29].

As of the time of writing, the first two planned modules will utilise the current

### Deep Underground Neutrino Experiment

One of four detector modules in South Dakota

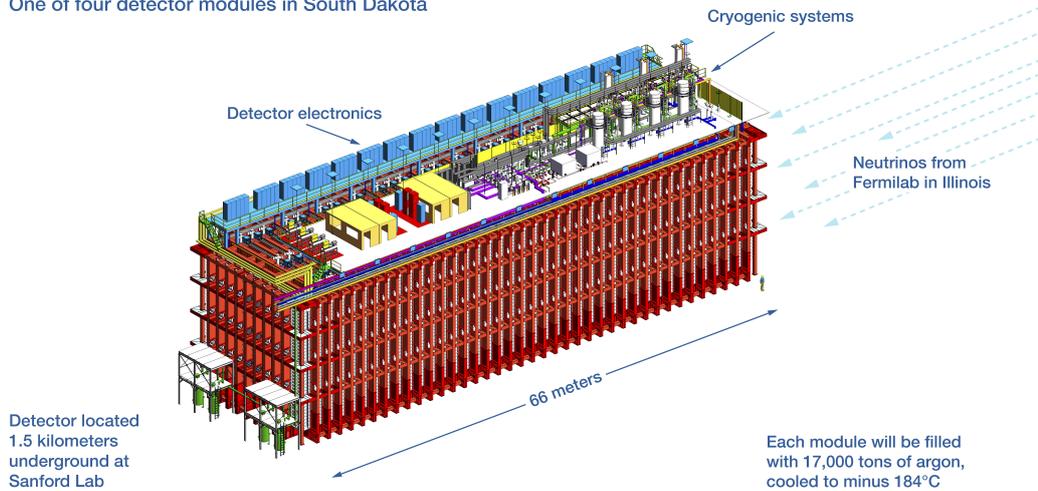


Figure 3.3: A graphic showing the size of the cryostat housing for a LArTPC FD module [38].

single-phase (Sec. 3.2.2) LArTPC technology being prototyped at various current and past experiments. The third module will likely use dual-phase (Sec. 3.2.3) technology if proven to work effectively. The design of the final module is left open to new technological developments in the coming years.

#### 3.2.1 Liquid Argon Time Projection Chambers

The time projection chamber was originally proposed by David Nygren in the late 1970's [39], as a gas based drift chamber amalgamated with a multiwire proportional chamber, designed to sit wrapped around the beam pipe of particle colliders. Carlo Rubbia then iterated upon the design, and suggested using a liquid argon medium instead of a gaseous one, as the liquid argon functions as both the target and detection medium [40].

Currently, LArTPC technology comes in two flavours - a single-phase (SP) and a dual-phase (DP) design. The SP design operates by drifting ionisation electrons, that have been liberated from their parent nuclei as a charged particle travels through the LAr medium, laterally towards an instrumented anode plane that is also submerged in the LAr volume. The ionisation electrons drift is instigated by a uniform  $\vec{E}$ -field, typically of strength in the order of a few hundred volts per cm - DUNE will nominally operate at  $500 \text{ V cm}^{-1}$ . The anode plane will itself consist of multiple planes of finely pitched wires, where the signal induced and collected on these planes will form

the basis for the reconstruction process. The DP technology works by drifting the ionisation electrons vertically, extracting them across the boundary with a gaseous argon volume where they are multiplied by an electron multiplier plane. The anode plane of the DP differs from the SP by using two-dimensional PCBs with gold-plated copper strips as the readout material.

In both technologies a prompt time signal is also provided by the VUV scintillation light of LAr, which is collected by a photon detection system installed on the frames of the SP anode planes and on the bottom of the DP TPC. The technology of SP LArTPCs has already been demonstrated by a number of prototyping experiments - ICARUS [41], ArgoNeuT [42], MicroBooNE [43], LArIAT [44], and ProtoDUNE [45] - and is illustrated in Fig. 3.4.

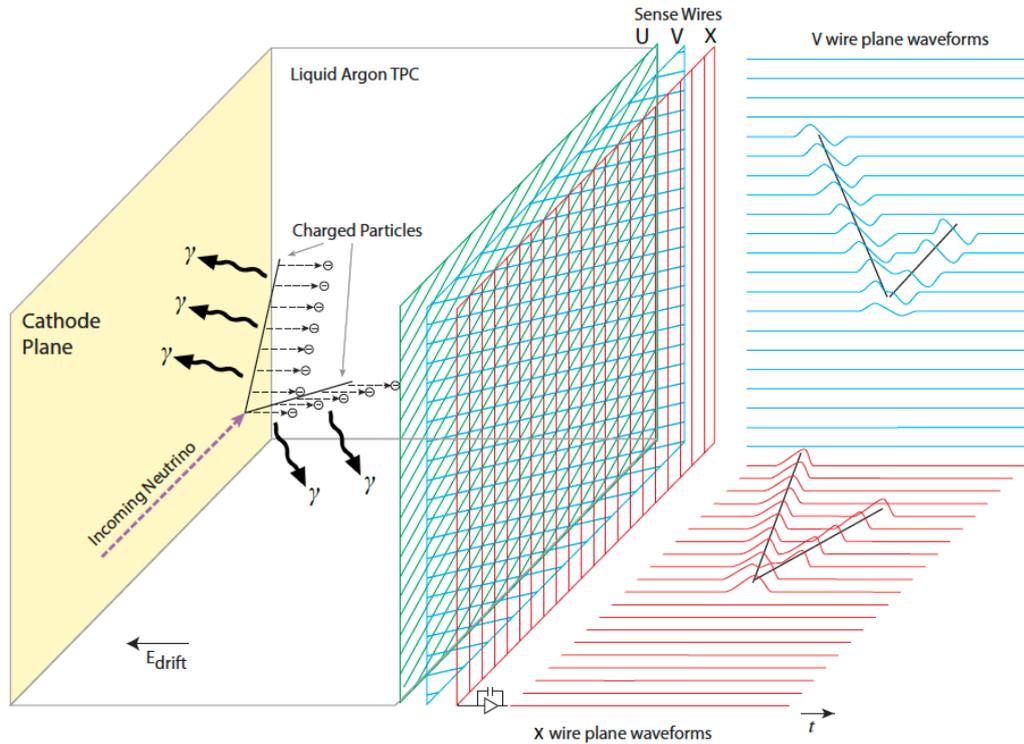


Figure 3.4: A cartoon illustrating the principals of how a single-phase LArTPC operates. Electrons are liberated from their parent nuclei as charge particles traverse the LAr volume. These ionisation electrons are then drifted, under a large electric field, towards a set of instrumented wire planes where the charge information is read out [46].

### 3.2.2 Single-Phase

Each DUNE-FD SP LArTPC module will have a total mass of 17.5 kt, with a fiducial mass of 10 kt. The fiducial volume will be divided down the length of the LArTPC by an alternating set of anode and cathode walls, as shown in Fig. 3.5. Each area located between one anode wall and the adjacent cathode wall is known as a drift volume. The individual anode walls will each be made from two rows of 25 anode plane assembly (APA) units stacked on top of each other, while the cathode walls will be made from three rows of 50 cathode plane assembly (CPA) units. A field cage is then also placed on the top, bottom, and ends of the LArTPC to ensure a uniform  $\vec{E}$ -field throughout the volume. Attached to the frame of the APAs will also be the photon detection system's X-Arapuca bars - with each APA having 10 bars attached. An Arapuca works by using a dichroic filter to trap light inside a reflective cell mounted with silicon photo-multiplier devices used to measure the light signal [46].

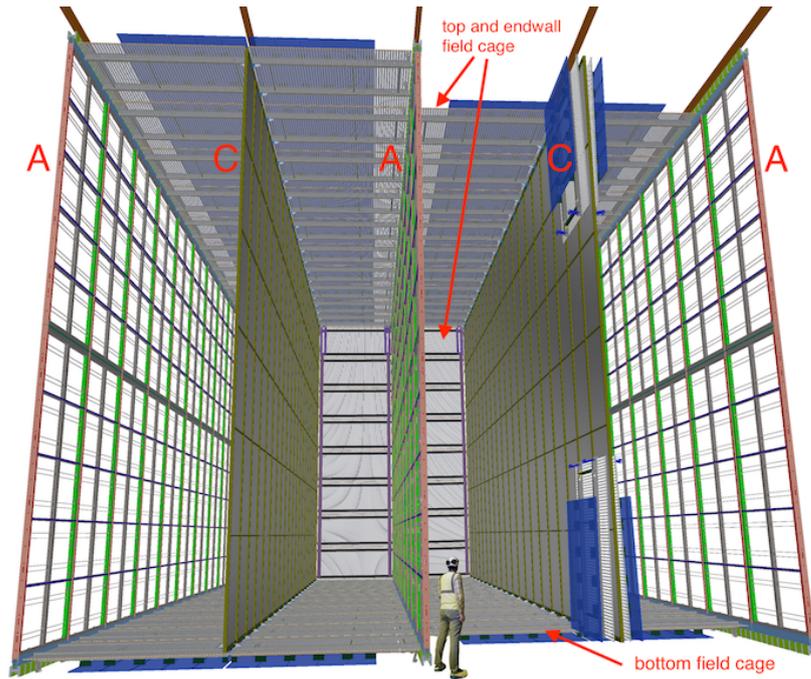


Figure 3.5: A schematic of the layout of a SP LArTPC. The anode (A) will be 2 APAs high, and the cathode (C) 6 CPA units high. The field cage can also be seen covering the top, bottom and end of the LArTPC [46].

The LAr will be cooled to a temperature of around 87 K, and must maintain a high purity. Oxygen contamination must be kept below 100 ppt (parts per trillion) to ensure an ionisation electron lifetime of greater than 3 ms, as oxygen and water impurities can absorb these drifting electrons. ProtoDUNE-SP's operational per-

formance has shown it is possible to achieve ionisation electron lifetimes exceeding 6 ms. Nitrogen contamination must also be kept below 25 ppm (parts per million), as nitrogen absorbs the scintillation photons. To maintain this purity the LAr is continuously cycled through a purification system. A summary of the required specifications for the single-phase LArTPC can be seen in Table 3.1.

Table 3.1: Single-phase LArTPC Specifications [46].

Item	Quantity
TPC size	12.0 m $\times$ 14.0 m $\times$ 58.2 m
Nominal fiducial mass	10 kt
APA size	6 m $\times$ 2.3 m
CPA size	1.2 m $\times$ 4 m
Number of APAs	150
Number of CPAs	300
Number of X-ARAPUCA PD bars	1500
X-ARAPUCA PD bar size	209 cm $\times$ 12 cm $\times$ 2 cm
Design voltage	- 180 kV
Design drift field	500 Vcm <sup>-1</sup>
Drift length	3.5 m
Drift speed	1.6 mm $\mu$ s <sup>-1</sup>

### 3.2.2.1 Anode Plane Assemblies

The APAs consist of 6 m tall by 2.3 m wide stainless steel frames, with multiple planes of wires covering the large flat faces. The APAs on the top row of the anode walls will have the readout electronics along the top edge, while the bottom row of APAs will be inverted with the readout electronics along the bottom. The steel frames are constructed from hollow bars to allow cables to be run through, and are mounted with a grounding mesh to prevent ionisation from within the APA creating signals on the instrumented wires. The wire planes are wrapped around the APA, creating four planes of wires on each side of the APA. The wires are made from 152  $\mu$ m diameter copper-beryllium (CuBe) alloy, chosen for its high durability and yield strength (the maximum amount of stress applied before plastic deformation occurs). The maximum tensile strength of the CuBe wires used in the ProtoDUNE-SP APAs has been recorded as being higher than 1380 MPa, and a yield strength of over 1100 MPa. This far exceeds the operational stress of around 340 MPa [46].

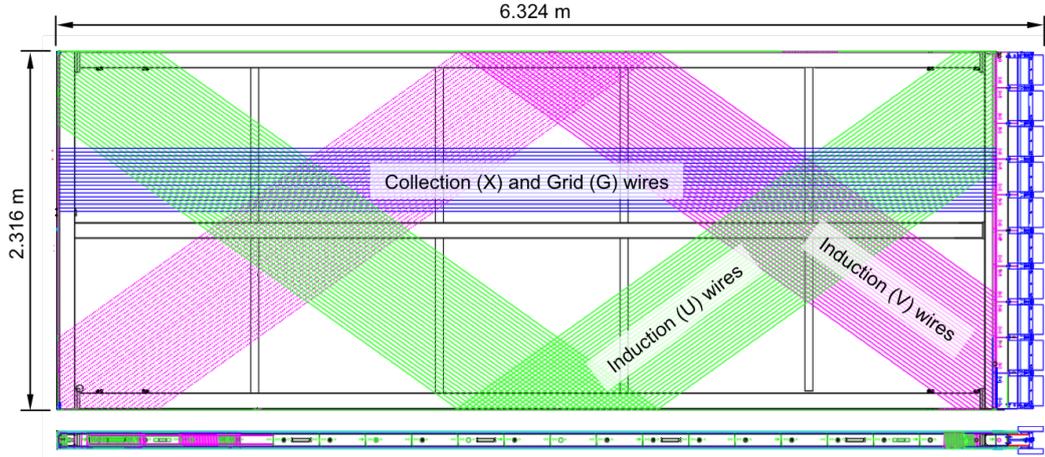


Figure 3.6: A face on schematic of an APA. The blue boxes on the right hand edge are the readout electronics. The green and magenta lines represent the two induction planes, and the blue lines represent the collection and shielding wires [46].

The wire planes are layered with the top and bottom layers vertical, and the two sandwiched layers at an angle of  $\pm 35.7^\circ$  to the vertical, as shown in Fig. 3.6 and Fig. 3.7. The two sandwiched layers, U and V, are held at voltages rendering them transparent to the drifting ionisation charge. This transparency allows the charge to completely drift past both of the planes, inducing<sup>1</sup> signals of positive and negative polarity as they do so. Once the charge has passed these two induction planes, it is collected on the electrically opaque X plane of wires - the collection plane - where a single unipolar signal is recorded. The outermost wire plane, G, is there to shield the U induction plane from seeing a long leading edge to the signal pulses, thus providing a cleaner measurement. The voltages of the wire planes can be seen in Table 3.2, and have been calculated by COMSOL software [46].

Table 3.2: Bias voltages for the wire planes of an APA [46].

Wire Plane	Voltage
G - Shield	-665 V
U - Induction	-370 V
V - Induction	0 V
X - Collection	820 V
Grounding Mesh	0 V

<sup>1</sup>Hence their names as ‘induction planes’

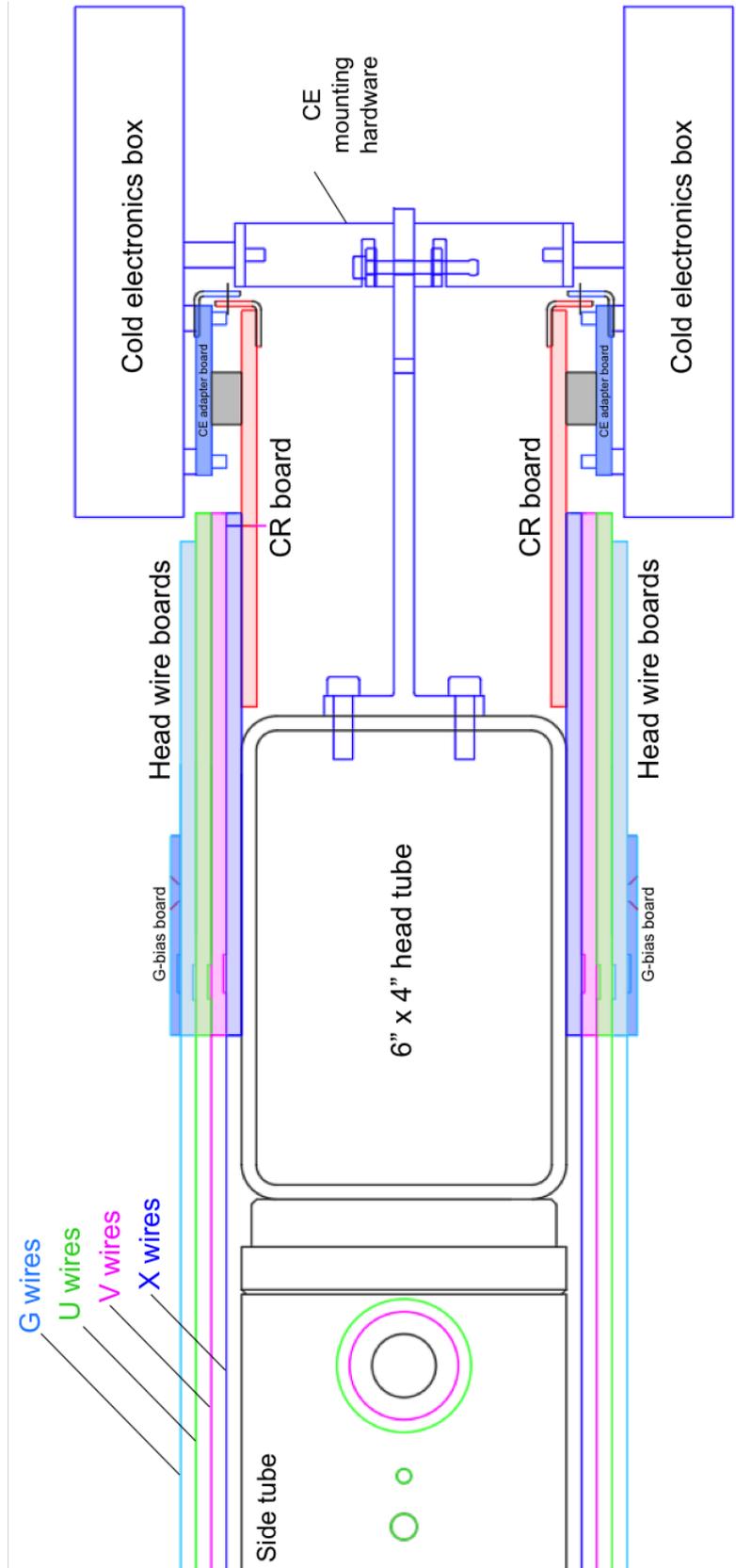


Figure 3.7: A side on schematic of the top of an APA frame, showing the order of the G, U, V, and X wire planes [46].

### 3.2.2.2 Cathode Plane Assemblies

Key to the success of a good LArTPC is a strong and highly uniform  $\vec{E}$ -field facilitating the effective transport of ionisation charge across the detector. To achieve this in DUNE there is the high voltage system, CPA plane, and the field cage. The high voltage system is formed of an external power supply system, and the required cabling and electrical filters to provide a stable and uniform voltage to the CPA plane. The CPA plane is formed from smaller CPA units made from carbon-impregnated Kapton<sup>2</sup> laminated to FR-4<sup>3</sup> sheets. The CPA plane will be 6 CPA units high, and 25 CPA units long, and will be kept at a bias voltage of -180kV, providing the  $500\text{ Vcm}^{-1}$   $\vec{E}$ -field. The field cage is constructed from extruded aluminium bars designed to ensure uniformity of the  $\vec{E}$ -field at the edges of the LArTPC. The CPA and field cage arrangement can be seen in Fig. 3.5.

The chosen nominal value for  $\vec{E}$ -field strength is a trade-off between detector performances that improve or degrade with  $\vec{E}$ -field values. For example,  $dE/dx$  measurements are strongly affected by electron-ion recombination, which is suppressed with a high value  $\vec{E}$ -field. However, the number of scintillation photons is inversely proportional to the strength of the  $\vec{E}$ -field, potentially increasing the difficulty of reconstructing a  $t_0$  for reconstructed particles. Through much operational experience from preceding prototype experiments, it has been shown that  $500\text{ Vcm}^{-1}$  is the most suitable  $\vec{E}$ -field value when considering potential trade-offs [46].

### 3.2.3 Dual-Phase

An alternative detector technology to the SP LArTPC is the DP LArTPC. The DP design differs from the SP design by drifting the ionisation charge vertically before extraction into a gaseous phase, as seen in Fig. 3.8. While in the gaseous phase, the ionisation charge is amplified using large electron multipliers and collected on instrumented PCBs as opposed to instrumented wire planes. This ionisation charge amplification will provide an enhanced signal to noise ratio compared to the SP, allowing for a lower threshold on the charge required to form a reconstructed particle object. Another benefit of the DP design is the use of one singular readout plane at the top of the TPC, reducing the amount of material in the LAr, providing a larger homogeneous volume compared with the SP.

---

<sup>2</sup>DuPont<sup>TM</sup>, Kapton<sup>®</sup> polyimide film, E. I. du Pont de Nemours and Company <http://www.dupont.com/>

<sup>3</sup>NEMA grade designation for flame-retardant glass-reinforced epoxy laminate material, multiple vendors, National Electrical Manufacturers Association<sup>TM</sup>, <https://www.nema.org/pages/default.aspx>

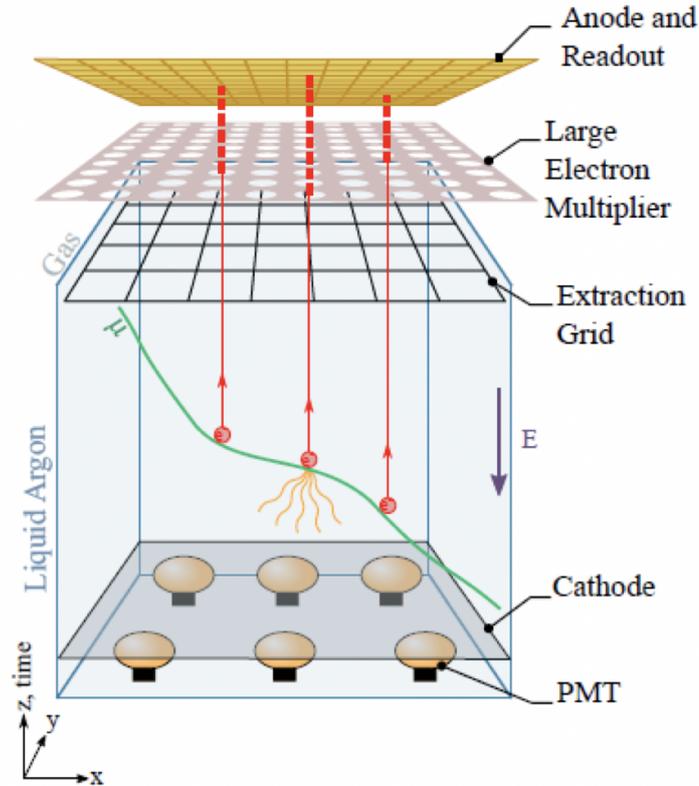


Figure 3.8: The operational principle of the dual-phase design. The ionisation electrons are drifted vertically to be extracted and multiple in a gaseous phase to increase signal to noise ratio [29].

The DP design, however, faces a number of challenges such as; maintaining an ultra-pure gaseous layer, a much higher cathode voltage (600 kV, compared to 180 kV for the SP design), build up of space charge in the liquid-gas interface layer, stability at the top of the liquid layer with Argon boiling off. The ability to overcome these challenges is currently being tested with the ProtoDUNE-DP located at the Neutrino Platform Facility<sup>4</sup> at CERN. While the construction of ProtoDUNE-DP was unable to be completed before CERN’s LS2, it is now operational and has seen its first particle tracks from cosmic rays [47, 48].

### 3.3 The LBNF Beamline and Target

The LBNF neutrino beam will be the world’s most intense neutrino beam, with an assumed  $1.1 \times 10^{21}$  protons-on-target per year [29], once in operation [49]. The Fermilab Main Injector proton accelerator will provide a 1.0-1.2 MW proton beam to

---

<sup>4</sup>Located in the EHN1 hall at CERN’s Preveessin site

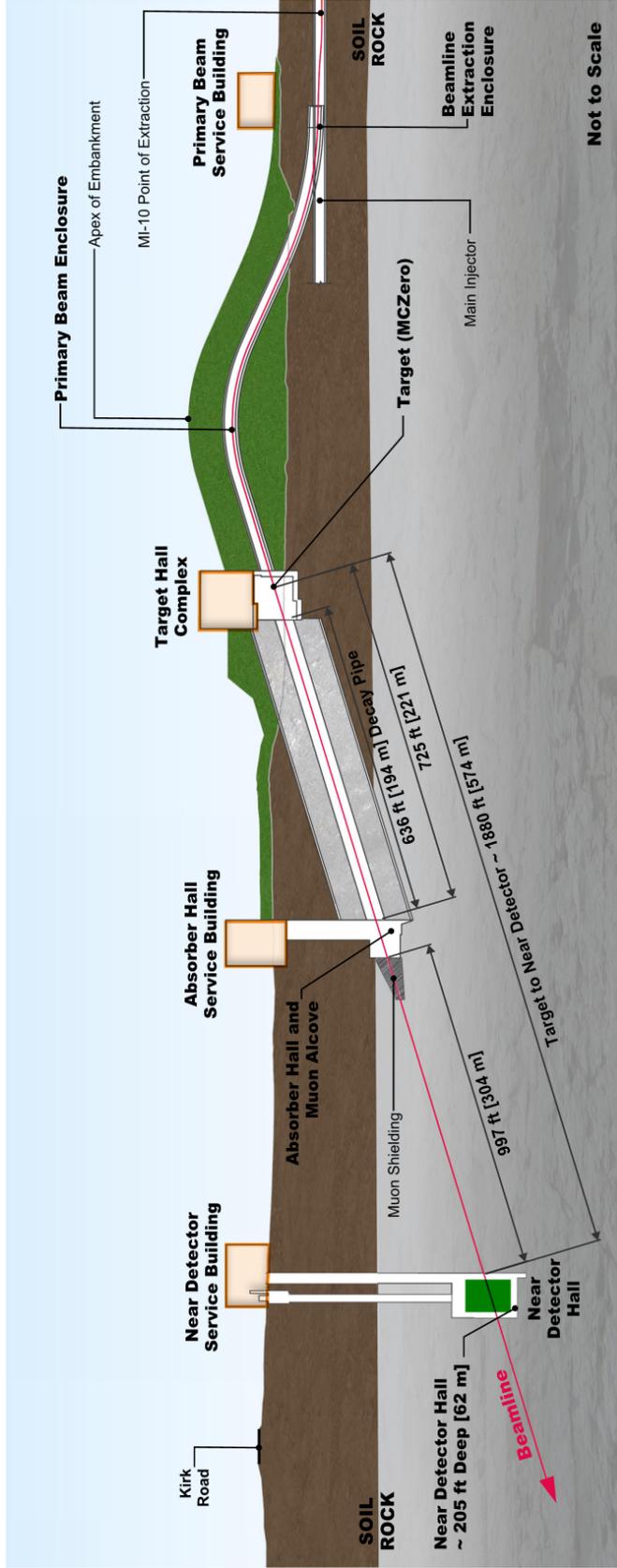


Figure 3.9: A side view of the beamline, target, and ND complexes. The proton beam comes from the right, provided by the Fermilab Main Injector, before being taken up through an embankment and directed towards the target. After striking the target, decay products from proton interactions are directed down a decay pipe, where they will decay into neutrinos before continuing on towards the ND and FD facilities [29].

LBNF, where the beam will be focused onto a target, creating a wide-band on-axis neutrino beam in the direction of the DUNE-FD. The current Fermilab roadmap also includes upgrading the beam power to 2.4 MW by 2030.

A side on view of the proposed LBNF beam site and ND are shown in Fig. 3.9. As the beam is leached from the Fermilab Main Injector, it is taken over an embankment before being directed towards the target hall complex, and subsequent near site facilities. The inclusion of the embankment allows the target hall complex to be above ground, removing the need to excavate more material than is strictly necessary, as well as minimising target installation and maintenance efforts going forwards.

Once the proton beam collides with the target, a beam of secondary charged particles (mainly kaons and pions) are directed down a decay pipe, where the particles decay into the neutrinos for the ND and FD detectors to measure. The proposed target is a 1.8 m long graphite rod with a 16 mm diameter. This choice of target results from considering a number of desired properties such as thermo-mechanical and radiological damage, as well as a trade off between converting enough protons into mesons without them simply absorbing the produced particles. The secondary particles produced by the protons in the target are then charge selected by a series of magnetic horns, that will direct the selected particles down the aforementioned decay pipe. The target system will also have to incorporate a cooling jacket for the target, as well as considerations for facilitating the replacement of the target material as it deteriorates over prolonged use. A cartoon of the discussed facilities can be seen in Fig. 3.10.

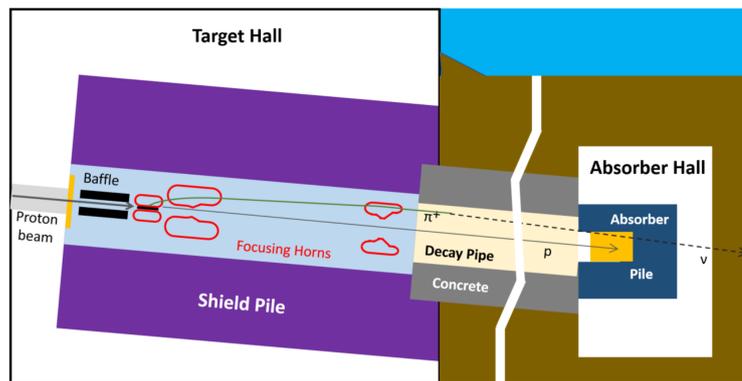


Figure 3.10: A cartoon showing the layout of the target hall and decay pipe at the near site. The proton beam comes in from the left before hitting the graphite target, creating a secondary set of charged particles which are directed towards the decay pipe by the focusing horns located immediately downstream of the target. The neutrinos produced in the decay pipe then travel on to the far site [50].

### 3.4 Near Detector

An important part of long-baseline neutrino experiments is understanding the neutrino properties at the source of the beam. To accomplish this DUNE will have a near detector (ND) located 304 m downstream of the end of the decay pipe. The DUNE ND will be a complex of three technologically different detectors<sup>5</sup>. The three detectors will be arranged linearly downstream of the neutrino beam, the planned layout of the ND complex can be seen in Fig. 3.11.

The most upstream detector, ArgonCube, will use a modular LArTPC design, similar to the FD but modified in areas to accommodate the much larger intensity expected so close to the neutrino source. This detector will be used to reduce detector driven systematic errors, and nuclear effects affecting the oscillation signal at the FD. The size of the detector will be sufficient to provide high statistics, and good hadron containment, but will not be able to contain muons with momenta greater than 0.7 GeV/c.

To measure the momenta of the muons escaping ArgonCube, there will be the multi-purpose detector (MPD) located directly downstream. The MPD will consist of a high-pressure gaseous argon TPC (HPgTPC), surrounded by an electromagnetic calorimeter (ECAL) inside a 0.5 T magnetic field. The lower density and high pressure system provides a number of benefits in tracking resolution and lower momentum acceptances, and will help identify the particles produced in the primary interactions in ArgonCube.

Both ArgonCube and the MPD will be able to move off-axis through a system called DUNE Precision Reaction-Independent Spectrum Measurement (DUNE-PRISM). Being able to take off-axis measurements allows the ND to create a neutrino energy distribution closely mimicking the spectrum at the FD by combining different flux measurements at different off-axis degrees.

---

<sup>5</sup>Affectionately known by collaborators as LArgon, GArgon, and NArgon ('Liquid argon', 'gaseous argon', and 'not argon')

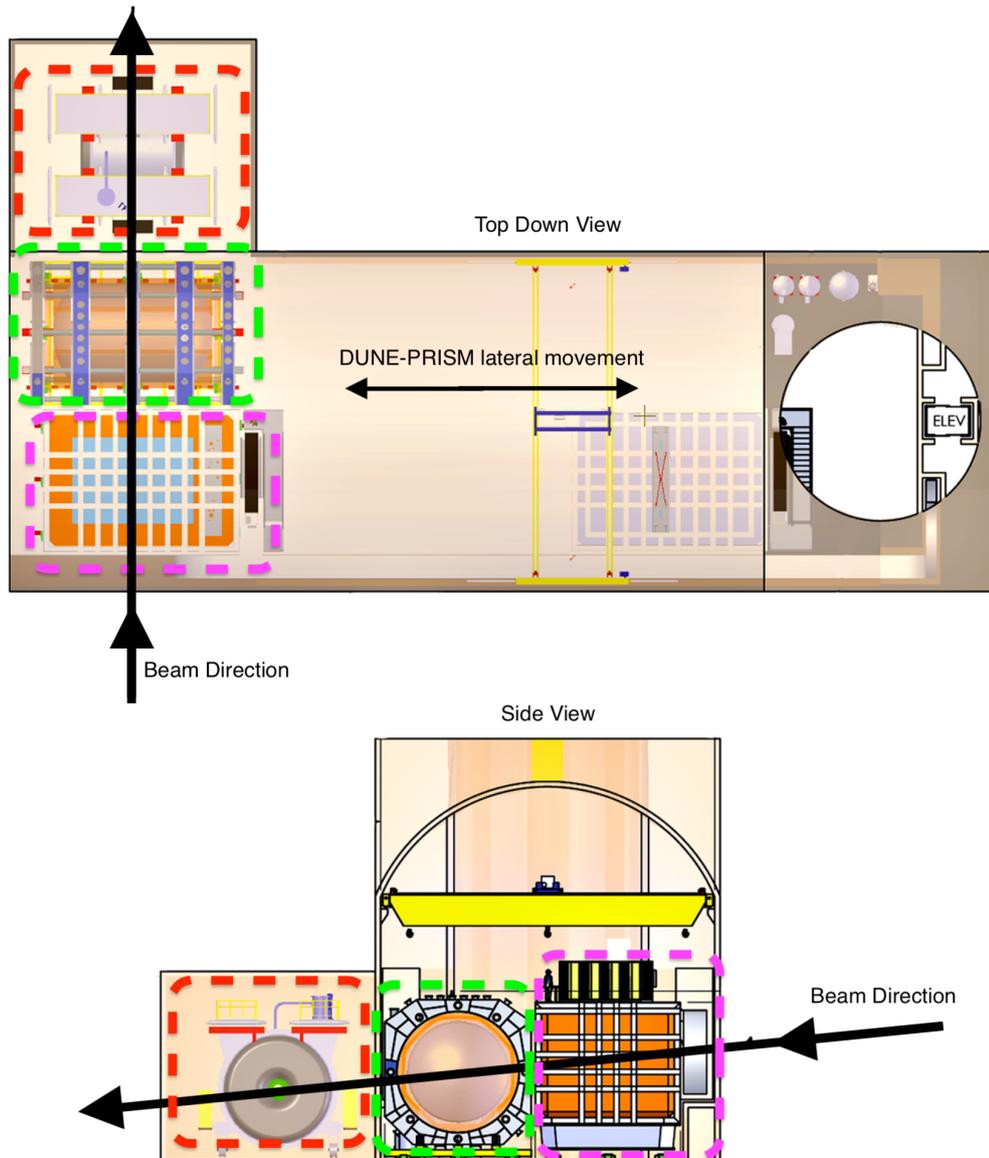


Figure 3.11: The layout of the DUNE ND complex. ArgonCube (pink box) can be seen to be the most upstream detector, SAND (red box) the most downstream, and the MPD (green box) sandwiched between the two. In the top down view you can also see the area where the DUNE-PRISM mechanism can move ArgonCube and the MPD off-axis [29].

The final detector in this complex will be the System for on-Axis Neutrino Detection (SAND). Consisting of  $1 \text{ cm}^3$  plastic scintillator cubes inside an ECAL and normal pressure TPCs, this will act as the neutrino spectrum monitor determining the on-axis neutrino flux, and will provide a good cross-check for flux measurements made with ArgonCube.

### 3.5 ProtoDUNE

The DUNE project's R&D phase has resulted in a number of LArTPC prototype devices [51–53], with the latest being the two sister ProtoDUNE experiments located at CERN - ProtoDUNE Single-Phase and ProtoDUNE Dual-Phase. The two ProtoDUNE detectors are located in the EHN1 building extension at CERN's Preveessin site, and serve to test all aspects of the technology to be deployed at the DUNE-FD. A photo of the two prototype detectors can be seen in Fig. 3.12.



Figure 3.12: The EHN1 building extension at CERN's Preveessin site - home to the two ProtoDUNEs. ProtoDUNE-SP can be seen in the foreground, while ProtoDUNE-DP is in the background. Located between the two detectors, surrounded by the large concrete blocks, is the end of the H4 beamline pointing towards ProtoDUNE-SP.

Both detectors use full size components following the FD design specifications, but only represent a small section of a full scale FD module. For example, ProtoDUNE-SP is only two drift volumes wide, with one cathode wall equidistant between two anode walls. The anode walls are also only one APA tall, and six APAs deep. ProtoDUNE-DP represents a similar portion of a full size DP FD module. The production of the full size components used in the prototype detectors has been crucial in understanding the needs required for mass fabrication of components in the years to come, as well as the methods for transportation and installation of the components.

The construction of the two ProtoDUNEs, including that of the EHN1 extension itself, was completed at breakneck speed between October 2015 and July 2018. ProtoDUNE-SP was fully completed, and then filled with liquid Argon and commissioned in August 2018, in time to receive test beam data between early September and mid

November. ProtoDUNE-DP unfortunately suffered from a number of setbacks, causing it to miss the test beam data taking opportunity offered before CERN's planned long shutdown 2. It is however now filled, and taking data from cosmic ray interactions along with ProtoDUNE-SP. The continued running after the test beam period is an important verification of the long term stability, and durability, of detector components. There are also further planned test beam data taking periods for 2022 and beyond.

### 3.5.1 H4-VLE Beamline

The beamline for the ProtoDUNE detectors is not a neutrino beam, but is instead a charged particle beam operating in two different particle modes - hadron mode, and electron mode. The purpose of the test beam is to test the response of the LArTPC volumes to a selection of charged particle types across a range of momenta. CERN's Super Proton Synchrotron (SPS) provides a 400 GeV/c proton beam which is directed towards a beryllium target. This produces a 80 GeV/c secondary mixed hadron beam, which is then subsequently directed at a secondary target of interchangeable material. The resultant tertiary beam is in the 0.3-7 GeV/c momentum range, and is transported for injection into the TPC. The materials used for the secondary target are tungsten and copper, where the former is used for sub 4 GeV/c momenta to enhance the beam's hadron content. The beamline has a three types of monitoring instruments; profile monitors, trigger counters, and Cherenkov counters. The profile monitors provide individual particle momentum measurements, the trigger counters provide a "beam trigger", and particle identification is provided by a combination of the trigger counters and the Cherenkov counters.

Over the test beam operating period, a total of four million beam triggers were recorded, at momenta ranging from 0.3 GeV  $c^{-1}$  to 7 GeV  $c^{-1}$ , as shown in Table 3.3. Below a beam momentum of 1 GeV  $c^{-1}$ , electrons are the dominant particle species, while above 1 GeV  $c^{-1}$  pions are dominant. At 1 GeV  $c^{-1}$ , protons marginally dominate over pions. A very small kaon contamination also exists at momenta above 1 GeV  $c^{-1}$ .

### 3.5.2 ProtoDUNE-SP

As the majority of the work done in this thesis uses data and Monte Carlo simulations from ProtoDUNE-SP, extra detail for the internal layout and terminology for detector components is laid out here. Fig. 3.13 and Fig. 3.14 provide diagrams that

Table 3.3: Number of beam triggers recorded at different energies provided by the test beam [54].

Momentum (GeV/c)	Total Recorded Beam Triggers
0.3	269 000
0.5	340 000
1.0	1 089 000
2.0	728 000
3.0	568 000
6.0	702 000
7.0	477 000
All Momenta	4 173 000

complement the following text.

As mentioned above, ProtoDUNE-SP consists of three APAs either side of a central CPA wall. The APAs are numbered from 1 through 6, and also have labels matching their location within the cryostat. First, the detector is split into three areas along the detector’s length: upstream, midstream, and downstream. Upstream represents the area covering the two APAs that are closest to the beamline, downstream represents the area covering the two APAs at the opposite side of the detector, and midstream is the two APAs between upstream and downstream. Secondly, the detector is split along the CPA wall into two halves: beam left, and beam right, from the perspective of facing towards the upstream face of the cryostat. This is shown in the diagram of Fig. 3.13. Each APA can also be split into two faces, a TPC face, and a cryostat face. The TPC face is the face that is on the side of the APA facing the CPA wall, while the cryostat face faces the internal wall of the cryostat.

The  $x$ ,  $y$ ,  $z$  coordinate system for ProtoDUNE-SP has the origin located at bottom corner of the upstream edge of the central CPA wall, with  $x$  increasing towards the beam left side,  $y$  increasing vertically upwards, and  $z$  increasing towards the downstream face of the cryostat. The test beam enters the cryostat slightly off centre towards the beam right side of the detector, and about three-quarters of the way up the upstream face. The test beam is also angled by about  $13^\circ$  towards the beam right side in the  $x$ - $z$  plane, and by roughly the same amount downwards in the  $y$ - $z$  plane.

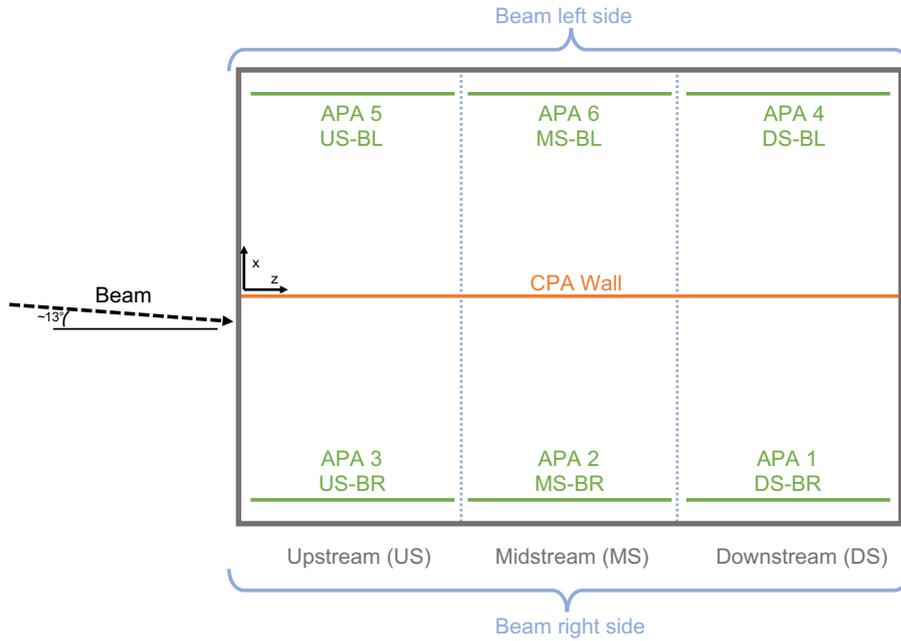


Figure 3.13: Diagram showing a top down view of the detector with the upstream, midstream, downstream, beam left, and beam right areas labelled. The  $x-z$  plane origin is located on the upstream edge of the CPA wall, with  $x$  increasing towards the beam left side of the detector. The beam is shown as entering the detector with a slight offset towards the beam right side, and with an angle of  $\sim 13^\circ$  to the central CPA wall.

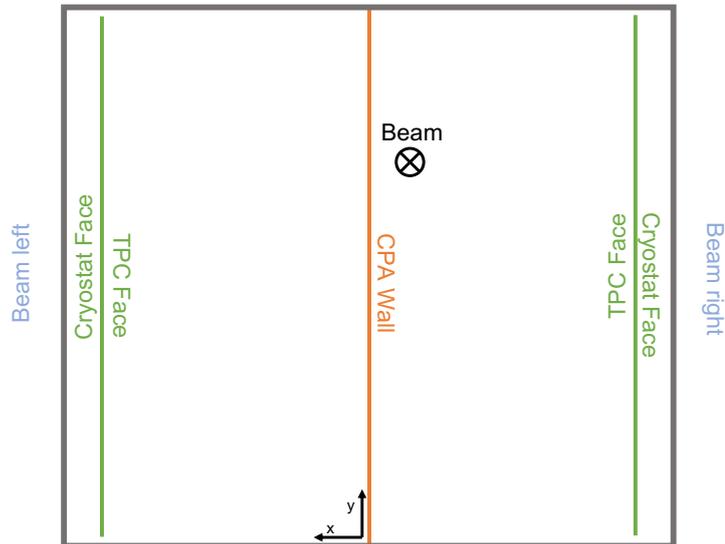


Figure 3.14: Diagram showing the upstream face of the detector, labelling the TPC and cryostat faces of the APAs. The entry point of the beam is also shown slightly offset to the beam right side, and about three-quarters of the way up the upstream detector face. The  $x-y$  plane origin is located at the bottom of the CPA wall, with  $y$  increasing vertically, and  $x$  increasing towards the beam left side.

### 3.5.2.1 Online Monitor

An important system for any detector is that of the online monitoring system used during data taking runs. Such a system must provide a suite of informative plots that provide any collaborator working in the control room with a good overview of the detector's current state. A particular aspect of the online monitor that was used very frequently for ProtoDUNE-SP is that of the event displays. A particularly important set of plots are the event displays for the wire planes of the APAs.

Each wire plane of an APA has its own event display in the online monitor, displaying a 2D histogram of a wire channel number against the detector time tick, with the intensity of each bin being the amount of charge collected. Alongside the event displays for each wire plane of the APAs are a number of special event displays, for example Fig. 3.15 shows the beam window event display. The beam window event display is a zoomed in look at the wires and time window of APA-3 that correspond to the volume of the TPC where the beam particles enter. As seen in Fig. 3.15, this particular event display can show very clean images of the interactions of the beam particles, and informs collaborators in the control room that the detector is correctly measuring the charge. The saw like structure seen on the left hand side of the event display is due to a bad timing clock on the motherboard that the wires are connected to, and is easily correctable in the later data reconstruction stages.

Some maintenance, however, is required to keep the event displays in a useable condition. As part of the event display creation, a pedestal value is subtracted from the collected charge of each wire to reduce noise on the wires. The pedestal values are calculated from a recorded time window when no test beam is being provided to the detector, and then stored locally. Over time, the pedestal values will drift from the ones previously calculated and stored. Fig. 3.16 shows what happens to the event display as the pedestal values drift. This is easily correctable by performing recalculations of the pedestal values on a frequent basis - ideally it should be at the start of every run, or once a day, whichever is more frequent.

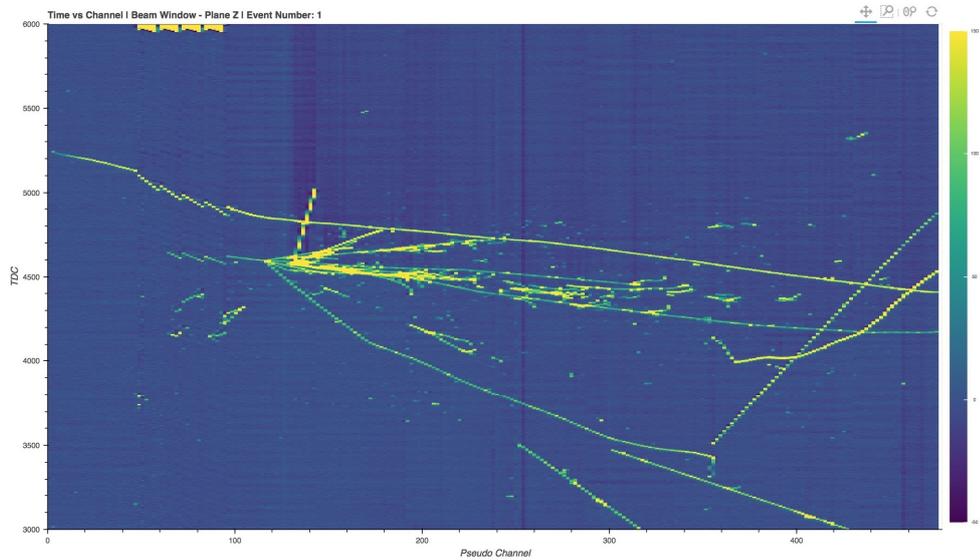


Figure 3.15: An example event display as seen on the online monitor’s webpage. This particular display shows the area of the TPC where particles from the test beam enter the detector. Collaborators find such a display useful as they can easily see the quality of data being recorded in the most important area of the detector. The colour scheme indicates collected ionisation charge.

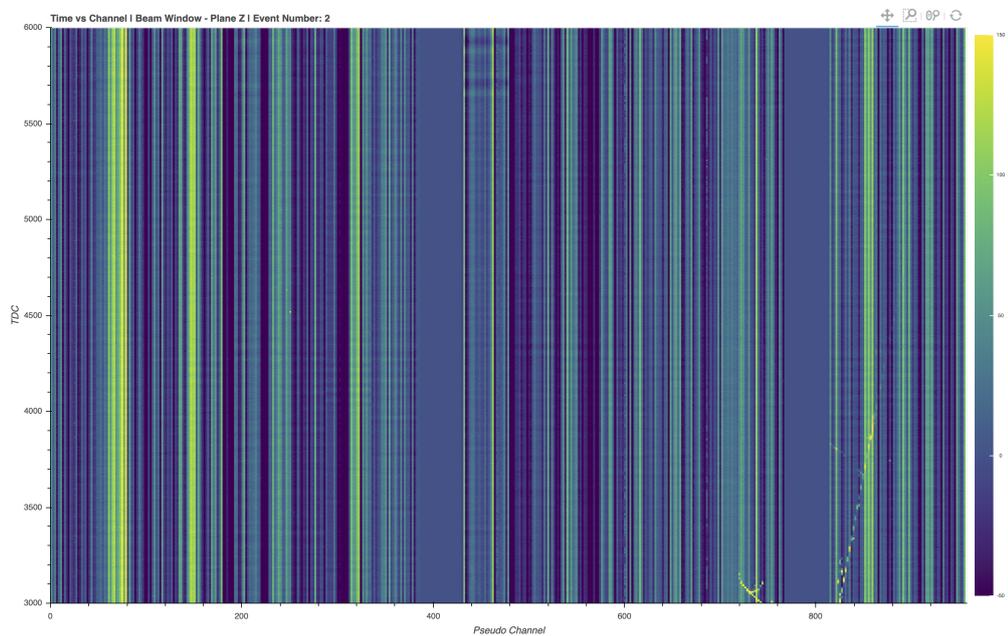


Figure 3.16: An example of what can happen to the event display if the pedestal values for the online monitoring system are not updated regularly. Each vertical line of colour represents a single wire. The colour scheme indicates collected ionisation charge.



# Particle Reconstruction and Identification

---

---

*“From the top of the mountain, you cannot see the mountain.”*

---

Frank Herbert - **Dune** - 1965

**I**n all modern particle physics experiments, the design, building, and measurement taking capabilities of a detector is only half the battle. The ability to reconstruct particle trajectories, and calculate their relevant kinematics, using fully automated algorithms is a battle of constantly increasing complexity as the detection techniques advance even further.

## 4.1 Monte Carlo Simulations

A key concept in experimental high energy physics is that of the Monte Carlo (MC) Simulation. MC simulations are very powerful methods of estimating the outcomes of the stochastic and probabilistic interactions that occur in the realms of particle physics. When simulating experimental data with the MC method it is often done in two stages. The first stage, event generation, sees a set of kinematic properties given to a number of primary particles. These properties will follow probabilistic physical laws, such as scattering angles, and will provide an initial state for each particle in the second MC stage. The second stage will use the initial particle states

as an input for a simulated detector, where MC methods are used to determine the interactions and detector responses to the initial particles.

The simulated MC data is then used in many different ways. In the first stage of an experiment, it can be used as a proof-of-concept and test-bed for detector designs. In late stages, MC simulations can be used to assess detector performance, data analysis techniques, or removing irreducible backgrounds. Experimental data can also be fed back into the MC simulations to improve them further, as no MC simulation is a perfect description of all physics. Often MC generators are specialised to a particular niche of particle physics interactions, and sometimes multiple different generators cover the same area.

For the DUNE FD MC, the event generator used is GENIE [55], a universal neutrino event generator, and GEANT4 [56–58] is used to simulate the detector response. For ProtoDUNE, a custom event generator [59] is used that generates beam particles after the SPS extracted proton beam is impacted on the beryllium target, as per sec 3.5.1, using information based on previous results [60]. The second stage is then split in two, with the GEANT-4 based framework G4BeamLine [61] simulating the secondary beam target and the particle transport in the beam pipe, and then a separate GEANT-4 simulation for detector response.

## 4.2 Data Reconstruction

The ability to reconstruct particles from measuring charge drifting by, or being collected on, a wire is a task with no one single solution. However, all solutions require algorithms that are adept at the innate human ability to recognise patterns. While being able to tell the difference between a cat and a dog can be instantaneous for any human, it is a process that requires the recognition of a series of features and patterns that distinguished the two.

Asking any individual to explain how they know a cat from a dog, and you will likely receive an answer similar to ‘Well, one looks like a cat, and one looks like a dog’, without them realising that they looked for a series of subtle patterns that distinguish them. It is this decomposition of a problem into a series of smaller pattern recognition aspects that sits at the heart of data reconstruction in particle physics. Building an algorithm that looks at all the data at once, and identifies

individual particles would be hard to build, and likely make lots of mistakes<sup>1</sup>. It would be better to have a suite of algorithms, that slowly build upon each other until the full image is complete. This is known as the multi-algorithm paradigm, and is followed by the leading LArTPC reconstruction software, Pandora [62–64] (Sec. 4.5).

Before being able to begin pattern recognition however, an event data model is required to store all the collected experimental data.

### 4.3 The *art* Framework

The *art* framework [65, 66] was developed to provide a universal base on which current and future Fermilab experiments could build on, greatly reducing start-up efforts and allowing greater cohesion between experiments. The design of *art* has been largely driven by the CMS framework, which many of the *art* developers have been involved with designing also. The *art* framework provides a method for running code modules, often written by collaborations or individual users, in order to simulate, reconstruct, or analyse data. The framework also provides a number of utility classes that can connect data across stages of an *art* program. The running of an *art* program is controlled by the custom made Fermilab Hierarchical Configuration Language (FHiCL). FHiCL is a very accessible language, and allows for complete control over the sequence and configuration of all invoked modules and utilities.

Importantly, *art* also provides a way to encapsulate all related data products into an *event*. An *event* is usually defined as a particular period of time by each individual experiment, but usually follows the time before and after some form of trigger. For ProtoDUNE-SP an *art event* corresponds to a 3 ms time window surrounding the beam trigger. Data products contained in an *event* can also be bi-directionally linked by an *art association*.

### 4.4 The LArSoft Framework

LArSoft is a software framework, maintained by the LArSoft Collaboration [67], providing a set of tools, algorithms, and utilities for LArTPC based detectors. LArSoft also provides all the data products that are contained within an *art event*. While LArSoft contains many hundreds of different data products, used by a multitude of different algorithms through any complete reconstruction chain, there are a number that are key for analysis work. Table. 4.1 lists and describes the data products that are of particular relevance.

---

<sup>1</sup>Deep learning algorithms attempt to combat this, and a lot of research is being done into their applications in particle physics.

Table 4.1: Important LArSoft data products, and their descriptions.

Data Product	Description
<i>MC Particle</i>	An object containing the truth information for simulated particles.
<i>Hit</i>	A 2D representation of the charge deposited on a wire.
<i>Cluster</i>	A collection of <i>hits</i> that are grouped together by pattern recognition algorithms.
<i>Space Point</i>	A 3D point created from matching <i>hits</i> across wire planes.
<i>PFPparticle</i>	An object made from groups of <i>clusters</i> , and the primary output of Pandora. A <i>PFPparticle</i> contains hierarchical information connecting it to a flow of <i>PFPparticles</i> .
<i>Track</i>	A high level fitted object created from a <i>PFPparticle</i> that has track like properties.
<i>Shower</i>	A high level fitted object created from a <i>PFPparticle</i> that has shower like properties.

The base data product for each event is the 2D *hit*, which serves as the input for the Pandora reconstruction chain - an example of which can be seen in Fig. 4.1. Each *hit* contains information regarding the charge measured, time of measurement, and wire number and plane. From this starting point, pattern recognition can be used to group hits into clusters, which can then be formed into *PFPparticles*. The *PF* in *PFPparticle* stands for ‘Particle Flow’, which is a method for linking particles through hierarchies to provide more information about the flow of particles in an *event*. Comparing 2D *hits* across all three planes of an APA, it is possible to form 3D hits called *Space Points*. This process requires 2D *hits* on multiple planes to be correlated, and so the number of *Space Points* does not always match the number of hits for a particular particle. Once a *PFPparticle* is made, it is possible to perform some form of track/shower identification, and then fill the corresponding data product with relevant particle properties. The *PFPparticle*, *track*, and *shower* are the primary data products that an analyser will likely access.

The *MC Particle* data object comes from the generation stage of an MC simulation. They are the objects that are propagated through the detector simulation, and generate the drift charge used for the creation of *hits*. As such, they contain the truth information for what was actually simulated by the MC simulation.

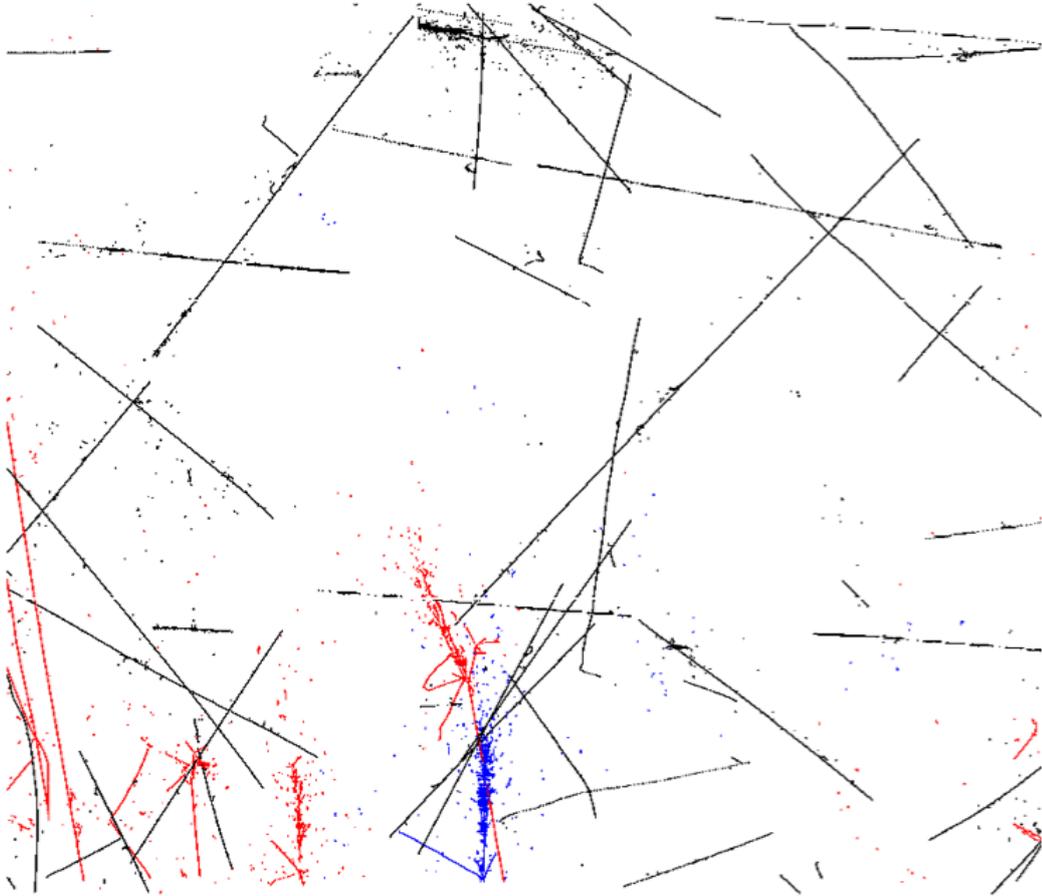


Figure 4.1: An example of the input hits from the collection planes of the APAs that get passed to Pandora. The hits in blue, red, and black represent hits caused by the test beam particle, test beam halo particles, cosmic-ray particles respectively [64].

## 4.5 The Pandora SDK

The Pandora software development kit (SDK) [62–64] is a solution to the problem of complex pattern recognition posed above by using a multi-algorithm approach to slowly build up a full picture of the particles, and interactions going on, inside a detector at any given moment. Initially developed for use on data from the International Linear Collider, Pandora has grown over the past 13 years to become the leading reconstruction software used in LArTPCs. The extreme fine granularity of LArTPCs, and the subsequent photograph like images produced, provides a challenging, yet rewarding, medium with which to really showcase the power of the Pandora multi-algorithm approach. The move from collider to LArTPC based experiments is also a real testament to how flexible the framework can be.

Pandora’s multi-algorithm approach to pattern recognition and event reconstruction is designed so that no downstream algorithm should have to correct errors of previous algorithms. This, in ways, mirrors the Unix philosophy of creating simple algorithms that attempt to do one thing well, and to not become bloated and overcomplicated. To achieve this, Pandora uses up to 75 algorithms on the 2D images that are created from the charge induced or collected on a LArTPC’s wires. The order that the algorithms are applied is completely configurable if needed<sup>2</sup>.

### 4.5.1 Pandora Reconstruction Chains

The Pandora team have created a consolidated reconstruction chain that emphasises the power of the multi-algorithm approach, by applying algorithm chains in a logical order designed to ensure the correct reconstruction for a particle. To facilitate this, two algorithm chains have been devised - Pandora Cosmic, and Pandora Test Beam.

#### 4.5.1.1 Pandora Cosmic

Pandora Cosmic is an algorithm chain tuned to reconstructing cosmic-ray like particles. The reconstruction chain targets track-like particles, and assumes that they are cosmic-ray muons. This means that the primary interaction vertex of each track-like particle is assumed to be the y-point nearest the top of the detector, and that all shower-like particles are assumed to be delta ray daughters of a parent track-like particle.

#### 4.5.1.2 Pandora Test Beam

Pandora Test beam is a slight modification to the Pandora Neutrino algorithm chain. The Pandora Neutrino algorithm is designed to find a primary neutrino interaction vertex, and then reconstruct the daughter particles emerging from it. The Pandora Test Beam algorithm adds an additional algorithm at the end of the chain that identifies the incident test beam particle by using knowledge of the location and direction of the test beam, and sets it as the primary particle for that hierarchy. This can be seen in Fig. 4.2.

#### 4.5.1.3 Particle Stitching

All particles are initially reconstructed assuming a drift start time that corresponds to the triggered test beam particle, from which two useful scenarios are possible when trying to identify true cosmic-ray particles. The first immediately obviously

---

<sup>2</sup>However, some algorithms must come before others. For example 2D reconstruction must come before 3D reconstruction.

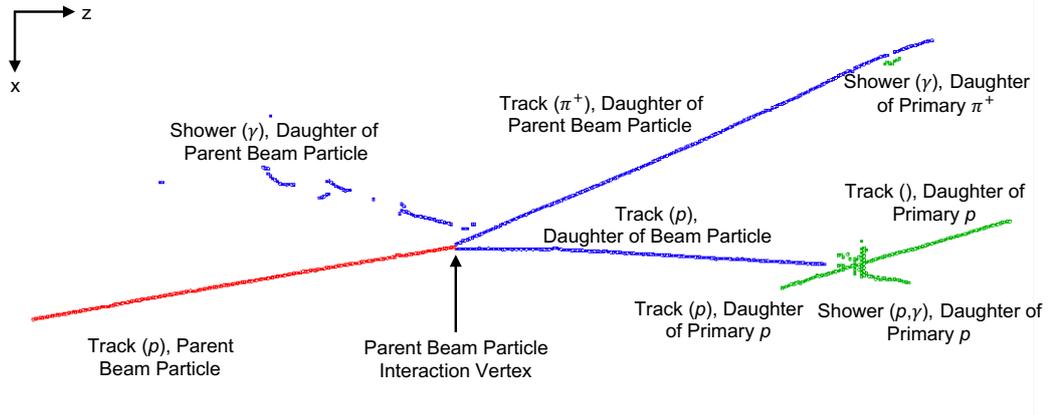


Figure 4.2: An example of a reconstructed test beam particle. The colours of particles represent the hierarchy, with red being the initial incident test beam particle, blue the daughters of the test beam particle, and green the granddaughters [64].

situation is where a particle track appears outside of the physical drift volume of the detector. This is caused from a cosmic-ray traversing the detector out of sync with the triggered beam particle, making it appear that the charge from the cosmic-ray particle is outside of the detector. The second scenario is where a CPA crossing particle can become split, as the direction of drift is mirrored between both sides of the CPA. This concept is illustrated in Fig. 4.3.

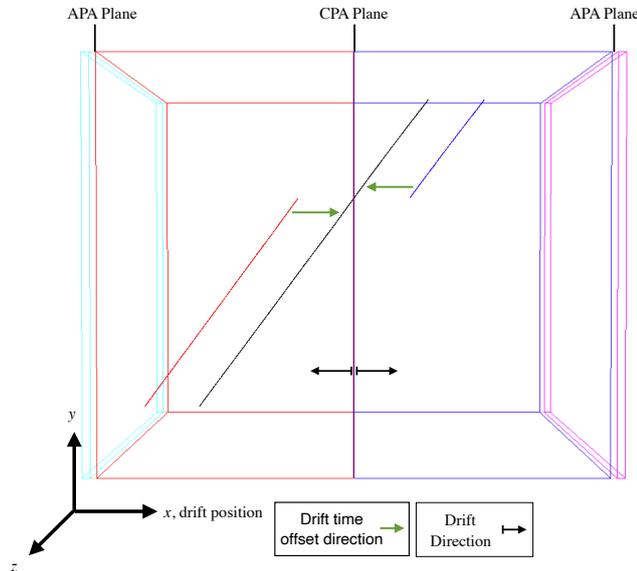


Figure 4.3: An example of how the stitching works for CPA crossing cosmic-rays. The red and blue tracks are the initial reconstructed particles, the green arrows show the direction of the drift time offset, and the black track represents the stitched together cosmic-ray [64].

When a particle crossing the CPA plane is split in this way, it is possible to attempt to ‘stitch’ the two individual tracks back together. If two reconstructed particles are found to have parallel direction vectors pointing towards or away from the central CPA wall, then a drift time offset is applied in opposite directions to each particle. If the separate particle tracks are found to line up at the CPA wall boundary, with drift time offsets of equal magnitude, then the two particles can be stitched together into one CPA crossing particle.

#### 4.5.1.4 Pandora Consolidated reconstruction

Together, the Pandora Cosmic and Pandora Test beam algorithm chains are combined into a consolidated reconstruction shown in Fig. 4.4. The consolidated reconstruction chain makes use of the multi-algorithm approach by walking the input hits through a decision tree like architecture to ensure the final result is the most appropriate reconstruction.

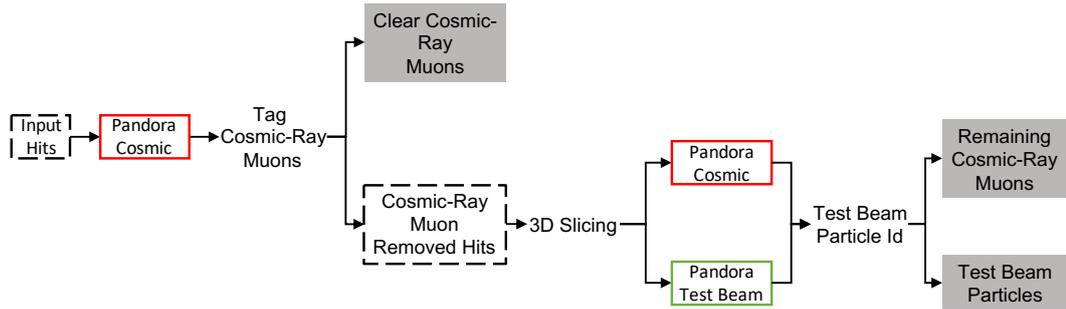


Figure 4.4: The Pandora consolidated reconstruction. The consolidated reconstruction first removes clear cosmic-ray particles by reconstructing all input hits under the Pandora Cosmic hypothesis. The non clear cosmic-ray hits are separated into 3D slices before being reconstructed under both the Pandora Test Beam and Pandora Cosmic hypotheses. The best outcome of which is decided by a boosted decision tree (BDT) [64].

The first algorithm chain to be run by the consolidated reconstruction is the Pandora Cosmic chain, reconstructing all particles under the cosmic-ray hypothesis. At the end of the Pandora Cosmic algorithm chain, three criteria are used to label every particle as being a ‘clear cosmic-ray’ or not. A particle is labelled as a clear cosmic-ray if [64]s:

- it crosses the CPA plane and can be stitched with a drift time offset of greater than  $6.2 \mu\text{s}$ , corresponding to a shift of 1 cm.

- using the triggered test beam timing, a particle’s hits appear outside of the detector.
- it enters the detector through the top face, and exists through the bottom face.

Any particle labelled as a clear cosmic-ray will have its constituent hits removed from the hits that are passed onto the following stage of the reconstruction.

The next stage of the consolidated reconstruction starts before a 3D slicing on the hits labelled as not clear cosmic-ray hits. The 3D slicing works by running a fast reconstruction where reconstructed 3D particles are used to isolate hierarchically linked particles into their own slices of the detector. Thus, one 3D slice will contain a single primary particle and its daughter particles, as determined by the fast reconstruction. The 3D slices are then passed onto another Pandora Cosmic instance, as well as a Pandora Test Beam instance. This provides two reconstructed outputs for each 3D slice, and a BDT (Boosted Decision Tree) is used to determine which reconstructed outcome is the most sensible to persist. The BDT uses features that are justified by the well understood position and direction of the test beam’s entry into the TPC [63, 64].

#### 4.5.2 Track/Shower Separation

Once Pandora has created a *PFP* it attempts to categorise it as either shower-like or track-like using a cut based approach. To do this, a sliding linear fit is performed on the clusters constituting the *PFP*. This is used to assess how well a straight line approximates the particle’s trajectory, and how the width - the particle’s spread perpendicular to the direction vector - of the particle changes along its length. By comparing ratios of the change in width, or the distance to the particle’s vertex, to the length of the linear fit with a predetermined optimised cut value, a reasonable track/shower separator can be made.

A new BDT approach is also being developed for Pandora. This new approach takes features such as length, like in the cut based approach, but also incorporates hierarchical information, like number of daughters, and calorimetric information, such as various charge ratios. The BDT approach has been shown to quickly outpace the performance of the cut based approach. This can be seen in Fig. 4.5, where accuracy is defined as ratio of correctly identified track-like and shower-like particles to total number of particles.

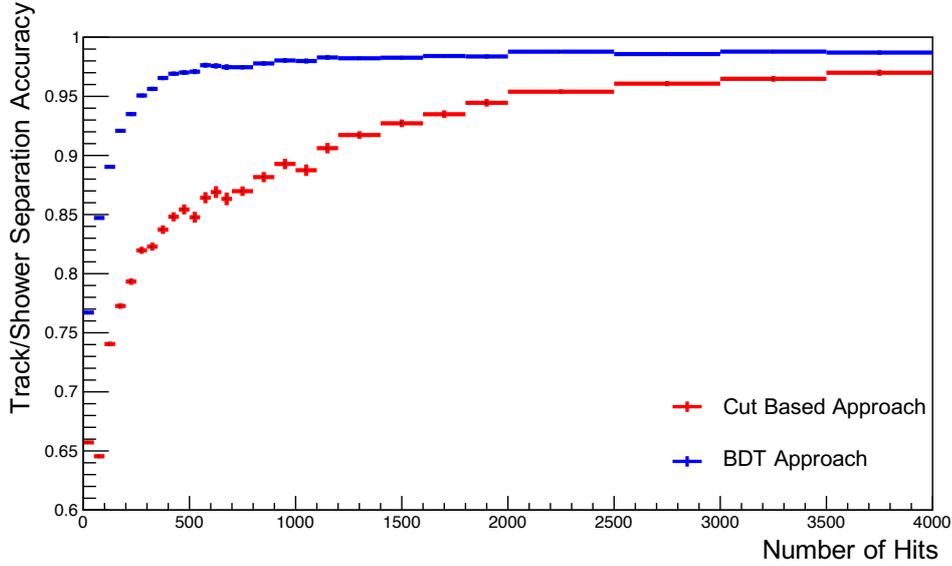


Figure 4.5: A comparison of Pandora’s original cut based approach (red) with the newer BDT approach (blue) to track separation. The BDT approach clearly has a much better accuracy at lower numbers of hits.

### 4.5.3 Performance

Pandora primarily uses three metrics to determine its reconstruction performance for simulated data:

$$Efficiency = \frac{Number\ of\ MC\ particles\ matched\ to\ at\ least\ one\ reco\ particle}{Total\ number\ of\ MC\ particles}$$

$$Purity = \frac{Number\ of\ hits\ shared\ between\ matched\ MC\ and\ reco\ particle}{Total\ number\ of\ hits\ in\ reco\ particle}$$

$$Completeness = \frac{Number\ of\ hits\ shared\ between\ matched\ MC\ and\ reco\ particle}{Total\ number\ of\ hits\ in\ MC\ particle}$$

When calculating the efficiency, a reconstructed particle is matched to the MC particle with which it shares the largest number of hits, provided the purity is above 50% and the completeness above 10%. All matches are used when assessing the purity and completeness. Fig. 4.6 shows the efficiency of Pandora’s reconstruction when presented with hits from the beam particle only (black), beam particle and beam halo particles (red), and beam particle, beam halo particles, and cosmic-rays (blue).

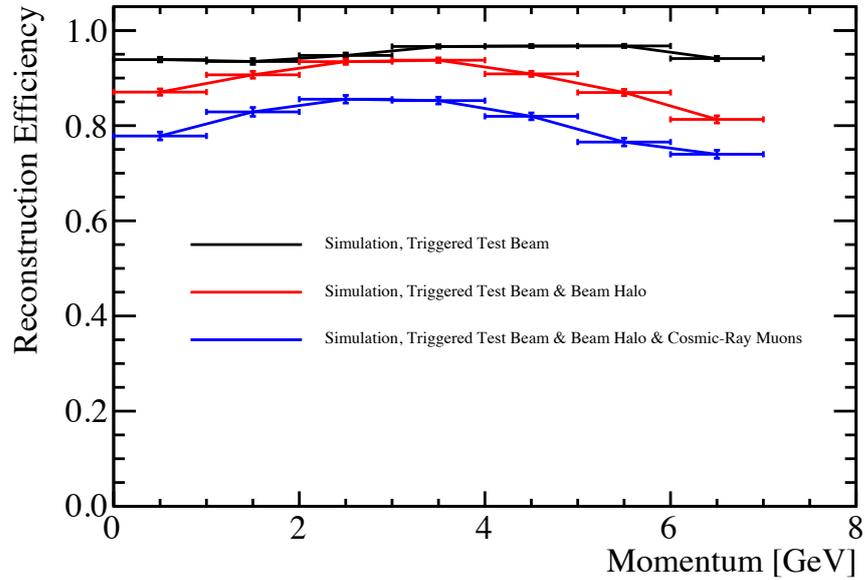


Figure 4.6: A breakdown of Pandora’s reconstruction efficiencies, for all beam particles, with increasing simulation complexity. Removing cosmic-ray muons and beam halo particles shows the scale of effect they have on decreasing the reconstruction efficiency. Across all momenta, the cosmic ray-muons decrease efficiency by a constant amount, while the beam halo particles have a much more pronounced effect at the higher momenta [64].

The figure shows how at the highest momenta, the reconstruction efficiency is mostly degraded by the beam halo particles, while around 3 GeV the beam halo particles have almost no effect. The cosmic-rays are shown to degrade the reconstruction efficiency by an almost uniform amount across the whole momentum spectrum.

When assessing the reconstruction efficiency of experimental data it is necessary to construct a different efficiency to the one described above, due to no truth information existing for experimental data. The efficiency definition is thus modified to count the number of beam triggers (from the beamline information in data, and from the MC particle hierarchy in simulation) that have a reconstructed beam particle, as a fraction of the total number of beam triggers. Using this definition of efficiency produces Fig. 4.7 for data (red) and simulation (black). The decreased efficiency in data for low momenta is due to a triggering beam particle not leaving a strong enough signal in the TPC for Pandora to realistically be able to successfully reconstruct it.

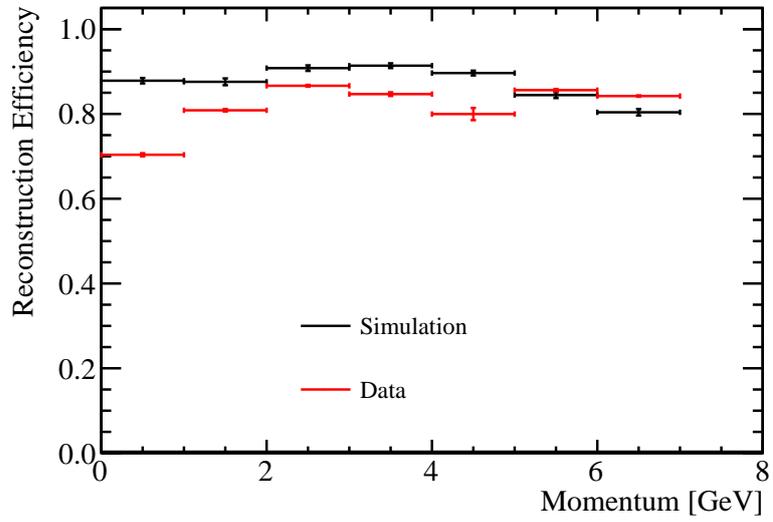


Figure 4.7: Pandora’s reconstruction efficiency against true particle momenta for real (red) and simulated (black) data. The simulated data’s efficiency differs from the blue line in Fig. 4.6 because of the different definition used for efficiency as described in the final paragraph of Sec. 4.5.3 [64].

# 5

## Characterisation of EM Showers

---

*“A process cannot be understood by stopping it. Understanding must move with the flow of the process, must join it and flow with it.”*

**Frank Herbert - Dune - 1965**

---

**E**lectromagnetic (EM) showers are QED processes where a large cascade of propagating particles is induced by a single initial particle. The cascade is initiated either by a high energy electron\* undergoing bremsstrahlung, or a high energy photon annihilating through pair production. As the cascade of particles propagates through the material, undergoing further pair production and bremsstrahlung, the energy of each individual particle slowly decreases until the dominant energy loss mechanism becomes ionisation, where the EM shower will slowly die out.

EM showers are typically discussed using two parameters: radiation length ( $X_0$ ) and Molière radius ( $R_M$ ). The radiation length is defined as the amount of material a high-energy electron will pass through before losing all but  $\frac{1}{e}$  of its energy through bremsstrahlung, and is often used as the length scale when describing EM showers. The Molière radius describes a cylinder in which one would expect to find 90% of an EM shower’s energy, and is defined relative to the radiation length as  $R_M = X_0 E_s / E_c$ , where  $E_c$  is the critical energy of the material, and  $E_s \approx 21 \text{ MeV}^\dagger$  [34]. The critical energy is the energy below which ionisation becomes the dominating energy loss mechanism. For LAr,  $E_c = 30.5 \text{ MeV}$  [68].

---

\*Electron is used here to represent both electrons and positrons.

†Material independent scaling factor.

An example diagram depicting how a shower initiated by an electron might propagate is shown in Fig. 5.1. As depicted, the electron will travel for one radiation length, shown by the vertical dashed lines, before radiating a photon through bremsstrahlung. After another radiation length, one would expect the bremsstrahlung photon to pair produce an electron and a positron, and the initial electron would radiate another photon. This process would repeat until the ionisation becomes dominant.

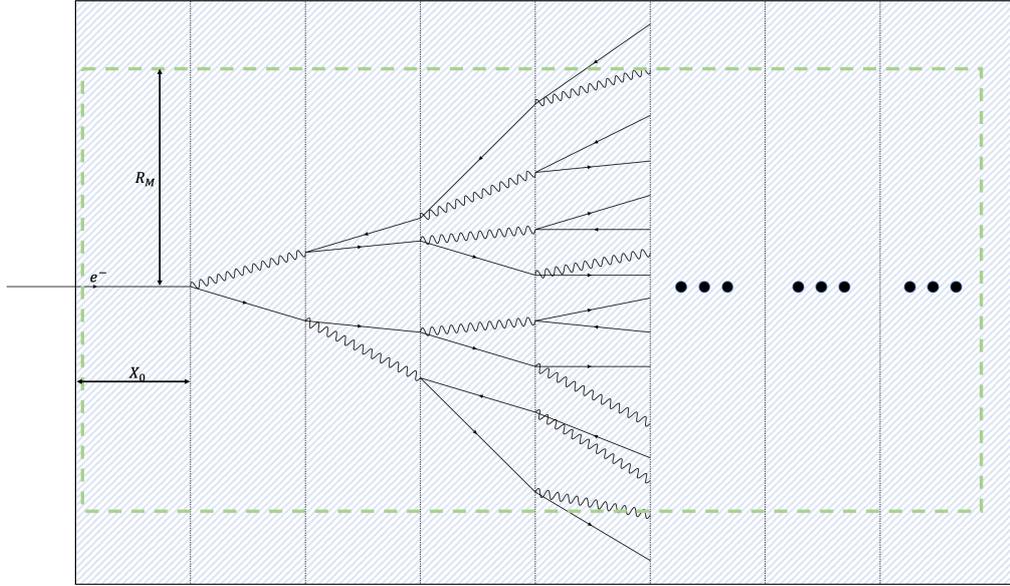


Figure 5.1: An example diagram of an electromagnetic shower induced by an incoming electron. After traversing one radiation length ( $X_0$ ) of a medium, indicated here by the blue hatched region, the electron radiates a bremsstrahlung photon. This process repeats until the EM shower dies out due to energy losses through ionisation.

## 5.1 Datasets

Three datasets have been used for the analysis presented in this chapter; a single particle gun MC dataset, a ProtoDUNE-SP beam simulation MC, and real ProtoDUNE-SP beam data.

The ProtoDUNE-SP beam data (PDSP-BD) used in this analysis are listed in Table 5.1. These data taking runs have been selected as they are free from instabilities in the ProtoDUNE-SP hardware that occurred during some other data taking runs. These runs provide a statistically significant sample of events per momentum interval. The ProtoDUNE-SP beam simulation MC (PDSP-MC) used in this chapter are the ‘Production 2’ MC found under the Fermilab SAMWeb definitions in Table 5.2.

Table 5.1: Beam data run numbers.

Momentum (GeV/c)	Run Numbers
1.0	5809, 5816, 5817
2.0	5824
3.0	5777, 5779, 5785, 5786
6.0	5770, 5771
7.0	5143, 5145, 5204

Table 5.2: ProtoDUNE-SP beam simulation MC datasets.

Momentum (GeV/c)	SAMWeb Definition
1.0	PDSPProd2_MC_1GeV_reco_sce_datadriven
2.0	PDSPProd2_MC_2GeV_reco_sce_datadriven
3.0	PDSPProd2_MC_3GeV_reco_sce_datadriven
6.0	PDSPProd2_MC_6GeV_reco_sce_datadriven
7.0	PDSPProd2_MC_7GeV_reco_sce_datadriven

As briefly discussed in section 4.1, the PDSPP-MC generates beam particles 2 m upstream of the TPC, including beam halo particles, and using the G4BeamLine framework transports the particles to the detector where GEANT-4 is used to simulate interactions and energy deposits in the TPC. Two LArSoft algorithms are then used to simulate ionisation charge and scintillation-photons. A cosmic ray background is simulated using either CORSIKA [69], or CRY [70]. The space charge effects SCE are applied using a map, made using SPaCE [71], of distortions to true position.

A single particle gun MC dataset is also used throughout parts of this chapter. The beam particles in this MC consist solely of electrons, and are generated at the beam window position on the upstream face of the TPC. The initial position, direction, and momentum of each initial electron is also subjected to a 5% smearing. This MC also does not contain any cosmic ray background, or any SCE. This MC is generated using the v08\_40\_00 version of dunetpc [72].

## 5.2 Space Charge Effect

The  $\vec{E}$ -field in a TPC should, nominally, be uniform throughout the detector. The uniformity of the  $\vec{E}$ -field is critical in ensuring accurate reconstruction of a particle's position and trajectory. However, as particles traverse the detector, slow moving positive ions build up inside the TPC leading to distortions in the electric field. As the build up of positive ions is caused by particles ionising the LAr, the effect is very notable in ProtoDUNE-SP as a surface detector. As the ProtoDUNE-SP beam is also not pure, each beam spill comprises of multiple particles and particle species, worsening the effect around the beam's entry position into the TPC. As such, the most upstream section of reconstructed particles is the most severely affected. As the positive charge is attracted to the central cathode, and constantly replenished by continual cosmic-ray interactions, the ionisation drift electrons are pulled towards the cathode slightly as they proceed to the anode. This can cause a bowing towards the cathode of subsequent reconstructed particles, as well as a squeezing in the  $z$ -direction. Using the SCE map it is then possible to correct the position of reconstructed hits. Fig. 5.2 shows the effect of correcting the position for

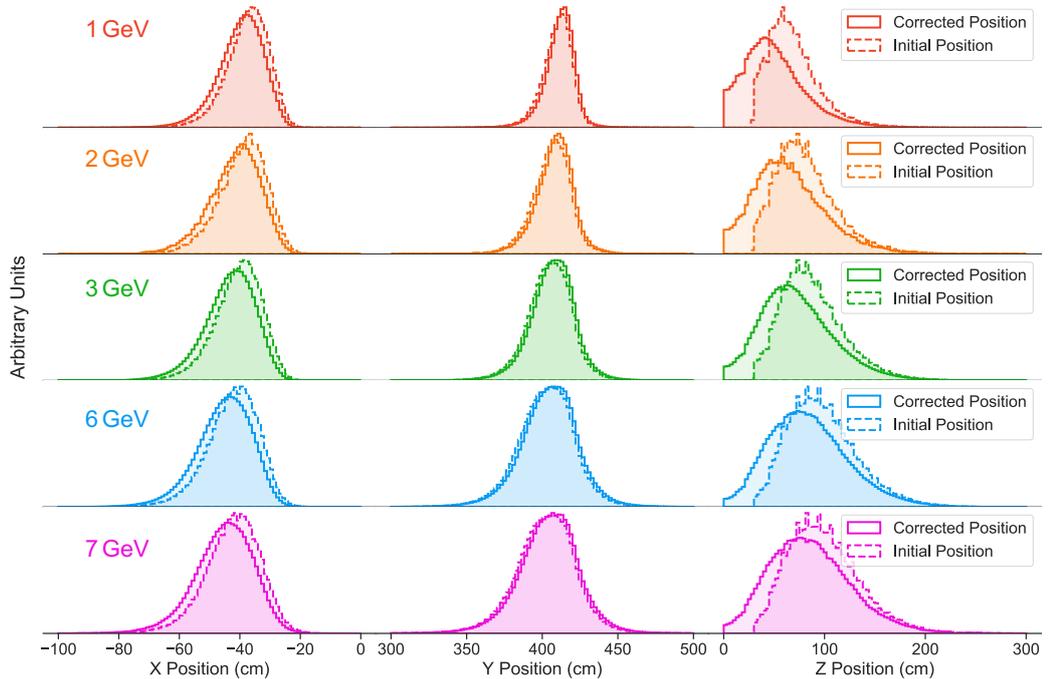


Figure 5.2: The uncorrected and corrected positions for 3D space points in the PDSP-MC dataset. The impact of SCE is most notable in the  $z$ -positions due to the squeezing of the drift charge.

the different momentum in the PDSP-MC dataset. The difference in the  $z$ -positions

demonstrates how large the z-direction squeezing is, and how the effect is greater nearer the edge of the TPC.

### 5.2.1 Event Selection

A LArSoft analyser is used to select events where the primary reconstructed particle is the initial electron. Doing so in the single particle gun MC is trivial, but the other two MCs require a more in-depth solution. The PDSP-MC requires matching reconstructed particles to the GEANT4 simulation, and the PDSP-BD requires using beam matching tools that utilise information from the various beam monitors operated by CERN’s beam group. The beam’s time of flight monitors and Cherenkov detectors are used to select electron candidates from the recorded events. The reconstructed events are matched back to information from the beam monitoring system by a LArSoft utility that compares the beam trigger’s timestamp with time of flight information. The analyser is written with the v08\_39\_00 protoduneana library of analysis tools.

Beyond selecting the electron candidates in the PDSP-BD dataset, it is important to also select EM showers that start showering inside the TPC. As the single particle gun MC electrons are generated within the TPC, they can serve as a benchmark for what a complete shower looks like. As seen in Fig. 5.3, two particularly useful

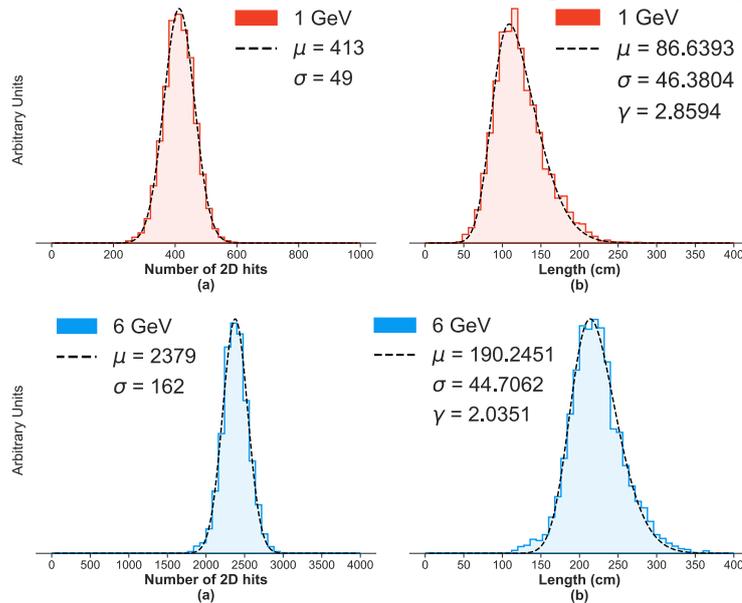


Figure 5.3: Number of reconstructed 2D hits (a) and length (b) for 1 GeV/c (red) and 6 GeV/c (azure) single particle gun MC events.

properties in determining a complete shower are the number of hits associated to, and the length of the EM shower. The distributions in Fig. 5.3 are for 1 and 6 GeV/c single particle gun MC events, and can provide a good benchmark for an understanding of whether selected electron events in the PDSP-MC and PDSP-BD datasets can be considered as fully contained showers. The **(a)** plots of Fig. 5.3 are fitted with a Gaussian function

$$f(x; A, \mu, \sigma) = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (5.1)$$

while the **(b)** plots are fitted with a skewed Gaussian function

$$f(x; A, \mu, \sigma, \gamma) = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left\{ 1 + \mathbf{erf} \left[ \frac{\gamma(x-\mu)}{\sigma\sqrt{2}} \right] \right\} \quad (5.2)$$

where  $A$  is the amplitude of the fit,  $\mu$  the mean,  $\sigma$  the standard deviation,  $\gamma$  the skew, and  $\mathbf{erf}$  the error function.

Using the number of 2D hits distributions in Fig. 5.3, and similar distributions for other momenta, a minimum number of 2D hits threshold can be set on the PDSP-MC and PDSP-BD datasets. The threshold being that all reconstructed electron candidates with less 2D hits than this threshold will be considered as not fully reconstructed (due to showering before the TPC), and will be rejected from the selection. The value for these hit thresholds can be seen in Table 5.3. The effect of applying a 2D hit threshold to the 1 and 6 GeV/c PDSP-MC and PDSP-BD can be seen in Fig. 5.4, where the two Fig. 5.4(a)s show a peak at lengths of 0 cm due to electrons that start showering before they reach the TPC. The two Fig. 5.4(b)s show how after applying the 2D hit threshold, the peaks near 0 cm are removed, and the samples are left with EM showers that start within the TPC.

Table 5.3: Number of 2D hit thresholds for PDSP-MC and PDSP-BD datasets.

Momentum (GeV/c)	Hit threshold
1.0	250
2.0	500
3.0	900
6.0	1800
7.0	2000

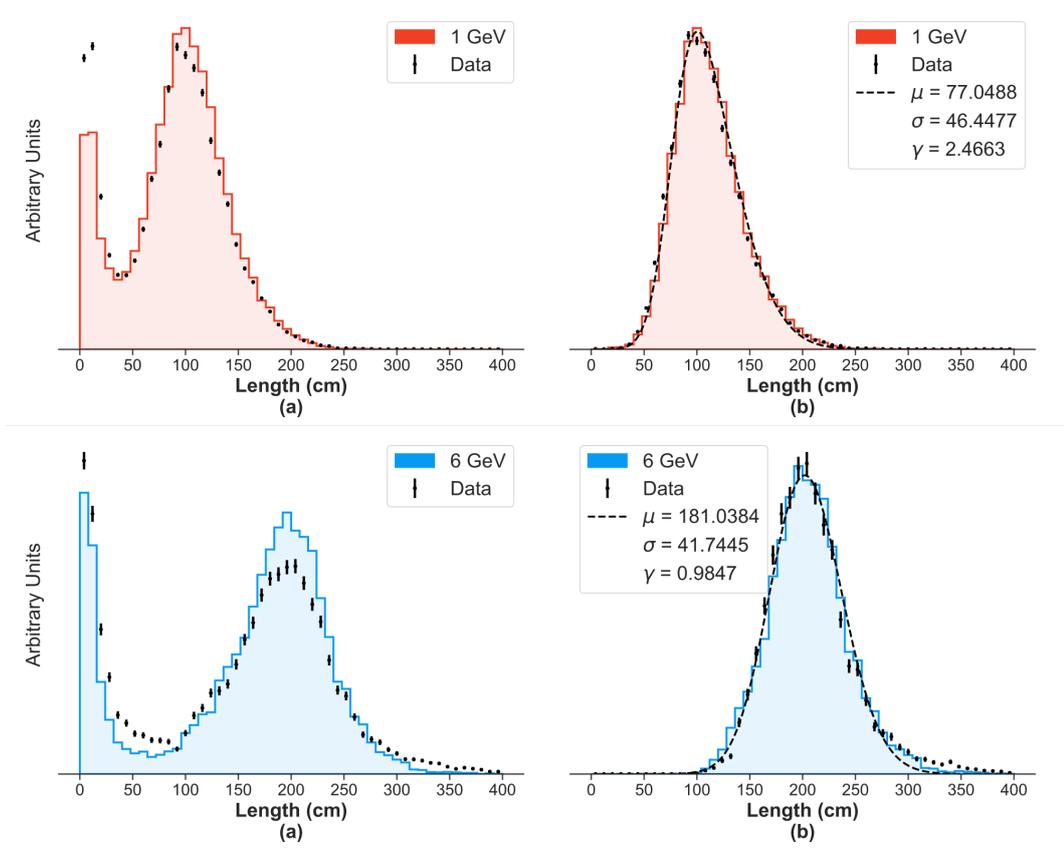


Figure 5.4: Length distribution before (a) and after (b) applying a 2D hit threshold to events in the PDSP-MC and PDSP-BD datasets. The fits in (b) are computed for the PDSP-MC distributions, and use the skewed Gaussian function of Eqn. 5.2.

### 5.3 Shower Principal Components Analysis

During the Pandora reconstruction chain, the 3D space points are used to perform a principal components analysis (PCA). The PCA is performed using the C++ template library *Eigen* [73], the outputs of which are three eigenvector-eigenvalue pairs. These eigenvector-eigenvalue pairs are used to form a coordinate system unique to each individual reconstructed shower object, which is used as the basis for calculating many kinematic variables (such as direction and length) when analysing EM showers. An example of the eigenvectors calculated by a PCA of a toy 2D dataset can be seen in Fig. 5.5.

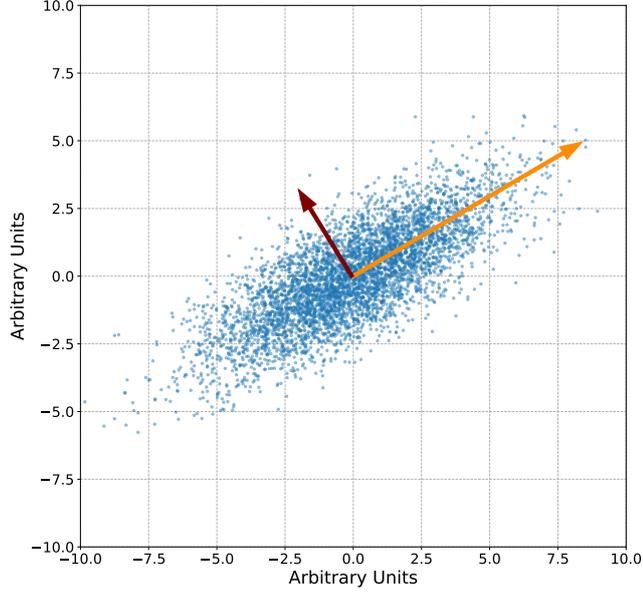


Figure 5.5: An example of the resultant eigenvectors, orange and maroon arrows, for a distribution of points, with the length of the arrow being three times the square root of the eigenvector's corresponding eigenvalue.

### 5.3.1 Pandora's PCA

The PCA is performed by the method of eigenvalue decomposition of the covariance matrix for the three spatial coordinates of the shower's space points. First assume that  $N$  space points are described by vectors  $\vec{s}_i = (x_i, y_i, z_i)$ , where  $i = 0 \rightarrow N$ . The space point vectors  $\vec{s}_i$  are then arranged into an  $N \times 3$  matrix,  $\mathbf{S}$ . The mean for each column of  $\mathbf{S}$  can be calculated and formed into a mean vector using

$$\vec{\mu} = \frac{1}{N} \sum_j^N \mathbf{S}_j \quad (5.3)$$

where  $j$  is the row index of the matrix  $\mathbf{S}$ . The mean vector is then subtracted from all row vectors in  $\mathbf{S}$  such that

$$\mathbf{P} = \mathbf{S} - \mathbf{h}\vec{\mu} \quad (5.4)$$

where  $\mathbf{h}$  is an  $N \times 1$  column vector of all 1s, and  $\mathbf{P}$  contains the  $N$  row vectors  $\vec{s}_i - \vec{\mu}$ , translated such that their mean position is now the origin. The covariance matrix of the space points is then calculated as

$$\mathbf{C} = \frac{1}{N} \mathbf{P}^T \mathbf{P} \quad (5.5)$$

where  $\mathbf{P}^T$  is the transpose of  $\mathbf{P}$ . The covariance matrix  $\mathbf{C}$  will then be a  $3 \times 3$  matrix with eigenvectors and eigenvalues such that

$$\mathbf{C} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \quad (5.6)$$

where  $\mathbf{Q}$  is a  $3 \times 3$  matrix whose columns,  $\vec{\mathbf{q}}_k$  with  $k = 1, 2, 3$ , are the three eigenvectors of  $\mathbf{C}$ , and  $\mathbf{\Lambda}$  is a diagonal matrix whose diagonal elements are the corresponding eigenvalues,  $\Lambda_{kk} = \lambda_k$ . The eigenvector-eigenvalue pairs,  $\vec{\mathbf{q}}_k$  and  $\lambda_k$ , are then ordered by magnitude of the eigenvalue and are called, in decreasing magnitude, the primary, secondary, and tertiary components of the PCA. As the components from the PCA are used to describe the coordinate system of a real 3D shower, they will be referred to as axes. Because the eigenvalue for each pair describes the variance along each axis, the spread of space points along each axis can be estimated by

$$\ell_{\lambda k} = 6\sqrt{\lambda_k} \quad (5.7)$$

where  $3\sqrt{\lambda_k}$  represents  $3\sigma$  along the corresponding axis, but because the axis originates from the mean position of the shower it is multiplied by two. An example of a reconstructed shower's 3D space points and their PCA axes are shown in Fig. 5.6, with the PCA axes emanating from the centroid of the shower and of width  $\ell_{\lambda i}$ . In this example it can be seen that the primary axis (red) appears alongside what a human analyser would select as the length of the shower's 3D space points. The secondary (green) and tertiary (blue) axes would be much harder for a human analyser to select, but a PCA is able to identify them. It is worth noting that in Fig. 5.6, and other 3D figures, the secondary and tertiary axes may not appear orthogonal to the reader because of the scaling of the plot's own axes and the perspective chosen by the author.

A useful operation to perform when looking at these 3D images is to project the space points onto two 2D axes. This has been done for the example shown in Fig. 5.6, where the projections of the secondary and tertiary axes are plotted against the projections of the primary axis, and can be seen in Fig. 5.7. Doing this creates a more easily digestible representation of the distribution of the Shower's space points. Fig. 5.7 shows that the secondary axis of the shower has a width roughly three times (ignoring the three potentially anomalous points at 50 cm from the primary axis in the secondary-primary plane) that of the tertiary axis. It is also evident that while the minimally ionising start of the shower lies along the primary axis in the primary-secondary axes plane, it is at a very noticeable angle in the primary-tertiary axes plane. A possible explanation is that the PCA might have settled inside a local minima during computation - but this was found not to be the case. The angle

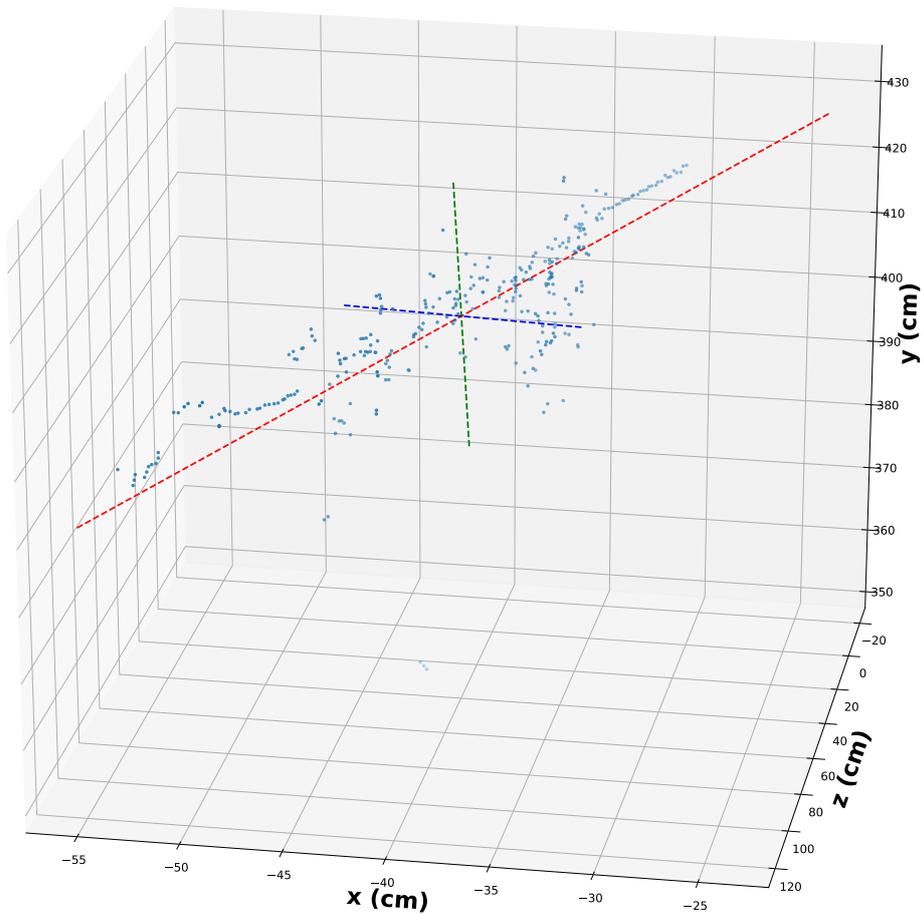


Figure 5.6: An example of a reconstructed 3D shower with its PCA primary (red), secondary (green), and tertiary (blue) axes overlaid. The secondary and tertiary axes do not appear orthogonal to the primary axis because of the different scale plot axes and the perspective of the image.

between the primary axis and the minimally ionising start of the shower is due to the diffuse stage of the developed shower, as all space points are given the same weight in the PCA computation. While this is the desired outcome of a traditional PCA, it is much more desirable from a physics perspective to have the primary axis lie along the initial track-like section of the shower. This is because the angle at which the particle leaves a vertex is an important quantity for various physics analyses, and the direction attributed to each Shower object is derived from the PCA primary axis - as discussed later, in section 5.4. Thus, it is prudent to create a PCA that does bias the primary axis to being along the minimally ionising start of the shower.

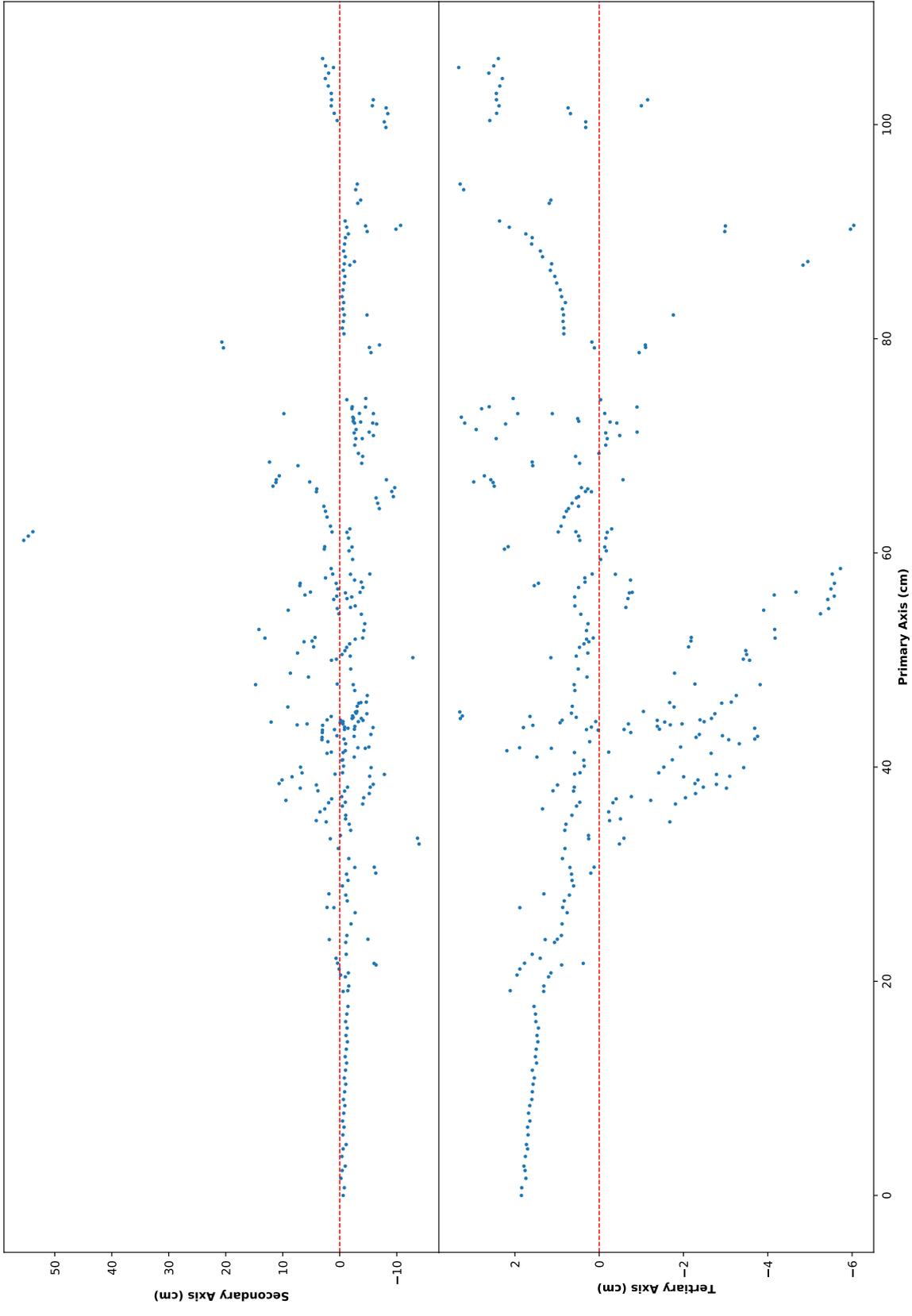


Figure 5.7: A shower that has had its 3D space points projected into the PCA axes coordinate system. The top plot shows the secondary axis projections against the primary projections, and the bottom plot shows the tertiary axis projections against the primary projections.

### 5.3.2 Recursive Weighted PCA

It is possible to bias a reconstructed Shower's PCA by providing a weight to each 3D space point, with the weights being designed in a way that would prioritise certain features of the Shower. For the reasons described in the previous section, it is desirable to produce a weight that would prioritise the initial minimally ionising section of the Shower, and a method using a recursive weighted PCA has been created to do so.

This method builds off of the initial unweighted PCA performed by the Pandora reconstruction software, and recursively performs subsequent PCAs with changing weights until one of two cut-off conditions are met. The recursive weighted PCA method works by calculating a weight for each 3D space point based off of the previously calculated PCA - and so the process must start with an unweighted PCA. The weight for a 3D space point is calculated using

$$w_i = \frac{1}{z_i r_i} \quad (5.8)$$

where  $z_i$  is the projection along, and  $r_i$  the radial distance to, the primary axis of the previous PCA. The origin for the projection is set as being  $\ell_{\lambda_i}/2$  away from the PCA's average position backwards<sup>‡</sup> along the primary axis. The PCA is then recomputed using the new weighted 3D points, and done so recursively until the average normalised weight ceases to change. The average normalised weight for any given PCA is calculated as

$$W = \frac{1}{N w_m} \sum_{i=0}^{i=N} w_i \quad (5.9)$$

where  $w_i$  are the weights as calculated in Eqn. 5.8,  $w_m$  is the maximum of the weights  $w_i$ , and  $N$  the number of weights. The value of  $W$  for a PCA is compared against the value of  $W$  for the previous PCA, and if they are the same (or similar to within 0.00001) then the recursion is halted and the last computed PCA is returned. The recursion is also stopped, and the last computed PCA returned, if the recursion depth reaches 50 without the two values of  $W$  becoming within 0.00001 of each other - this is to prevent any possible case of infinite recursion.

The result of running the recursive weighted PCA method on the same example Shower as shown previously can be seen in Fig. 5.8 and Fig. 5.9. The effect of the recursive weighted PCA can be readily seen in the 2D projections of Fig. 5.9, where

---

<sup>‡</sup>Shower direction is discussed further in the next section.

the minimally ionising part of the Shower in the primary-tertiary plane now runs parallel to the primary axis. The fact that the secondary and tertiary axes are now mirrored along the primary axes is down to the nature of a PCA, and is completely unproblematic as the specific direction of the secondary and tertiary axes are not used.

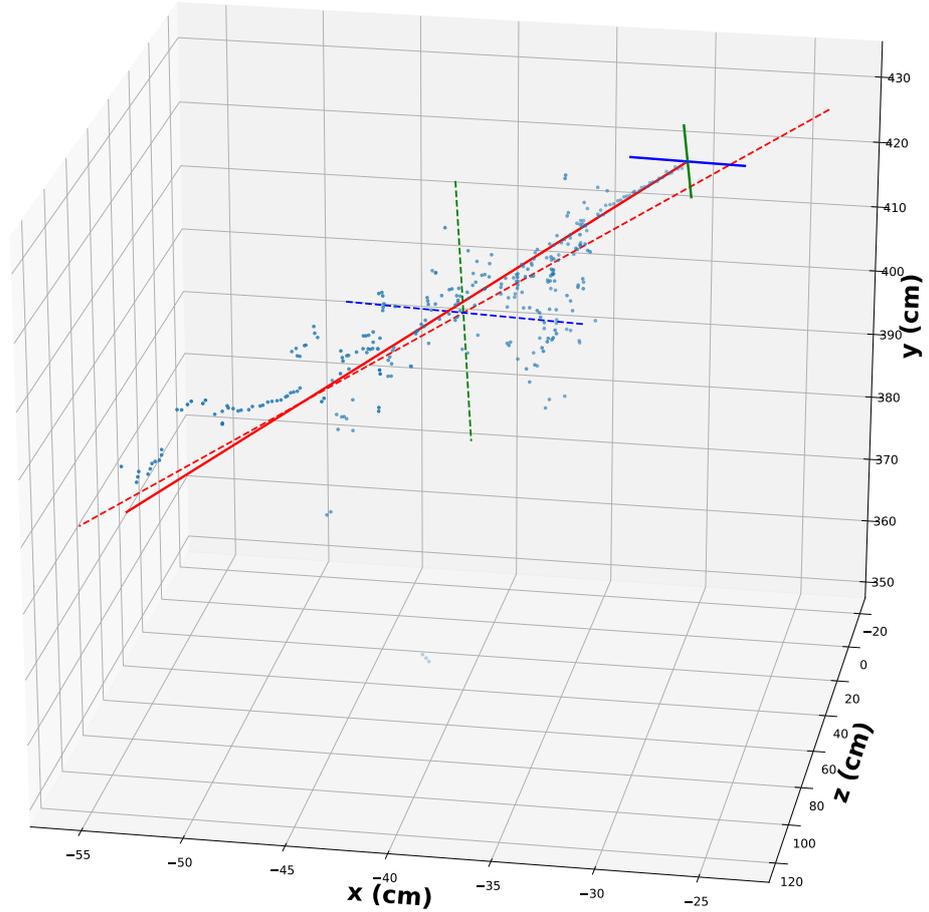


Figure 5.8: An example of a reconstructed 3D shower with its unweighted PCA and recursive weighted PCA axes overlaid. The unweighted PCA axes are the dotted lines, and the recursive weighted PCA axes are the solid lines.

## 5.4 Shower Direction

The direction vector of a Shower is set by Pandora as the direction along the primary axis from the PFParticle’s vertex (as determined by Pandora’s vertex finding algorithm) to the centroid calculated from the unweighted PCA. As mentioned above, this direction is not well suited to physics analyses that will rely on the angle at which a particle leaves an interaction vertex, and thus a new method is needed to

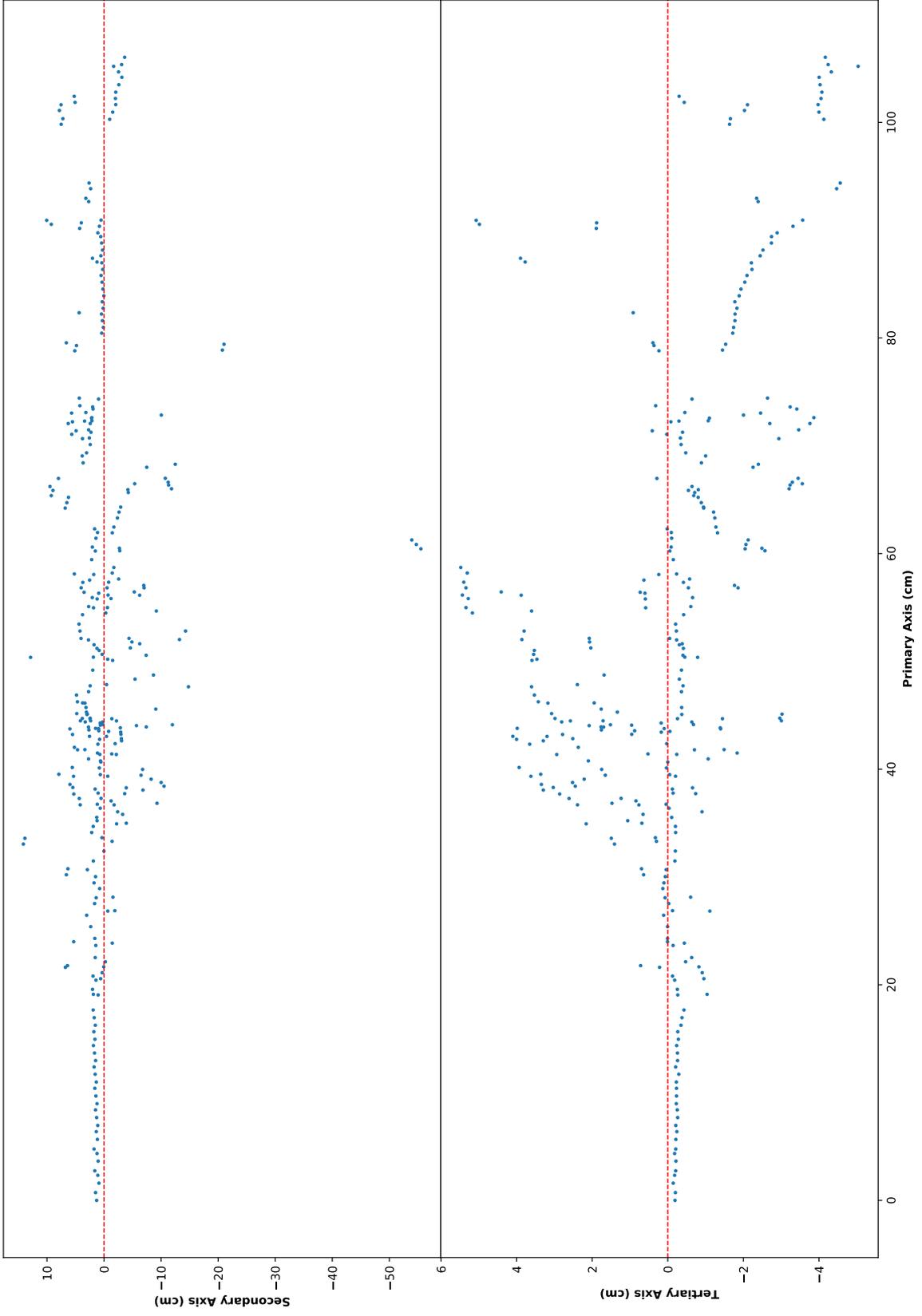


Figure 5.9: A shower that has had its 3D space points projected into the PCA axes coordinate system.

provide a direction more closely matching this.

To see if the new recursive weighted PCA gives a better estimation of the direction of the particle along the minimally ionising part of the Shower, the angle between the primary axis and the Monte Carlo truth can be measured. The truth value for the direction vector is taken as the direction of the Monte Carlo true particle at the simulation step where it starts to interact. Fig. 5.10 shows two box plots for each simulated momentum interval, the blue box plot showing the angle between the MC true particle’s direction and the unweighted PCA, and the orange box plot showing the angle between the MC true particle’s direction and the recursive weighted PCA. Comparing the orange box plot to the blue box plot for each momentum interval, it can be seen that the recursive weighted PCA provides a shower direction that is much closer to the true direction. Fig. 5.10 is made using electrons from the single particle gun MC dataset.

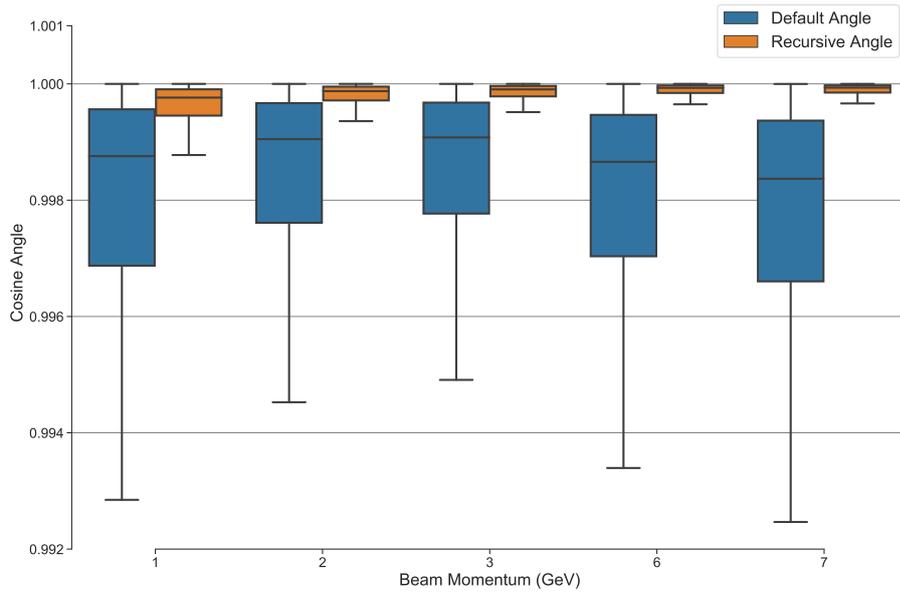


Figure 5.10: Plots comparing the angle between the direction of the Monte Carlo true particle and the primary axis from the unweighted PCA (blue) and the recursive weighted PCA (orange), for electrons from the single particle gun MC dataset.

Applying the recursive weighted PCA direction to the PDSP-MC and PDSP-BD datasets can be seen in Fig. 5.11. The value of true direction for the PDSP-MC dataset remains the same as the single particle gun MC, but the truth value for PDSP-BD comes from the last measurements made by the beam monitors. The cosine of the angle between the truth and the reconstruction is larger in PDSP-BD

as the truth measurement is made a few metres upstream of the detector, and so any interactions in the cryostat wall will cause a greater deviation between the true value and the reconstructed value.

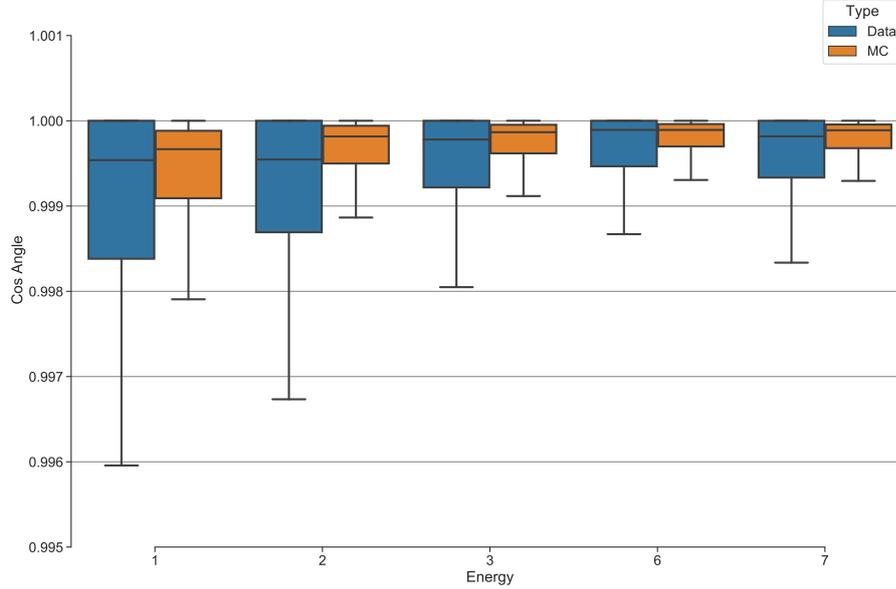


Figure 5.11: Comparison of the angle between true and reconstructed direction using the recursive weighted PCA on PDSP-MC and PDSP-BD. The truth value for PDSP-BD is taken from the final beam monitors.

## 5.5 Fractional Explained Variance

It is also possible to learn about the relative spread of a shower, along each axis, from the eigenvalues of the PCA. When taking the ratio of one eigenvalue calculated by the PCA to the total of all eigenvalues, this gives a quantification of how much of the total variation of the data is described by that axis. This is often known as the fractional explained variance<sup>§</sup>. The fractional explained variance of the unweighted PCA is more interesting from a physics perspective, as it will describe the whole Shower better, while the recursive weighted PCA's fractional explained variance is likely to be largely uninteresting by design of being biased to the minimally ionising section.

Fig. 5.12 shows the fractional explained variances for the axes of the unweighted PCA at each momentum interval for PDSP-MC and PDSP-BD. It can be seen in both datasets, that at higher energies the shower's width increases at a faster rate than the length of the shower as energy increases. It can also be seen that Showers

<sup>§</sup>Explained variance is another term for the PCA's eigenvalues, as the eigenvalues are a measure of the variance along any given axis.

in PDSP-BD appear wider relative to their length, as compared with PDSP-MC Showers. The fractional explained variances of the recursive weighted PCA for PDSP-MC and PDSP-BD events can be seen in Fig. 5.13. This figure shows that the recursive weighted PCA is being heavily weighted by a very linear set of 3D space points, which would be consistent with it having found the minimally ionising section of the EM shower. Again however, like in Fig. 5.12, these minimally ionising sections of events for PDSP-BD do appear to be slightly less linear than for PDSP-MC.

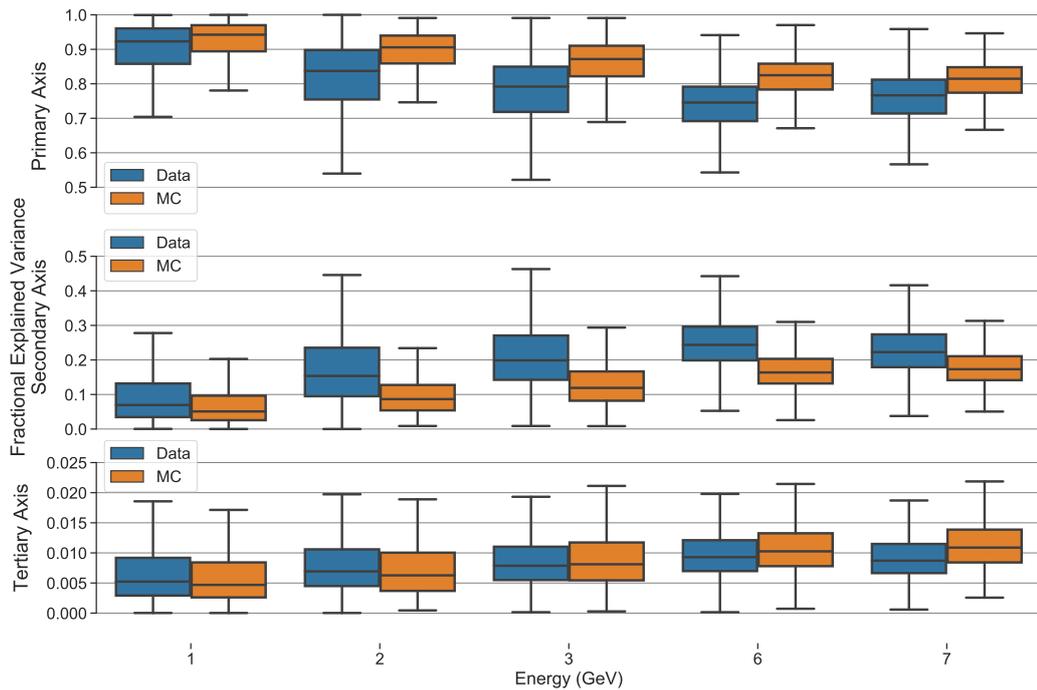


Figure 5.12: Box plots for the fractional explained variance of the three axes from the unweighted PCAs at each momentum interval for the PDSP-MC (orange) and PDSP-BD (blue) datasets. The decrease in the values for the primary axis (blue boxes) shows that EM showers of higher momentum are more spread out laterally relative to their length.

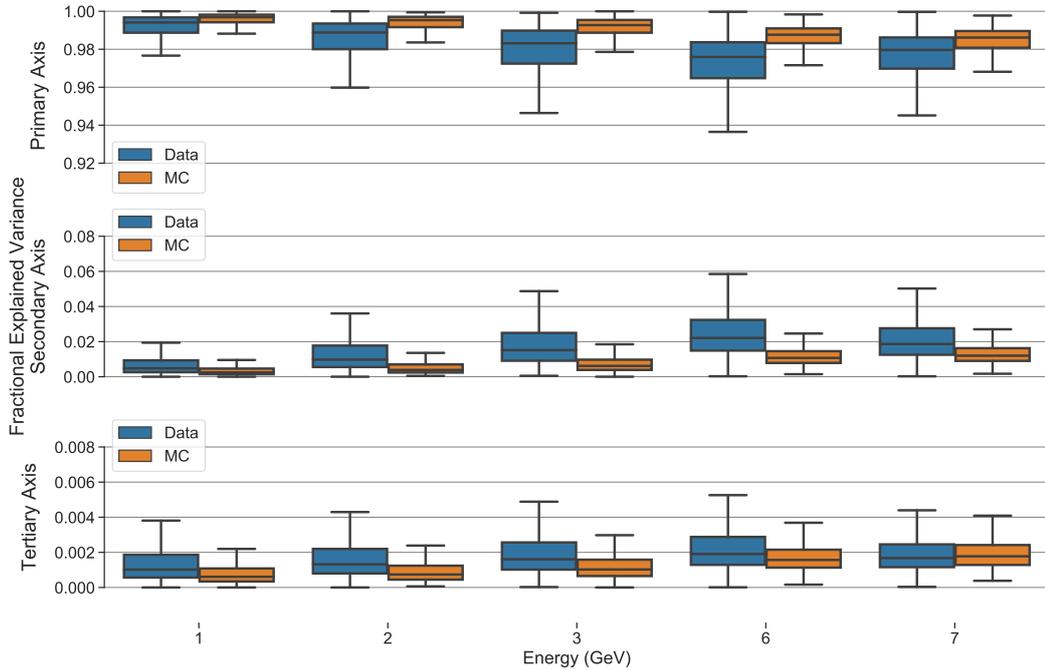


Figure 5.13: Box plots for the fractional explained variance of the three axes from the recursive weighted PCA at each momentum interval for the PDSP-MC (orange) and PDSP-BD (blue) datasets.

## 5.6 Shower Length

Because showers are cascades of multiple particles, there is no well defined value for their length. Pandora assigns the width of the primary axis, calculated using Eqn. 5.7, to the length variable of a reconstructed shower object. This is a reasonable measure of the length of a shower, however, it is subtly different from what one might intuitively expect for a quantity labelled as the length of the shower. The axis width as per Eqn. 5.7 is actually just twice the value used to commonly quantify the variance of a principal component, which is normally measured in one direction along the component's axis from the PCA centroid. So while this usually describes the width of a PCA axis well, it can appear to overestimate the length of a shower if measuring from the initial hit from a shower. This subtle difference can be seen in Fig. 5.14, where the PCA primary axis width is labelled as the Eigenvalue length. While the Eigenvalue length encapsulates the whole of the shower, it is rather severely overestimating the length of the shower as measured from the most upstream point.

On the same figure is another measure of length, the Projection length, which is defined as the projection of the most downstream 3D space point onto the primary

axis, relative to the most upstream 3D space point. This definition of length provides a value which is closer to the shower length one might conclude ‘by-eye’ given the distribution of 3D space points. A downside to this length definition is that it will be much more sensitive to any extreme downstream 3D space point outliers.

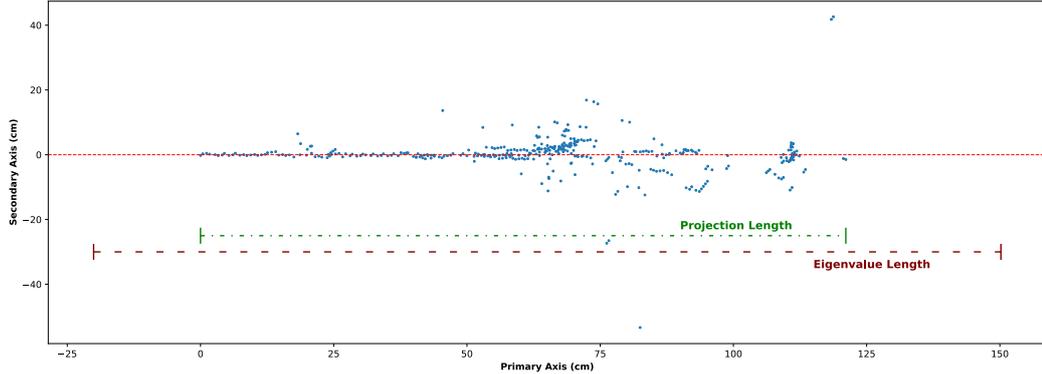


Figure 5.14: A 1 GeV/c single particle gun MC shower that has had its 3D space points projected into the PCA axes coordinate system. Overlaid are the two different definitions of length for the shower, with the eigenvalue length being about 50 cm longer than the projection length.

A comparison between the Eigenvalue length and the ‘Projection Length’ for each momentum interval of the single particle gun MC can be seen in Fig. 5.15. This figure shows how the Projection length is generally shorter than the Eigenvalue length, with the difference between the two lengths decreasing as the particle’s momentum increases. Although both methods of length estimation converge for high momenta, the conclusion from this study is that at lower momenta the projection length is preferred due to a tendency of the eigenvalue method to overestimate. Applying the projection length to the PDSP-MC and PDSP-BD can be seen in Fig. 5.16, which shows the lengths of the MC and data showers are the same at each momentum interval.

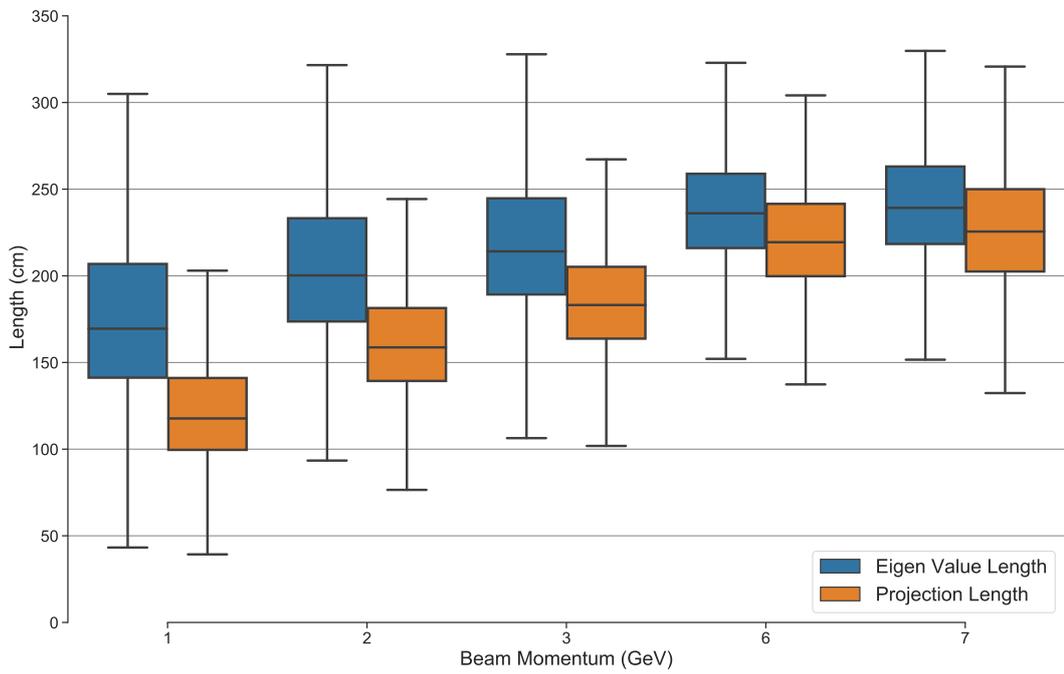


Figure 5.15: Comparison of eigenvalue length and max projection length for single particle gun MC electrons.

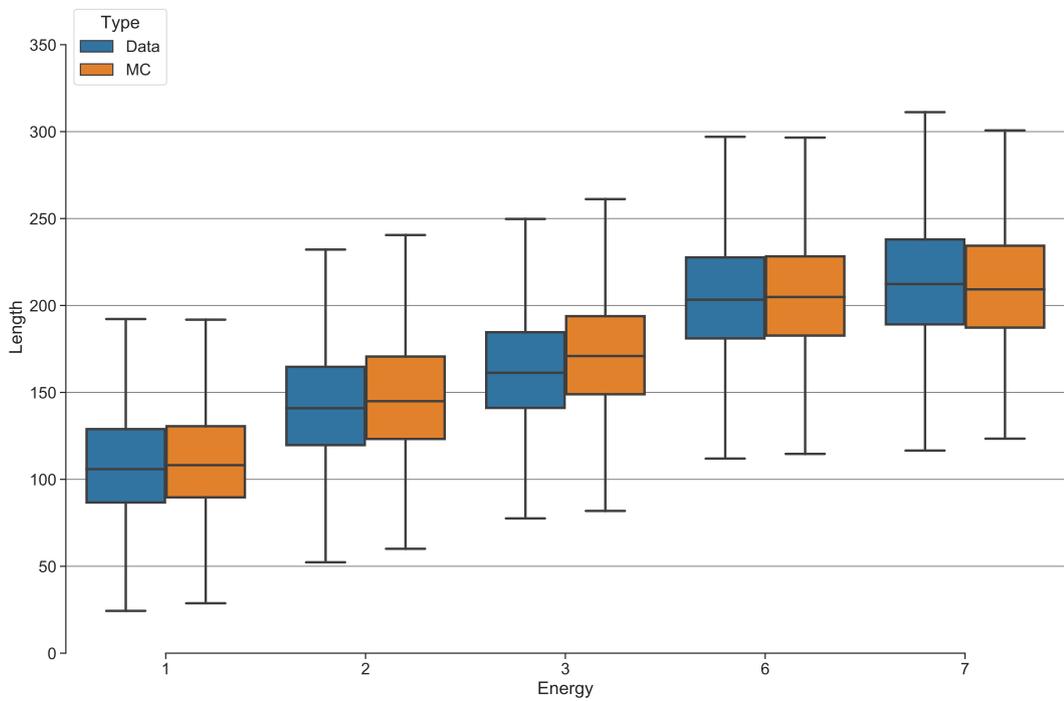


Figure 5.16: Comparison of the projection lengths for PDSP-MC (orange) and PDSP-BD (blue).

## 5.7 Shower Energy Estimation

Estimating the energy for all reconstructed particles is a very important part of any analysis. For EM showers in LAr this is done by summing over individual 2D hit estimations using [74]

$$E_{shower} = \sum_{i=0}^{i=N} \frac{\epsilon(X, YZ) \cdot N \cdot Q_i}{C_{cal} \cdot R} \cdot \frac{1}{4.237 \times 10^7} \text{ GeV} \quad (5.10)$$

The energy for each individual 2D hit is estimated by taking the collected charge  $Q_i$ , applying a SCE based correction  $\epsilon(X, YZ)$ , and normalising it values at the anode using the normalisation factor  $N$ . This corrected and normalised charge is then converted from ADC, the unit of charge as measured by the detector, into an estimate for the number of ionisation electrons. As ionisation electrons drift from the location of the initial travelling particle, some electrons are reabsorbed by argon atoms and so the number of ionisation electrons is corrected by the recombination factor  $R$ . Now that the number of ionisation electrons has been estimated, they can be converted to a measure of the energy of the particle based on the minimum ionisation energy of argon. This is represented by the  $\frac{1}{4.237 \times 10^7}$  factor in Eqn. 5.10.  $C_{cal}$ ,  $N$ , and  $\epsilon(X, YZ)$  are all run dependant.

The recombination factor,  $R$ , in Eqn. 5.10 is dependant on both the energy loss per unit length ( $dE/dx$ ), see section 5.8, and the electric field strength applied across the LArTPC. As these values would differ between individual charge deposits, and  $dE/dx$  is ill-defined for the bulk of an EM shower, a single global value is used. The recombination factor used in the Monte Carlo simulation is calculated using the Modified Box recombination model [74], and the values of a simulated 1 GeV/c electron can be seen in Fig. 5.17. The average value of the recombination factor in

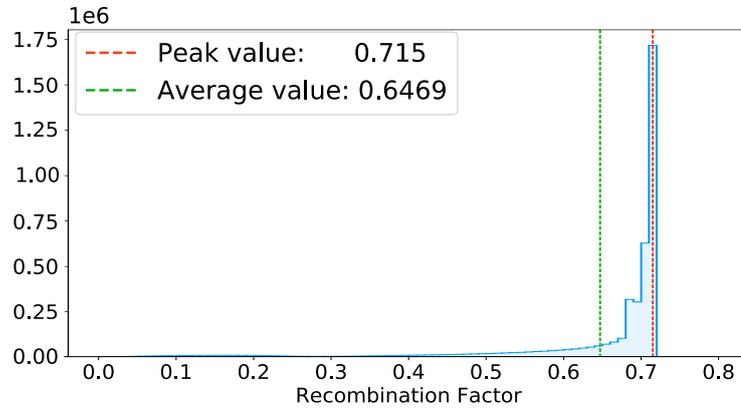


Figure 5.17: Recombination factors for simulated 1 GeV/c electrons.

Fig. 5.17 has a value of 0.6469, however, due to the long tail of the distribution, 77.3% of values are larger than the average. Therefore the average of this distribution is not a good estimator for the recombination factor, and the peak value of 0.715 is used instead for Eqn. 5.10. The results are similar for momenta up to 7 GeV/c.

The true momentum, from the PDSP-MC, and the measured beam momentum, from the PDSP-BD, is shown in Fig. 5.18. It can be seen that the PDSP-MC production

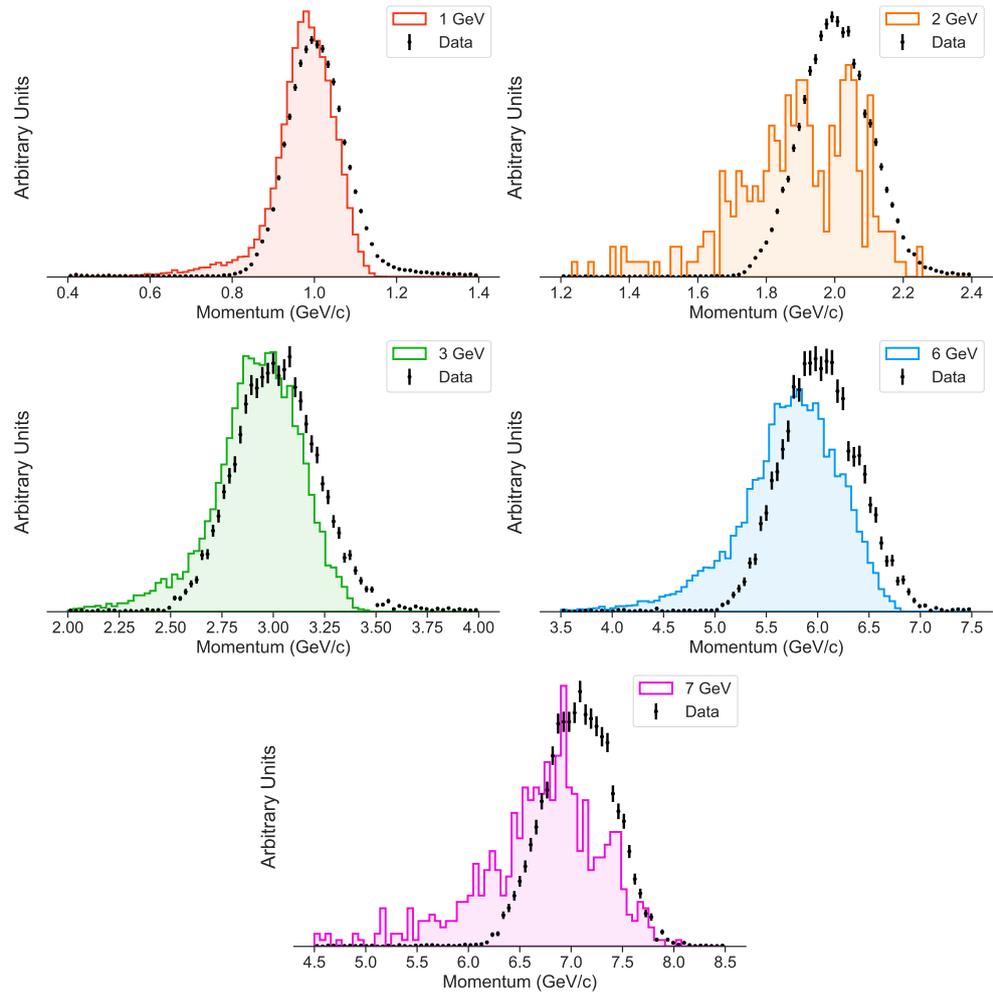


Figure 5.18: True momentum for PDSP-MC (histograms), and beam momentum for PDSP-BD (points). The measured beam momenta for data appears to be more accurately reconstructed than anticipated by the MC, and this has been corrected for in the next MC production.

used in this analysis assumed a worse distribution in momenta from the beamline

than was actually provided. This is to be fixed in the next PDSP-MC production run. The results of applying Eqn. 5.10 to reconstructed Showers, separated by their true momentum, can be seen in Fig. 5.19. While the momentum estimations of PDSP-MC and PDSP-BD appear to match better than their true momentum, this is not particularly surprising as both datasets have different  $C_{cal}$ ,  $N$ , and  $\epsilon(X, YZ)$  constants. A notable issue is the underestimation of the momentum. While the absolute underestimation grows with momentum, the relative peak underestimation remains around 20%.

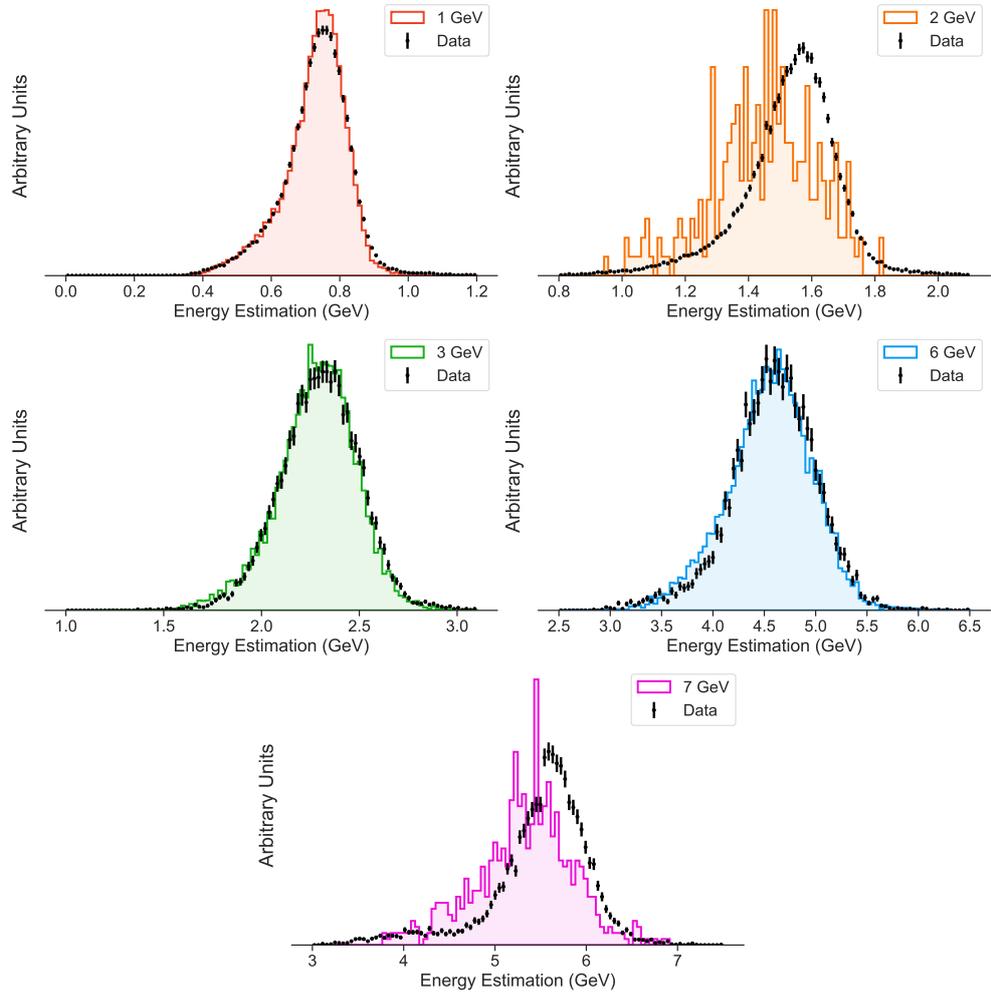


Figure 5.19: Energy estimations using Eqn. 5.10 for electrons generated at different momenta. It can be seen that there is a systematic underestimation across all momenta.

## 5.7.1 Energy Estimation Corrections

To investigate the deficit in the estimated momentum of electrons, as shown in Figs. 5.19, five estimation corrections were identified: deposition correction, missed hits correction, no hits correction, contamination correction, and hit energy correction. These corrections are discussed in the following sections, and plots showing the ratio of the correction to mean expected momentum - the mean energy of which the Monte Carlo beam particles are generated around - are used to show the relative effect of each correction. These corrections are discussed in terms of the PDSP-MC dataset.

### 5.7.1.1 Deposition Correction

This correction encapsulates any energy from the initial particle that is not deposited in the TPC. Energy may not be deposited in the TPC due to some parts of the shower reaching beyond the fiducial volume, or any interactions the particle may undergo between the beam pipe and the TPC (for example in the cryostat wall). To calculate this value, the sum of the energy associated to all ionisation deposits made by the true particle are subtracted from the energy with which the true particle was created. Fig. 5.20 shows that this correction accounts for less than 2% of the mean expected energy, and decreases with momentum as expected.

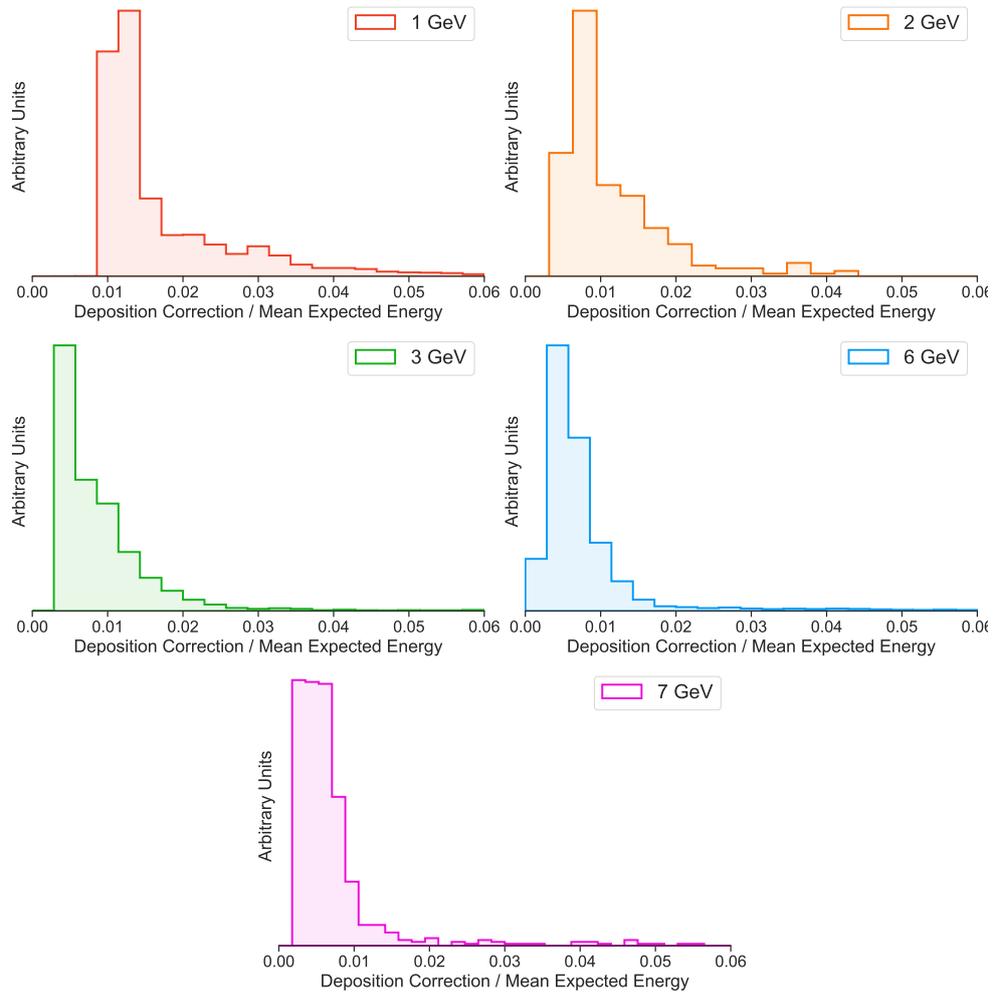


Figure 5.20: The ratio of the deposition correction to the mean expected energy for different momenta.

### 5.7.1.2 Missed Hits Correction

The missed hits correction accounts for energy associated to 2D hits that have been successfully reconstructed, but have not been associated with the correct particle. This value is calculated by summing the energy for 2D hits that are resultant from ionisation electrons caused by the simulated beam particle, but have not been included in the reconstructed particle tagged as the beam particle by Pandora. As can be seen in Fig. 5.21 this correction is highly skewed for each momentum, though each peak between 2-4% with the width decreasing as momentum increases.

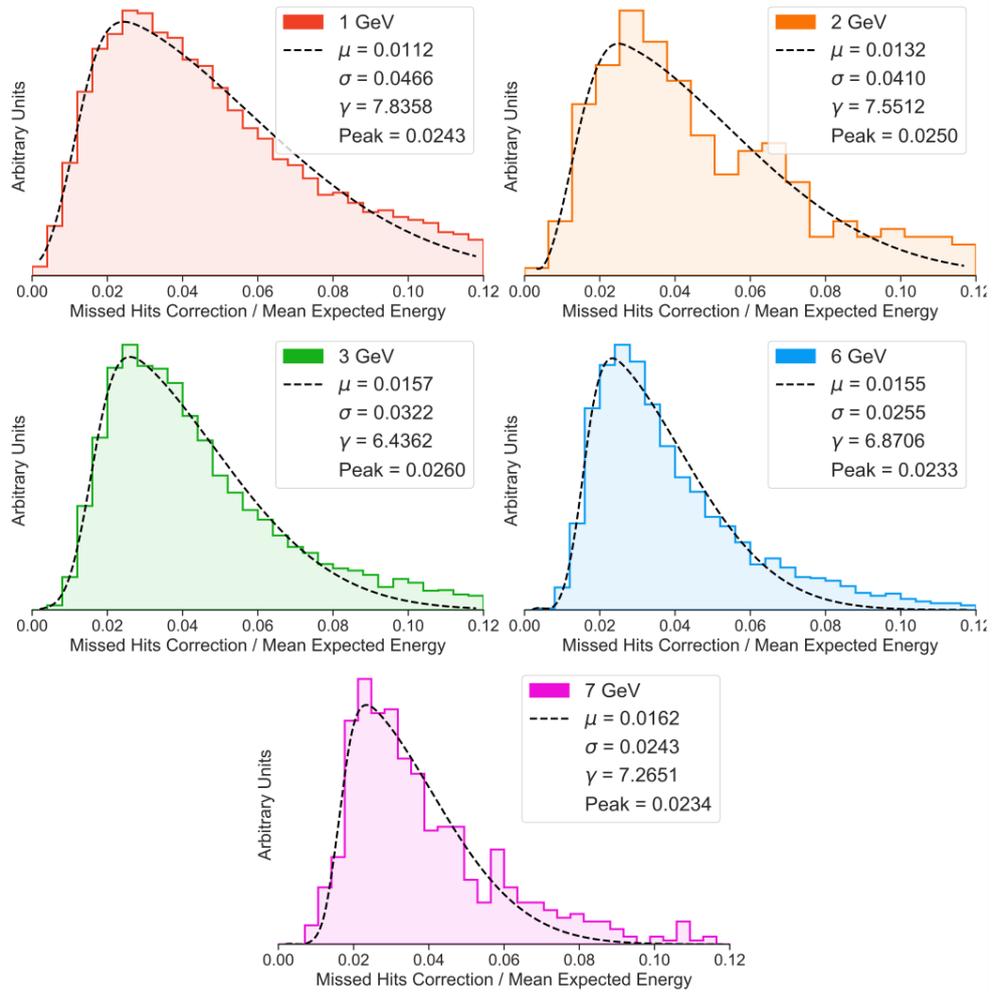


Figure 5.21: The ratio of the missed hits correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2.

### 5.7.1.3 No Hits Correction

This correction accounts for all the energy that was deposited inside the TPC by the particle, but did not get incorporated into any reconstructed 2D hits. The reason some drift electrons are not incorporated into any 2D hit is simply because the 2D hit fitting function has a threshold to reduce noise, and sometimes drift electrons will not create signals on the sense wires that meet this threshold. This value is calculated by summing all the energy of any ionisation electrons that have no associated reconstructed 2D hit. This correction is the largest contributor, as seen in Fig. 5.22, to the momentum deficit in the estimator, with mean values approaching 25% of the mean expected energy.

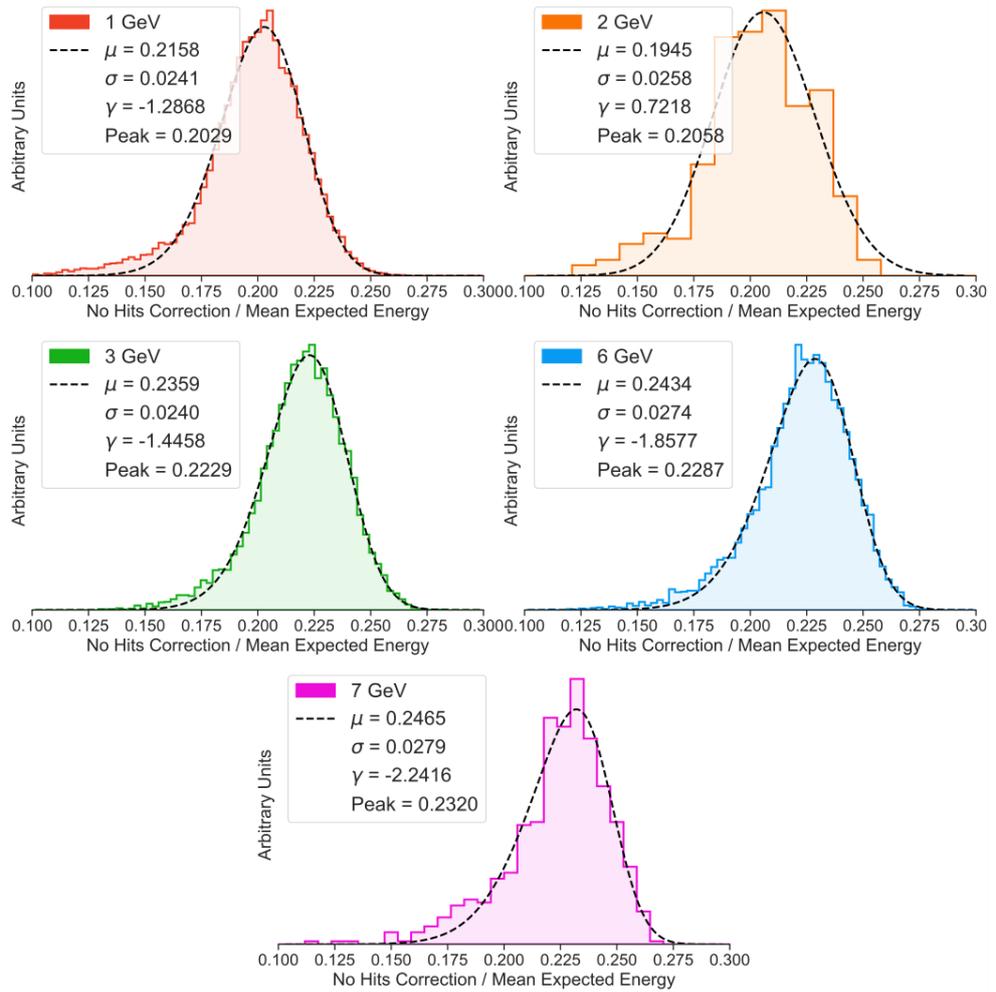


Figure 5.22: The ratio of the no hits correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2.

#### 5.7.1.4 Contamination Correction

The contamination correction attempts to correct for any 2D hits that have been incorrectly associated to the selected particle. Such incorrectly associated 2D hits could come from nearby activity from cosmic ray particles, or beam halo particles. This value is calculated by summing the energy associated to any 2D hits that have been associated to the Pandora tagged reconstructed beam particle, but are not associated to the true MC beam particle. Looking at Fig. 5.23 it can be seen that this correction is peaked at zero for all momentum, testifying to the reconstruction abilities of Pandora, with tails out to 8% of the mean expected energy.

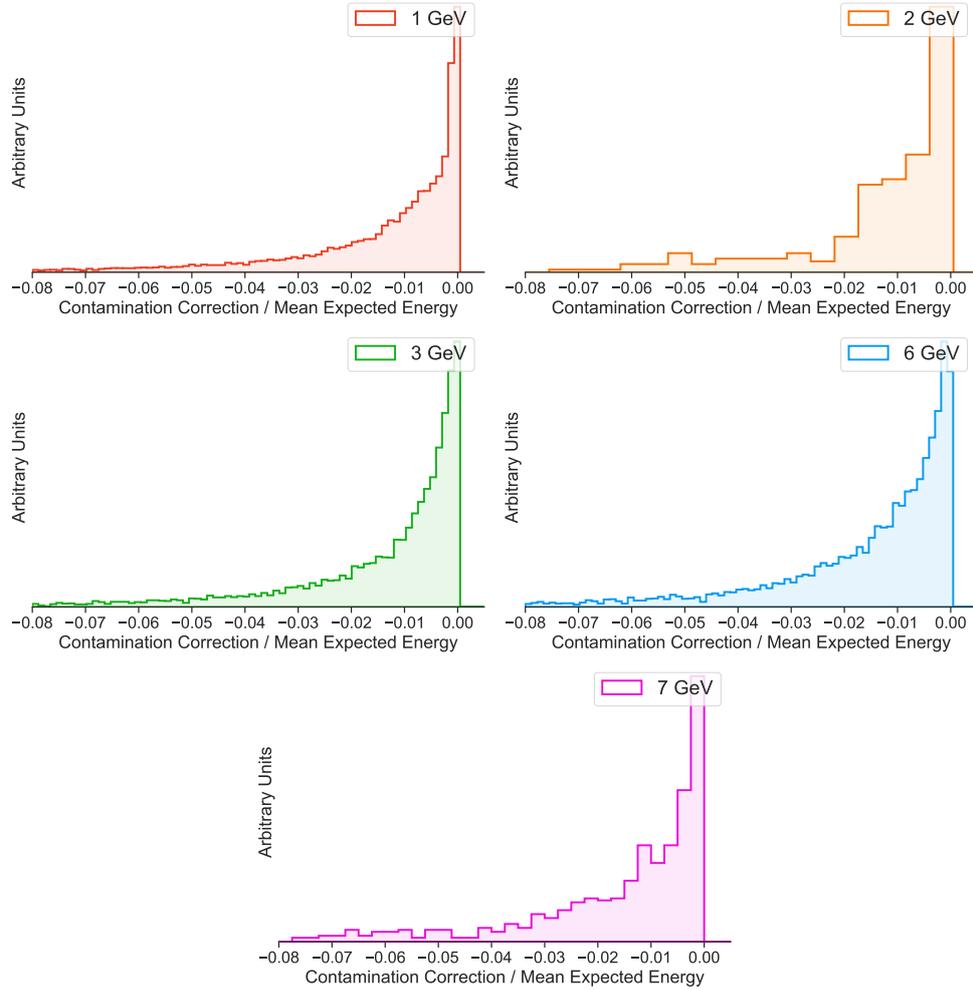


Figure 5.23: The ratio of the no hits correction to the mean expected energy for different momenta.

### 5.7.1.5 Hit Energy Correction

This correction looks at how accurate the energy estimator is for the 2D hits used in the estimator itself. Fig. 5.24 shows that for all momenta, the energy estimator used is over estimating the energy of the 2D hits included in the estimation. Thus this correction is a negative correction. The reason the energy estimator overestimates is due to using a singular value for the recombination factor,  $R$ , in Eqn. 5.10. If an  $R$  factor was calculated for each individual hit, this correction would likely be much smaller and centred around zero. This value is calculated by summing the true energy associated to the reconstructed 2D hits, and comparing it to the estimated energy as per Eqn. 5.10.

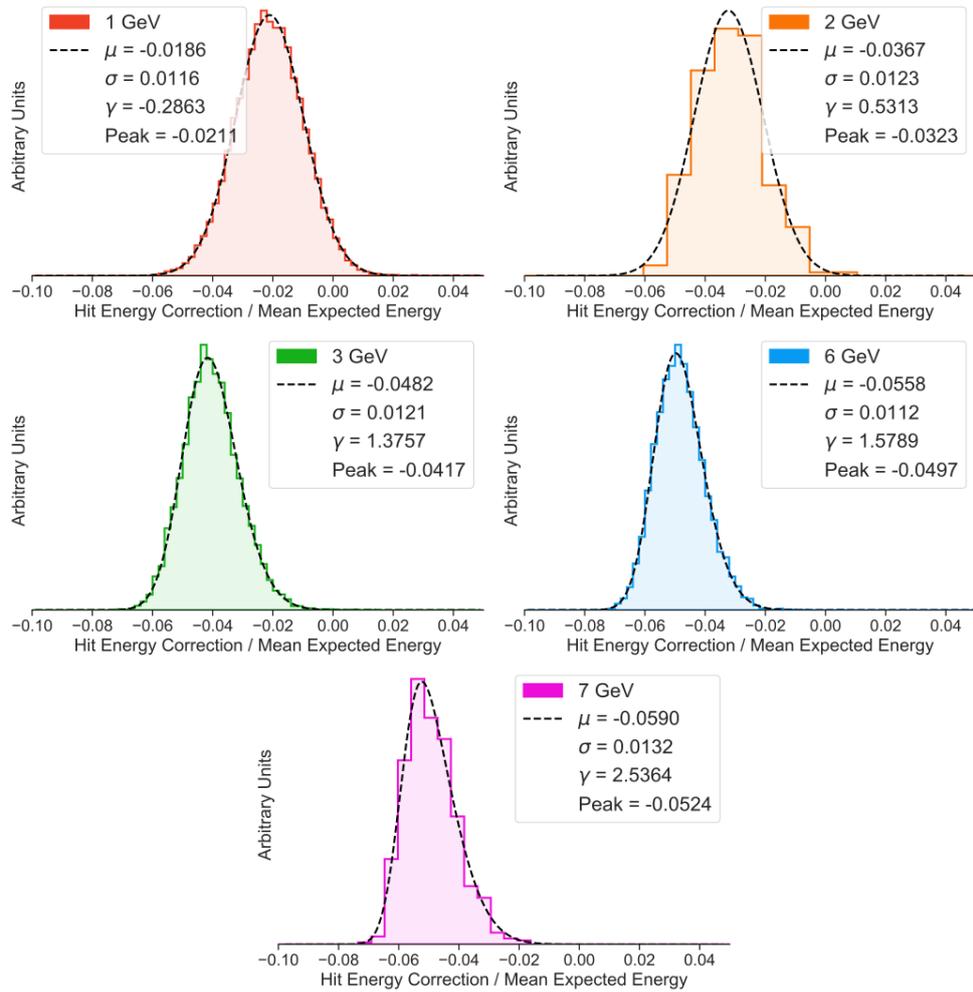


Figure 5.24: The ratio of the hit energy correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2.

### 5.7.1.6 Total Correction

Summing all of the above corrections leads to the results in Fig. 5.25, and shows that the total correction as a ratio to mean expected energy remains mostly consistent as the initial true energy increases. The distributions do, however, have a notable positive skew. This makes estimating a global correction factor slightly more complicated, as any chosen value will invariably cause a subsequent skew in the final momentum estimation value.

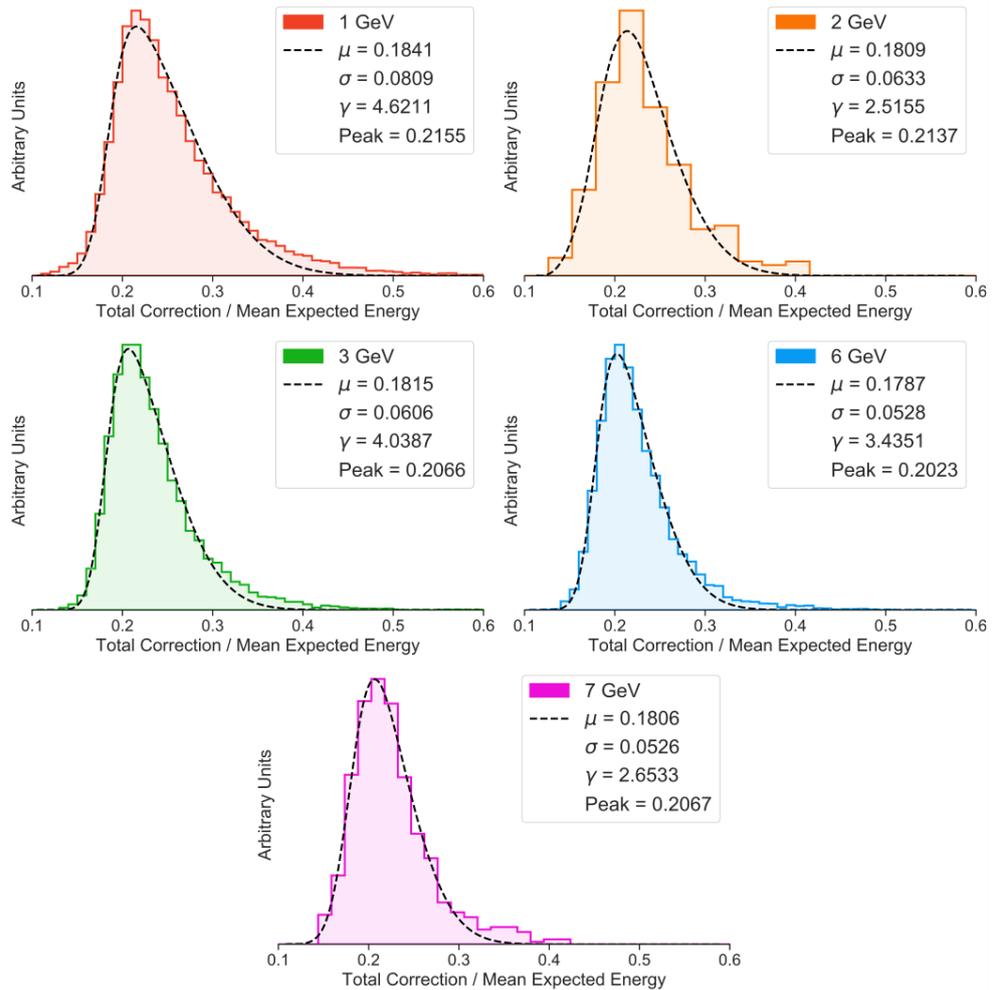


Figure 5.25: The ratio of the total correction to the mean expected energy for different momenta. The fit used is Eqn. 5.2.

However, as this is Monte Carlo data it is possible to apply the exact correction for each event, allowing for a confirmation of whether or not the total energy deficit has been accounted for. The outcome of doing so is shown in Fig. 5.26, and it can be seen that the corrections appear to be correcting the energy estimations towards their true values. There is however some slight jitter that increases with higher energies, and this is down to the complexity of matching reconstructed objects with the truth information from the Monte Carlo simulation. Being confident that these corrections recover the correct energy estimation in the PDSP-MC dataset, the peak values for each momentum interval in Fig. 5.25 are used as a correction factor for the PDSP-BD dataset. The results of correcting the PDSP-BD energy estimations can be seen in the residual plots shown in Fig. 5.27. By fitting a Gaussian to each

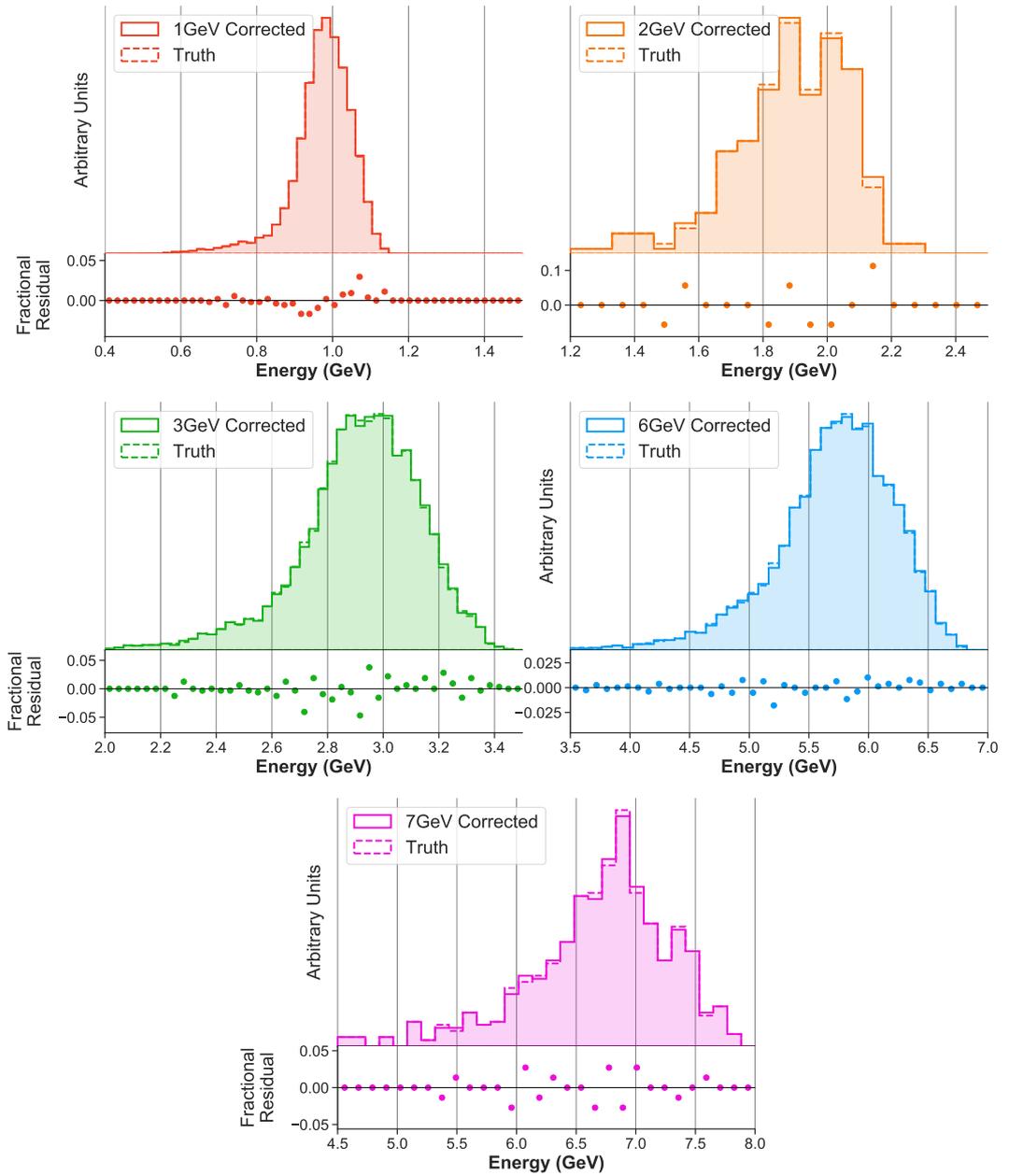


Figure 5.26: The corrected energy estimation for the PDSP-MC dataset, using an event-by-event correction.

peak in Fig. 5.27, the energy resolution achieved can then be calculated using:

$$\frac{\sigma}{\langle E \rangle} = a \oplus \frac{b}{\sqrt{\langle E \rangle}} \oplus \frac{c}{\langle E \rangle} \quad (5.11)$$

where,  $a$ ,  $b$ , and  $c$  are the constant, stochastic, and noise terms respectively,  $\oplus$

denotes a summation in quadrature, and  $\langle E \rangle$  is the mean expected energy. The right hand plot of Fig. 5.28 shows the achieved energy resolution of 1%-6%, an expected range for calorimeters. The left hand plot of Fig. 5.28 shows good linear agreement between the corrected estimated energy and the true energy from the beamline. Fit values for Fig. 5.28 are listed in Table 5.4.

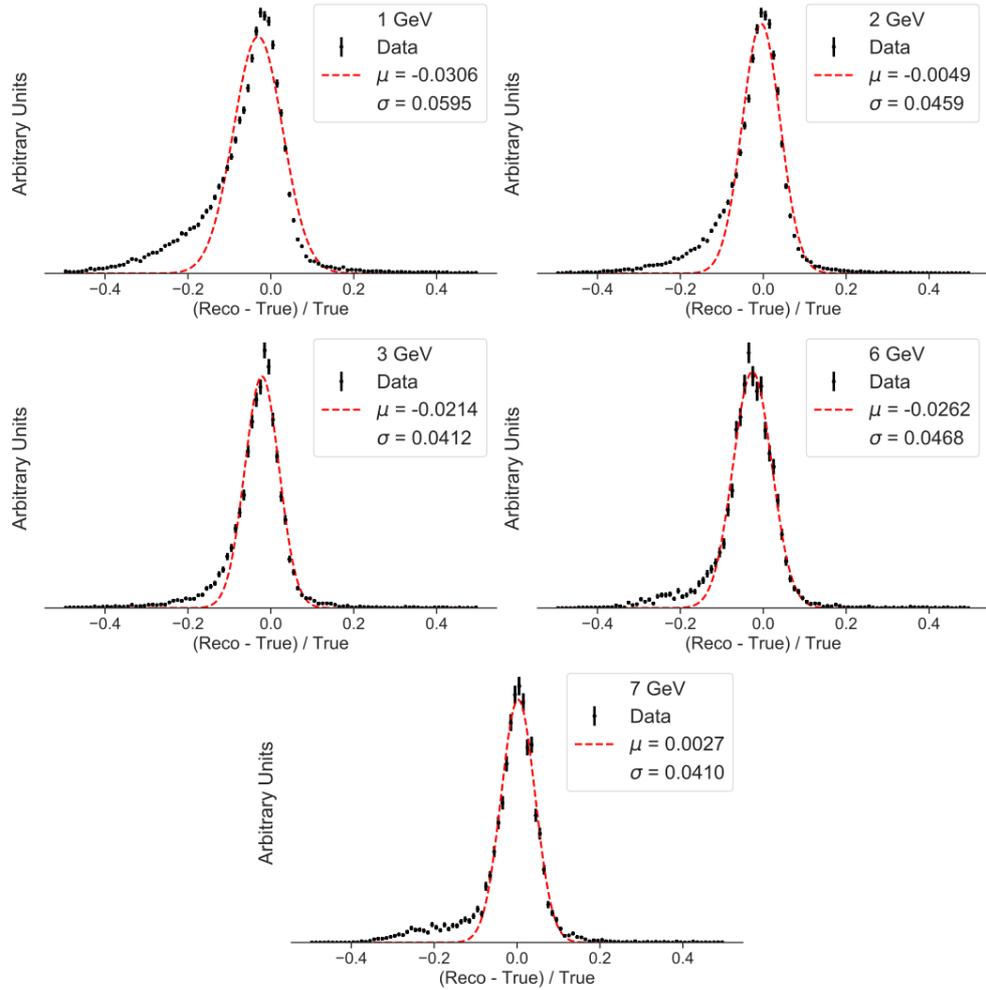


Figure 5.27: Residuals for corrected energy estimations of the PDSP-BD dataset. The correction factor is calculated as the peak values from Fig. 5.25. The fits are Gaussians as per Eqn. 5.1.

Table 5.4: Fit parameters for Fig. 5.28

Figure	Parameter	Value
5.28 (a)	Gradient	$0.99 \pm 0.011$
5.28 (a)	Intercept	$-0.05 \pm 0.005$
5.28 (b)	a	$0.050 \pm 0.001$
5.28 (b)	b	$0.15 \pm 0.016$
5.28 (b)	c	$-0.20 \pm 0.011$

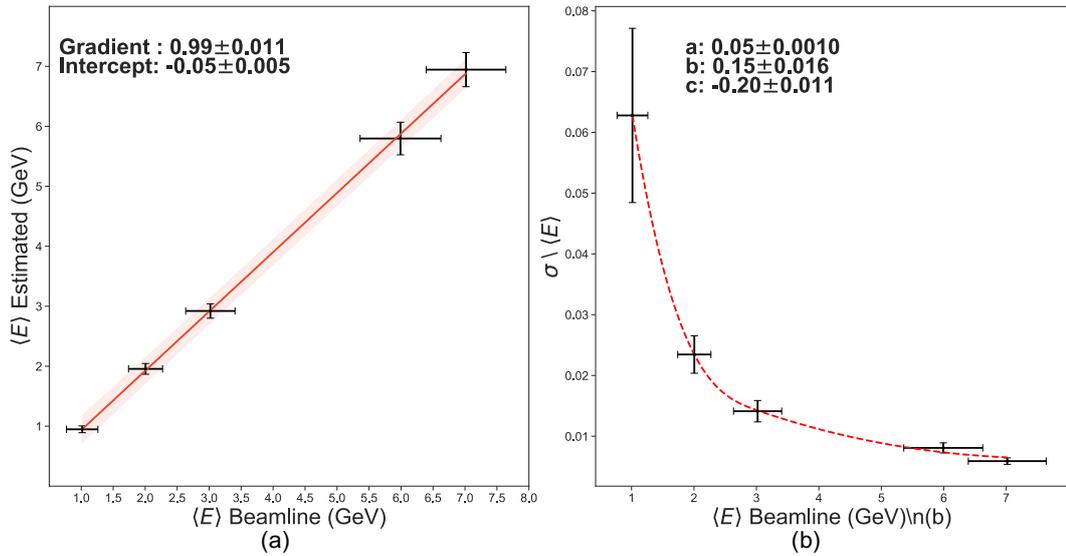


Figure 5.28: Energy resolution for the PDSP-BD dataset. The left plot shows good agreement between the corrected estimated energy and the true energy. The right plots shows that, as expected, the resolution improves as the expected energy increases.

## 5.8 Shower $dE/dx$

An important quantity to measure with EM showers is the energy loss per unit length,  $dE/dx$ , at the start of the shower. The expected energy loss per centimetre for a minimally ionising particle in ProtoDUNE-SP is 2.1 MeV/cm [75]. The method employed in this thesis to calculate the  $dE/dx$  of a Shower attempts to utilise the energy estimation of each individual 2D hit, as calculated in the previous section. The recursive weighted PCA is used to determine the minimally ionising section of the shower, and a 3D cylinder, of radius and length of 5 cm, is built around the primary axis. The energy located within this cylinder is then used as the value of  $dE$ .

Fig. 5.29 shows the  $dE/dx$  measurements made on showers in the PDSP-MC and PDSP-BD datasets. It can be seen that the peak of each distribution is congruent with the expected value for  $dE/dx$  in ProtoDUNE-SP. The data, however, is less strongly peaked than the MC. This is likely due to the method for  $dE/dx$  estimation used here, and could be refined further to improve the estimation in data.

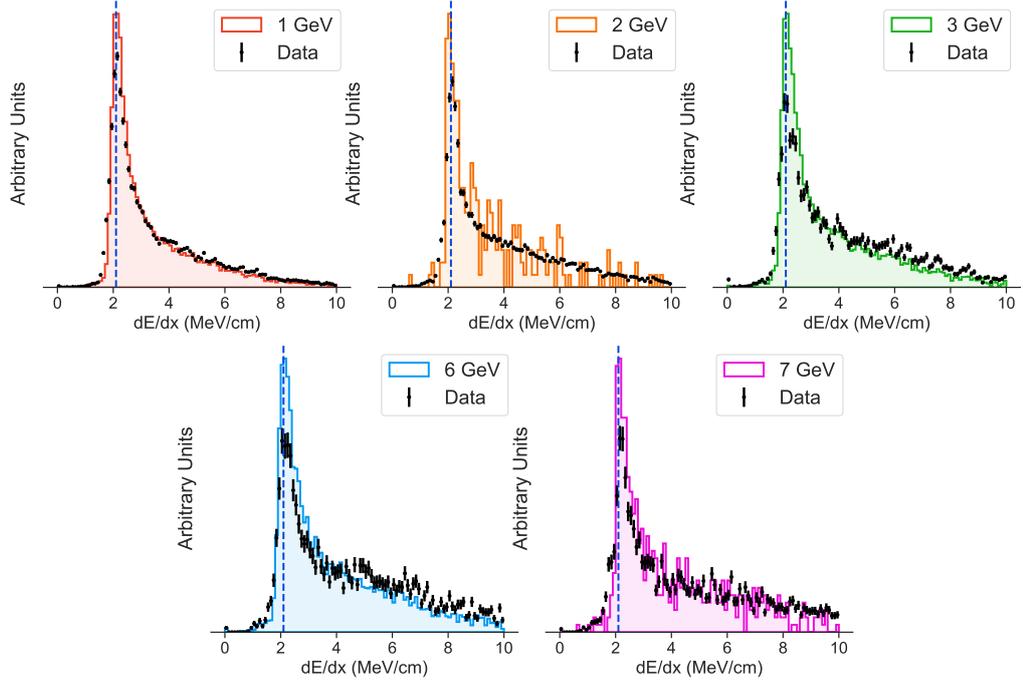


Figure 5.29: Comparison of measured  $dE/dx$  for the first 5 cm of electron showers in the PDSP-MC (histograms) and the PDSP-BD (points) datasets. Each distribution is area normalised.

## 5.9 Shower Longitudinal Profile

The energy loss per centimetre can be generalised for the whole length of a shower to energy loss per radiation length,  $dE/dt$ , where  $t = x/X_0$ . This is known as the longitudinal profile of a shower, and should be reasonably well described by the gamma function [34]:

$$\frac{dE}{dt} = \langle E \rangle b \frac{(bt)^{a-1} e^{-bt}}{\Gamma(a)} \quad (5.12)$$

where  $\langle E \rangle$  is the mean expected energy, and  $a$  and  $b$  are constants describing the rise and fall of the function. The longitudinal profiles shown in Fig. 5.30 shows how the longitudinal profiles for each momentum interval are reasonably well described by the above gamma function. It can also be seen that the peak of the distribution moves

further down the length of the shower, and the width of the distribution increases, as the momentum increases.

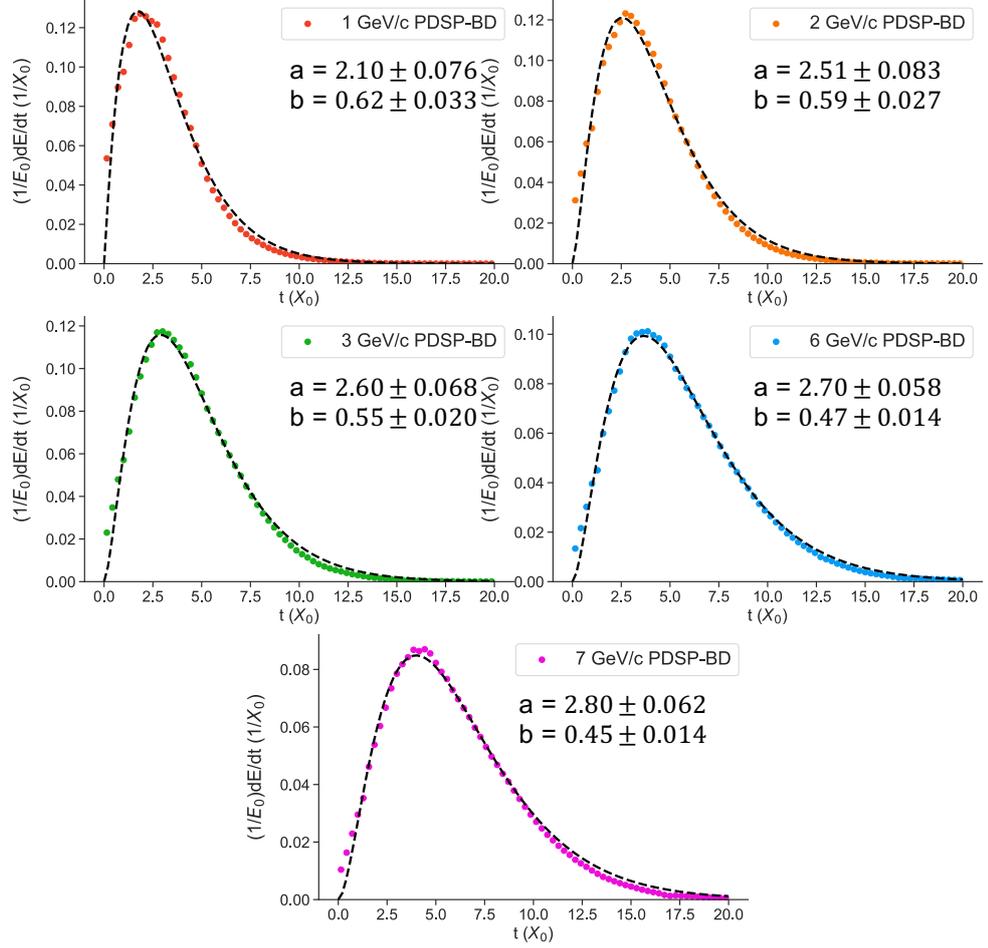


Figure 5.30: Energy deposition per radiation length for PDSP-BD dataset events.

The increase in width of the Showers can also be seen in Fig. 5.31. These plots show the standard deviations ( $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  going outwards from the  $x$ -axis) of the cross-sectional area of the 3D space points along the secondary and tertiary PCA axes. It can also be noted that the showers in the PDSP-BD dataset appear to be wider than those found in the PDSP-MC dataset.

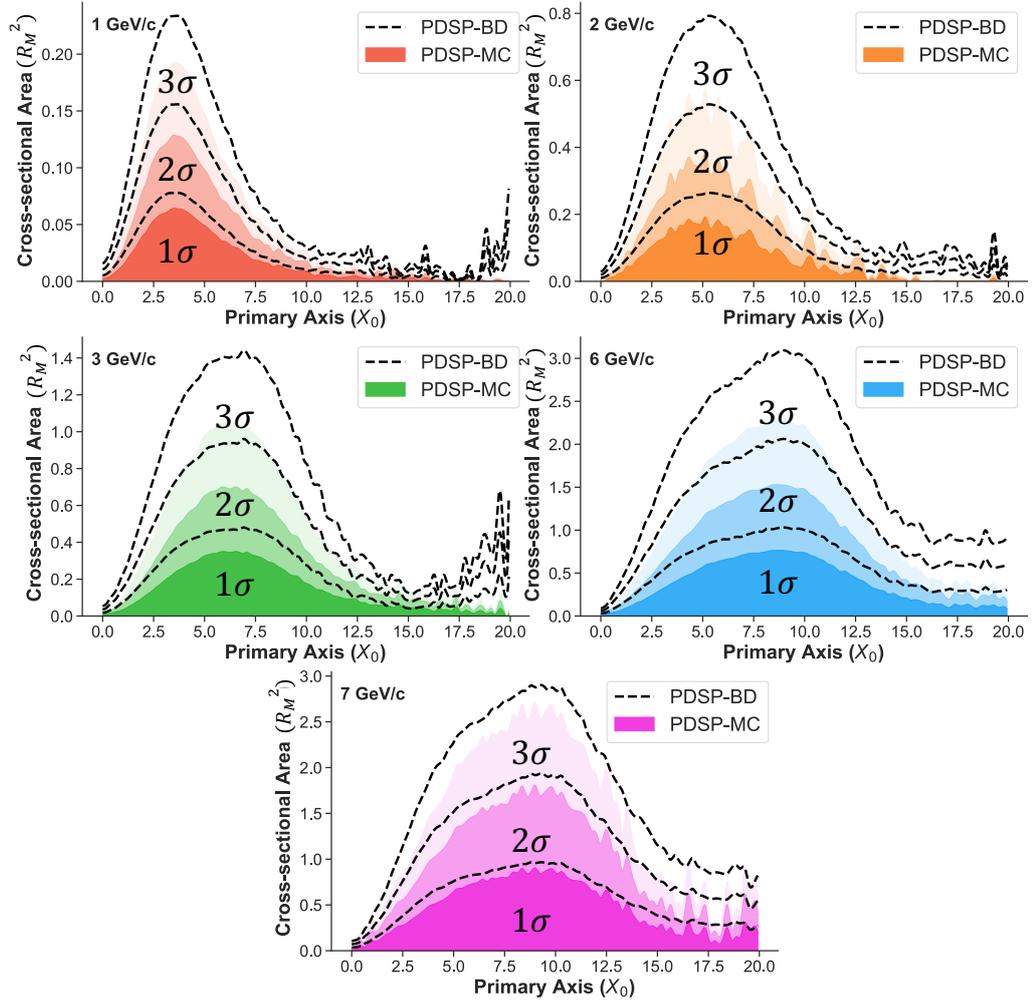


Figure 5.31: Averages of the binned cross-sectional areas for the PDSP-BD dataset (dashed lines), and the PDSP-MC dataset (shaded regions). As seen in Fig. 5.12, the PDSP-BD EM showers appear to be wider than the EM showers in the MC.

### 5.9.1 Cascade Start

As shown in Fig. 5.1 at the start of the chapter, an electron is expected to start cascading after travelling through one radiation length of material, but with the incredible positional resolution of LArTPCs it is possible to make an actual measurement of when the cascading section of the EM shower starts?

To attempt to solve this, individual shower profiles like the one seen in Fig. 5.32 were used to identify a method of finding the start of the cascading section of the EM shower. The plots of Fig. 5.32 are made by:

- Rotating the 3D space points onto the recursive weighted PCA axes.
- The 3D space points are binned into 1 cm bins.
- The standard deviation of the 3D space point along the secondary and tertiary axes for each bin are calculated.
- A cross-sectional area is formed by multiplying the two standard deviations for each bin.

The cross-sectional area from the final step is then the value used to attempt to find the start of the cascading section. By walking along each bin from the start of the shower, the bin where the cross-sectional area goes above a threshold is selected as the start of the cascade. The bin that has been selected by this method in Fig. 5.32 is denoted by the green dashed line in the bottom plot. Through visual inspection of this plot, and many others, the currently selected threshold value of  $0.005 R_M$  appears to do a reasonable job. However, no attempt has yet been made to find a more optimal value through an analytical method. Applying the above method to the PDSP-BD dataset yields the cascade start distribution in Fig. 5.33. While one radiation length is within  $1\sigma$  for all momenta when fitted with a skewed Gaussian, as per Eqn. 5.2, the width of each distribution is too wide to draw any conclusions about the effectiveness of this method. A deeper investigation would need to be made into this subject, and a possible exploration into the use of deep learning algorithms could yield improved results.

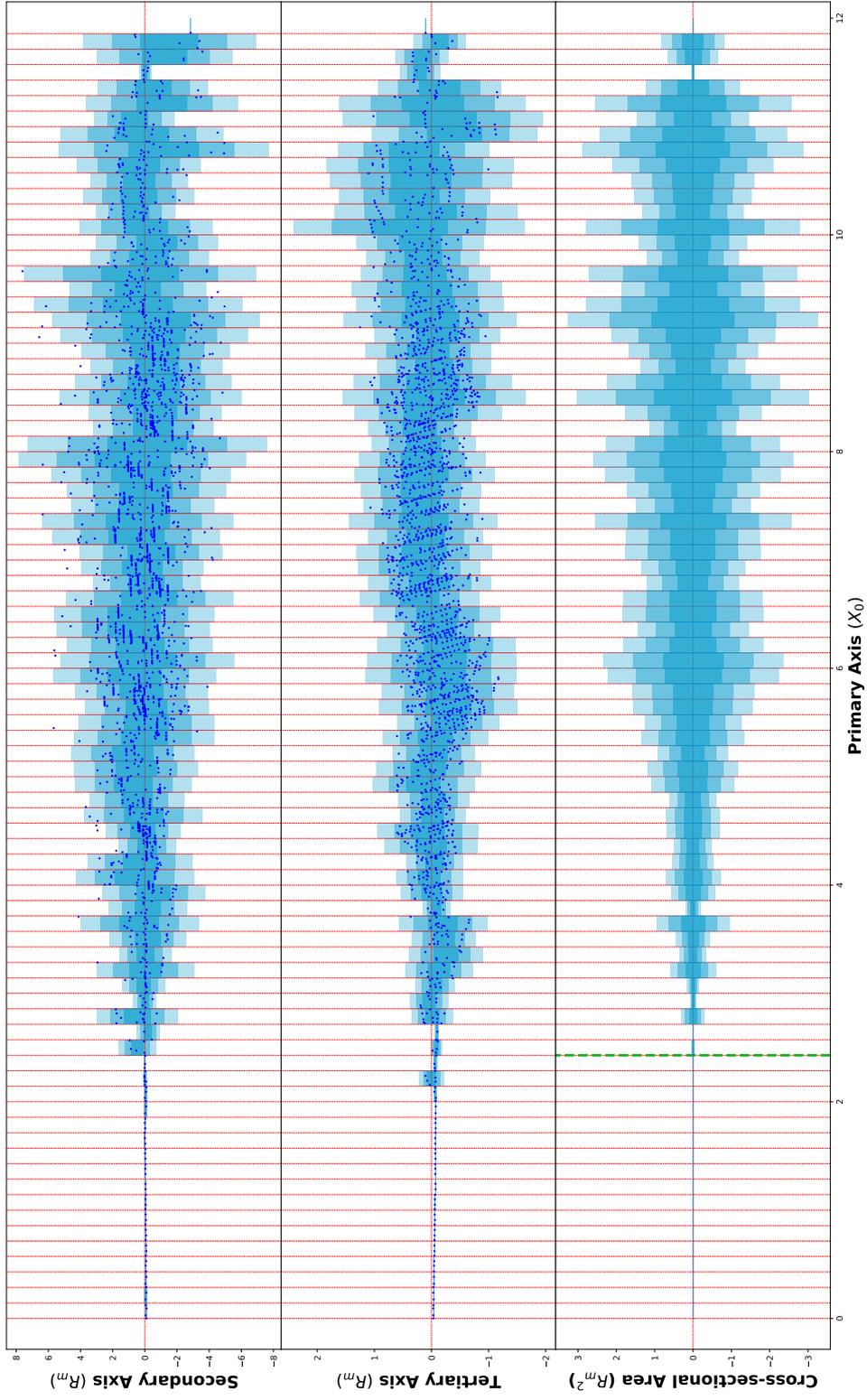


Figure 5.32: An example plot of the binning method used to identify the start of an EM shower's cascading section. The 3D space points have been rotated onto the PCA axes, and placed into 1 cm bins (red lines). The blue shaded areas then represent the standard deviations for the 3D Space points in that bin, along each axis. The bottom plot shows the product of the standard deviations for each bin. The green dotted line shows the bin that went above the threshold selection for the start of the EM cascade.

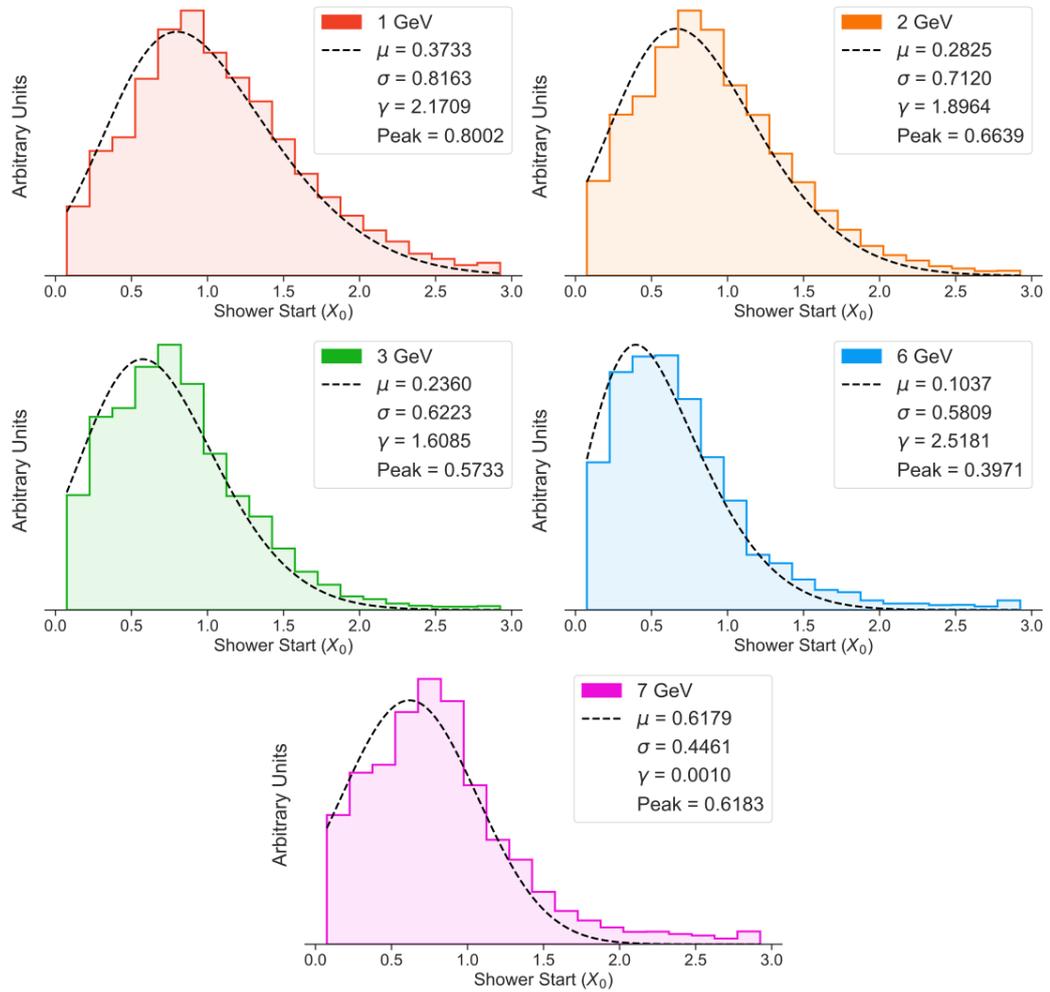


Figure 5.33: The start positions, in radiation lengths, of the cascading part of the EM showers from the PDSP-BD dataset.

## 5.10 Shower Transverse Profile

A method for calculating the transverse profile of an EM shower has been devised. The reconstructed hits of a Shower object are projected onto the plane defined by the secondary and tertiary axes of the weighted recursive PCA, and then binned in concentric rings centred on the origin. The radii of the concentric rings are such that the area of each ring remains constant with a value of  $\pi(0.2R_M)^2$ . This then allows an averaging to be done across a momentum range, and for different momentum ranges to be comparable. The ADC density of each bin is then calculated, and the results of doing so are shown in Fig. 5.34, where the coloured dots represent the PDSP-MC dataset and the black represent the PDSP-BD dataset. The tails for each

momentum range in Fig. 5.34 shows good agreement between both datasets, however at the core of the shower the PDSP-MC is more energy dense than PDSP-BD at all momenta. The reason for this is currently thought to be due to the space charge effect, especially in the region directly around where the beam particles enter the TPC. This will likely improve as modelling of the space charge effect is improved, but a more in-depth investigation is planned if this difference persists in updated Monte Carlo simulations.

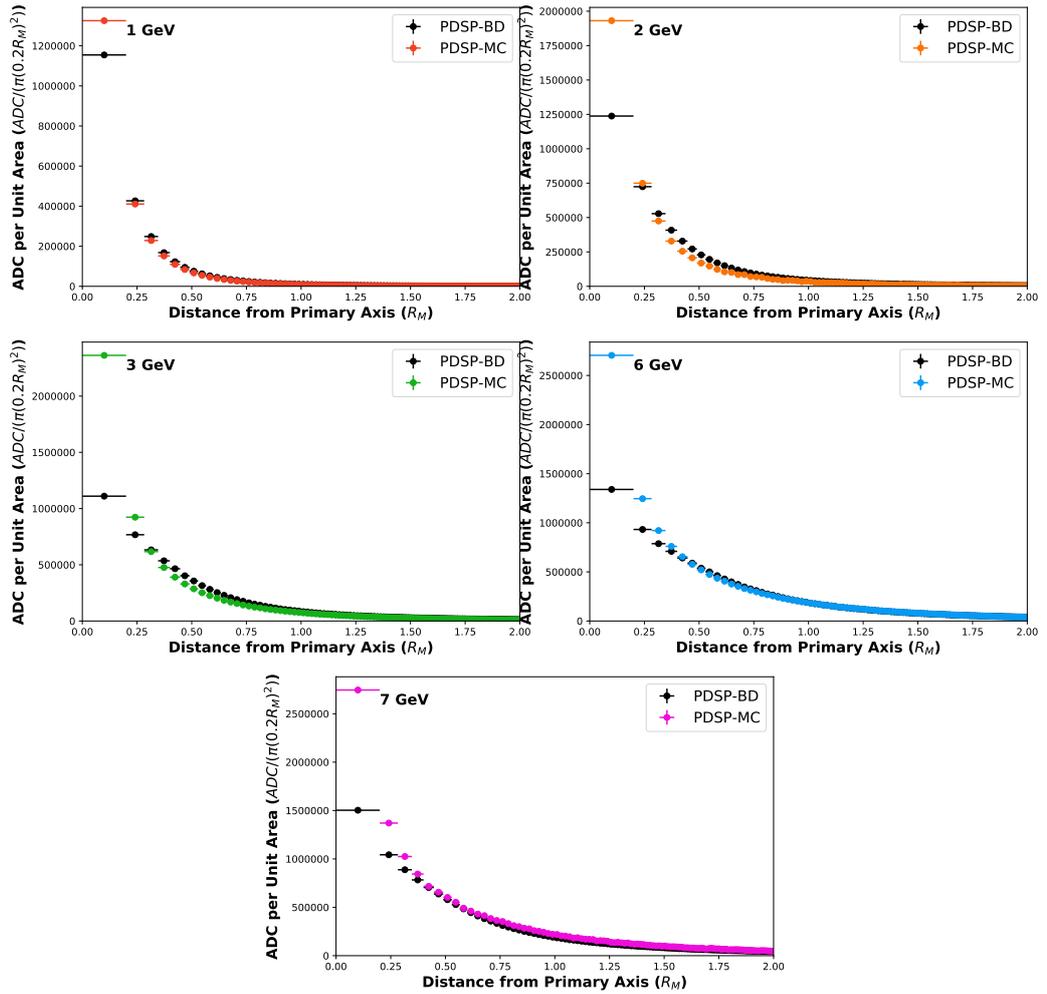


Figure 5.34: Transverse profiles for EM showers in the PDSP-MC and PDSP-BD datasets.

## 5.11 Summary

In summary, a new recursive weighted PCA has been created to better align the primary axis, and thus direction, of a Shower with its minimally ionising section. The aim of which was to improve the accuracy of angular measurements with respect to interaction vertices for showering particles. This method was seen in Fig. 5.10 to improve the angle between the direction of the reconstructed particle, and the direction of the truth information from the MC.

A new length measurement based on the projection of 3D space points onto the PCA axis has been suggested as an alternative to the using the eigenvalue length of Eqn. 5.7. This alternative method appears to have better agreement with ‘by-eye’ inspection of lower momenta events, while converging with the current method at higher momenta, as seen in Fig. 5.15.

An energy estimator has been constructed, and five correction factors have been identified to deal with a systematic underestimation of a Shower’s energy. Once the corrections are applied to data from the PDSP-BD dataset, the resolution of the energy estimator was calculated to be between 1%-6%, with better resolution for higher momenta.

A method for calculating the  $dE/dx$  for the minimally ionising section of a EM shower has been demonstrated, and provides results expected for minimally ionising particles in LAr. The longitudinal profile of EM showers has also been seen to follow the expected gamma function.

A method for estimating the starting point of the particle cascade for an EM shower has also been suggested, and with further investigations could prove a useful parameter for the future.

# 6

## Proton Decay

---

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

**Frank Herbert - Dune - 1965**

**T**he stability of the proton is a rather important aspect of the universe, but is not explicitly ensured by the standard model. The local symmetries that give birth to the photon, and the conservation of electric charge, do not exist for baryon number - the quantity behind the apparent stability of the proton. In fact, it is predicted to not be a conserved quantity by most Grand Unified Theories (GUTs), where nucleon decay is often a low energy observable phenomena [76–82]. At a high level, current GUTs can be categorised in two ways; supersymmetric, and non-supersymmetric models. While each category of GUT can be subdivided further, supersymmetric models predominantly predict the  $p \rightarrow K^+\bar{\nu}$  decay mode as being most common, while non-supersymmetric models predict  $p \rightarrow e^+\pi^0$  as the most common decay mode. The current lifetime limits for these decay modes can be seen in Fig. 6.1.

Fig. 6.1 also compares a number of GUT model predictions with the current limits, which completely reject some of the models, and highly constrain a number of the others. Fig. 6.1 also shows that the limits for  $p \rightarrow e^+\pi^0$  have been set much more stringently than for the other decay mode. This is in part because the  $p \rightarrow e^+\pi^0$  has a larger predicted branching fraction, and the kaon's energy resulting from  $p \rightarrow K^+\bar{\nu}$  is regularly below the Cherenkov threshold for water, and so cannot be seen by Super-Kamiokande very easily. LArTPCs, however, are able to reconstruct particles

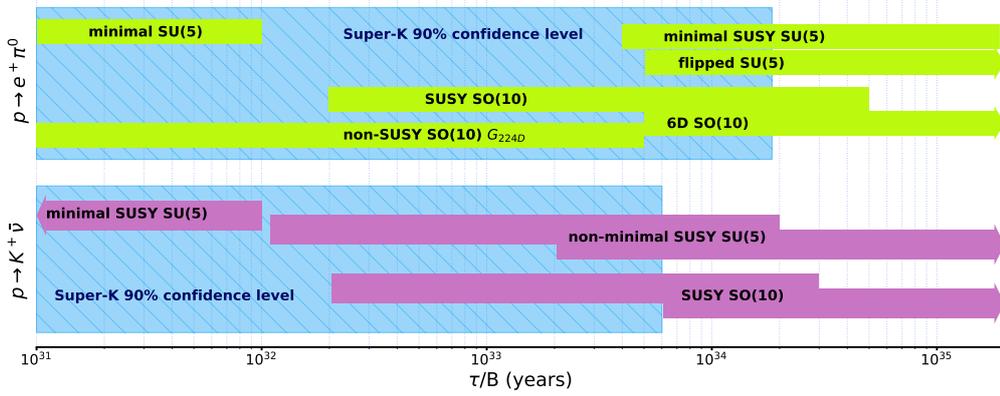


Figure 6.1: Summary of the 90% confidence level lower limits from Super-K for  $p \rightarrow K^+\bar{\nu}$  and  $p \rightarrow e^+\pi^0$ , and predicted lifetime from various GUT models. Adapted from [28].

with

much lower energies, and so DUNE will be able to more often see the signature kaon from proton decay, as well as the kaon's decay products - the dominant six of which can be seen in Table 6.1.

Table 6.1: The six most probable decay modes for a  $K^+$  [34].

Decay Mode	Branching Fraction
$K^+ \rightarrow \mu^+\nu_\mu$	$63.56 \pm 0.11\%$
$K^+ \rightarrow \pi^+\pi^0$	$20.67 \pm 0.08\%$
$K^+ \rightarrow \pi^+\pi^+\pi^-$	$5.583 \pm 0.024\%$
$K^+ \rightarrow \pi^0e^+\nu_e$	$5.07 \pm 0.04\%$
$K^+ \rightarrow \pi^0\mu^+\nu_\mu$	$3.352 \pm 0.033\%$
$K^+ \rightarrow \pi^+\pi^0\pi^0$	$1.760 \pm 0.023\%$

As such, DUNE will be a major driving force behind lifetime limits set on the  $p \rightarrow K^+\bar{\nu}$  decay channel. It will however, of course, be able to also push the current lifetime limits of the  $p \rightarrow e^+\pi^0$  channel set by Super-Kamiokande. The expected sensitivities to the two discussed proton decay channels can be seen in Fig. 6.2 for DUNE and the future Hyper-Kamiokande. These two experiments will make a great pairing for increasing the overall lifetime limit on proton decay.

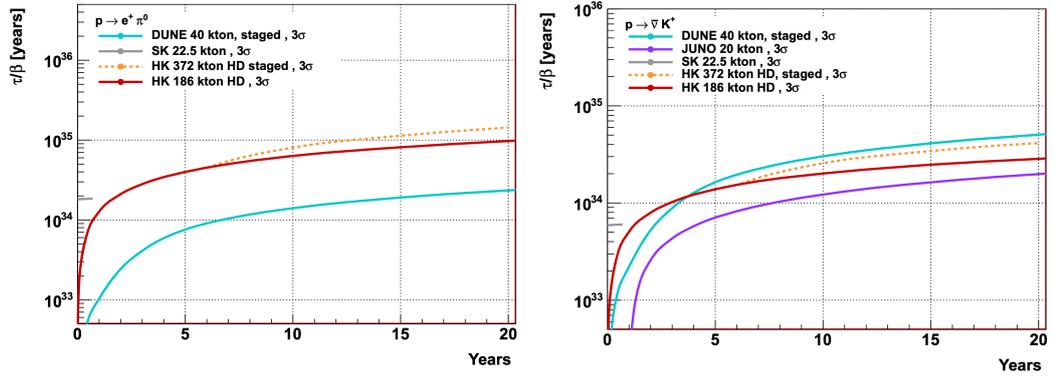


Figure 6.2: DUNE (cyan) and Hyper-Kamiokande’s (red) expected sensitivity to  $p \rightarrow e^+ \pi^0$  (left) and  $p \rightarrow K^+ \bar{\nu}$  (right) [83].

The major background for the  $p \rightarrow K^+ \bar{\nu}$  decay mode are interactions caused by atmospheric neutrinos. A CC  $\nu_\mu$  interaction has a proton and a muon emerging from the interaction vertex, which very closely mimics the dominant decay of  $K^+$ s, where a kaon decays into a muon. As seen in Table 6.2, CC  $\nu_\mu$  interactions are the most commonly expected interactions in DUNE. The event rates shown in Table 6.2 are calculated using the Bartol model of atmospheric neutrino flux [84], and neutrino interaction cross-sections from GENIE. Neutrino oscillation effects are not included in these calculations. An example signal event and a CC  $\nu_\mu$  background event can be seen in Fig. 6.3. The colour of each pixel represents the total charge of hits contained within the pixel. The background event of Fig. 6.3 is a particularly challenging type of background to remove as it is very similar to the signal event in a lot of aspects. The key to removing this background event is in identifying the location in the rise in charge deposition of the kaon and proton, in relation to the vertex with the muon.

For proton lifetime measurements where no signal is observed, any background should be reduced as much as possible as it is a dominant factor in calculations. The DUNE TDR [28] suggests an optimal rate of one event per megaton year (Mt · year), which requires the percentage of background events that need to be rejected be equal to

$$\left(1 - \frac{1}{2886 \times 100}\right) \times 100 = 99.9997\%$$

where the denominator’s factor of 100 is from scaling the number of events expected in Table 6.2 to 1 Mt · year.

Table 6.2: Expected rate of atmospheric neutrino interactions in  $^{40}\text{Ar}$  for a 10 kt · year exposure (not including oscillations) [28].

10 kt · year	CC	NC	Total
$\nu_\mu$	1038	398	1436
$\bar{\nu}_\mu$	280	169	449
$\nu_e$	597	206	83
$\bar{\nu}_e$	126	72	198
Total	2014	845	2886

Final state interactions (FSI) complicate the identification of signal events even further. As the kaon from  $p \rightarrow K^+ \bar{\nu}$  propagates out of the nucleus, it can undergo scattering processes causing it to lose energy, or to cause the emission of a nucleon alongside it. The scattering processes have the effect of reducing the average kinetic energy of the kaons, making them immediately more difficult to reconstruct. Furthermore, the presence of associated nucleons blurs the  $dE/dx$  measurement at the start of the kaon track, complicated particle identification.

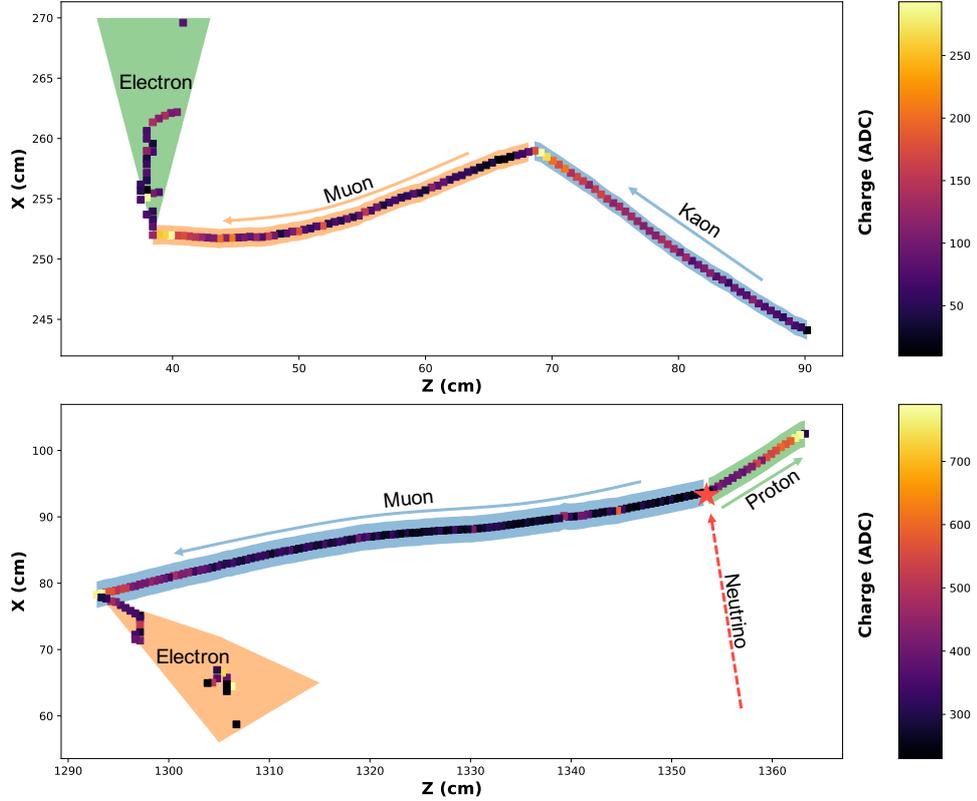


Figure 6.3: Images for a  $p \rightarrow K^+ \bar{\nu}$  signal event (top), and a CC  $\nu_\mu$  background event (bottom).

## 6.1 Event Reconstruction and Classification

The aim of this analysis is to develop further the proton decay sensitivity for the  $p \rightarrow K^+\bar{\nu}$  channel using the Pandora reconstruction software, and a new event selection based on some of the tools developed earlier. While reading the results of this analysis it is important to keep in mind that this is the first time Pandora has been used on such events, and has not yet received any sort of tuning. Fig. 6.4 shows

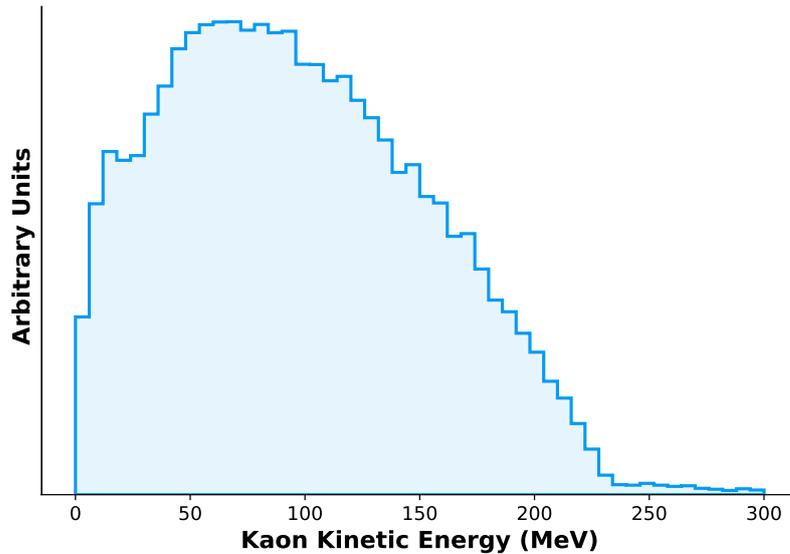


Figure 6.4: True kaon kinetic energy for  $p \rightarrow K^+\bar{\nu}$

the kinetic energy of the kaons in the signal MC events. The nucleon decay signal events are generated using GENIE v2.12.10, and are generated using the DUNE FD geometry.

This analysis is done using Pandora’s output PFParticles, as opposed to the track and shower objects provided by LArSoft, as these provide hierarchy information for the reconstructed particles. The PFParticles provide *art associations* to the reconstructed 2D hits, which are used as the basis for all features used in the BDT for signal and background separation. The efficiency for reconstructing a particle of type  $x^\pm$  can be defined as [28]

$$\epsilon_{x^\pm} = \frac{\text{events where true } x^\pm \text{ can be matched to a reconstructed track}}{\text{events with } x^\pm \text{ particle}} \quad (6.1)$$

The efficiency can be viewed as a function of either the true length of the kaon, or the kinetic energy of the kaon. The reconstruction efficiencies as functions of both

variables can be seen in Fig. 6.5. The red points show the current performance of the

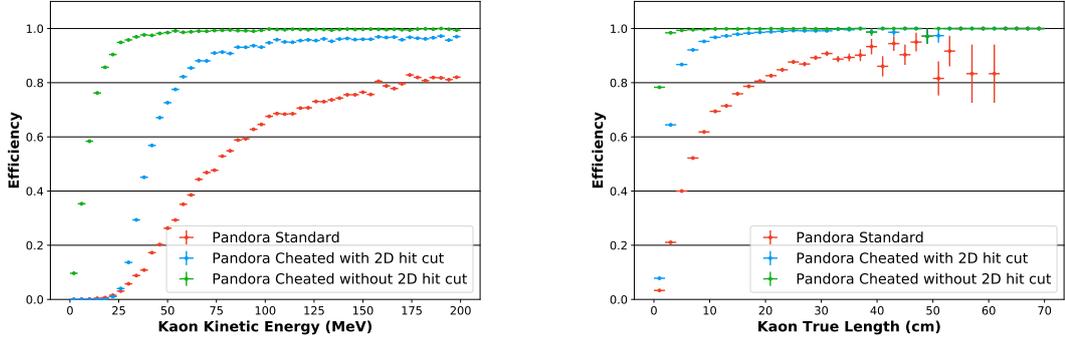


Figure 6.5: Efficiency for reconstructed kaons using Pandora (top).

standard version v03\_11\_01g of the Pandora reconstruction software. The efficiency as a function of the kinetic energy tops out at 80%, and has a relatively slow turn on. The efficiency as a function of true length performs better, and tops out at around 90%. The overall efficiency for the standard Pandora reconstruction is 53%. However, it is possible to glimpse at the full potential of Pandora by ‘cheating’ the reconstruction. When the reconstruction is cheated, the hierarchy information, connecting the true MC particles together, is passed in parallel with the normal reconstructed 2D hits to Pandora. Pandora can then use this extra hierarchy information to more accurately group 2D hits together. The rest of the reconstruction chain remains identical to the un-cheated case. The result of cheating the reconstruction is shown by the green points on Fig. 6.5, which shows almost perfect reconstruction efficiency - and any inefficiencies are from very low energy particles simply not having any reconstructable 2D hits. This is of course a theoretically perfect scenario, and a more realistic outcome for when Pandora is appropriately tuned in this regime, can be seen by applying a threshold on the number of 2D hit for the cheated reconstruction. This is seen in Fig. 6.5 with the blue points, where a requirement of four 2D hits is imposed on whether or not a kaon is considered to have been reconstructed by Pandora or not.

To discriminate between signal and background events a BDT is used, where every reconstructed event (signal or background) is reduced to a number of key features. Firstly, the two largest PFParticles for each event are selected, and all other PFParticles are disposed of - though the total number of PFParticles is retained. The size of a PFParticle is determined by the number of associated 2D hits. For each of the two remaining PFParticles, the features listed in Table 6.3 are then recorded, and the largest PFParticle is labelled as ‘particle 1’, and the smaller labelled as ‘particle 2’. Fig. 6.6 shows two example features for the BDT. The left image shows

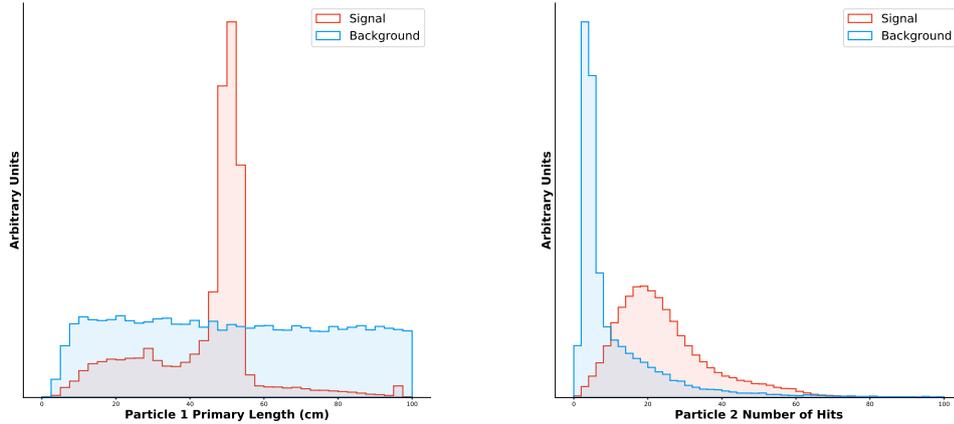


Figure 6.6: Two example BDT features.

the length of the largest particle, with a peaked signal distribution from primary kaons and a uniform background distribution cutting off at 100 cm (as per selection requirements discussed below). The right image shows the number of hits for the smaller particle. The correlation matrix for the BDT features can be seen in Fig. 6.7.

Table 6.3: Features used in the BDT for separating nucleon decay signal,  $p \rightarrow K^+ \bar{\nu}$ , from atmospheric neutrino background.

Feature	Description
Total Particles	The total number of PFParticles in an event's hierarchy
dE/dx	The dE/dx at the start of the PFParticle
End dE/dx	The dE/dx at the end of the PFParticle
No. Hits	The number of collection plane 2D hits in the PFParticle
Primary Length	The length of the primary axis from a principal component analysis performed on the particle
Secondary Length	The length of the secondary axis from a principal component analysis performed on the particle
Energy	The estimated energy of the PFParticle
Direction	The direction of the PFParticle from the recursive weighted PCA

Three pre-BDT selection requirements are then used to remove obvious background events. The first requirement is a minimal threshold on the number of particles in an event, requiring at least two PFParticles - preferentially selecting events with a kaon and a decay product. The second requirement is a limit on the length of the longest PFParticle to be less than 100 cm - removing events from high energy

neutrino interactions. The third requirement is a limit of 400 2D hits for the particles with the most 2D hits - removing events with large EM showers. The pre-BDT selection requirements leave 17.5% of the background, and 74.85% of the signal remaining.

The BDT is implemented using scikit-learn's [85] ensemble.AdaBoostClassifier. The initial sample size is  $\sim 82\,000$  signal events, and  $\sim 166\,000$  background events. After the pre-BDT selection requirements above are applied,  $\sim 28\,500$  background events, and  $\sim 61\,000$  signal events are left. As it is important for there to be equal number of background and signal events when training a BDT, 20 000 events are taken from each remaining sample. While a slightly larger training set would have been more ideal, this is certainly adequate for an initial foray.

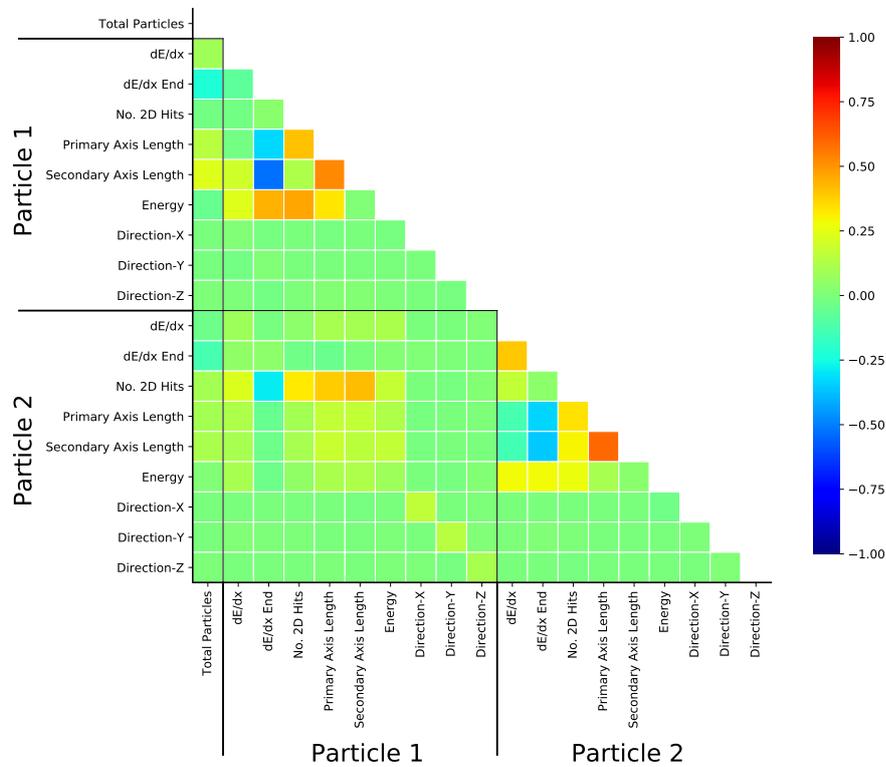


Figure 6.7: Correlation matrix for the features of the BDT. The colour scheme denotes the level of correlation between two variables.

The BDT responses can be seen in Fig. 6.8. A BDT response threshold is then selected to give the suggested background rejection of 99.9997% when combined with the pre-BDT selection requirements. The signal efficiency from the training sample is then 23%.

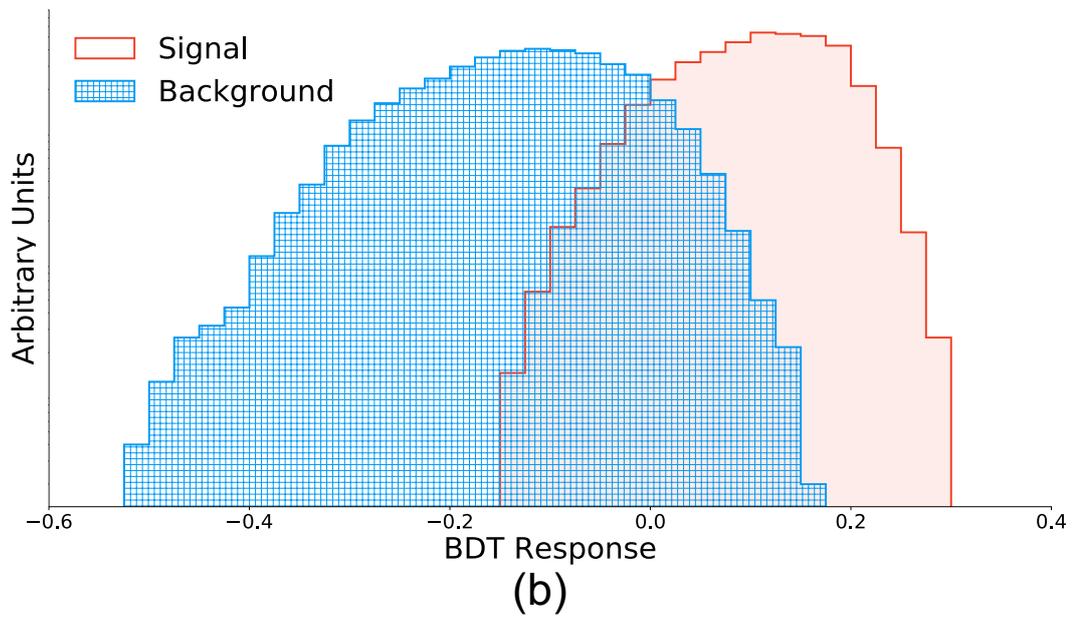
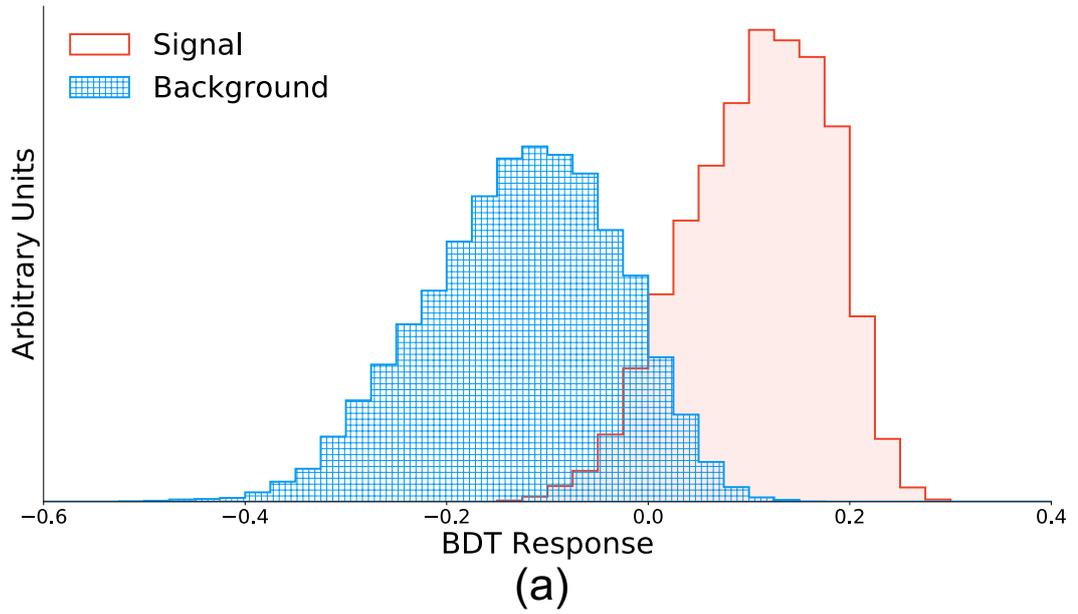


Figure 6.8: BDT response for  $p \rightarrow K^+ \bar{\nu}$  for background (blue hatch) and signal (red), with linear (a) and log (b) axes.

The remaining signal and background events are then used to test the BDT model. However, because the number of events remaining is relatively low, once the remaining events have been processed by the BDT their response is Poisson fluctuated to approximate a larger dataset. The results of doing so, and calculating their signal

efficiencies and background rejections for the chosen BDT response threshold from the training sample, can be seen in Fig. 6.9. The signal efficiency across all samples remains consistent with the training sample, with a mean signal efficiency of 23.64%. However, the background rejection falls slightly below the required value to a peak rejection of 99.99922%, corresponding to 2.25 events per Mt · year.

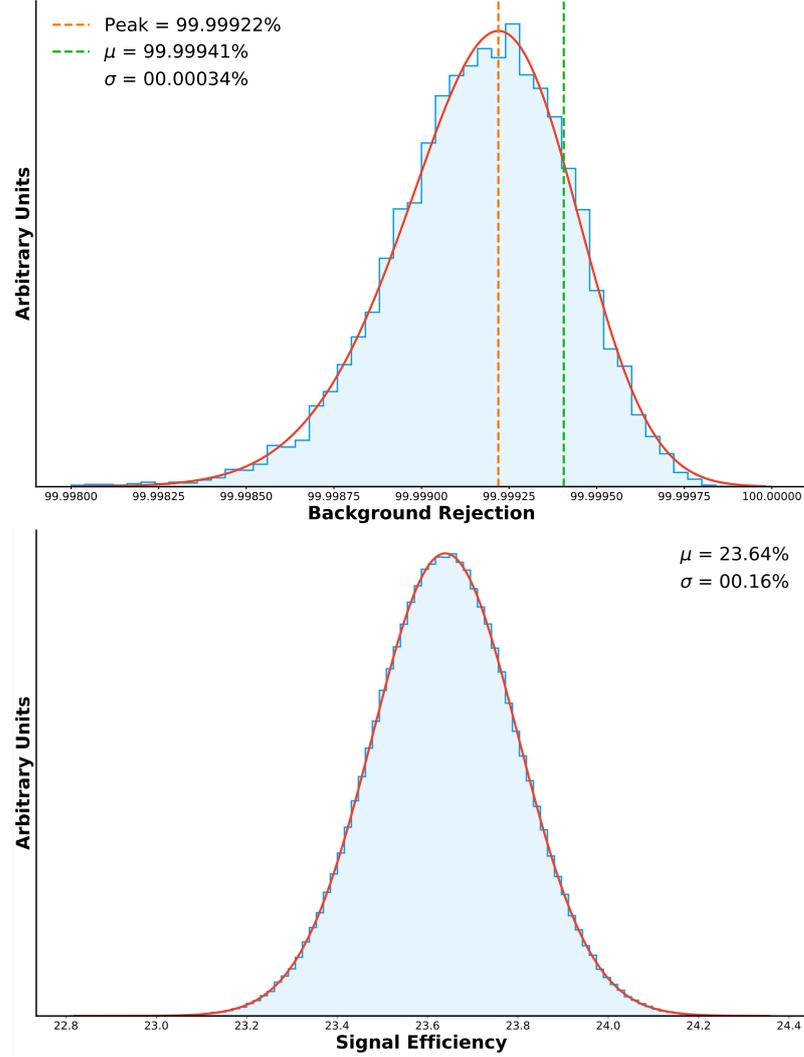


Figure 6.9: Background rejection (left) and signal efficiency (right) for Poisson fluctuated BDT responses to the non-training sample of events.

Assuming no signal events are to be observed for a given exposure, and a constant signal efficiency and background rejection, then the partial lifetime lower limit can be calculated using [86]

$$\tau/B(p \rightarrow K^+\bar{\nu}) > \lambda \cdot N_p \cdot \epsilon_{sig} \cdot \frac{1}{S} \quad (6.2)$$

where  $\lambda$  is the exposure in  $\text{kt} \cdot \text{years}$ ,  $N_p$  the number of protons in one kiloton of argon ( $2.7 \times 10^{32}$ ),  $\epsilon_{sig}$  the signal selection efficiency, and  $S$  the upper limit for 90% confidence level for the expected number of background and signal events [86]. Assuming ten years of running, and DUNE's complete 40 kt fiducial mass, an exposure of 400  $\text{kt} \cdot \text{years}$  can be reached. The value of  $S$  can be obtained from Feldman-Cousins [87] for an experiment measuring a Poisson distributed variable that is expecting no signal events but is expecting background events. For the case presented here, with 0.9 background events expected in 400  $\text{kt} \cdot \text{years}$ ,  $S = 3.2$ . Combined with the current signal efficiency of 23.64%, a lower limit on the partial lifetime can be set as  $8.0 \times 10^{33}$  years. While this exceeds the current limit set by Super-Kamiokande of  $5.9 \times 10^{33}$  years [88], it is less than the sensitivity presented by the method used in the DUNE TDR [28]:  $1.3 \times 10^{34}$  years.

DUNE is sure to play a prominent role in the future of proton lifetime limit measurements. While the reconstruction software used here and in the DUNE TDR differ, Pandora and Projection Matching Algorithm (PMA) respectively, both are currently unoptimised for the proton decay regime but show good signs for performance improvements. A study done for the DUNE TDR showed PMA's current overall tracking efficiency being increased from 58% to 80% - enhancing the final selection performance, which resulted in the  $1.3 \times 10^{34}$  years lower lifetime limit. If the cheated reconstruction for Pandora is considered, as discussed above, then Pandora's overall tracking efficiency can be improved from 53% to 80% also. This could see the lower lifetime limit of the method presented here reaching as high as  $1.6 \times 10^{34}$  years with the same background rejection. Further advancements to the reconstruction techniques and abilities employed at DUNE over the next five years, and full utilisation of the Pandora particle flow hierarchy, will certainly bring many other unforeseen improvements before the DUNE program starts taking data at the full size far detectors.



## Summary

---

*"We change our past by what we learn."*

**Frank Herbert - Dune - 1965**

**T**his thesis has explored a number of interesting areas around the capabilities of LArTPCS, but the full power of the technology is still yet to be utilised.

In Chapter 5 an in-depth analysis of the reconstruction of electron induced electromagnetic showers was performed. The current PCA method employed by the Pandora reconstruction software was developed further to utilise a weighted recursion technique that provides a better estimate for the direction of an electromagnetic shower. Better estimations on the direction of an electromagnetic shower will help with analyses that require angular information of particles as they emerge from interaction vertices. The subsequent primary axis from the weighted recursive PCA was also used repeatedly in determination of other electromagnetic shower properties. A new measure for the length of the electromagnetic shower based on 3D reconstructed hit's projections onto the primary axis was shown to better describe an electromagnetic shower's length at lower momenta, while converging to the currently employed method at higher momenta. The longitudinal profile of electromagnetic showers along their primary axes was shown to be well described by the gamma function presented in the PDG review [34]. The transverse profiles in terms of charge density of concentric rings of equal area in the secondary-tertiary PCA axes plane highlighted a wider dense central core for the 6 and 7 GeV/c PDSP-BD dataset as compared to showers in the PDSP-MC dataset. The energy estimator was shown to

systematically underestimate a shower's energy by 20-21%. This was recovered by identifying five energy loss mechanisms using the PDSP-MC dataset, and a correction factor based on the mean expected energy was found. Once the correction factors were applied to energy estimates for the PDSP-BD dataset it was shown that an energy resolution of 1-6%, typical for calorimeters, can be achieved. The energy estimates for individual hits in the first 5 cm of the shower were used to calculate a  $dE/dx$  value for the minimally ionising section, and was shown to be the expected value. Finally, an attempt was made to find the position along the shower's primary PCA axis at which the electromagnetic cascade initiates. While initial results are promising, further investigation will be needed to improve upon the currently used threshold limit for defining the start of the cascade.

A sensitivity estimate of the  $p \rightarrow K + \bar{\nu}$  decay channel was also presented using a BDT based on some features developed in Chapter 5, and utilising the power of the Pandora reconstruction software. The current BDT performance stands at a 23.64% signal efficiency and a 99.99922% background rejection rate. This gives a sensitivity of  $8.0 \times 10^{33}$  years for the  $p \rightarrow K + \bar{\nu}$  decay channel. With potential tuning to the Pandora reconstruction software, a sensitivity of  $1.6 \times 10^{34}$  years could be achieved, even if the background rejection rate remains the same. It will be interesting to see how this is developed further, especially if a powerful CNN predictor is introduced.

Some interesting years certainly lay ahead as more research is done to unlock the full potential of the LArTPC technology, with exciting developments to surely be made in the area of deep learning which greatly benefits from the high resolution particle reconstruction.

# Bibliography

- [1] E. D. Bloom *et al.*, “High-Energy Inelastic  $e - p$  Scattering at  $6^\circ$  and  $10^\circ$ ,” *Phys. Rev. Lett.* **23** (Oct, 1969) 930–934.
- [2] M. Breidenbach *et al.*, “Observed Behavior of Highly Inelastic Electron-Proton Scattering,” *Phys. Rev. Lett.* **23** (Oct, 1969) 935–939.
- [3] CMS Collaboration, S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B* **716** no. 1, (2012) 30 – 61.
- [4] ATLAS Collaboration, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B* **716** no. 1, (2012) 1 – 29.
- [5] Q. R. Ahmad, R. C. Allen, T. C. Andersen, J. D. Anglin, J. C. Barton, E. W. Beier, M. Bercovitch, J. Bigu, S. D. Biller, R. A. Black, and *et al.*, “Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory,” *Physical Review Letters* **89** no. 1, (Jun, 2002) 16 – 26. <http://dx.doi.org/10.1103/PhysRevLett.89.011301>.
- [6] J. Chadwick, “The intensity distribution in the magnetic spectrum of beta particles from radium (B + C),” *Verh. Phys. Gesell.* **16** (1914) 383–391.
- [7] W. Pauli, “Letter to Lise Meitner,”  
[http://cds.cern.ch/record/83282/files/meitner\\_0393.pdf](http://cds.cern.ch/record/83282/files/meitner_0393.pdf).
- [8] F. Reines and C. L. Cowan, “Neutrino physics,” *Physics Today* **10** (1957) 12–18.
- [9] F. L. Wilson, “Fermi’s Theory of Beta Decay,” *Am. J. Phys.* **36** no. 12, (1968) 1150–1160.
- [10] E. Fermi, “Versuch einer theorie der  $\beta$ -strahlen. i,” *Zeitschrift für Physik* **88** no. 3, (Mar, 1934) 161–177.

- [11] F. Reines, C. L. Cowan, F. B. Harrison, A. D. McGuire, and H. W. Kruse, “Detection of the Free Antineutrino,” *Phys. Rev.* **117** (Jan, 1960) 159–173. <https://link.aps.org/doi/10.1103/PhysRev.117.159>.
- [12] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, “Detection of the Free Neutrino: a Confirmation,” *Science* **124** no. 3212, (1956) 103–104.
- [13] N. Cooper, “Los Alamos Science, Number 25 – 1997: Celebrating the neutrino,” <https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-97-2534-02>.
- [14] P. Curd, “Anaxagoras,” in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, ed. Metaphysics Research Lab, Stanford University, winter 2015 ed., 2015.
- [15] J. N. Bahcall, “Solar Neutrinos. I. Theoretical,” *Phys. Rev. Lett.* **12** (Mar, 1964) 300–302. <https://link.aps.org/doi/10.1103/PhysRevLett.12.300>.
- [16] J. N. Bahcall and R. K. Ulrich, “Solar models, neutrino experiments, and helioseismology,” *Rev. Mod. Phys.* **60** (Apr, 1988) 297–372. <https://0-link-aps-org.pugwash.lib.warwick.ac.uk/doi/10.1103/RevModPhys.60.297>.
- [17] G. Danby, J.-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger, “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos,” *Phys. Rev. Lett.* **9** (Jul, 1962) 36–44. <https://0-link-aps-org.pugwash.lib.warwick.ac.uk/doi/10.1103/PhysRevLett.9.36>.
- [18] B. Pontecorvo, “Inverse beta processes and nonconservation of lepton charge,” *Sov. Phys. JETP* **7** (1958) 172–173.
- [19] Z. Maki, M. Nakagawa, and S. Sakata, “Remarks on the Unified Model of Elementary Particles,” *Progress of Theoretical Physics* **28** no. 5, (11, 1962) 870–880. <https://doi.org/10.1143/PTP.28.870>.
- [20] B. Pontecorvo, “Neutrino Experiments and the Problem of Conservation of Leptonic Charge,” *Sov. Phys. JETP* **26** (1968) 984–988.
- [21] **Kamiokande** Collaboration, Y. Fukuda *et al.*, “Solar neutrino data covering solar cycle 22,” *Phys. Rev. Lett.* **77** (1996) 1683–1686.

- [22] Y. Fukuda *et al.*, “Measurements of the Solar Neutrino Flux from Super-Kamiokande’s First 300 Days,” *Phys. Rev. Lett.* **81** (Aug, 1998) 16 – 26. <https://link.aps.org/doi/10.1103/PhysRevLett.81.1158>.
- [23] **GALLEX** Collaboration, W. Hampel *et al.*, “GALLEX solar neutrino observations: Results for GALLEX III,” *Physics Letters B* **388** no. 2, (1996) 384 – 396. <http://www.sciencedirect.com/science/article/pii/S0370269396011215>.
- [24] **GALLEX** Collaboration, M. Altmann *et al.*, “GNO solar neutrino observations: results for GNO I,” *Physics Letters B* **490** no. 1, (2000) 16 – 26. <http://www.sciencedirect.com/science/article/pii/S0370269300009151>.
- [25] **SAGE** Collaboration, J. Abdurashitov *et al.*, “Results from SAGE (The Russian-American gallium solar neutrino experiment),” *Physics Letters B* **328** no. 1, (1994) 234 – 248. <http://www.sciencedirect.com/science/article/pii/0370269394904545>.
- [26] **SAGE** Collaboration, V. Gavrin *et al.*, “Solar neutrino results from SAGE,” *Nuclear Physics B - Proceedings Supplements* **91** no. 1, (2001) 36 – 43. <http://www.sciencedirect.com/science/article/pii/S0920563200009208>. Neutrino 2000.
- [27] K. Zuber, *Neutrino Physics: Second Edition*. CRC Press, 2012.
- [28] **DUNE** Collaboration, B. Abi *et al.*, “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II DUNE Physics,” [arXiv:2002.03005](https://arxiv.org/abs/2002.03005) [hep-ex].
- [29] **DUNE** Collaboration, B. Abi *et al.*, “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE,” [arXiv:2002.02967](https://arxiv.org/abs/2002.02967) [physics.ins-det].
- [30] **T2K** Collaboration, K. Abe *et al.*, “Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations,” *Nature* **580** no. 7803, (2020) 339–344.
- [31] L. Wolfenstein, “Neutrino Oscillations in Matter,” *Phys. Rev.* **D17** (1978) 2369–2374. [,294(1977)].
- [32] S. P. Mikheyev and A. Y. Smirnov, “Resonant amplification of oscillations in matter and solar-neutrino spectroscopy,” *Il Nuovo Cimento C* **9** no. 1, (1986) 17–26. <https://doi.org/10.1007/BF02508049>.

- [33] E. Majorana, “Teoria simmetrica dellelettrone e del positrone,” *Nuovo Cim.* **14** (1937) 171–184.
- [34] **Particle Data Group** Collaboration, M. Tanabashi *et al.*, “Review of particle physics,” *Phys. Rev. D* **98** (Aug, 2018) 030001.  
<https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [35] F. Piquemal, “The supernemo project,” *Physics of Atomic Nuclei* **69** (2006) 2096–2100.
- [36] K.-H. Ackermann *et al.*, “The gerda experiment for the search of  $0 \nu\beta\beta$  decay in  $76 \text{ ge}$ ,” *The European Physical Journal C* **73** no. 3, (2013) 1–29.
- [37] M. Auger *et al.*, “The exo-200 detector, part i: detector design and construction,” *Journal of Instrumentation* **7** no. 05, (2012) P05010.
- [38] “Fermilab creative services.”  
<https://vms.fnal.gov/asset/detail?recid=1961263>.
- [39] D. R. Nygren and J. N. Marx, “The Time Projection Chamber,” *Physics Today* **31** (1978) 46–53.
- [40] C. Rubbia, “The Liquid Argon Time Projection Chamber: A New Concept for Neutrino Detectors,” *CERN-EP-INT-77-08*, *CERN-EP-77-08* (1977) .  
<https://inspirehep.net/literature/857394>.
- [41] S. Amerio *et al.*, “Design, construction and tests of the icarus t600 detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **527** no. 3, (2004) 329 – 410. <http://www.sciencedirect.com/science/article/pii/S0168900204004966>.
- [42] C. Anderson *et al.*, “The ArgoNeuT detector in the NuMI low-energy beam line at fermilab,” *Journal of Instrumentation* **7** no. 10, (Oct, 2012) P10019–P10019.
- [43] **MicroBooNE** Collaboration, R. Acciarri *et al.*, “Design and Construction of the MicroBooNE Detector,” *JINST* **12** no. 02, (2017) P02017,  
[arXiv:1612.05824](https://arxiv.org/abs/1612.05824) [[physics.ins-det](https://arxiv.org/abs/1612.05824)].
- [44] **LArIAT** Collaboration, F. Cavanna, M. Kordosky, J. Raaf, and B. Rebel, “LArIAT: Liquid Argon In A Testbeam,” [arXiv:1406.5560](https://arxiv.org/abs/1406.5560) [[physics.ins-det](https://arxiv.org/abs/1406.5560)].
- [45] **DUNE** Collaboration, B. Abi *et al.*, “The Single-Phase ProtoDUNE Technical Design Report,” [arXiv:1706.07081](https://arxiv.org/abs/1706.07081) [[physics.ins-det](https://arxiv.org/abs/1706.07081)].

- [46] **DUNE** Collaboration, B. Abi *et al.*, “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume IV Far Detector Single-phase Technology,” arXiv:2002.03010 [physics.ins-det].
- [47] K. Riesselmann, “Tests start at CERN for large-scale prototype of new technology to detect neutrinos,” <https://news.fnal.gov/2019/10/tests-start-cern-large-scale-prototype-new-technology-detect-neutrinos/>.
- [48] K. Riesselmann, “DUNE scientists see particle tracks with dual-phase technology,” <https://www.symmetrymagazine.org/article/dune-scientists-see-particle-tracks-with-dual-phase-technology>.
- [49] P. Derwent *et al.*, “Proton Improvement Plan-II,” <https://projectx-docdb.fnal.gov/cgi-bin/ShowDocument?docid=1232>.
- [50] **DUNE** Collaboration, P. Adamson *et al.*, “Optimized beam cdr,” [https://docs.dunescience.org/cgi-bin/ShowDocument?docid=4559&fbclid=IwAR3QgySKjwGLPKvZQqQ5uUB\\_OK1oVtTV7anFi\\_juljvjM4ldzgODDK01Tsg](https://docs.dunescience.org/cgi-bin/ShowDocument?docid=4559&fbclid=IwAR3QgySKjwGLPKvZQqQ5uUB_OK1oVtTV7anFi_juljvjM4ldzgODDK01Tsg).
- [51] CERN, “First particle tracks seen in prototype for international neutrino experiment,” <https://home.cern/news/press-release/experiments/first-particle-tracks-seen-prototype-international-neutrino>.
- [52] L. Biron, “Success after a three-year sprint,” <https://news.fnal.gov/2019/01/success-after-a-three-year-sprint/>.
- [53] L. Biron, “ProtoDUNE in pictures,” <https://www.symmetrymagazine.org/article/protodune-in-pictures>.
- [54] **DUNE** Collaboration, F. Cavanna, “NP04/ProtoDUNE-SP Status Report 2019,” Tech. Rep. CERN-SPSC-2019-015. SPSC-SR-252, CERN, Geneva, Mar, 2019. <https://cds.cern.ch/record/2669536>.
- [55] C. Andreopoulos *et al.*, “The GENIE Neutrino Monte Carlo Generator,” *Nucl. Instrum. Meth. A* **614** (2010) 87–104, arXiv:0905.2517 [hep-ph].
- [56] S. Agostinelli *et al.*, “Geant4a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250 – 303. <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [57] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Transactions on Nuclear Science* **53** no. 1, (2006) 270–278.

- [58] J. Allison *et al.*, “Recent developments in geant4,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835** (2016) 186 – 225. <http://www.sciencedirect.com/science/article/pii/S0168900216306957>.
- [59] F. Tellander and N. Charitonidis, “Tertiary particle production and target optimization of the H2 beam line in the SPS North Area,” [arXiv:1610.09066](https://arxiv.org/abs/1610.09066) [physics.acc-ph].
- [60] H. W. Atherton *et al.*, “Precise measurements of particle production by 400 GeV/c protons on beryllium targets,” <https://cds.cern.ch/record/133786>.
- [61] “Muons, Inc., G4Beamline.” <http://www.muonsinternal.com/muons3/G4beamline>.
- [62] J. S. Marshall and M. A. Thomson, “The Pandora software development kit for pattern recognition,” *The European Physical Journal C* **75** no. 9, (Sep, 2015) . <http://dx.doi.org/10.1140/epjc/s10052-015-3659-3>.
- [63] **MicroBooNE** Collaboration, R. Acciarri and Others, “The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector,” [arXiv:1708.03135](https://arxiv.org/abs/1708.03135) [hep-ex].
- [64] A. Blake, L. Escudero Sanchez, S. Green, J. Marshall, and L. H. Whitehead, “The Pandora multi-algorithm approach to automated pattern-recognition of cosmic-ray muon and test beam particle interactions in the ProtoDUNE-SP detector,” *In preparation* .
- [65] C. Green, J. Kowalkowski, M. Paterno, M. Fischler, L. Garren, and Q. Lu, “The Art Framework,” *J. Phys. Conf. Ser.* **396** (2012) 022020.
- [66] “The art Event Processing Framework.” <https://art.fnal.gov>.
- [67] **LArSoft** Collaboration. <https://larsoft.org>.
- [68] “Liquid Argon Properties (Tables and Calculators).” <https://lar.bnl.gov/properties/>.
- [69] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. N. Capdevielle, “corsika: A monte carlo code to simulate extensive air showers,” <https://www.ikp.kit.edu/corsika/index.php>.

- [70] C. Haggmann, D. Lange, J. Verbeke, and D. Wright, “proton-induced cosmic-ray cascades in the atmosphere,”  
[http://nuclear.llnl.gov/simulation/doc\\_cry\\_v1.7/cry.pdf](http://nuclear.llnl.gov/simulation/doc_cry_v1.7/cry.pdf).
- [71] M. Mooney, “The MicroBooNE Experiment and the Impact of Space Charge Effects,” in *Meeting of the APS Division of Particles and Fields*. 11, 2015.  
 arXiv:1511.01563 [physics.ins-det].
- [72] “dunetpc.” <https://cdcvs.fnal.gov/redmine/projects/dunetpc>.
- [73] G. Guennebaud, B. Jacob, *et al.*, “Eigen v3.” [Http://eigen.tuxfamily.org](http://eigen.tuxfamily.org), 2010.
- [74] S. Amoruso and others, “Study of electron recombination in liquid argon with the icarus tpc,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **523** no. 3, (2004) 275 – 286. <http://www.sciencedirect.com/science/article/pii/S0168900204000506>.
- [75] **DUNE** Collaboration, B. Abi *et al.*, “The Single-Phase ProtoDUNE Technical Design Report,” arXiv:1706.07081.
- [76] H. Georgi and S. L. Glashow, “Unity of All Elementary-Particle Forces,” *Phys. Rev. Lett.* **32** (Feb, 1974) 438–441.  
<https://link.aps.org/doi/10.1103/PhysRevLett.32.438>.
- [77] J. C. Pati and A. Salam, “Is Baryon Number Conserved?,” *Phys. Rev. Lett.* **31** (Sep, 1973) 661–664.  
<https://link.aps.org/doi/10.1103/PhysRevLett.31.661>.
- [78] S. Dimopoulos, S. Raby, and F. Wilczek, “Proton decay in supersymmetric models,” *Physics Letters B* **112** no. 2, (1982) 133 – 136. <http://www.sciencedirect.com/science/article/pii/0370269382903136>.
- [79] P. Langacker, “Grand unified theories and proton decay,” *Physics Reports* **72** no. 4, (1981) 185 – 385. <http://www.sciencedirect.com/science/article/pii/0370157381900594>.
- [80] W. De Boer, “Grand unified theories and supersymmetry in particle physics and cosmology,” *Progress in Particle and Nuclear Physics* **33** (1994) 201 – 301.  
<http://www.sciencedirect.com/science/article/pii/0146641094900450>.

- [81] G. Altarelli and D. Meloni, “A non supersymmetric SO(10) grand unified model for all the physics below MGUT,” *Journal of High Energy Physics* **2013** no. 8, (2013) 21. [https://doi.org/10.1007/JHEP08\(2013\)021](https://doi.org/10.1007/JHEP08(2013)021).
- [82] P. Nath and P. Fileviez Pérez, “Proton stability in grand unified theories, in strings and in branes,” *Physics Reports* **441** no. 5, (2007) 191 – 317. <http://www.sciencedirect.com/science/article/pii/S0370157307000683>.
- [83] **Hyper-Kamiokande** Collaboration, K. Abe *et al.*, “Hyper-kamiokande design report,” [arXiv:1805.04163](https://arxiv.org/abs/1805.04163) [physics.ins-det].
- [84] G. D. Barr, T. K. Gaisser, P. Lipari, S. Robbins, and T. Stanev, “Three-dimensional calculation of atmospheric neutrinos,” *Phys. Rev. D* **70** (Jul, 2004) 023006. <https://0-link-aps-org.pugwash.lib.warwick.ac.uk/doi/10.1103/PhysRevD.70.023006>.
- [85] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research* **12** (2011) 2825–2830.
- [86] C. Alt, B. Radics, and A. Rubbia, “Neural-network-driven proton decay sensitivity in the  $p \rightarrow \bar{\nu}K^+$  channel using large liquid argon time projection chambers,” [arXiv:2010.06552](https://arxiv.org/abs/2010.06552) [hep-ex].
- [87] G. J. Feldman and R. D. Cousins, “Unified approach to the classical statistical analysis of small signals,” *Phys. Rev. D* **57** (Apr, 1998) 3873–3889. <https://link.aps.org/doi/10.1103/PhysRevD.57.3873>.
- [88] K. Abe *et al.*, “Search for proton decay via  $p \rightarrow \bar{\nu}K^+$  using 260 kiloton $\cdot$ year data of Super-Kamiokande,” *Physical Review D* **90** no. 7, (Oct, 2014) . <http://dx.doi.org/10.1103/PhysRevD.90.072005>.

*“And I say: Look! I have hands!  
But the people all around me say: What are hands?”*

---

**Frank Herbert - Dune - 1965**