A Blessing in Disguise

Characterisations of ProtoDUNE photon showers for neutrino measurements in DUNE

Milo Alexander Vermeulen
of the University of Amsterdam

ISBN: 978-90-831893-0-7

A Blessing in Disguise – Characterisations of ProtoDUNE photon showers for neutrino measurements in DUNE
Thesis, Universiteit van Amsterdam

Cover: A pion track and two photon showers in the ProtoDUNE detector.
Printed by Proefschriftenprinten.nl in the Netherlands

This work is part of the research programme of the Dutch Research Council (NWO) and has been performed at the National Institute for Subatomic Physics (Nikhef).

A Blessing in Disguise

Characterisations of ProtoDUNE photon showers for neutrino measurements in DUNE

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 9 november 2021, te 15.00 uur

door Milo Alexander Vermeulen
geboren te Amsterdam

Promotiecommissie

Promotor:	prof. dr. S.C.M. Bentvelsen	Universiteit van Amsterdam
Copromotor:	dr. F. Filthaut	Radboud Universiteit Nijmegen
Overige leden:	prof. dr. S.J. de Jong	Radboud Universiteit Nijmegen
	prof. dr. ir. P.J. de Jong	Universiteit van Amsterdam
	prof. dr. A.P. Colijn	Universiteit van Amsterdam
	prof. dr. M. Vreeswijk	Universiteit van Amsterdam
	dr. M.E.J. Postma	Nikhef
	dr. G. Lehmann Miotto	CERN
	prof. dr. W. Verkerke	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Contents

Introduction

Neutrinos are fascinating particles because of their superlative status in high energy physics. Of all the elementary particles that we know to have mass, neutrinos are by far the lightest. Along with photons and possible dark matter candidates they are among the most abundant particles in the universe. Of all known particles they have the weakest coupling to other particles, giving them the ability to pass through entire planets and stars with ease. At the same time, the neutrino might hold the answers to some outstanding mysteries in particle physics. Great effort has been made over the last few decades to construct large-scale neutrino detectors so that a few of these elusive particles may be investigated.

This thesis was written in the context of DUNE: the Deep Underground Neutrino Experiment. Being an upcoming precision neutrino experiment, it aims to measure neutrino properties to an unprecedented degree using an intense neutrino source and a set of large yet sensitive detectors. A prototype of DUNE's largest detector was built at CERN and operated in late 2018 under the name of ProtoDUNE. Too small and too noisy with other particles to detect neutrinos, the prototype instead recorded interactions of energetic charged particles, demonstrating the effectiveness of the technology behind DUNE. Despite their differing particle sources, many of the interactions expected in DUNE were seen in ProtoDUNE and could be studied in preparation for eventual neutrino collisions.

The detector technology behind DUNE and ProtoDUNE is a relative newcomer in the field and promises high-resolution images of particle interaction events at a higher rate than ever before. For all its capabilities, however, this detection mechanism still faces various challenges in identifying and characterising neutrinos. Solid conclusions about the nature of neutrinos require vast volumes of data, all of which need to be processed and saved in a sustainable manner. Neutrino processes must be identified precisely, since there are various ways in which one type of neutrino may disguise itself as another. The energy of an incoming neutrino, being an important piece of the

physics puzzle, must be pieced together from the tracks and showers of the secondary particles it creates in an interaction. The work in this thesis touches on these three concepts: data access and compression, neutrino identification and energy calibration.

Thesis outline and contributions of the author

Chapter 1 starts with a brief introduction of the Standard Model and goes into detail on neutrino physics, particularly the phenomenon of neutrino oscillation. It also discusses outstanding questions in the field and provides an overview of various historic and currently operational neutrino experiments.

Chapters 2 and 3 expand on the upcoming neutrino experiment DUNE and its prototype ProtoDUNE. Since much of the technology is shared between the two, chapter 2 focuses on DUNE's physics goals and experimental setup, whereas chapter 3 discusses the technical details of both experiments in detail. The latter also gives an overview of ProtoDUNE's period of data taking in 2018.

Chapter 4 considers the data acquisition system of ProtoDUNE. In particular, it discusses a prototype data acquisition called FELIX. I contributed to the host software of FELIX, providing raw detector data access methods to downstream systems. I also investigated software-based data compression algorithms for use in the FELIX data acquisition chain.

Chapter 5 discusses the complex process of reconstruction, with special focus on the main reconstruction chain in DUNE and ProtoDUNE: Pandora. It expands on the analysis design used in this thesis, as well as the used data samples.

Chapter 6 considers a procedure to suppress a major background in DUNE, performed on ProtoDUNE data. I implemented a method to determine the energy deposition per unit length, or $\mathrm{d}E/\mathrm{d}x$, of photon showers. I found that a clear distinction could be made from positron showers, which had already been characterised in a previous study.

Chapter 7 discusses a calibration method performed on ProtoDUNE data. Observing a similar photon shower sample to that used in chapter 6, I reconstructed the π^0-particles that produced them. I made use of the π^0-particle rest mass as a standard candle to correct the energy reconstruction of electromagnetic showers. The method is foreseen to be applied to DUNE data in much the same manner.

Chapter 1

Neutrino oscillation

Ancient cultures from Greece to Tibet described the world around them using just a handful of elements, most commonly water, earth, fire and air. The idea that nature could be made up of indivisible particles was not widespread, however. Atomism as a concept was debated by ancient Greek and Indian philosophers, but lacked a solid scientific base. Natural philosophy in the middle ages produced an expanded list of alchemical elements that included many modern metals such as lead and gold, but it was only in the 19th century that chemicals were observed to react with each other in certain ratios, sparking theories that they were comprised of particles called atoms, "indivisible" in Greek. For instance, water and ammonia would each split into two distinct components, suggesting that they were really compound particles. Thus the theory that all matter is composed of indivisible atoms gained scientific traction. Scientists such as Alexandre-Émile Béguyer de Chancourtois and Dmitri Mendeleev organised the then-known atoms into a periodic table which has expanded to a total of 118 elements today, several tens of which are purely synthetic.

As indivisible as atoms were thought to be, more divisions could be made still. The electron was discovered to be one part of the atom, and the nuclear proton and neutron particles were discovered shortly after. Nuclear physics looked tidy with only three particles making up all matter in the universe, but the proton and neutron would turn out to be divisible themselves, and multiple additional particles were discovered that did not consist of those found in atoms. With the rise of high-energy physics through particle accelerators and cosmic ray observatories, more particles were observed that had markedly different properties from the particles known so far. As these particles were grouped and classified, quarks were thought to be the constituents of hadrons, just like atoms had been theorised to be the building blocks of molecules. Particle

accelerators and detectors continued to improve throughout the 20th century, allowing for more energetic and more numerous particles to be produced and detected. One by one, the currently known quarks and leptons were discovered and characterised. Together with the force-mediating bosons, they were grouped into a theory called the Standard Model, seen in figure 1.1. All of these particles are currently considered elementary, but time will tell if they truly are.

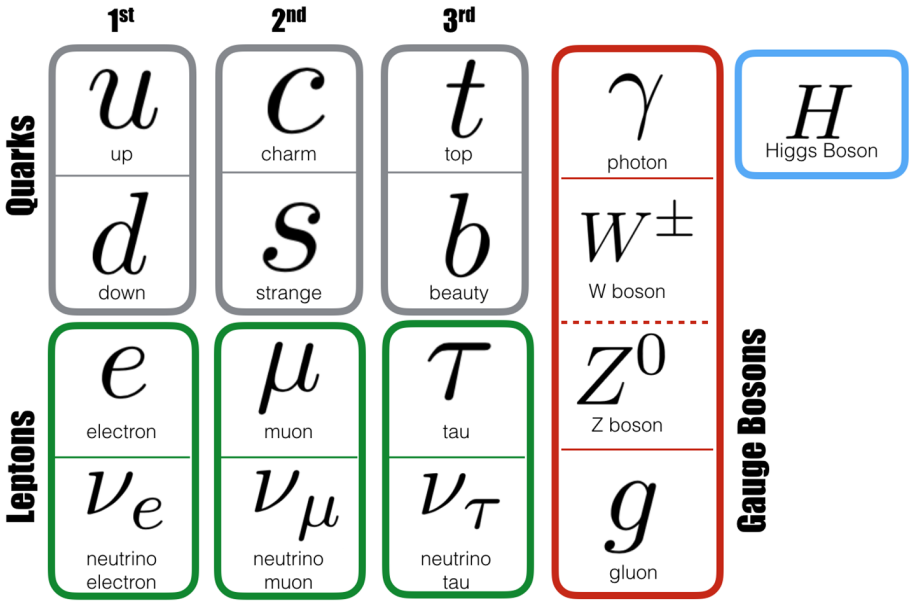

Figure 1.1: The current standard model of particle physics, describing all known elementary particles. These are grouped by generation ("1st, 2nd, 3rd") and are further divided into quarks, leptons and gauge bosons. Taken from [1].

1.1 The Standard Model

The Standard Model consists of bosons and fermions. The former comprise the carriers for every fundamental force except gravity: the gluon mediates the strong nuclear force, the photon the electromagnetic force and the W and Z bosons the weak nuclear force. Additionally, the Higgs boson is an excitation of the Higgs field, which gives the elementary particles their mass. The fermions of the Standard Model are made up of quarks and leptons, which can each be divided into three generations. The most common quarks are the first-generation up and down variety which make up protons and neutrons. Quarks from the second and third generations

can be found in unstable exotic matter created by powerful particle accelerators and cosmic phenomena. Leptons consist of the charged leptons: the electron, muon and tau particle, and the neutrinos. Each of the three neutrinos has a charged lepton counterpart for which they are named: the electron neutrino, muon neutrino and tau neutrino.

Whether elementary particles interact with one another depends on their fundamental properties. Electromagnetic interaction only occurs between electrically charged fermions and the force-mediating photon. This includes all quarks and all non-neutrino leptons. The strong nuclear force, on the other hand, connects particles with 'colour,' which is a three-dimensional analogue to electrical charge, being able to take on the values of (anti-)red, (anti-)green and (anti-)blue. Aside from the force-mediating gluon, only quarks possess this quantum property. The weak nuclear force, mediated by the W and Z bosons, depends on the intrinsic chirality, or handedness, of particles. As will be explored in section 1.3, only particles that are left-handed have been observed to participate in this type of interaction, breaking a fundamental symmetry of physics.

Despite the apparent completeness of the Standard Model, questions remain. The mechanism that gives particles their mass is now known, but it remains unclear what causes the wide range of masses among the elementary particles. It is not known whether generations of particles beyond the currently recognised ones exist. In its current state, the Standard Model also does not explain matter-antimatter asymmetry in the universe, nor does it cover the fourth fundamental force: gravity.

1.1.1 Neutrinos

Postulated by Wolfgang Pauli in 1930 and discovered experimentally in 1956 [2], the neutrino was originally proposed as a solution to the apparent non-conservation of momentum in beta decay. Without knowledge of the neutrino, the beta decay process

$$(A, Z) \rightarrow (A, Z+1) + e^{-} + \bar{\nu}_e \tag{1.1}$$

is expected to produce electrons at a momentum uniquely set by the initial and final nucleon mass, here denoted as (A, Z) and $(A, Z+1)$ respectively. Here, A is the mass number of the nucleus, which counts its total number of nucleons, and Z is the atomic number, or number of protons within the nucleus. The observation of the electron's

continuous kinetic energy spectrum in this decay process implied the existence of a second particle that invisibly carries a variable portion of the reaction energy away: the neutrino. Since its first discovery in inverse beta decay, this lepton has been under intense study. Multiple neutrino flavours were discovered: the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ), which are so-called weak eigenstates of the neutrino. Counting solar neutrinos, the Homestake experiment [3] consistently observed a third of the expected solar electron neutrino flux. It was later confirmed by the Sudbury Neutrino Observatory (SNO) [4] that the solar neutrino flux contains a non-electron flavour, suggesting that a fraction of the electron neutrinos changed flavour while travelling between the sun and the earth. This so-called neutrino oscillation will be discussed in detail in this chapter.

Neutrinos have no electrical or colour charge, but they do interact via the weak nuclear force. In order to produce a detectable signal, neutrinos must exchange either a W or Z boson with particles in matter, which can then be observed via much stronger electromagnetic interactions. Electrons, protons and neutrons are all valid targets for interaction via a Z boson, which is called neutral current interaction due to the electric neutrality of the force carrier. W bosons, through their charged nature, mediate charged current interactions between neutrinos and atomic matter. Although charged current interactions are possible between all neutrino flavours and matter particles, some interactions have an energy threshold that is too high to consider here. In particular, charged current interactions between electrons in matter and ν_μ/ν_τ are only kinematically allowed for $E_{\nu_\mu} > 11\,\text{GeV}$ and $E_{\nu_\tau} > 3090\,\text{GeV}$ [5], for which reason these interactions will be disregarded. The most common interactions between neutrinos and matter are shown diagrammatically in figure 1.2.

1.2 Neutrino oscillation in vacuum

One of the neutrino's more unique features is its ability to change flavour over time. For instance, a neutrino created with the electron flavour may be observed as a muon neutrino after it has travelled some distance. The underlying mechanism of this phenomenon is the principle that the flavour states of neutrinos ($|\nu_\alpha\rangle$) are not pure mass eigenstates ($|\nu_i\rangle$) but are instead superpositions of these. Transforming between

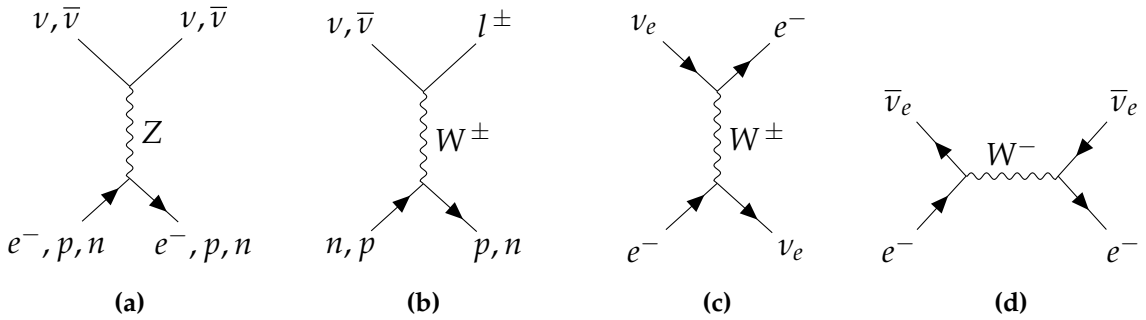

Figure 1.2: The main interactions of neutrinos with ordinary matter. All (anti-)neutrinos undergo neutral current interaction with quarks and electrons (a) and charged current interaction with quarks (b). In the latter, an (anti-)lepton $l^{\pm}$ is emitted corresponding to the neutrino flavour. At neutrino energies < 11 GeV, only electron neutrinos (c) and anti-electron neutrinos (d) undergo charged current interactions with electrons.

the flavour and mass bases is done with the neutrino mixing matrix:

$$|\nu_\alpha\rangle = U^\dagger |\nu_i\rangle\,, \quad |\nu_i\rangle = U |\nu_\alpha\rangle\,, \tag{1.2}$$

where the dagger operation ($U^\dagger$) refers to Hermitian conjugation and U refers to the PMNS (Pontecorvo-Maki-Nakagawa-Sakata) mixing matrix. In the three-neutrino model, the neutrino flavour α can be any of e, μ, or τ while the index i refers to the mass eigenstates and can take on the values 1, 2 and 3. In order to preserve probability amplitudes between basis transformations, the mixing matrix is unitary so that $UU^\dagger = U^\dagger U = I$. In other words, transforming from the flavour basis to the mass basis and back returns the original state.

Neutrino oscillation can be derived from the combination of two principles: the existence of separate flavour and mass bases, and the mass differences between the neutrino mass eigenstates. Mass eigenstates propagate through space as plane waves, making it relatively simple to describe their propagation through the time-dependent Schrödinger equation:

$$\begin{aligned} i\hbar \frac{\mathrm{d}\,|\nu_i(t)\rangle}{\mathrm{d}t} &= H\,|\nu_i(t)\rangle = E_i\,|\nu_i(t)\rangle \\ \Rightarrow |\nu_i(t)\rangle &= e^{-iE_i t/\hbar}\,|\nu_i(0)\rangle\,, \end{aligned} \tag{1.3}$$

where $|\nu_i(0)\rangle$ describes the initial state of the pure mass eigenstate and $|\nu_i(t)\rangle$ describes its final state after travelling for a time t. E_i is the energy of the mass eigenstate as

given by the Hamiltonian H. In natural units, the reduced Planck constant and the speed of light $\hbar = c = 1$. Therefore for highly relativistic particles $x = ct = t$, and the time derivative in equation 1.3 can be replaced with a derivative over distance travelled x. A simplification can be made if the momentum of the neutrino is much greater than the mass of the mass eigenstates, which is always the case for detectable neutrinos:

$$\begin{aligned} E_i = \sqrt{p^2 + m_i^2} \;&\approx\; p + \frac{m_i^2}{2p} \;\approx\; p + \frac{m_i^2}{2E} \\ \Rightarrow \quad |\nu_i(t)\rangle = e^{-i\left(p + \frac{m_i^2}{2E}\right)x} |\nu_i(0)\rangle &= S\,|\nu_i(0)\rangle\,, \end{aligned} \tag{1.4}$$

where m_i is the mass of the mass eigenstate, p is the total momentum and E is the total energy of the neutrino. The evolution operator S governs the propagation of neutrino states, and in the case of pure neutrino mass eigenstates in vacuum, as in equation 1.4, this operator is diagonal.

The propagation of mass eigenstates through empty space is a simple affair, but in reality neutrinos undergo interactions based on their flavour eigenstate. An electron neutrino may undergo a charged current interaction with an electron, but never with a muon. Likewise, neutrinos are produced according to their flavour eigenstate, which is always a superposition of all mass eigenstates. A charged pion will almost always decay to a muon and muon neutrino, for example. As both production and detection depend on flavour and not mass, it is more relevant to speak of neutrino propagation in terms of the flavour basis. The mixing matrix can be used to transform between the two bases, as shown in equation 1.2. This needs to be done twice to describe the journey of a neutrino from creation to interaction: once to transform the initial flavour state into the mass basis and once after propagation of the mass eigenstates, when the neutrino interacts with matter at some distance x:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) = |S_{\beta\alpha}|^2 &= |\left\langle \nu_\beta \right| S \left| \nu_\alpha \right\rangle|^2 \\ &= \left| \sum_{ij} \left\langle \nu_j \right| U_{\beta j}\, S\, U^*_{\alpha i} \left| \nu_i \right\rangle \right|^2 = \left| \sum_i U^*_{\alpha i} U_{\beta i}\, e^{-i\frac{m_i^2}{2E}x} \right|^2 . \end{aligned} \tag{1.5}$$

Notably, the term $-ipx$ has been omitted from the propagator since it affects all eigenstates equally and thereby does not participate in the description of neutrino oscillation.

The rest of the derivation is expanded in appendix A. The result is the following neutrino oscillation probability:

$$P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} - 4\sum_{i<j} \mathrm{Re}\left(J^{ij}_{\alpha\beta}\right) \sin^2\left(\frac{\Delta m^2_{ij}}{4E}x\right) - 2\sum_{i<j} \mathrm{Im}\left(J^{ij}_{\alpha\beta}\right) \sin\left(\frac{\Delta m^2_{ij}}{2E}x\right), \tag{1.6}$$

where $\Delta m^2_{ij} = m^2_i - m^2_j$ signifies the difference between squared neutrino masses and $\mathrm{Im}\left(J^{ij}_{\alpha\beta}\right) = U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j} = \pm J$ is the Jarlskog invariant. The invariant depends exclusively on the parameters of the neutrino mixing matrix and is only nonzero in the case that neutrinos violate the CP symmetry [6], making it an experiment-independent metric of CP violation. This is explained in more detail in section 1.3.

From equation 1.6, it becomes clear why neutrino mass differences are necessary for neutrino oscillation. Without them, no distance travelled x would let the oscillation probability deviate from the initial $\delta_{\alpha\beta}$: there would be no oscillation at all. This also implies that all but the lightest neutrino must have a nonzero mass in order to participate in oscillation.

1.2.1 Two-neutrino oscillation

For the purpose of simplicity, one can start by picturing a universe in which just two neutrino flavour and mass eigenstates exist. Due to constraints related to unitarity, the mixing matrix can be parametrised as a simple rotation matrix with only one variable: the angle θ:

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{1.7}$$

Entering this matrix into equation 1.6 then yields the following neutrino oscillation probability:

$$P(\nu_\alpha \to \nu_\beta)|_{\alpha\neq\beta} = \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2_{12}}{4E}L\right) \tag{1.8}$$

and $P(\nu_\alpha \to \nu_\alpha) = 1 - P(\nu_\alpha \to \nu_\beta)$. Notably, $J^{ij}_{\alpha\beta}$ has been simplified to $\pm\frac{1}{4}\sin^2(2\theta)$ and Δm^2_{ij} has been replaced with Δm^2_{12}, since in the two-neutrino case this is the only squared mass difference present. Additionally, the imaginary second part of equation 1.6's right-hand side has been omitted, since the two-dimensional mixing matrix is fully real. The neutrino path length L is synonymous with x.

Figure 1.3 shows the two-neutrino oscillation probability as a function of $(\Delta m^2_{12}/4E)L$ for various mixing angles. As the mixing angle decreases from the maximal value $\pi/4$, so does the amount of mixing between the two neutrino states. The squared neutrino mass difference and the neutrino energy dictate the oscillation frequency over distance.

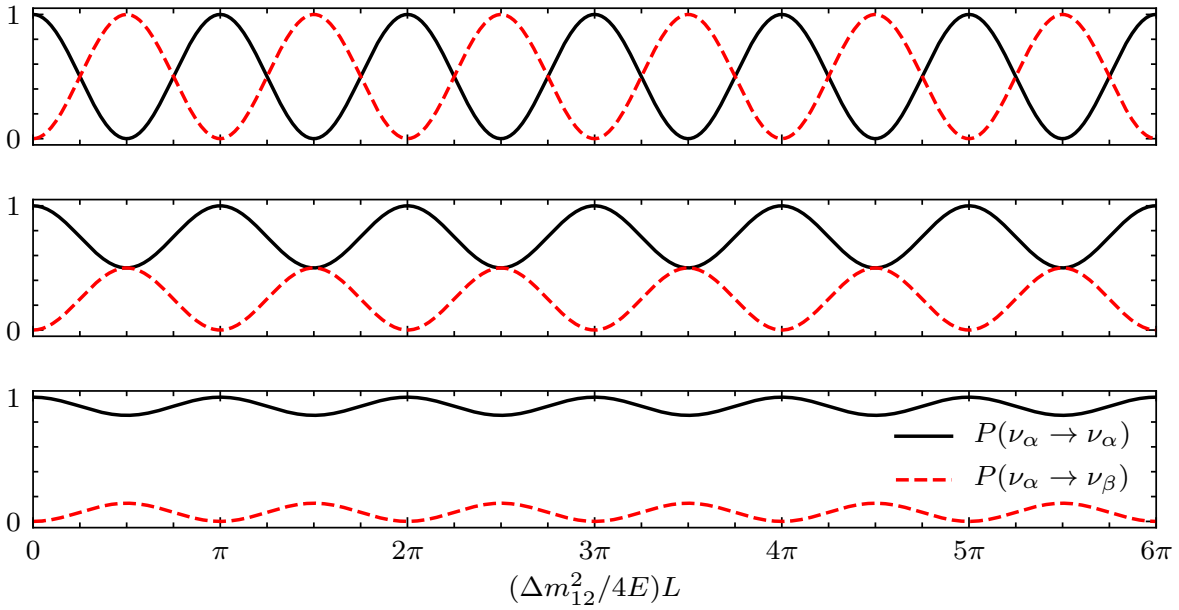

Figure 1.3: Two-neutrino oscillation for maximal mixing at $\theta = \pi/4$ (top), $\theta = \pi/8$ (middle) and $\theta = \pi/16$ (bottom).

1.2.2 Three-neutrino oscillation

The Standard Model possesses three generations of fermions and the neutrino sector is no exception. The three-dimensional mixing matrix is considerably more complex than its two-dimensional counterpart and can be parametrised using no fewer than

six parameters: three mixing angles and three phases:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{21} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\eta_1} & 0 & 0 \\ 0 & e^{i\eta_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.9}$$

where $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$ refer to the three mixing angles, δ_{CP} is the CP-violating phase (expanded upon in section 1.3) and $\eta_{1,2}$ are the Majorana phases. The latter are only nonzero if neutrinos are Majorana particles, which is an ongoing field of study. These phases are not visible in oscillation phenomena, leaving just three angles and one phase to parametrise the three-dimensional mixing matrix. Combining everything into a single matrix yields

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{CP}} & c_{13}c_{23} \end{pmatrix}, \tag{1.10}$$

which is considerably more intricate than its two-dimensional analogue. There are unfortunately no significant simplifications that can be applied when entering this mixing matrix into the general oscillation probability of equation 1.6.

Table 1.1: Best-fit values for neutrino oscillation parameters in the normal mass ordering (NO) and inverted mass ordering (IO) assumptions. Taken from [7].

		Best fit $\pm 1\sigma$	
Parameter	Description	NO	IO
θ_{12} [°]	Solar oscillation angle	$33.44^{+0.77}_{-0.74}$	$33.45^{+0.78}_{-0.75}$
θ_{23} [°]	Atmospheric oscillation angle	$49.2^{+0.9}_{-1.2}$	$49.3^{+0.9}_{-1.1}$
θ_{13} [°]	Reactor oscillation angle	$8.57^{+0.12}_{-0.12}$	$8.60^{+0.12}_{-0.12}$
δ_{CP} [°]	CP-violating phase	197^{+27}_{-24}	282^{+26}_{-30}
Δm^2_{21} [10^{-5} eV^2]	Long range oscillation	$7.42^{+0.21}_{-0.20}$	$7.42^{+0.21}_{-0.20}$
Δm^2_{31} [10^{-3} eV^2]	Short range oscillation	$2.517^{+0.026}_{-0.028}$	$-2.498^{+0.028}_{-0.028}$

The current best-fit values for the neutrino oscillation parameters are shown in table 1.1. They are obtained by a global fit that makes use of the results of a wide variety of experiments, discussed in more detail in section 1.5. Using these values, the mixing matrix can be filled in and described as follows in terms of its absolute squared components:

$$\begin{pmatrix} |U_{e1}|^2 & |U_{e2}|^2 & |U_{e3}|^2 \\ |U_{\mu 1}|^2 & |U_{\mu 2}|^2 & |U_{\mu 3}|^2 \\ |U_{\tau 1}|^2 & |U_{\tau 2}|^2 & |U_{\tau 3}|^2 \end{pmatrix} = \begin{pmatrix} 0.676 \pm 0.036 & 0.301 \pm 0.036 & 0.022 \pm 0.002 \\ 0.135 \pm 0.091 & 0.328 \pm 0.121 & 0.500 \pm 0.096 \\ 0.162 \pm 0.096 & 0.352 \pm 0.122 & 0.469 \pm 0.096 \end{pmatrix}. \quad (1.11)$$

Equation 1.11 shows, within a 3σ error, what fraction of each flavour eigenstate is composed of each mass eigenstate and vice versa. For instance, considering the first row, a ν_e flavour eigenstate is composed for $\sim 68\%$ of mass eigenstate ν_1, $\sim 30\%$ of ν_2 and $\sim 2\%$ of ν_3. The columns likewise describe the constituents of the mass eigenstates in terms of flavour eigenstates. As expected of a unitary matrix, all rows and all columns add up to 1 within experimental error. It is worth noting that the neutrino mixing matrix is far from diagonal, with relatively large values in off-diagonal terms. This stands in contrast to the quark mixing matrix, the so-called Cabibbo-Kobayashi-Masakawa (CKM) matrix, which has a similar function as the neutrino mixing matrix but is much closer to the identity [8].

The presence of three neutrino mass eigenstates implies the existence of two independent squared mass differences Δm^2_{21} and Δm^2_{31}, and a third squared mass difference $\Delta m^2_{32} = \Delta m^2_{31} - \Delta m^2_{21}$. The general equation for neutrino oscillation (equation 1.6) receives extra terms relative to the two-neutrino case containing these squared mass differences. Since the mixing matrix contains a complex phase, the imaginary part of equation 1.6 also contributes to the final oscillation probability. Three-neutrino oscillation in vacuum is visualised in figure 1.4 for all three initial neutrino flavour states. A long-range effect can be seen in which Δm^2_{21} dominates whereas the much larger Δm^2_{31} causes oscillation on a shorter length scale.

An ambiguity arises with the existence of three squared mass differences: there are multiple possible orderings for the neutrino masses. Through oscillation experiments, it has been found that $\Delta m^2_{21} \ll \Delta m^2_{31}$, placing m_1 and m_2 close together and m_3 further apart, but it is not known whether m_3 is much smaller or much larger than m_1 and m_2.

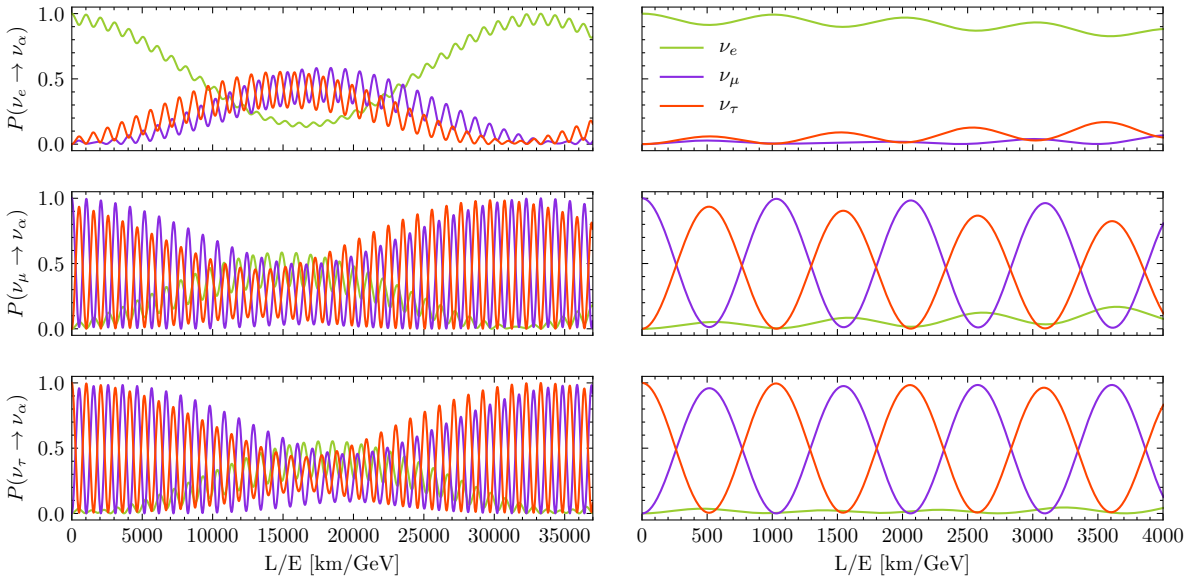

Figure 1.4: Three-neutrino oscillation in vacuum for long-range (left) and short-range (right) baselines. The initial neutrinos are ν_e (top), ν_μ (middle) and ν_τ (bottom).

This is visualised in figure 1.5, where $\Delta m^2_{31} > 0$ signifies the so-called normal mass ordering (NO) and $\Delta m^2_{31} < 0$ signifies inverted ordering (IO).

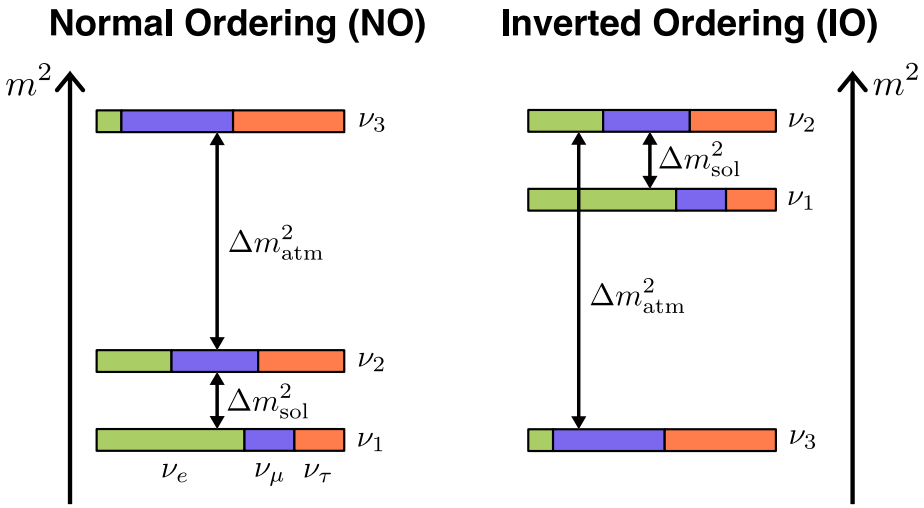

Figure 1.5: A visual representation of the possible neutrino mass hierarchies given two squared mass differences. The coloured bars indicate the flavour components of each mass eigenstate.

Although nature features at least three oscillating neutrinos, the two-neutrino case discussed in section 1.2.1 is often a good approximation. This is due to the fact that the two independent squared mass differences in the neutrino mass ordering are of very different magnitudes: following the conventions of figure 1.5, $\Delta m^2_{21}/\Delta m^2_{31} \approx 3.3 \cdot 10^{-2}$. Considering the real part of neutrino oscillation as described in equation 1.6, at small L/E, the term $\sin^2(\Delta m^2_{21} L/4E)$ is negligible compared to the terms $\sin^2(\Delta m^2_{31} L/4E) \approx \sin^2(\Delta m^2_{32} L/4E)$. Conversely, at large L/E, the oscillation of the $\sin^2(\Delta m^2_{21} L/4E)$-term becomes apparent whereas the terms $\sin^2(\Delta m^2_{31} L/4E) \approx \sin^2(\Delta m^2_{32} L/4E)$ average out. The imaginary part of the neutrino oscillation equation behaves similarly. For observable neutrino energies, short-range oscillation is visible over atmospheric distances whereas long-range oscillation is visible over much longer path lengths, such as the distance between the sun and the earth. For this reason, $\Delta^2_{31} \approx \Delta m^2_{32}$ is also called Δm^2_{atm} and Δm^2_{21} is called Δm^2_{sol}, as indicated in figure 1.5.

1.2.3 From natural units to useful units

This chapter's equations have made use of natural units, in which $\hbar = c = 1$. In long-baseline neutrino experiments, more intuitive units are the kilometre for distance travelled, eV^2 for the squared neutrino mass difference and GeV for the neutrino energy. In the neutrino oscillation probability (for example equation 1.6), all variables to be converted are contained in sin terms. They can be converted as follows:

$$\frac{\Delta m^2_{ij} L}{4E} = 1.267 \cdot \left(\frac{\Delta m^2_{ij}}{\text{eV}^2}\right) \left(\frac{L}{\text{km}}\right) \left(\frac{\text{GeV}}{E}\right), \tag{1.12}$$

where use was made of the fact that $\hbar c = 1.973 \cdot 10^{-7}$ eV m.

1.3 CP violation

Much of physics is built on the presence of symmetry, generally described as invariance under a certain transformation. Taking spatial translation as an example, one can expect the speed of light to be unchanged when moving one metre to the left, or for that matter one parsec to the left. Translation invariance leads to conservation of momentum in classical mechanics, making it a useful construct to consider. Other ex-

amples of symmetry-conservation pairs are rotational symmetry leading to conserved angular momentum and time symmetry leading to energy conservation.

Although nature seems to be trivially symmetric, the weak nuclear force in particle physics violates some surprising members of the symmetry family. Parity (P) symmetry concerns mirrored systems, stating that a process will occur just as often as its mirror image. This assumption was famously disproved in 1956 by measuring the radioactive decay of cobalt-60 [9] and finding that beta radiation was emitted preferentially in the opposite direction of the parent atom's spin. The reason behind this discrepancy is in the nature of weakly interacting particles such as neutrinos. In nature, fermions are only observed to undergo weak nuclear interactions if they are left-handed, defined as having a negative chirality. This is an intrinsic quantum parameter that manifests itself physically as helicity, which determines whether particles are right or left-handed based on whether their chirality points along or opposite their momentum vector. The helicity of a particle is identical to its chirality if it is massless. The same goes for practically all directly detectable neutrinos, as these are always highly relativistic. As the neutrino is created and interacts via the weak force, it has therefore always been observed to be left-handed. This perplexing violation of a seemingly trivial symmetry has only been observed in the weak nuclear force, with the other three fundamental forces of physics so far conserving the symmetry.

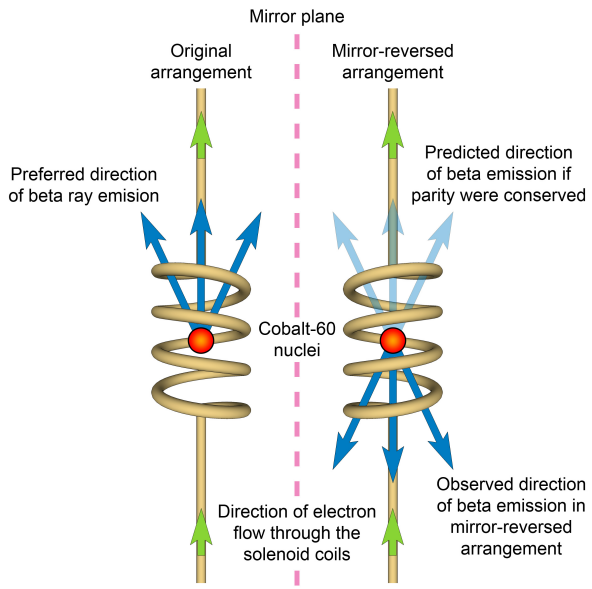

Figure 1.6: A schematic of the Wu experiment, in which cobalt-60 decay was observed under two opposite states of polarisation. In reality, the experiment was 'mirrored' by reversing the current through the solenoid that generated the polarising magnetic field. The outcome of the experiment was indicative of leptonic CP violation. Taken from [10].

Like P symmetry, charge (C) symmetry seems trivial but is likewise violated by the weak force. C symmetry is the conservation of physics under a charge conjugation, switching between particles and antiparticles. It holds under influence of the electro-

magnetic, strong nuclear and gravitational forces, but the weak nuclear force is once again the exception. Whereas weakly-interacting fermions are always observed to be left-handed, antifermions are always observed to be right-handed. A charge conjugation of a left-handed neutrino would produce a left-handed antineutrino, something never observed in nature.

The violation of the P and C symmetries suggest that a combined CP symmetry might be conserved. A charge conjugation transforms a left-handed fermion to left-handed antifermion, after which a parity flip transforms it into being right-handed. However, this symmetry was also found to be violated, albeit to a much lesser degree than its constituent symmetries. Kaon decay processes, specifically $K_L \rightarrow \pi^+\pi^-$, violate the CP symmetry, displaying a small but clear bias between matter and antimatter [11]. The consequences of a fundamental asymmetry between matter and antimatter are large: one of the biggest open questions in physics is the observed imbalance between the abundance of matter and antimatter in the universe. A CP-violating effect that occurred just after the Big Bang could form an explanation for this. Practically, leptonic CP violation can be measured directly by examining neutrino and antineutrino oscillation in the same experiment, explained in further detail in section 2.3. In beam experiments, this can be done by measuring the outcome of both a neutrino and an antineutrino beam.

The Jarlskog invariant, first introduced in equation 1.6, provides one method to quantify to which degree the CP symmetry is violated. In the three-neutrino case, it can be described in terms of the mixing angles and CP-violating phase in the following manner:

$$J = \cos\theta_{12}\sin\theta_{12}\cos\theta_{23}\sin\theta_{23}\cos^2\theta_{13}\sin\theta_{13}\sin\delta_{CP} = J^{\max}\sin\delta_{CP}, \tag{1.13}$$

where the parametrisation described in section 1.2.2 was used. Given the uncertainty connected to δ_{CP}, it is useful to consider the maximum possible value of the invariant outside of this parameter, here named $J^{\max}$. Combining the relevant mixing angles returns $J^{\max} = 0.03359 \pm 0.0019$ at a 3σ significance. Entering the best-fit value of δ_{CP} then results in $J = -0.019$ [8]. By contrast, the Jarlskog invariant in the quark sector has a much smaller value of $J_{CKM} = (3.18 \pm 0.15) \cdot 10^{-5}$ [8].

1.4 Neutrino oscillation in matter – the MSW effect

Neutrinos are famously unaffected by other forms of matter. Being electrically neutral colourless leptons, they only interact through the weak nuclear force which, as the name suggests, does not provide a strong coupling between interacting particles. Matter can, however, affect the oscillation of neutrinos. Thanks to the abundance of electrons in ordinary matter, the ν_e component of any neutrino is affected disproportionately compared to its other flavour components. This can be seen in figure 1.2: whereas all (anti)neutrinos interact weakly with matter through the Z-boson, only the electron-(anti)neutrino undergoes charged current interactions with electrons in ordinary matter by exchanging a $W^{\pm}$-boson.

1.4.1 Matter potential

When the time evolution of neutrino mass eigenstates was first described in equation 1.3, an implicit simplification was made. The real solution to the Schrödinger equation involves a matrix exponential $|\nu_i(x)\rangle = e^{-iH_0}\,|\nu_i(0)\rangle$ that can easily be written in terms of its components E_i because the vacuum Hamiltonian H_0 is diagonal in the mass basis. The exponential of a diagonal matrix is simply that same matrix with its components exponentiated:

$$H_0 = \mathrm{diag}(E_0, \ldots, E_n) \quad \Rightarrow \quad e^{-iH_0} = \mathrm{diag}(e^{-iE_0}, \ldots, e^{-iE_n}). \tag{1.14}$$

When considering matter effects, an extra potential V is added to the Hamiltonian. Since only electron-(anti)neutrinos participate in charged-current interactions with ordinary matter, this potential is quite simple in the flavour basis: $V = \mathrm{diag}(\pm\sqrt{2}G_F n_e, 0, 0)$, where G_F is the Fermi coupling constant and n_e is the number density of electrons in the path of the neutrino. The sign of the matter potential is positive for neutrinos and negative for antineutrinos. The mixing matrix U can be used to transform this potential to the mass basis and add it to the vacuum potential H_0:

$$H = H_0 + U^{\dagger} V U. \tag{1.15}$$

H_0 and V are both diagonal in the mass and flavour basis respectively, but $U^\dagger VU$ and therefore H are not. H_0 and $U^\dagger VU$ also do not commute, so that $e^{-i(H_0+U^\dagger VU)} \neq e^{-iH_0}e^{-iU^\dagger VU}$. This makes it impossible to expand neutrino oscillation probabilities as described in appendix A. An approximate simplified form can be achieved, however, by making use of the Lie product formula and the fact that the matter potential is diagonal in the flavour basis:

$$\begin{aligned} \text{Lie product formula:} \quad & e^{A+B} = \lim_{n\to\infty}\left(e^{A/n}e^{B/n}\right)^n \\ V \text{ diagonal:} \quad & e^{-iU^\dagger VU} = U^\dagger e^{-iV}U \\ & e^{-iH} = e^{-i(H_0+U^\dagger VU)} = \lim_{n\to\infty}\left(e^{-iH_0/n}\cdot U^\dagger e^{-iV/n}U\right)^n, \end{aligned} \tag{1.16}$$

where in practice n should be chosen such that $n \gg \Delta m^2_{ij}x/E$ and $n \gg \sqrt{2}G_F n_e x$. Equation 1.16 is useful for speeding up computation because the diagonal matrices can once again be exponentiated componentwise according to equation 1.14. It does not, however, simplify the final oscillation probability formulae without further approximation.

Although the described approximation is useful in the exploration of neutrino oscillation, the figures presented in this chapter were generated using the Eigen library [12], which uses the Padé approximant [13] to compute matrix exponentials. Arbitrary exponentials of complex non-diagonal matrices can be computed using this method, such as the Hamiltonian of a neutrino travelling through matter. Unless specified otherwise, the oscillation parameters from table 1.1 were used, as well as a travel length of 1285 km and an average earth density of 2848 kg/m^3 along the neutrino path. These values match the experimental parameters of DUNE, detailed in chapter 2. In reality, its beamline reaches a depth of over 30 km and encounters a variable earth density, but using the average density instead is a good approximation, leading to errors in oscillation probability < 0.0006 [14].

In order to gain an understanding of three-neutrino oscillation, a small program called EigenNeut was developed to visualise generalised neutrino oscillation as a function of any parameter [15]. This includes mixing angles and neutrino mass differences, as well as experimentally dictated parameters such as the neutrino energy, travel length and matter density across the neutrino path. Its main method of visualisation is the ternary plot, an example of which is shown in figure 1.7. This type of plot shows the ratio between three components that combine to 100%. In this case, an initial muon

neutrino obtains other flavour components as it travels through time. Although shown here as a still image, EigenNeut produces interactive and animated ternary plots. It makes use of the fast Lie product formula approximation described in 1.16 for user interaction and animation, whereas the more accurate Padé approximation is used when exporting the oscillation data for the figures shown throughout this chapter.

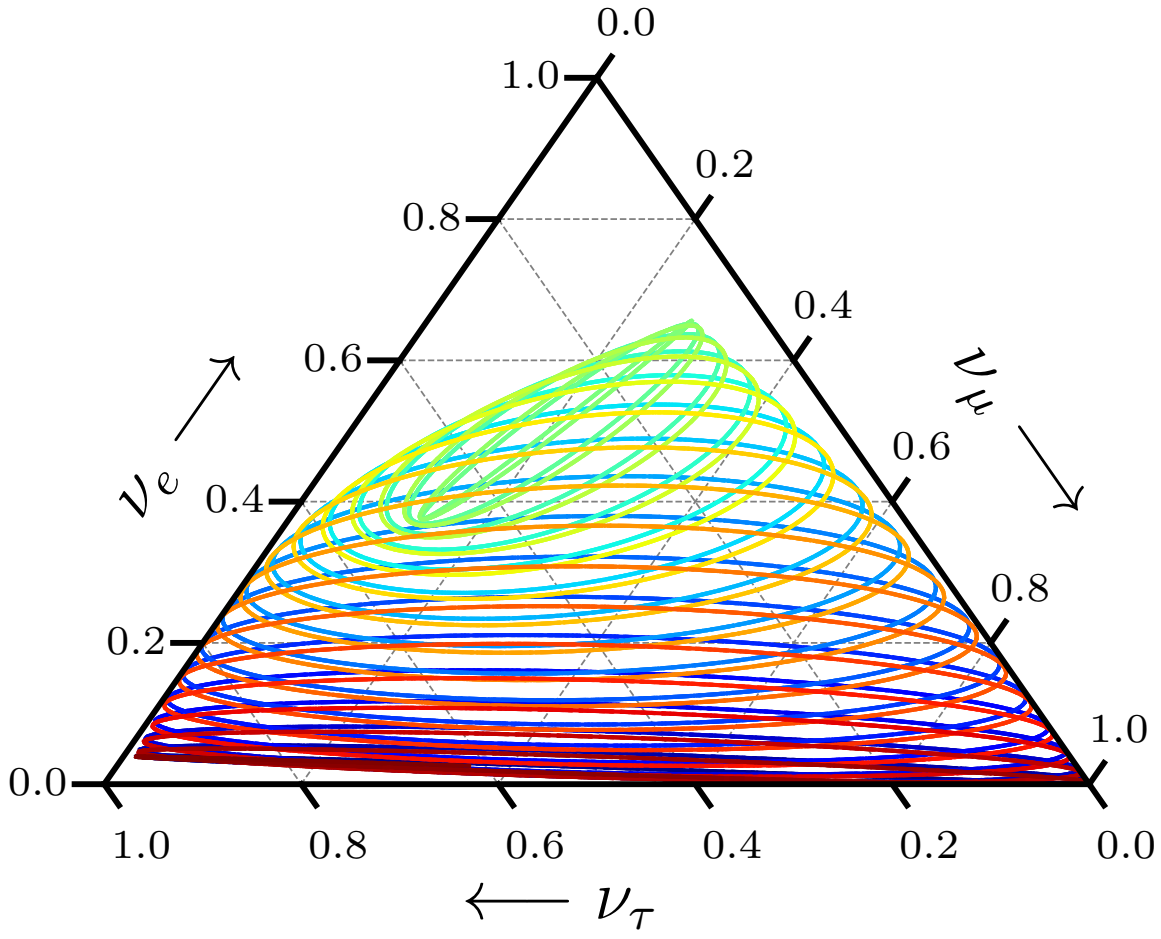

Figure 1.7: A ternary plot that shows the path of an initial 1 GeV muon neutrino through flavour space. The parameters shown in table 1.1 were used, with the exception of $\delta_{CP} = 0$. A path length of 33000 km through vacuum was chosen so that the flavour path nearly returns to the pure ν_μ state in the bottom right corner of the triangle. The colour scale indicates distance along the path length, starting with blue and shifting via green to red.

1

1.4.2 Two-neutrino oscillation in matter

Matter effects can be explored in relative simplicity in the two-neutrino oscillation case, as described in section 1.2.1. The two neutrinos in this case are the electron neutrino ν_e, which interacts with electrons in matter, and a combination of the other neutrinos forming the state ν_a. It is convenient to consider the Hamiltonian in the flavour basis, denoted by H_f. The two-dimensional mixing matrix U, as in the vacuum case, contains a single angle θ.

$$\begin{aligned} H_f &= UH_0U^\dagger + V \\ &= \frac{1}{2E}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} + \begin{pmatrix} \sqrt{2}G_F n_e & 0 \\ 0 & 0 \end{pmatrix} \\ &= \frac{1}{4E}\left[\begin{pmatrix} -\Delta m^2\cos 2\theta & \Delta m^2 \sin 2\theta \\ \Delta m^2 \sin 2\theta & \Delta m^2 \cos 2\theta \end{pmatrix} + (2m_1^2 + \Delta m^2)I\right] + \frac{G_F n_e}{\sqrt{2}}\left[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + I\right] \end{aligned} \tag{1.17}$$

As before, terms proportional to the identity I affect all eigenstates equally and are therefore invisible in oscillation measurements. Dropping these out of the equation yields the following:

$$H_f = \frac{1}{4E}\begin{pmatrix} A - \Delta m^2\cos 2\theta & \Delta m^2 \sin 2\theta \\ \Delta m^2 \sin 2\theta & -(A - \Delta m^2 \cos 2\theta) \end{pmatrix}, \tag{1.18}$$

where $A = 2\sqrt{2}G_F n_e E$. The eigenvalues $\pm\Delta\tilde{m}^2/4E$ and corresponding diagonalising rotation of the Hamiltonian can be described as follows:

$$\begin{aligned} H_f &= \begin{pmatrix} \cos\tilde{\theta} & \sin\tilde{\theta} \\ -\sin\tilde{\theta} & \cos\tilde{\theta} \end{pmatrix}\begin{pmatrix} -\Delta\tilde{m}^2/4E & 0 \\ 0 & \Delta\tilde{m}^2/4E \end{pmatrix}\begin{pmatrix} \cos\tilde{\theta} & -\sin\tilde{\theta} \\ \sin\tilde{\theta} & \cos\tilde{\theta} \end{pmatrix} = \tilde{U}\tilde{H}\tilde{U}^\dagger, \\ \text{where} \quad \Delta\tilde{m}^2 &= \Delta m^2 B, \quad \sin 2\tilde{\theta} = \frac{\sin 2\theta}{B}, \quad \cos 2\tilde{\theta} = \frac{\cos 2\theta - A/\Delta m^2}{B} \\ \text{and} \quad B &= \sqrt{(\cos 2\theta - A/\Delta m^2)^2 + \sin^2 2\theta}. \end{aligned} \tag{1.19}$$

The newly formulated Hamiltonian and mixing matrix in matter $\tilde{H}$ and $\tilde{U}$ return to the familiar vacuum form described in section 1.2.1, with the Hamiltonian a diagonal

matrix in the mass basis and the mixing matrix a rotation matrix with a single angle. Using equation 1.12, the two-neutrino oscillation probability in matter can therefore be described as

$$P(\nu_\alpha \to \nu_\beta)|_{\alpha\neq\beta} = \sin^2 2\tilde{\theta} \sin^2 \left(1.267 \frac{\Delta\tilde{m}^2_{12} x}{E} \right). \tag{1.20}$$

The relations between vacuum and matter variables in equation 1.19 are highly nonlinear, but can be captured in the multiplicative variable B. For $A = 0$, $B = 1$ and the vacuum oscillation probability is recovered. Figure 1.8 shows the behaviour of B as a function of the sole mixing angle θ and relative matter potential $A/\Delta m^2$. As can be deduced from equation 1.20, $B > 1$ implies weaker mixing ($\sin^2 2\theta$ is suppressed) and faster oscillation (Δm^2 is enhanced). $B < 1$ causes the opposite effect.

An example of matter effects in two-neutrino oscillation is shown in figure 1.9. It can be seen that matter effects increase the effective mixing angle, evidenced by the larger amplitude of the oscillations in the bottom plot. On the other hand, the period of the oscillations is longer, suggesting a smaller squared mass difference.

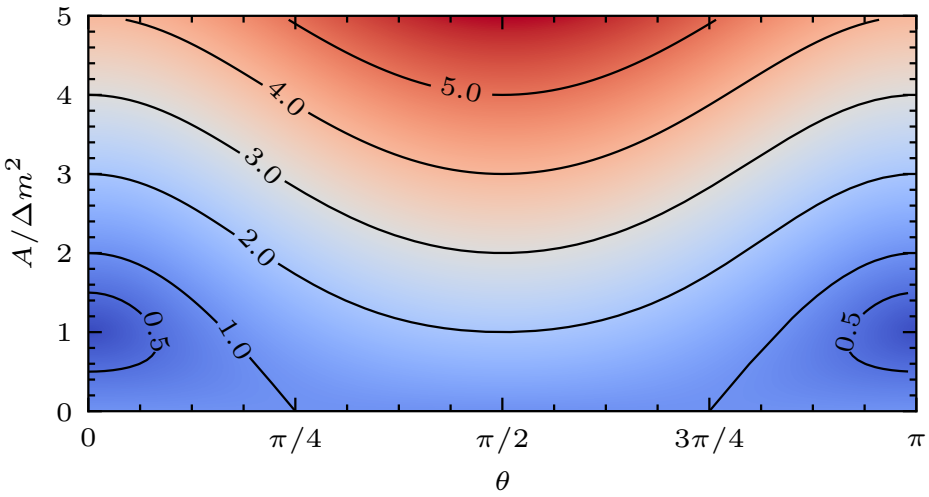

Figure 1.8: B as a function of θ and $A/\Delta m^2$ as it appears in equation 1.19.

1.4.3 Three-neutrino oscillation in matter

Matter effects complicate three-neutrino oscillation probabilities to the point where exact analytical solutions are impossible. Analytical approximations exist that are

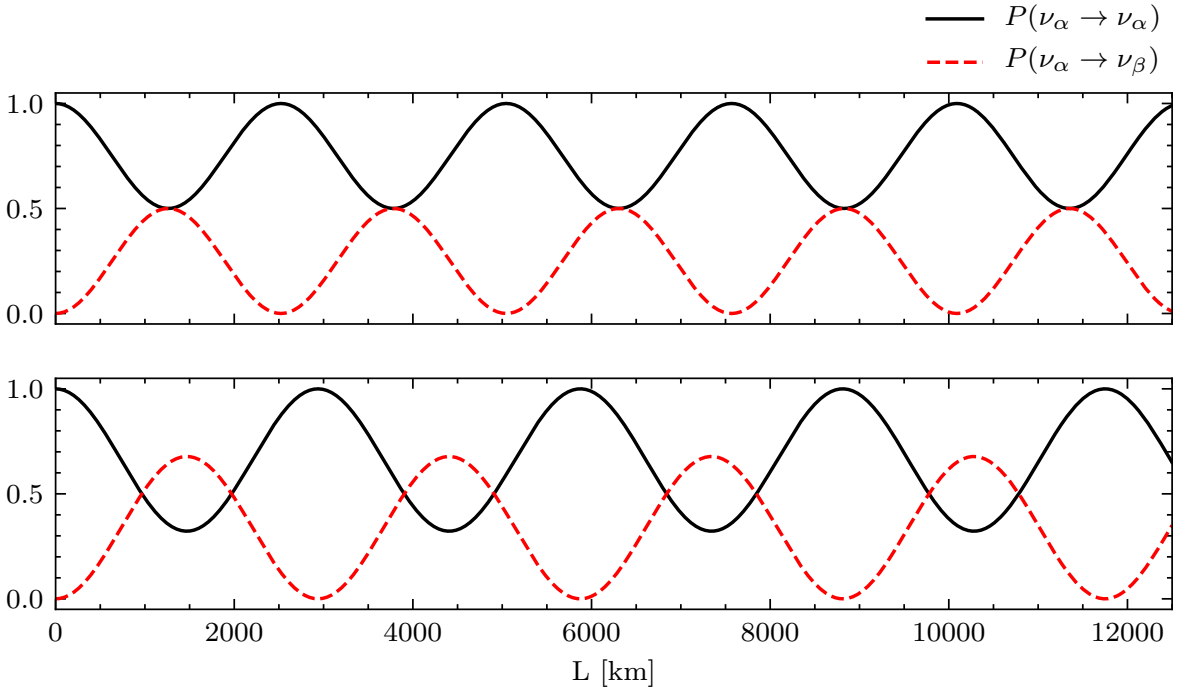

Figure 1.9: Two-neutrino oscillation for an initial ν_α of 2.5 GeV as a function of distance with the vacuum mixing angle $\theta = \pi/8$ and squared mass difference $\Delta m^2_{12} = 2.5 \cdot 10^{-3}\,\mathrm{eV}^2$, visualised in vacuum (top) and matter (bottom, matter density $\sim 2800\,\mathrm{kg/m^3}$). Δm^2_{12} was chosen in accordance with table 1.1 and the matter density is representative of the earth's crust.

applicable in a variety of cases, but are too lengthy to expand on in this chapter [16][17][18]. One approximate oscillation probability that demonstrates the complexity of such solutions can be formulated as follows [19]:

$$\begin{aligned} P(\nu_\mu \to \nu_e) \approx & \sin^2\theta_{23} \sin^2 2\theta_{13} \frac{\sin^2(\Delta_{31} - aL)}{(\Delta_{31} - aL)^2} \Delta^2_{31} \\ & + \sin 2\theta_{23} \sin 2\theta_{13} \sin 2\theta_{12} \frac{\sin(\Delta_{31} - aL)}{\Delta_{31} - aL)} \Delta_{31} \frac{\sin(aL)}{aL} \Delta_{21} \cos(\Delta_{31} + \delta_{CP}) \\ & + \cos^2\theta_{23} \sin^2 2\theta_{12} \frac{sin^2(aL)}{(aL)^2} \Delta^2_{21}, \end{aligned} \tag{1.21}$$

where $\Delta_{ij} = \Delta m^2_{ij} L/4E$ and $a = A/4E = G_F n_e/\sqrt{2}$.

As can be seen in figure 1.10, the $P(\nu_\mu \to \nu_e)$ oscillation probability is heavily suppressed over very long ranges. The vacuum Hamiltonian is inversely proportional to the neutrino energy ($H_0 \sim 1/E$) whereas the matter potential is independent of the

energy scale. For this reason, matter effects are particularly pronounced in high energy neutrinos. Even on length scales on the order of 1000 km, neutrinos around 1 GeV show a deviation due to the MSW effect.

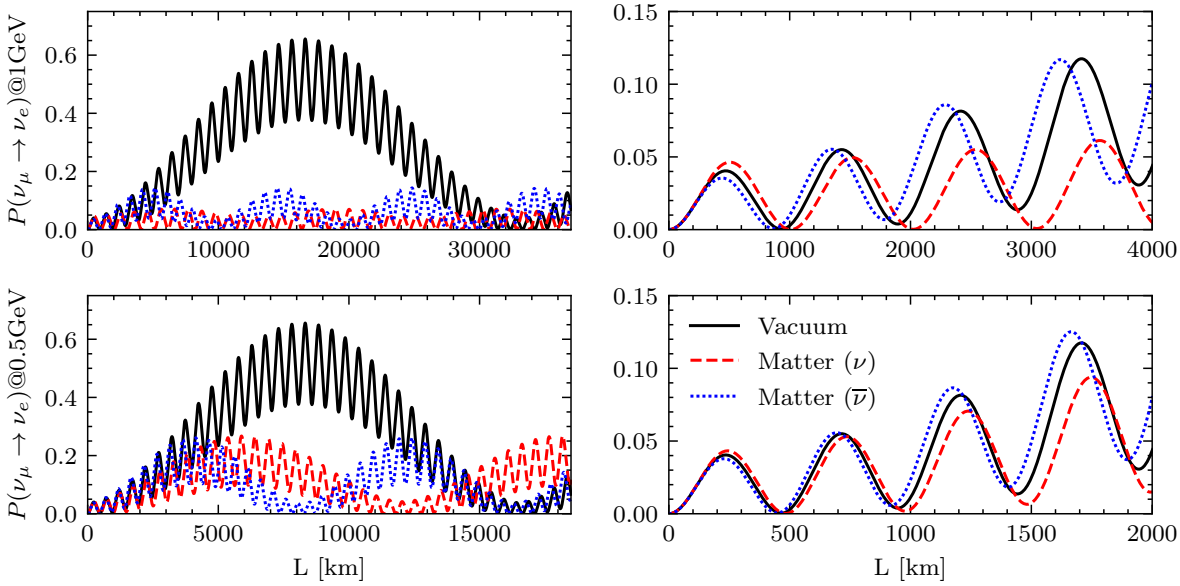

Figure 1.10: The oscillation probability $P(\nu_\mu \to \nu_e)$ for a 1 GeV (top) and 0.5 GeV (bottom) neutrino as a function of distance travelled. Assuming normal ordering and $\delta_{CP} = 0$, vacuum and matter oscillations are compared for very long-range (left) and short-range (right) travel distances. Also shown is the antineutrino oscillation probability $P(\bar{\nu}_\mu \to \bar{\nu}_e)$ in matter. Note that the horizontal range for the 0.5 GeV neutrino has been set to be half that of the 1 GeV neutrino in order to make the vacuum oscillations correspond.

Figure 1.10 shows the matter effect in neutrinos as well as antineutrinos. It can be seen that the oscillation probability in matter is suppressed to a similar degree for both neutrinos and antineutrinos when compared to the vacuum oscillation probability. The only difference between the two in terms of oscillation parameters is that both V and δ_{CP} are of opposite sign. The former flips sign due to the differing interactions that electron neutrinos and antineutrinos have with electrons in matter, as described in section 1.1.1. δ_{CP} is defined by its sign flip between neutrino and antineutrino oscillation, being the parameter that describes the asymmetry between the two. Note that $\delta_{CP} = 0$ in figure 1.10 in order to visualise the matter effect in isolation.

Beyond the simple parametric differences between neutrino and antineutrino oscillation in matter, a detailed description is outside the scope of this chapter. It is important to note, however, that neutrino experiments usually do not have the luxury

of a neutrino source that travels through vacuum as well as an equivalent source that travels through matter. Instead of comparing vacuum oscillations to those in matter, neutrino oscillation is often compared to antineutrino oscillation. It can be seen in figure 1.10 that the 1 GeV neutrino asymmetry becomes notable around 2500 km. Neutrinos of an even higher energy can be expected to show this asymmetry at smaller distances still.

1.5 Experimental determination of oscillation parameters

In order to measure the parameters that govern neutrino oscillation, a wide variety of experiments has been operated. Whereas these experiments measure the same oscillation mechanism, they differ in their source of neutrinos, experimental layout and method of detection. This section provides a short summary of the most notable neutrino experiment types and their contributions to the oscillation parameter space. The majority of the information shown in this section was taken from The Review of Particle Physics by the Particle Data Group [8] and global fits from the NuFIT group [7], both of which provide excellent summaries of the achievements in neutrino physics.

Special mention must be made of the first direct detection of the neutrino. This event occurred on the 13th of November 1970 in the 12-foot bubble chamber of the Zero Gradient Synchrotron at the Argonne National Laboratory. Shown in figure 1.11, the interaction products of the invisible neutrino can be seen seemingly appearing out of nowhere. Bubble chambers visualise particles with a superheated liquid. As high-energy charged particles pass through the detector material, liquid evaporates around the ionisation tracks left in the particles' wake. Cloud chambers work on a similar principle, using a supersaturated vapour in gas instead. Although the spatial resolution of these experiments is excellent, the event recording is necessarily two-dimensional, losing valuable information concerning the kinetics of the interaction system. Additionally, bubble and cloud chambers are ill-suited for observing large-scale high-frequency physics events due to their small size and long reset time.

Solar neutrino experiments rely on neutrinos that are created within the sun's core by thermonuclear reactions. As the sun fuses protons into helium and heavier elements, it creates enormous numbers of neutrinos: the total solar neutrino production

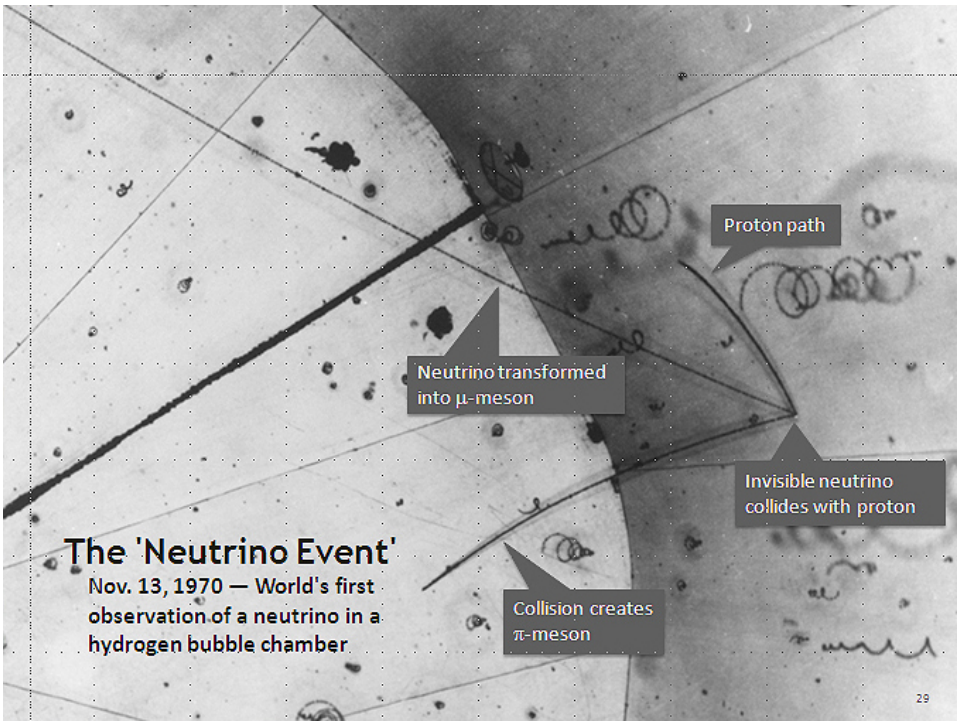

Figure 1.11: The first direct detection of a neutrino interaction, recorded on the 13th of November 1970 at the Argonne National Laboratory.

is estimated to be $2 \cdot 10^{36}\,\mathrm{s}^{-1}$, making for a flux of $7 \cdot 10^{10}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ on earth. Equally important is the energy of these neutrinos, which at $\mathcal{O}(0.1) - \mathcal{O}(10)$ MeV is relatively low compared to other experiment types [20]. For this reason, any detectors employed in the search for these particles must have a sufficiently low energy threshold. The Homestake experiment [3] is notable for first measuring the solar neutrino flux, detecting a deficiency that is now known to be caused by neutrino oscillation. Other notable experiments include (Super-)Kamiokande [21], SNO [4], KamLAND [22] and Borexino [23].

Another natural source of neutrinos is the collision of cosmic rays with the earth's atmosphere. Pions, kaons and muons created in high-energy cosmic ray showers decay to predominantly release electron and muon (anti)neutrinos. Because cosmic rays impinge on the earth isotropically, experiments may expect atmospheric neutrinos from all directions, giving the particles a path length ranging from 10 to 10^4 km. Combined with the variable density of the earth, the wide range of neutrino energies and path lengths make atmospheric neutrino detectors quite versatile. Since the neutrino flux is low at higher energies, atmospheric neutrino detectors need to be large. Notable examples include (Super-)Kamiokande, IceCube [24], ANTARES [25]

and KM3NeT [26]. Because of their great size, these detectors are also able to observe high-energy cosmogenic neutrinos.

Reactor neutrino experiments make use of the large antineutrino flux with an energy of $\mathcal{O}(1)$ MeV generated by nuclear reactors. Although the decay chains of the most common fission fuels are complex, it is estimated that a reactor produces approximately $2 \cdot 10^{20}\,\bar{\nu}_e/\mathrm{s}$ per GW of thermal power. The experiments simultaneously observe multiple reactors in the vicinity to total $\mathcal{O}(10)$ GW at various distances. Notable experiments include KamLAND, Double Chooz [27], Daya Bay [28], RENO [29] and JUNO [30].

In accelerator experiments, the particle source is a beam of neutrinos that is targeted at a detector. A distinction can be made between short baseline (SBL, $\mathcal{O}(100)$ m), medium baseline (MBL, $\mathcal{O}(1)$ km) and long baseline (LBL, $\mathcal{O}(100) - \mathcal{O}(1000)$ km) accelerator experiments, referring to the distance between the beam source and detector. Oftentimes, LBL experiments employ a near detector as well as a far detector so that both the initial and final state of the neutrino beam may be measured. Neutrino beams are conventionally generated by sending a proton beam into a fixed target, creating pions and kaons that are focused and decay primarily into muon neutrinos. Currently operational beams most commonly have a neutrino energy of $\mathcal{O}(1)$ GeV. Notable examples of currently operational accelerator experiments are T2K [31], MINOS [32], NOνA [33] and MicroBooNE [34]. Upcoming experiments most notably include Hyper-Kamiokande [35] and DUNE, the latter of which is discussed in chapter 2.

The most accurate neutrino oscillation parameter fits are performed using the observations made by all relevant experiments at once, resulting in a so-called global fit [7]. However, some types of experiment are more relevant to specific parameters than others. Section 1.2 demonstrates the link between the experimentally dictated value of L/E and the fundamental Δm^2 oscillation parameters in the vacuum limit. As a result, the various types of experiments are sensitive to differing portions of the squared mass difference scale. Taking the types of neutrinos being investigated into account, each experiment type is also sensitive to various other oscillation parameters, such as the mixing angles and CP-violating phase. As can be seen in table 1.2, the various types of neutrino experiments are largely complementary.

Table 1.2: A summary of the experimental contributions to the current understanding of the neutrino oscillation parameter space. Shown are the dominant and other important parameters that govern neutrino oscillation at the various types of experiments. Super-Kamiokande has been shortened to 'SK' and Double-Chooz to 'D-Chooz.' Taken from [8].

Experiment type	Notable experiments	Dominant	Important
Solar	SK, KamLAND, SNO, Borexino	θ_{12}	Δm^2_{21}, θ_{13}
Reactor LBL	KamLAND	Δm^2_{21}	θ_{12}, θ_{13}
Reactor MBL	Daya-Bay, RENO, D-Chooz	θ_{13}, $\lvert\Delta m^2_{31}\rvert$	
Atmospheric	SK, IceCube, ANTARES		θ_{23}, $\lvert\Delta m^2_{31}\rvert$, θ_{13}, δ_{CP}
Accel. LBL (ν_μ, $\bar{\nu}_\mu$ disapp.)	T2K, MINOS, NOνA	$\lvert\Delta m^2_{31}\rvert$, θ_{23}	
Accel. LBL (ν_e, $\bar{\nu}_e$ app.)	T2K, MINOS, NOνA	δ_{CP}	θ_{13}, θ_{23}

1

Chapter 2

The Deep Underground Neutrino Experiment

The Deep Underground Neutrino Experiment (DUNE) is an upcoming neutrino experiment that is a part of the neutrino physics community's strategy to venture further into precision measurement of neutrino properties. The experiment is to measure leptonic matter-antimatter asymmetry and to examine the underlying structures of neutrino physics by refining the current measurements of neutrino mass ordering and mixing angles. Additional goals are the detection and measurement of supernova neutrinos and the testing of models beyond the Standard Model by measuring the proton lifetime [36]. The experiment will be built at Fermilab and the Sanford Underground Research Facility, two research laboratories that are $\sim 1300\,\text{km}$ apart. The neutrino source and supporting facilities will be built under the name of the Long-Baseline Neutrino Facility (LBNF).

This chapter discusses the general design of DUNE and LBNF, as well as their research goals and their capability to meet them. Technical details regarding the detector technology used in the experiment is expanded upon in chapter 3.

2.1 Experiment overview

In its simplest form, the combined setup of LBNF and DUNE will consist of a neutrino beam, a near detector and a far detector. The beam and near detector will be constructed at Fermilab in Batavia, Illinois, with the near detector being placed roughly 500 m from the beam source. A modular far detector will be built at SURF in Lead,

South Dakota, almost 1300 km from Fermilab. To generate the required neutrinos, an upgradeable 1.2 MW proton beam will be constructed at Fermilab and put into operation by 2026 [37]. Two 10 kt liquid argon time projection chamber (LArTPC) modules are projected to be put into operation as part of the far detector by 2027, with two additional modules to be installed by 2029 [38][39]. These will be built deep underground at SURF at a distance of 1300 km from the neutrino beam source. DUNE is foreseen to accumulate at least 120 kt·MW·year of exposure by 2035 in order to achieve its physics goals [40].

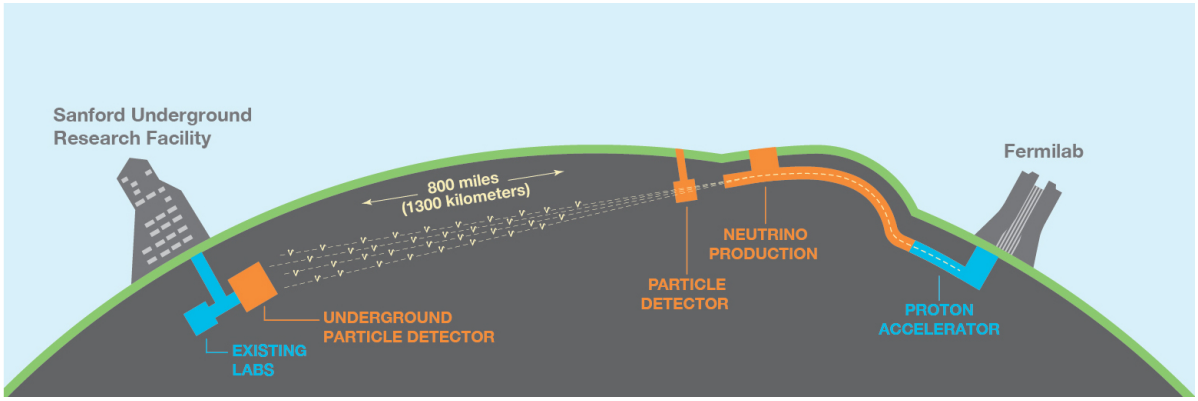

Figure 2.1: The basic setup of LBNF and DUNE. Protons are accelerated at Fermilab in Batavia, IL, and used to produce a neutrino beam that passes through a near detector as well as a far detector in Lead, SD, 1300 km from the neutrino source. Taken from [41].

2.1.1 Beam

As neutrinos are electrically neutral, they cannot be accelerated or directed by themselves. Neutrinos must therefore be obtained as secondary particles from a particle beam that can be manipulated. At Fermilab's accelerator complex, protons are accelerated to energies of 60 – 120 GeV before being directed into a graphite target. There, in highly energetic collisions, they generate electrically charged pions and kaons. After carefully focusing these secondary particles using a series of magnetic horns, they are left to fly through a long helium-filled decay pipe where they decay into neutrinos and charged by-products [36]. While the neutrinos fly freely through the beam absorber that caps the far end of the decay pipe, any other secondary particles and left-over protons are captured to prevent the radioactive activation of earth beyond the beam installation. The entire assembly, from the target to the end of the decay pipe, is angled

into the earth at 101 mrad ($\sim 5.79^{\circ}$) in order to accommodate the curvature of the earth and hit the far detector in Lead, SD [42].

Controlling the beam characteristics is vital for DUNE's research goals. Depending on the direction of the current running through the magnetic horns, they focus either positive or negative secondary particles which primarily decay to neutrinos or antineutrinos respectively. By controlling the direction of the current, it is thus possible to create either a neutrino or antineutrino beam [36].

Fermilab has a long history of particle experiments. It hosts its own accelerator complex that held the Tevatron between 1983 and 2011, famous in part for its discovery of the top quark. The complex's protons are extracted from a source and gain their first major energy boost from a linear accelerator (linac). This raises their energy up to 400 MeV and injects them into the booster ring of almost half a kilometre in circumference that gives them energies up to 8 GeV. It is from this booster that Fermilab extracts protons for use in the generation of low-energy neutrinos. The protons that continue through the complex are sent to the recycler, a ring more than three kilometres in circumference, where multiple bunches are collected and stacked on top of each other for an even denser beam. The last acceleration step comes in the form of the main injector, a synchrotron in the same tunnel as the recycler that boosts the protons from their previous 8 GeV to 120 GeV. Protons from this ring generate high-energy high-intensity neutrinos for NOvA, LBNF and DUNE. A schematic of the full Fermilab accelerator complex can be seen in figure 2.2.

In order to generate the required proton beams, Fermilab's accelerator complex is undergoing upgrades under the Proton Improvement Plan (PIP). Divided into multiple stages, the initial PIP-I programme aimed at increasing the proton output power to a consistent 700 kW at 120 GeV through small upgrades and the reduction of beam losses. This goal was achieved in early 2017, when the accelerator complex reached an average power output of 716 kW for over an hour. Space charge effects in the beam cause it to expand beyond acceptable bounds at higher numbers of protons per pulse. Denser beams therefore need to be accelerated faster. The PIP-II programme aims to replace the currently existing linac with a superconducting linac that is able to accelerate protons up to 800 MeV instead of 400. Higher initial energies allow for more protons to be bunched together in the final beamline, allowing for a power increase to 1.2 MW at 120 GeV by 2026. Further upgrades planned by 2030 will increase the power output further to 2.4 MW under the tentative name of PIP-III [36]. An overview of the beam design parameters and potential upgrades is shown in table 2.1.

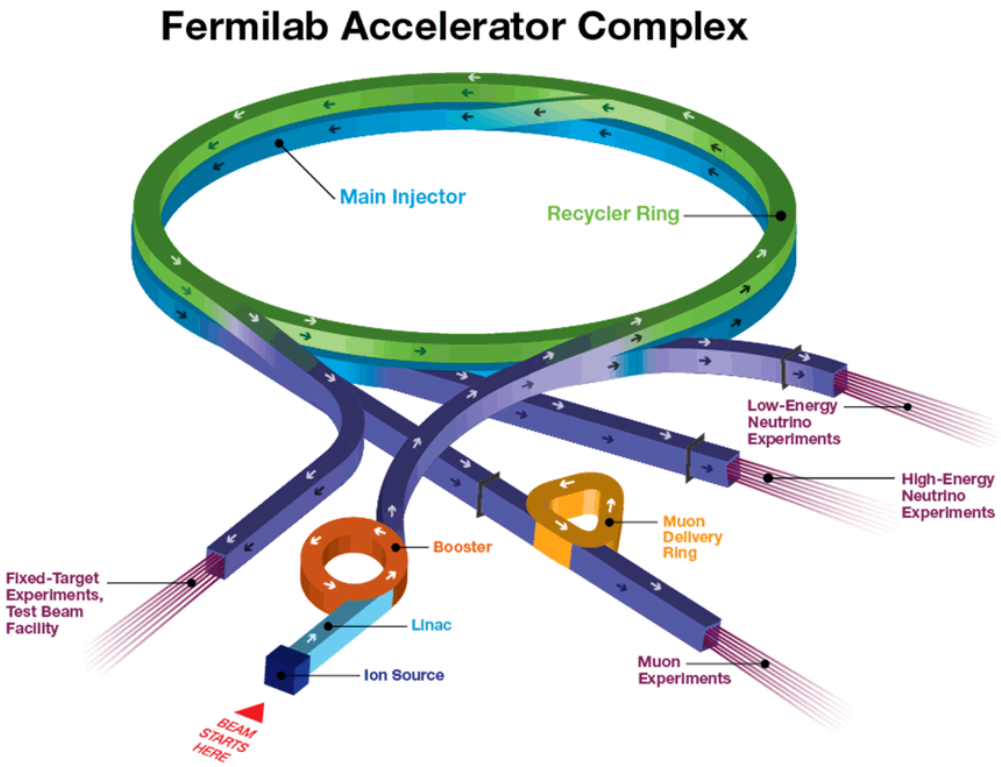

Figure 2.2: Fermilab's accelerator complex, excluding the decommissioned Tevatron synchrotron. The proton beam path is indicated with arrows. They traverse the linac, Booster, Recycler Ring and Main Injector, and are split off to various experiments along the way. Taken from [43].

The neutrino flux distributions at the near and far detectors are determined by the proton beam and the focusing of the magnetic horns. Depending on its energy, every proton on target (POT) produces a certain number of charged mesons, some of which are focused successfully into the decay pipe to decay into neutrinos. Taking the resulting neutrino production as input, simulations were performed using the GLoBES framework [45] to propagate these neutrinos along their paths and determine the average neutrino flux per POT both at the near and far detector sites. For every POT, the near detector can expect an average flux of $1.14 \cdot 10^{-3}\ \nu/\mathrm{m}^2$ whereas the far detector receives $1.65 \cdot 10^{-10}\ \nu/\mathrm{m}^2$ [46]. Given that a single extraction cycle ($\sim 1\,\mathrm{s}$) of the Main Injector deposits $7.5 \cdot 10^{13}$ protons onto the target and that the beam

Table 2.1: A summary of the LBNF proton beam design parameters, showing two distinct modes of operation in its initial configuration, with potential upgrades increasing the proton energy and beam power. Taken from [44].

Parameter	Value		Potential upgrade
Energy	60 GeV	120 GeV	150 GeV
Protons per cycle	$7.5 \cdot 10^{13}$	$7.5 \cdot 10^{13}$	
Spill duration	$1.0 \cdot 10^{-5}$ s	$1.0 \cdot 10^{-5}$ s	
Cycle time	0.7 s	1.2 s	
POT per year	$1.9 \cdot 10^{21}$	$1.1 \cdot 10^{21}$	
Beam power	1.03 MW	1.2 MW	2.4 MW

uptime and efficiency is expected to be 56% [38], the average neutrino flux is $4.8 \cdot 10^{10}$ and $6.9 \cdot 10^{3}$ $\nu/\mathrm{m}^2/\mathrm{s}$ at the near and far detector respectively. In neutrino mode, the beam consists primarily of muon neutrinos, but contains contamination from electron neutrinos ($\sim 0.5\,\%$) and various antineutrinos as well ($\sim 5\,\%$), as can be seen in figure 2.3. All components of the beam peak in the 0.5 – 4 GeV region.

The described beam contamination occurs mainly through the undesired decay of particles in the decay pipe. In neutrino mode, the magnetic horns select positively charged particles, such as π^+ and K^+, that decay into anti-muons and muon neutrinos according to $\pi^+ \to \mu^+ + \nu_\mu$ and $K^+ \to \mu^+ + \nu_\mu$. However, the kaon has many more decay modes, some of which include negatively charged pions that decay into $\bar{\nu}_\mu$ and some of which include $\bar{\nu}_e$. Additionally, some produced anti-muons may decay in the decay pipe according to $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$, further adding to the beam contamination. In antineutrino mode, beam contamination occurs along similar processes.

2.1.2 Near detector

The near detector of DUNE acts as an initial measurement of the Fermilab-generated neutrino beam. Because this beam is generated from particle decay, its composition and spread cannot be as tightly controlled as a beam of charged particles could be. It is therefore important that a reliable measurement of the initial state of the neutrinos is made. The near detector of DUNE aims to do exactly this. In order to create an accurate picture of the neutrino beam composition and spread, the near detector is

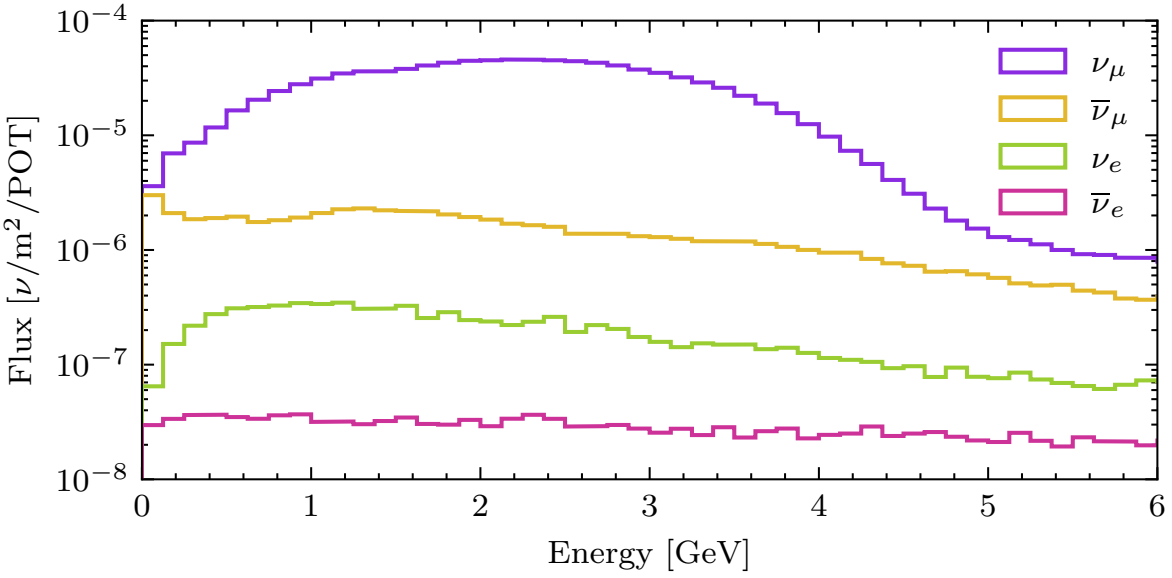

Figure 2.3: Expected near detector neutrino spectra with the neutrino beam in ν_μ-production mode. Identification of beam contamination is key in understanding oscillated neutrino spectra at the far detector. Data taken from [46].

situated sixty metres underground and over 300 metres from the end of the decay pipe. This, in addition to specialised muon shielding at the far end of the decay pipe, works to ensure that as few muons as possible reach the near detector hall, be they cosmic or originating from the beam. In total, the distance between the proton target and detector hall amounts to approximately 574 m [42]. The neutrino beam setup is visible in figure 2.4.

As implied by the name, the near detector is a relatively short distance away from its neutrino source. This allows it to fulfill its purpose while remaining compact, making its design more flexible than that of the far detector. The near detector design consists of three parts which are, in order along the beam trajectory, a 50t LArTPC named ArgonCube, a Multi-Purpose Detector (MPD) and a System for on-Axis Neutrino Detection (SAND), all of which are visible in figure 2.5. The first detector in line, ArgonCube, is meant to have an active material that is identical to that of the far detector so that an accurate comparison between the two can be made. The MPD uses a pressurised gaseous argon TPC which is surrounded by an electromagnetic calorimeter (ECAL). The entire system is enveloped in a magnetic field generated by superconducting Helmholtz coils. The gaseous argon TPC principle has seen extensive use at LEP and in the ALICE detector [47] and is therefore a very well-understood technology that can determine the efficiency of the near and far detectors'

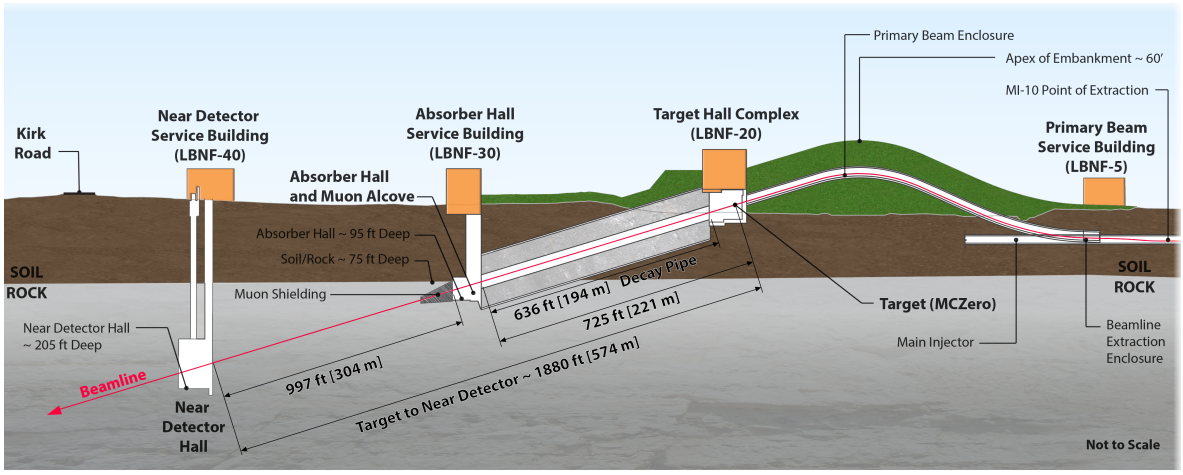

Figure 2.4: DUNE's near site facilities at Fermilab, including the proton beam target, decay pipe and near detector cavern. Taken from [42].

LArTPCs. In addition, since the MPD features a magnetic field and is located directly downstream from ArgonCube, it can identify the sign of charged particles produced by reactions in the liquid argon. SAND, lastly, consists of a three-dimensional scintillator system embedded in a gaseous TPC that is itself embedded in an electromagnetic calorimeter. All of these components are subsequently subjected to a magnetic field $> 0.6\,\mathrm{T}$. The primary function of SAND is to provide continuous monitoring of the neutrino beam using its substantial mass of $\sim 8.7\,\mathrm{t}$ [38]. The detection principle of TPC-based detectors is explained in detail in sections 2.1.3 and 3.2.2.

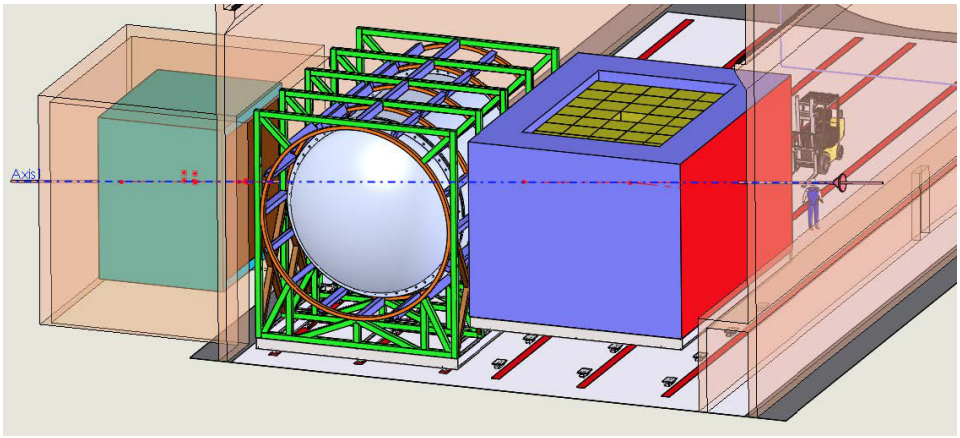

Figure 2.5: A side view of the near detector in its cavern. The beam enters from the right and passes through ArgonCube, the MPD and SAND, in that order. Taken from [36].

Determining the energy of a reconstructed neutrino accurately is vital for an experiment such as DUNE, which relies on a broad neutrino energy range to meet multiple physics goals at once. This energy determination would normally be based on interaction models of neutrinos with nuclei, but no complete set of models is available for such interactions. Another technique will therefore be used: in the DUNE-PRISM design, ArgonCube and the MPD of the near detector will be able to move off-axis while SAND stays on-axis to monitor the beam, as visible in figure 2.6. In what is called the off-axis effect, a neutrino takes less of its parent's energy at larger decay angles. It also concentrates the resulting neutrino energy spectrum into a smaller region, making for a more monoenergetic neutrino source at the expense of reduced intensity [48]. Both of these effects are shown in figure 2.7. The NOνA [33] and T2K [31] experiments make use of off-axis beams in order to obtain a narrow neutrino energy spectrum. Their beams deliberately "miss" their targets. Evaluating the neutrino beam at various off-axis angles ($0^\circ - 3.6^\circ$) by moving the set of detectors sideways by up to 36 metres thus gives various neutrino energy spectra. These spectra can be linearly combined to reproduce oscillated far detector measurements, thereby determining neutrino oscillation parameters in a model-independent manner. Because of the limitations in how far the near detector subdetectors can shift, the energy spectrum that can be fitted is bounded from below at approximately 0.6 GeV. Additionally, the fit may not extend far above the on-axis spectrum peak energy as it is not possible to shift the spectrum peak to higher energies [36][49].

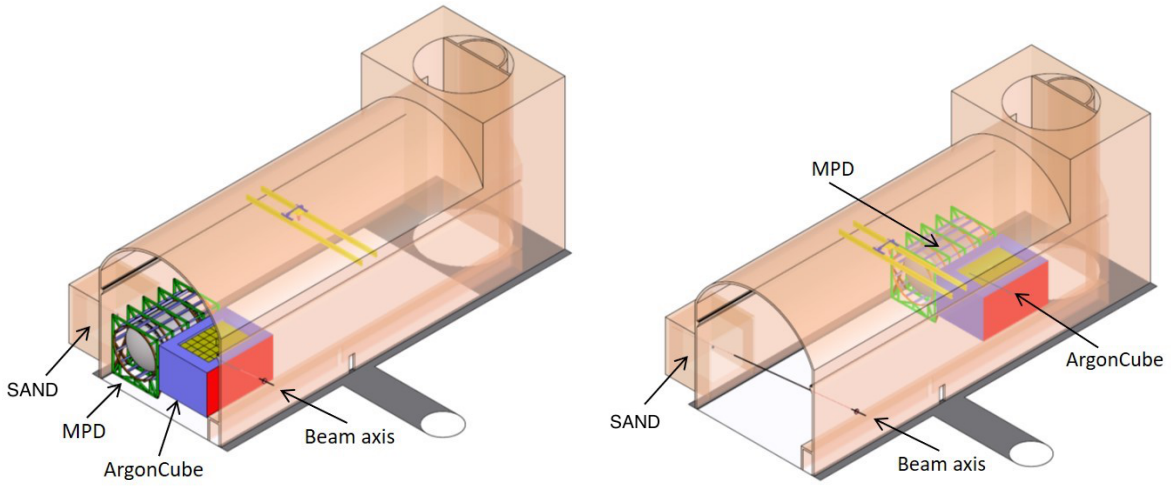

Figure 2.6: The ArgonCube and MPD of the near detector can move up to 36 m off-axis in their cavern. Taken from [36].

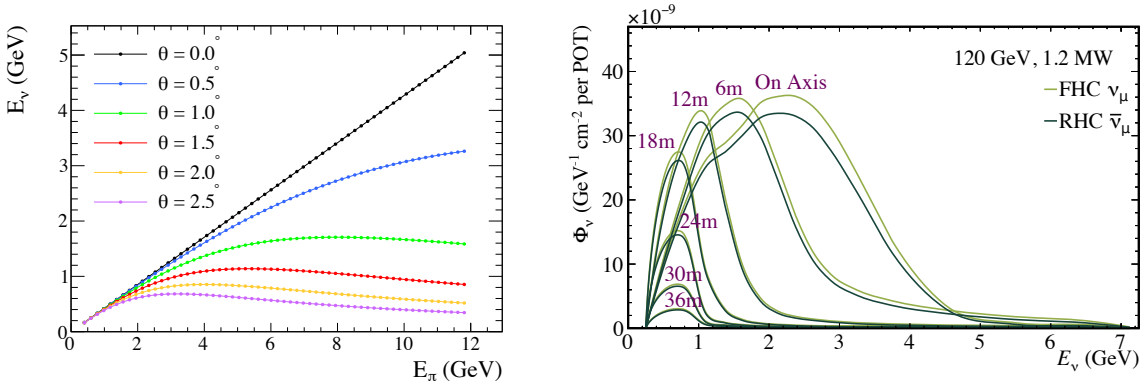

Figure 2.7: Neutrino energy as a function of its parent pion energy for various decay angles (left) and the resulting neutrino energy spectrum for various near detector off-axis positions (right). Neutrino energy spectra are shown for the beam in neutrino mode, or forward horn current (FHC), and antineutrino mode, or reverse horn current (RHC). The full deflection of 36 m corresponds roughly to an off-axis angle of 3.6°. Taken from [49].

2.1.3 Far detector

DUNE takes an initial and a final measurement of the neutrino beam. The final measurement is performed by the far detector, which is located 1300 km from the beam source and receives a neutrino flux of only $1.65 \cdot 10^{-10}\ \nu/\mathrm{m}^2$ per POT, or $5.8 \cdot 10^3\ \nu/\mathrm{m}^2/\mathrm{s}$ taking the expected beam intensity into account [44]. The expected neutrino flux as a function of energy is shown in figure 2.8. The main signal depends on $\nu_e/\bar{\nu}_e$ appearance and $\nu_\mu/\bar{\nu}_\mu$ disappearance, the expected interaction rates of which are listed in table 2.2. The total background from both modes combined is expected to originate primarily from beam contamination and neutrino misidentification. These numbers depend highly on the exact value of neutrino oscillation parameters and are given here for a normal-ordered hierarchy with $\delta_{CP} = 0$.

The number of detected neutrino events scales linearly with the detector mass, the beam power and the operational time of the experiment. In the case of a 10 kt (fiducial) far detector module and a proton beam of $\sim$1 MW, one can expect and average of 1.8 neutrino events per day, requiring 30 years of continuous operation to reach the desired 300 kt·MW·year. The far detector complex therefore needs to be outfitted with yet more massive detectors and that the beam be upgradeable to provide $\sim$2.4 MW of proton beam power to the target. The required large mass of the detector places

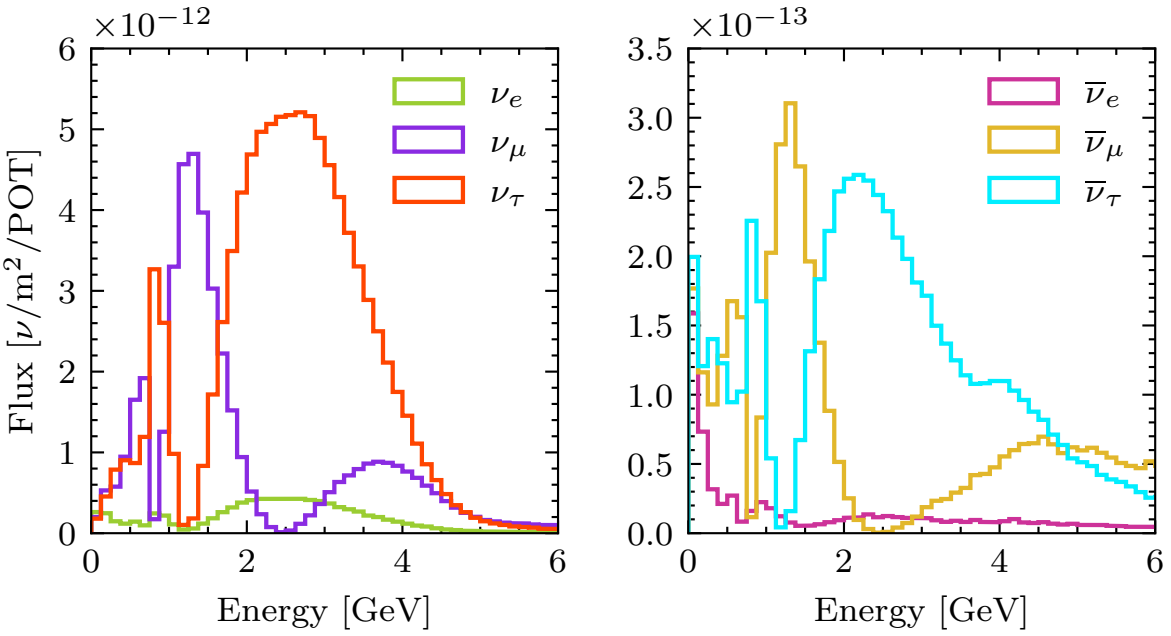

Figure 2.8: Expected far detector neutrino (left) and antineutrino (right) flux with the beam in neutrino mode. Neutrino oscillation simulations were performed with the GLoBES framework [45] using best-guess oscillation parameters. The scale difference between the plots shows the low flux of the expected antineutrino contamination. Data taken from [46].

Table 2.2: The expected number of signal and background events in the far detector for neutrino disappearance (ν_μ, $\bar{\nu}_\mu$) and appearance (ν_e, $\bar{\nu}_e$) charged-current interaction. Numbers are shown for the first 3.5 years of operation, with assumed normal ordering and $\delta_{CP} = 0$. Data taken from [36].

	ν_μ	$\bar{\nu}_\mu$	ν_e	$\bar{\nu}_e$
Signal	6200	2303	1110	300
Background	643	1259	317	180

limitations on its design. It must be composed of relatively cheap material and cannot feasibly feature a magnetic field to distinguish differently charged particles. Compared to the near detector, the modules of the far detector are therefore relatively monolithic.

The far detector of DUNE is planned to consist of four separate cryostat modules, each holding 17 kt of liquid argon of which 10 kt will be active detector material. This brings the liquid argon weight in the finalised detector to almost 70 kt. In order to accommodate all of this argon, the internal dimensions of the four cryostats will be 14.0 (W) × 14.2 (H) × 62.0 (L) m^3 and they will be equipped with sophisticated

insulation and argon purification systems [39]. The detector halls will be situated 1.5 km underground at SURF, in what used to be the Homestake goldmine. Each of the four cryostats will be identical in dimensions, but there is the possibility to use different detection techniques between the modules. The first module to be installed is planned to be a single-phase (SP) TPC, which uses exclusively liquid argon. Another module may be of the dual-phase (DP) TPC type, which employs a layer of gaseous argon in addition to its main body of liquid argon [38].

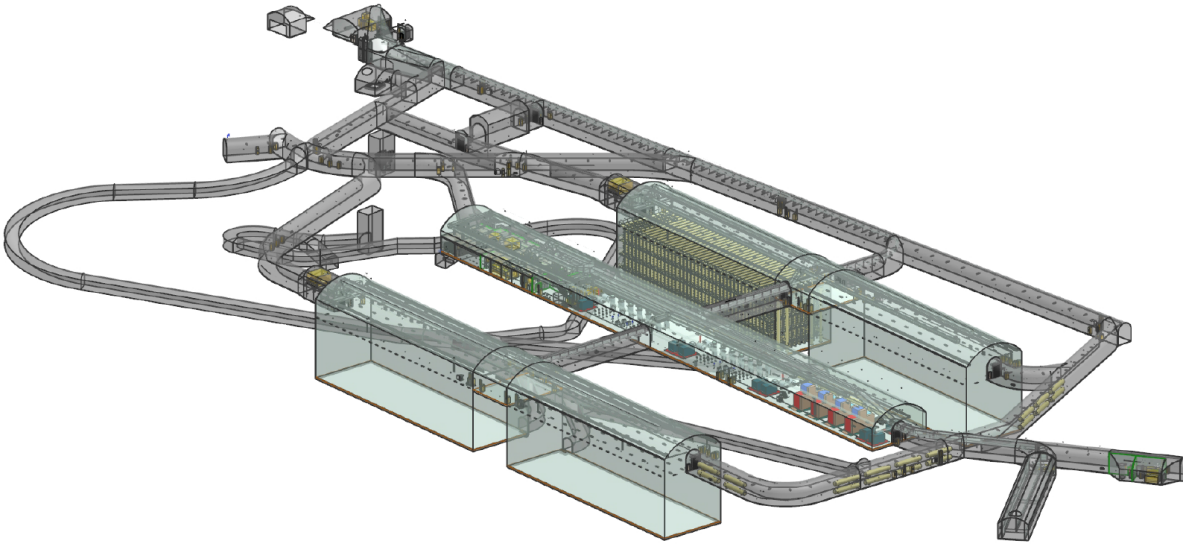

Figure 2.9: The DUNE far detector cavern system. Clearly visible are the four detector module caverns and central service cavern. One detector module is shown installed. Taken from [41].

The far detector will be installed in the Homestake mine, previously the largest and deepest gold mine in North America. This mine was previously host to another groundbreaking neutrino experiment: the Homestake (or Davis) experiment of the 1960s [3]. It has hosted a range of experiments since then that either require very low background radiation or observe the rock and life deep underground. Together, these laboratories form the Davis Campus of the Sanford Underground Research Facility. A set of caverns and shafts will be excavated for the installation of the DUNE far detector, requiring the removal of around 800,000 tons of rock [50]. Being 1.5 km underground, the rubble from this excavation will be transported to the surface in batches via the mine complex's shafts. Likewise, detector components must fit through these shafts to be transported to the detector caverns. They may be transported in the mineshaft transport cages if dimensions permit or slung under them if not.

The DUNE far detector is set up to be built like a ship in a bottle. The outer dimensions of a single module will be 19.2 (W) $\times$ 18.1 (H) $\times$ 66.1 (L) m^3, but its components must fit down shafts approximately 3.7 m $\times$ 1.4 m in size [51]. In particular, the anode and cathode planes of DUNE's TPCs are therefore necessarily modular: they are divided into pieces roughly 2.3 $\times$ 6 m wide. One such section of an anode plane is named an Anode Plane Assembly (APA) and a cathode plane module is named a Cathode Plane Assembly (CPA). Aside from logistics, the modularity of these components also allows for ease of construction. The APAs in particular are delicate instruments that require a large amount of labour to produce and are thus best spread over multiple institutes. Their modularity also allows for prototyping of the full-sized components in a far detector prototype called ProtoDUNE. Chapter 3 goes into more technical detail on this prototype and the detector components.

Time projection chamber

Time Projection Chambers work by collecting charge deposits in a volume of gas or liquid using an electric field generated between cathode and anode planes. As electrically charged particles travel through the detector medium, they leave trails of ionised particles, the electrons of which are driven to the anode plane by the electric field. This so-called electron drift must be well-understood, since the ionisation electrons form the main signal in the TPC. By measuring exactly when and where on the anode electrons arrive, a three-dimensional snapshot of the inside of the detector can be made. The TPC records distance in the drift direction in terms of time, which is where it derives its name from. Since the exact timing of particle interactions is not generally known from TPC data alone, a photon detection system can used to locate the interactions in time via their scintillation photons in the detector medium. These concepts are explained in more detail in sections 3.2.2 and 3.2.3. As can be seen in figure 2.10, the single-phase DUNE far detector TPC will contain three anode and two cathode planes in an alternating configuration, positioned vertically along the longest axis of the module.

Historically, TPCs were often filled with a gas in which high energy particles produce ionisation signals. The sparse nature of gases allows for a relatively clear path for the ionisation electrons, and gas TPCs are easier to fill and flush. However, gases are sparse detection media, being thousands of times less dense than liquids and solids. In attempting to detect neutrinos, where a high detector mass is vital, it is

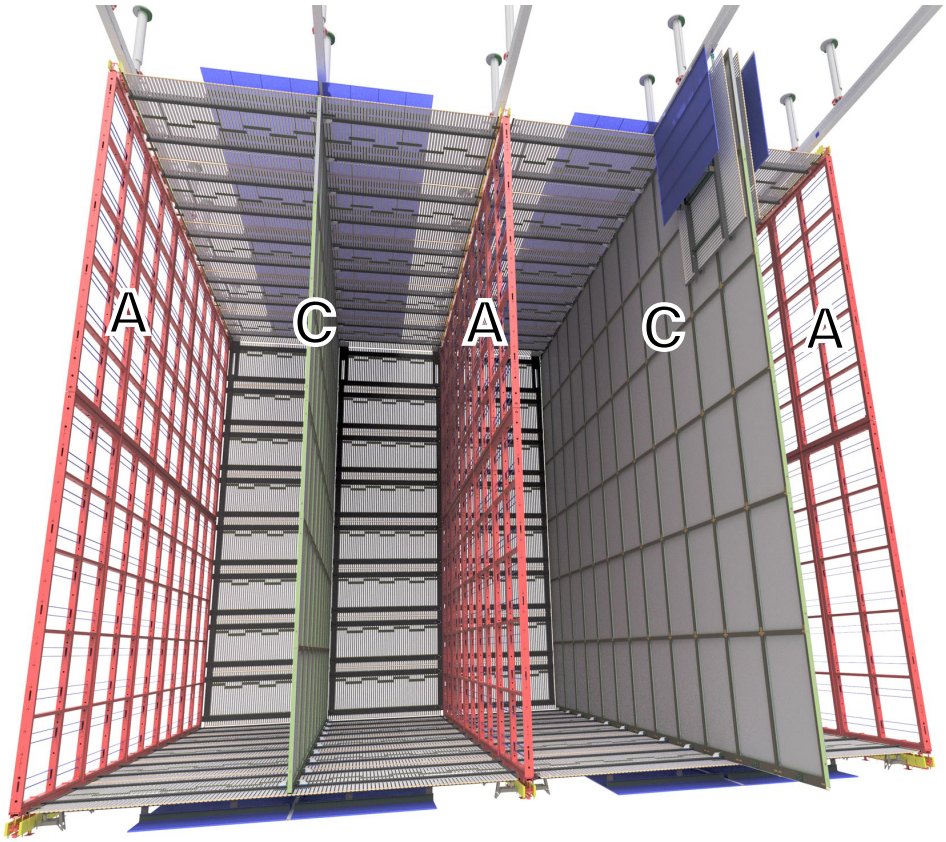

Figure 2.10: A cross-section of a single-phase DUNE far detector TPC. Anode planes (A) and cathode planes (C) are positioned in an alternating configuration, creating multiple drift volumes between them. Taken from [39].

therefore preferable to use a liquid as the active TPC material instead. Argon provides a good choice because it is inert, relatively cheap and abundant. In addition, it liquefies at 87 K, which is easily attainable, being above the boiling point of nitrogen (77 K). Argon also scintillates and is transparent to its own scintillation light, which aids in finding the absolute position of ionisation deposits in the detector volume, explained in more detail in section 3.2.3. In the single-phase DUNE far detector module and the ProtoDUNE prototype, electrons are driven through liquid argon by an electric field of 500 V/cm, reaching a velocity of around $1.6 \cdot 10^3$ m/s [52] and traversing a maximum of 3.6 m between the cathode and anode planes.

2.2 Research goals

The scale and configuration of DUNE make it qualified to perform several key measurements in physics. Some, such as the refinement of neutrino oscillation parameters and mass hierarchy are competitive with other experiments, whereas the detection of supernova neutrinos is more collaborative.

- DUNE will refine current knowledge of the parameters that govern neutrino oscillation, in particular θ_{23}, θ_{13}, Δm^2_{31} and δ_{CP}. Determining the sign of Δm^2_{31} is vital in determining the neutrino mass ordering and a nonzero CP-violating phase δ_{CP} would indicate leptonic CP violation. This CP violation is to be determined with a sensitivity of at least 3σ for 75% of the possible δ_{CP} range.
- In the event of a core-collapse supernova in the vicinity of Earth ($< \sim 100\,\text{kpc}$ [36]), the DUNE far detector is expected to observe several events, primarily from ν_e-argon interactions. Properties of these events, such as timing and direction, will provide valuable insights into the mechanics of the supernova process. In order to observe the entire supernova, the detector will need to switch to its continuous readout mode for the duration of the process, as opposed to its usual pulsed readout focusing on neutrinos from the neutrino beam.
- Required by the grand unification of forces in our current understanding, the measurement of proton decay can be performed in DUNE by observing the $p \to K^+\bar{\nu}$ and $p \to e^+\pi^0$ decay channels. This measurement is performed in the far detector, being a large sensitive detector shielded from most cosmic ray activity.
- Thanks to its intense neutrino beam and two capable sets of detectors, DUNE is capable of probing many additional fields of research, such as the measurement of atmospheric neutrino oscillation, neutrino-matter cross sections and dark matter searches [36].

2.2.1 Near detector requirements

The near detector of DUNE serves many functions, but its most important function is to serve as an accurate reference for the far detector. This includes flux measurements of the neutrino beam and all of its individual components, most importantly ν_μ, $\bar{\nu}_\mu$,

ν_e and $\bar{\nu}_e$. Unlike the far detector, the near detector should be able to distinguish neutrino and antineutrino interactions in order to measure the beam's contaminants. It will do so in magnetised portions of the detector. Another task of the near detector is to characterise event signatures that resemble signal events. These mimicking processes may come from beam contaminants, cosmic rays and nuclear decay, and will be identifiable through combined measurement from all of the near detector's subdetectors [36].

2.2.2 Far detector requirements

The primary goal of the far detector is to measure ν_μ ($\bar{\nu}_\mu$) disappearance and ν_e ($\bar{\nu}_e$) appearance with the neutrino beam in ν ($\bar{\nu}$) mode. The experimental setup fulfils some of the requirements to complete this goal. By building the far detector deep underground, it is shielded from cosmic rays to such a degree that they interfere with less than 1% of the beam neutrino events. The distance between the beam start and far detector is likewise of vital importance. In combination with a neutrino beam that has an energy spectrum between 0.5 and 4 GeV, the first and second maximum of the neutrino oscillation are clearly visible at 1300 km, as shown in section 2.3. Both the depth and distance requirements are satisfied by planning the DUNE far detector to be built at SURF, which is located at an appropriate distance from Fermilab and holds a research complex roughly 1.5 km underground. In terms of the number of recorded neutrino events, the mass of the detector plays a large role: roughly 900 kt·MW·years of exposure is required to determine CP violation at 3σ for 75% of the parameter space [36]. This corresponds to over five years of data taking even with the fully completed 70 kt far detector and upgraded 2.4 MW proton beam, none of which will be available initially.

Two trigger principles will be applied in the readout of the far detector. Firstly, high energy events (O(100 MeV)) originating from the beam, cosmic rays and potential nucleon decay will trigger a readout in the relevant section of the detector. The readout window is set at 5.4 ms, which is sufficient to record the full volume between the targeted anode and cathode planes in the single-phase TPC configuration [39]. Secondly, in the case of a supernova, a so-called burst trigger will be used. This type of trigger activates on events spread out over a relatively long time (~ 10 s) with an energy above the background level from 5 to a few tens of MeV. Although these events might be of a lower energy than the beam neutrinos that the far detector usually

detects, depending on the proximity of the supernova there could be thousands of neutrino interactions in the detector within ten seconds [36].

2.2.3 Reconstruction and calibration

Estimates of the exposure required to reach DUNE's physics goals are based in part on assumptions of reconstruction efficiencies. The experiment makes use of a broadband neutrino spectrum to make multiple neutrino measurements using a single beam. As such, it is vital that the neutrino energy is reconstructed well. Detailed Geant4-based [53] particle simulations are used to simulate particle propagation and interaction, and detector response simulations provide simulated detector output that reconstruction algorithms such as Pandora [54] can test their capabilities on as they are developed. The reconstruction chain involves hit finding, hit clustering and particle hierarchy construction, among other tasks. All of these are explained in more detail in chapter 5. Charged-current neutrino events are estimated to be reconstructed with an average energy resolution of 18 – 20% (ν_μ) and 13% (ν_e). These numbers are in part constructed from the resolutions shown in table 2.3. When averaging over the entire energy range, the leptonic energy resolution amounts to 8% and the hadronic energy resolution to 49% [36].

DUNE is an experiment with many unknowns: it uses cutting-edge detector technologies at an unprecedented scale to observe what will be the most intense neutrino beam in the world. In order to publish convincing physics results, the collaboration must therefore first make a case that the behaviour of these detectors is well-understood. Firstly, the physical properties of the detectors must be well-known in order for the sensitivity studies shown in section 2.3 to resemble reality. For example, the fiducial (active) volume of the far detector must be known to the 1% level, the position resolution of the detector must be within 1 – 2% and biases in leptonic and hadronic energy reconstruction must be kept small. The size of the far detector requires sophisticated corrections when it comes to electric field strengths and other position-dependent factors [36].

Events that have a characteristic energy signature, such as cosmic muons, π^0-particles (see chapter 7) and ^{39}Ar beta decays, can provide so-called standard candles. These help locate detector elements in the data, provide a source for $\mathrm{d}E/\mathrm{d}x$ calibration (see chapter 6) and give a sense of the electron lifetime in the detector volume, among

Table 2.3: An overview of the estimated far detector response to various particles. In some cases, the particle energy resolution depends on the track length, or whether it is contained and classified correctly. These numbers were used to generate sensitivity plots as shown in section 2.3 and were taken from [40].

Particle type	Detection Threshold (KE)	Energy / Momentum Resolution	Angular Resolution
$\mu^{\pm}$	30 MeV	Contained track: track length Exiting track: 30%	1°
$\pi^{\pm}$	100 MeV	μ-like contained track: track length π-like contained track: 5% Showering or exiting: 30%	1°
$e^{\pm}/\gamma$	30 MeV	$2\% \oplus 15\%/\sqrt{E[\text{GeV}]}$	1°
p	50 MeV	p<400 MeV/c: 10% p>400 MeV/c: $5\% \oplus 30\%/\sqrt{E[\text{GeV}]}$	5°
n	50 MeV	$40\%/\sqrt{E[\text{GeV}]}$	5°
other	50 MeV	$5\% \oplus 30\%/\sqrt{E[\text{GeV}]}$	5°

other subjects. Similar experiments that have paved the way for DUNE likewise provide valuable information and techniques, such as the ProtoDUNE prototypes (see chapter 3) [55] and MicroBooNE [34]. The far detector will also have its own monitoring and calibration systems which will include argon purity monitors, thermometers, an ionisation laser and possibly the addition of radioactive material to the detector volume with a known decay energy spectrum [36].

2.3 Measurement in DUNE

Two of DUNE's primary research goals are to measure leptonic CP violation and to determine the neutrino mass ordering. It is also designed to refine the current measurements of neutrino oscillation parameters. In practice, the energy spectra of far detector-measured neutrinos and antineutrinos ($\nu_\mu \to \nu_\mu$, $\bar{\nu}_\mu \to \bar{\nu}_\mu$, $\nu_\mu \to \nu_e$, $\bar{\nu}_\mu \to \bar{\nu}_e$)

are fitted simultaneously to obtain all neutrino oscillation parameters at the same time. However, some of the most important features can be distinguished from neutrino spectra by eye. A useful measure for this is the observed neutrino asymmetry, which is the normalised difference between neutrino and antineutrino appearance:

$$\mathcal{A}_{\mu e} = \frac{P(\nu_\mu \to \nu_e) - P(\bar{\nu}_\mu \to \bar{\nu}_e)}{P(\nu_\mu \to \nu_e) + P(\bar{\nu}_\mu \to \bar{\nu}_e)}. \tag{2.1}$$

In order to determine the quality with which DUNE will be able to achieve its research goals, one can speak of the sensitivity to a measurement. A useful quantity to describe this metric is the test statistic $\Delta\chi^2$, which is determined from the variation between outcomes of a set of simulations that follow varying hypotheses. These hypotheses consist of the oscillation parameters, some of which are only loosely constrained, as well as experimental parameters that find their origin in the beam and detectors. Experimental results are often given with a significance in σ, which can be obtained from the test statistic as being roughly equal to $\sqrt{\Delta\chi^2}$ when $\Delta\chi^2 > 1$ [36]. A significance of 3σ is traditionally called "evidence" and a significance of 5σ means a measurement is certain enough to be called a "discovery." Wherever possible, the figures in this section indicate these 3σ and 5σ significance levels.

2.3.1 Neutrino mass ordering

The neutrino mass ordering is measurable through the sign of Δm^2_{32} ($= m^2_3 - m^2_2$), as can be seen intuitively in figure 1.5, and is most apparent in combination with matter effects. Although the connection between the mass ordering and matter effects is not obvious in three-neutrino oscillation, it is visible analytically in the two-neutrino case. Equation 1.19 shows that the matter potential always appears in combination with the squared mass difference as $A/\Delta m^2$. A sign flip in the squared mass difference thus inverts the matter effects in oscillation phenomena, making the distinction between possible mass orderings especially clear.

The matter potential also flips sign between neutrino and antineutrino oscillation due to the differing interactions that electron neutrinos and antineutrinos undergo with matter, allowing for an even clearer distinction between mass ordering cases. The combination of these two effects in the DUNE baseline is shown in figure 2.11, where it can be seen that matter effects create a large asymmetry between electron neutrino

and antineutrino appearance. As explained in section 1.4.3, matter effects are most apparent at higher neutrino energies. In the case of DUNE's 1300 km baseline, they appear most clearly at the first oscillation maximum around 2.5 GeV.

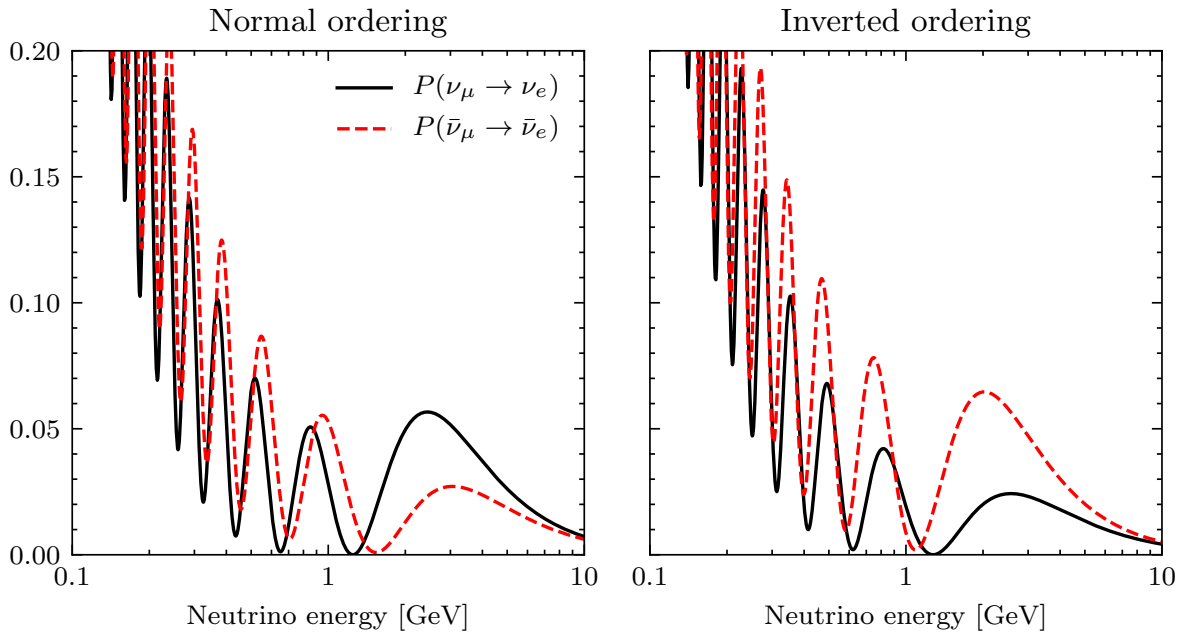

Figure 2.11: Neutrino and antineutrino appearance probabilities as a function of energy for normal (left) and inverted (right) mass ordering. Aside from $\delta_{CP} = 0$, parameters from table 1.1 were used, as well as the nominal DUNE baseline length and matter density.

One goal of the DUNE Collaboration is to determine the mass ordering with a significance of at least 5σ from beam neutrino observation alone. The exposure needed for this quantity depends on the final beam design, but also on other neutrino oscillation parameters, most notably δ_{CP} and the mass ordering itself. Figure 2.12 illustrates how these parameters influence the measurement of the mass ordering.

2.3.2 CP violation

The CP-violating phase δ_{CP} features directly in three-neutrino oscillation, as introduced in equation 1.9. In figure 2.13, it can be seen that the effect of CP violation is more pronounced at lower energies. This is in contrast to the matter effect, which manifests itself most strongly at the higher end of the energy range. The broad energy range

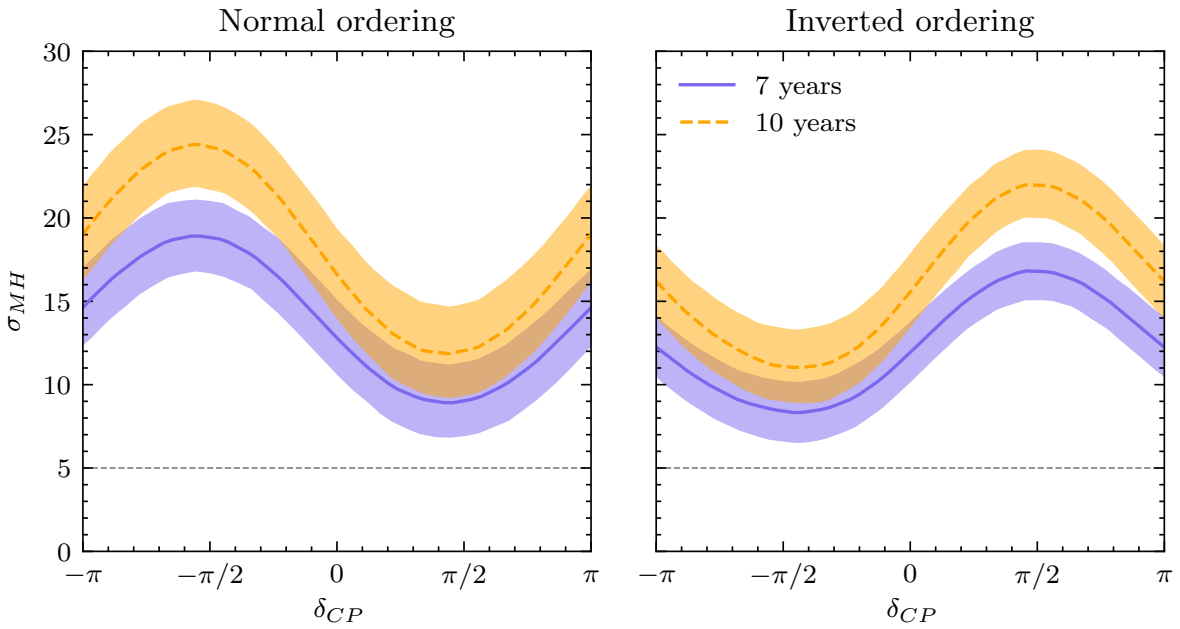

Figure 2.12: DUNE's sensitivity to the mass ordering as a function of the CP-violating phase in the case of true normal ordering (left) and true inverted ordering (right) for 7 and 10 years of equal neutrino and antineutrino exposure. The bands signify a 1σ uncertainty. Data taken from [36].

of the neutrino beam thus allows for a simultaneous measurement of both effects, as explored in more detail in section 2.3.3.

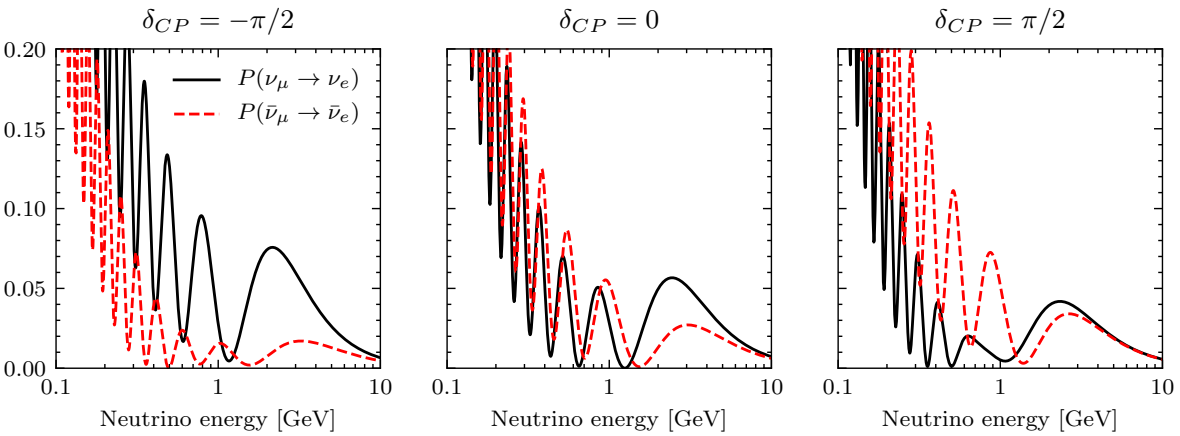

Figure 2.13: Neutrino and antineutrino appearance energy spectra for maximal (left, right) and minimal (middle) CP violation. Parameters from table 1.1 were used with assumed normal ordering, as well as the nominal DUNE baseline length and matter density.

DUNE's ability to determine CP violation ($\delta_{CP} \neq 0, \pm\pi$) is heavily dependent on the true value of δ_{CP}. In particular, CP-violating effects wane where the δ_{CP} parameter approaches 0 or $\pm\pi$. Figure 2.14 shows the resulting characteristic dips in the CP violation measurement, as well as the effect of different neutrino mass hierarchies. It can be seen that this measurement generally requires significantly more exposure than that of the mass ordering.

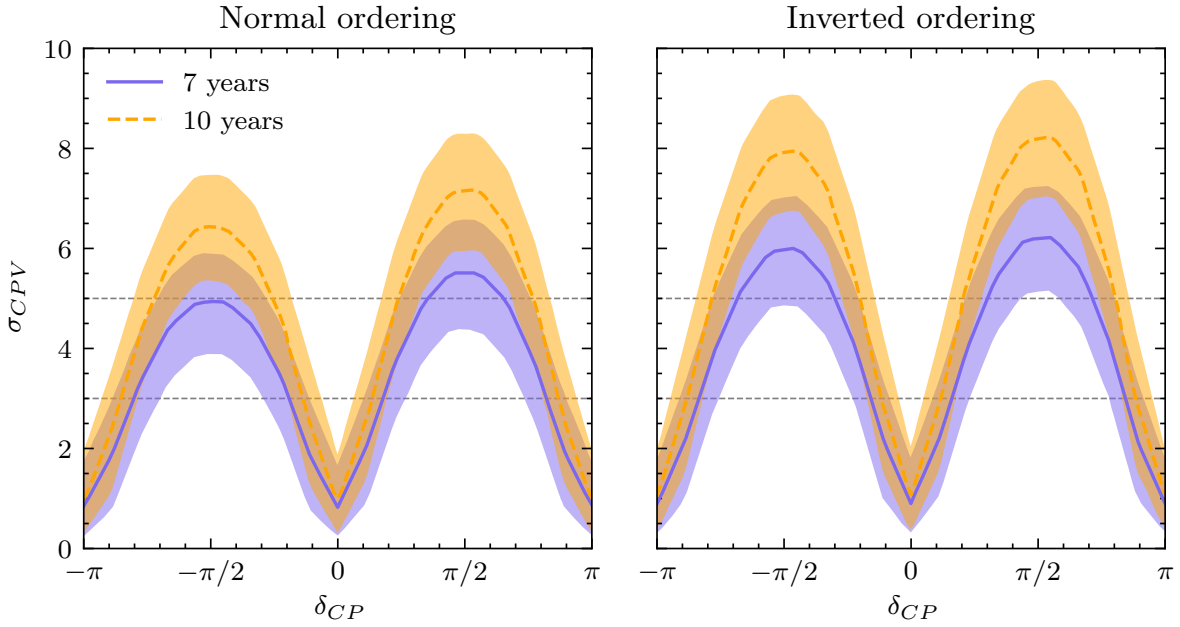

Figure 2.14: DUNE's sensitivity to CP violation as a function of the CP-violating phase in the case of true normal ordering (left) and true inverted ordering (right) for 7 and 10 years of equal neutrino and antineutrino exposure. The bands signify a 1σ uncertainty. Data taken from [36].

2.3.3 Distinguishing CP violation and the mass ordering

CP violation and the MSW effect can be distinguished thanks to the fact that the two effects manifest themselves most strongly in different energy regimes. Figure 2.15 shows the electron neutrino appearance asymmetry for various values of δ_{CP} and the two mass ordering possibilities. The vertical dashed lines visualise the energy of the two most prominent electron neutrino appearance peaks. These correspond to the peaks around 2.5 and 0.9 GeV in figures 2.11 and 2.13. The peaks morph and move slightly for different values of the neutrino oscillation parameters, for which reason

they might not appear exactly at the shown values when considering a different set of oscillation parameters.

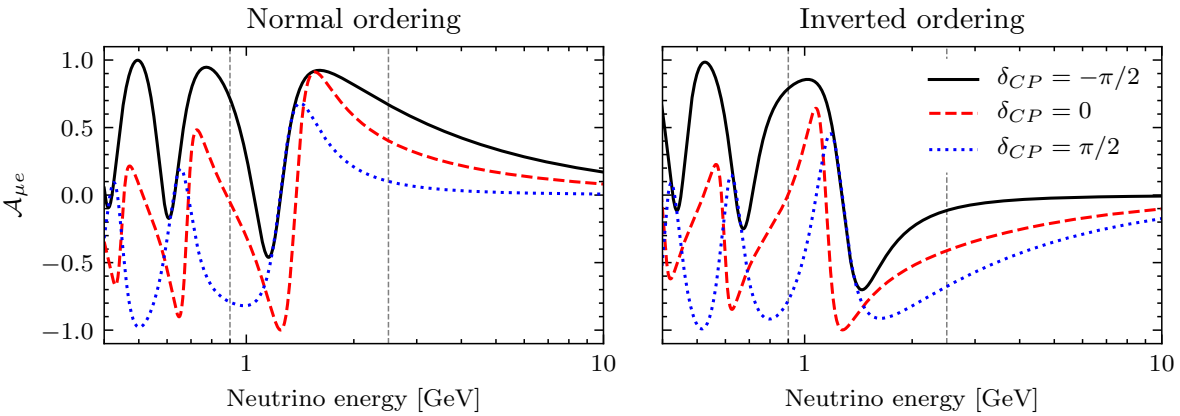

Figure 2.15: The neutrino appearance asymmetry (defined in equation 2.1) as a function of the neutrino energy for various values of δ_{CP}. Both the normal (left) and inverted (right) mass ordering are shown. The first and second oscillation maximums are indicated with vertical lines around 2.5 and 0.9 GeV respectively.

Figure 2.15 shows that the behaviour of $\mathcal{A}_{\mu e}$ is quite different for the two most important peaks in the neutrino energy spectrum. Around the first oscillation maximum at 2.5 GeV, the asymmetry flips sign between mass orderings but is relatively unchanged by δ_{CP}. Around the second oscillation maximum of 0.9 GeV, however, the asymmetry remains nearly unchanged between mass orderings and varies drastically for different values of δ_{CP}. These effects are shown in figure 2.16, where for 0.9 GeV neutrinos the two mass ordering cases overlap near-perfectly whereas they are completely separated in neutrinos around 2.5 GeV. As such, low-energy neutrinos give insight into the value of δ_{CP} regardless of mass ordering, while high-energy neutrinos reveal the mass ordering regardless of the value of δ_{CP}.

2.3.4 Exposure

In order for DUNE to reach its physics goals, it needs to record a large volume of oscillation data over a large period of time. The statistics of this process are complicated and DUNE will start recording data while it is only partially completed, being modular by design. Summarising the extensive studies published in [36], the significance of the mass ordering and CP violation measurements over time is shown in figure 2.17. The true value of δ_{CP} can be seen to affect the measurement significance significantly. This

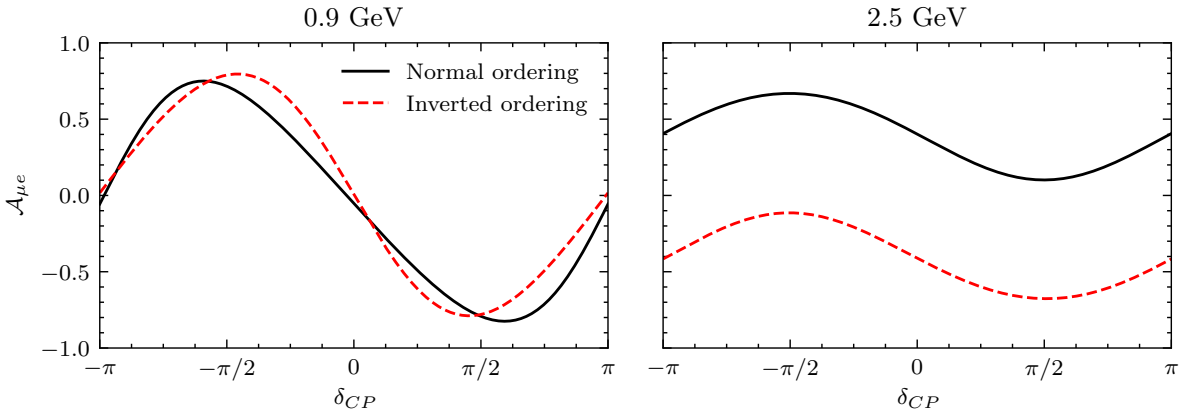

Figure 2.16: The neutrino appearance asymmetry $\mathcal{A}_{\mu e}$ as a function of δ_{CP} and the two mass ordering possibilities, shown both for a neutrino energy of 0.9 GeV (left) and 2.5 GeV (right).

agrees with figures 2.12 and 2.14, which show large disparity in expected significance for different values of δ_{CP}. As anticipated, the mass ordering measurement is expected to reach a conclusive state of $> 5\sigma$ long before the CP violation measurement.

2.3.5 Other oscillation parameters

In the previous sections, special attention was paid to the measurements that determine CP violation and the neutrino mass ordering, governed by δ_{CP} and the sign of $\Delta m^2_{32} \approx \Delta m^2_{31}$ respectively. DUNE is also poised to increase the resolution of these parameters and will surpass or approach previously attained resolutions of the mixing angles θ_{13} and θ_{23} [36]. Figure 2.18 shows the expected resolution of these parameters as a function of time. It can be seen that the measurement of the CP-violating phase is highly dependent on its true value, where CP-conserving values ($0, \pm\pi$) are able to be determined to a greater precision than CP-violating values. Of special note is the mixing angle θ_{23}, of which it is not known yet whether it is maximal ($\theta_{23} = 45°$). Maximal mixing in θ_{23} would imply that the ν_μ and ν_τ states contribute to the mass eigenstate ν_3 equally, indicating a new natural symmetry.

As mentioned, oscillation parameters are generally not determined in isolation, but are found through a fit involving all parameters. The expected resolutions shown in figure 2.18 were obtained assuming the current best fit values for all parameters.

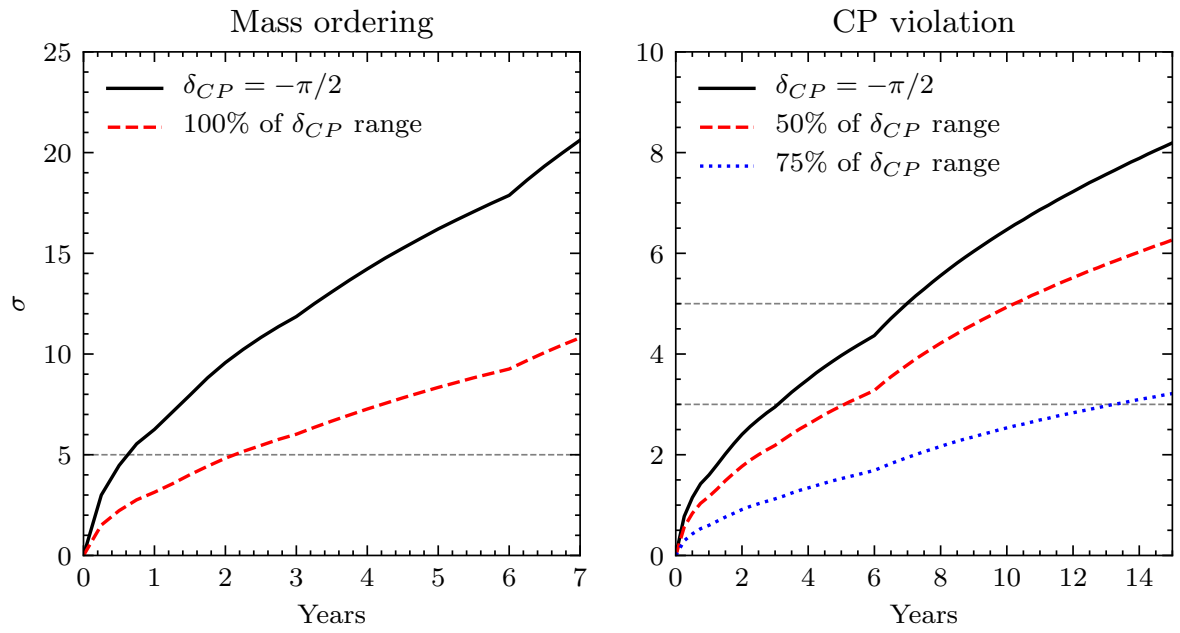

Figure 2.17: Significance of the mass ordering (left) and CP violation (right) measurements as a function of operating time. Normal ordering and equal neutrino and antineutrino exposure is assumed. The graphs are kinked because enhancements of the DUNE far detector and neutrino beam are expected over time. Note the large difference in both the time and significance range between the two plots. Data taken from [36].

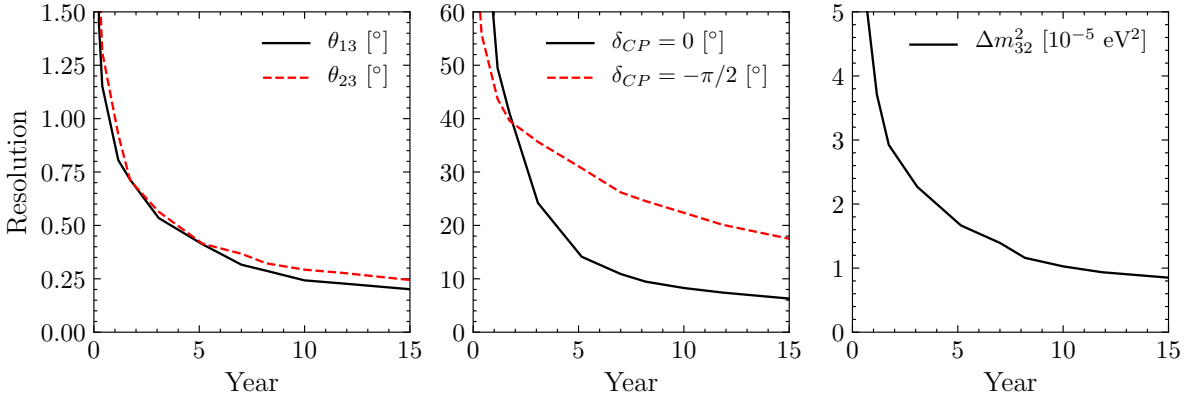

Figure 2.18: The resolution of θ_{13} and θ_{23} (left), δ_{CP} (middle) and Δm^2_{32} (right) over time. Data taken from [36].

Various dependencies of parameter resolutions upon the parameter space are shown in appendix B.

2.3.6 Near detector physics

Thanks to its close proximity to the neutrino beam source, the DUNE near detector receives an unprecedented neutrino flux and is thereby a capable experiment in its own right. The expected ν_μ event rate in its 50 t LArTPC alone is expected to be on the order of $1.2 \cdot 10^8$/year (4/s) [44]. It will be able to reduce uncertainties in neutrino-nucleus interactions and investigate rare neutrino interactions with cross-sections $\sigma < 10^{-44}\,\mathrm{cm}^2$. There is potential to perform precision measurements of a number of model parameters. One of them is the weak mixing angle, which is best accessible through νN deep-inelastic scattering and will be able to be determined down to an uncertainty of 0.2%. This is on par with the current standard set by NuTeV [56][57]. Isospin physics may also be probed by comparing neutrino and antineutrino scattering off hydrogen in polypropylene foils. In this manner, a conserved quantity known as the Adler-sum may be measured down to the percent level, making it possible to test its conservation [57]. Likewise, the contribution of the strange quark to the nucleon spin can be determined by neutrino scattering off nucleons in the detector material. Aside from performing precision measurements of known quantities, the Near Detector makes studies into BSM physics possible. Oscillations to sterile neutrinos can be probed as the Near Detector is at an $L/E \sim 1\,\mathrm{eV}^{-2}$ and is thereby sensitive to $\Delta m^2 \sim 1\,\mathrm{eV}^2$ oscillations. This could hint at one or more previously undiscovered sterile neutrinos. Additionally, the near detector may be sensitive to certain types of light dark matter. These particles may be created in large numbers in the proton-target collisions of the neutrino-generating beam, or as part of subsequent meson (π^0, η) decay. It is theorised that dark matter particles could interact with ordinary matter by undergoing neutral current-like interactions, exchanging a so-called dark photon with nuclei in the near detector [57].

2.3.7 Rare natural processes

The DUNE far detector is set to be a high-resolution and low-background particle detector and as such can listen for rare processes with faint signatures. Two of its goals are the detection of supernova neutrinos and observation of proton decay. Other

secondary goals exist that will not be discussed in further detail, such as the detection of atmospheric neutrinos [36].

Supernova neutrinos

Much has been learned through an event that occurred on the 3rd of February, 1987. A star in the Large Magellanic Cloud exploded in a core-collapse supernova, sending an estimated 10^{58} neutrinos into the far reaches of space. Neutrino detectors around the world observed the neutrino burst hours before the supernova's light reached the earth, the most sensitive of which was the Kamiokande II observatory, detecting 12 antineutrinos in a short period of time. The observation led to constraints on various neutrino properties, such as an upper limit to the electron neutrino mass. Aside from neutrino properties, the detection also gave insight into the supernova process itself [58].

The scientific value of a handful of neutrinos from a single source speaks to the importance of detecting more. One of DUNE's secondary goals is therefore to be ready to observe any supernovae that may occur within its lifetime and detection range. The DUNE Technical Design Report [36] lists several supernova phenomena that would be observable in the DUNE far detector, of which a selection is listed below:

- The supernova explosion mechanism is theorised to be driven in part by neutrino creation and interaction. Observing the neutrinos themselves would provide more insight into this mechanism.
- Neutrino emission would be abruptly cut off by the formation of a black hole, should the supernova star be massive enough.
- Various dynamic processes such as shock waves, so-called sloshing instabilities and turbulence effects would alter the flavour composition of the neutrino flux, making them observable by neutrino detectors.

Proton decay

Detecting proton decay would be an important step in the unification of nature's fundamental forces while at the same time having fundamental implications for the stability of matter. As nucleon decay is one of the pillars of grand unified theories,

great efforts have been made towards its detection. So far, only limits on the proton lifetime have been set, the most stringent of which were formulated by the Super-Kamiokande detector. It found lower limits of the proton lifetime of $\tau > 1.6 \cdot 10^{34}$ years for the $p \to e^+\pi^0$ channel and $\tau > 7.7 \cdot 10^{33}$ years for the $p \to \mu^+\pi^0$ channel at 90% confidence. Current and future limits on the most commonly theorised proton decay channels can be seen in figure 2.19.

DUNE, with its 10 kt/module fiducial mass, is comparable to Super-Kamiokande's 22.5 kt and is much smaller than Hyper-Kamiokande's 187 kt, but has some advantages when it comes to measuring charged particles that are too slow to produce a Cherenkov light flash in water-based detectors. In particular, the $p \to K^+\bar{\nu}$ channel can directly be probed with high efficiency in DUNE, whereas Hyper-Kamiokande must rely on the kaon's decay products to produce a signal, such as $K^+ \to \mu^+ + \nu_\mu$ (64% branching ratio) and $K^+ \to \pi^+ + \pi^0$ (21% branching ratio) [35]. Figure 2.20 shows the lifetime sensitivity of the proton through the positron and kaon channels for DUNE and Hyper-Kamiokande, as well as the previously established Super-Kamiokande limit.

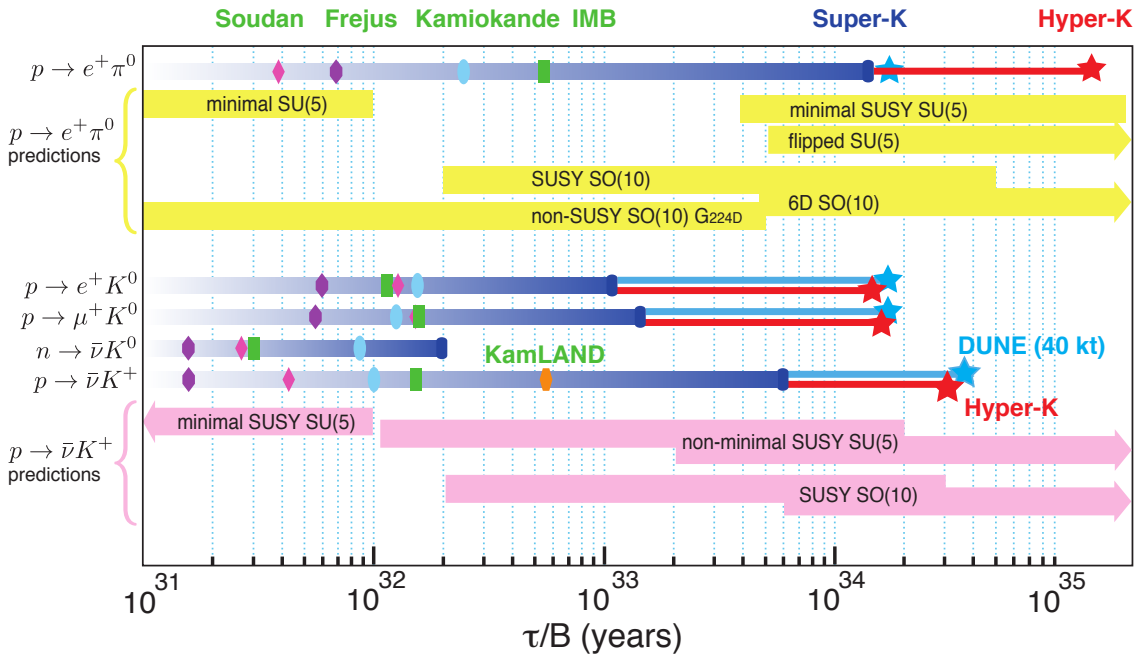

Figure 2.19: Current proton lifetime measurements through various decay channels. Also shown are the planned experiments of DUNE and Hyper-Kamiokande. Taken from [35][40].

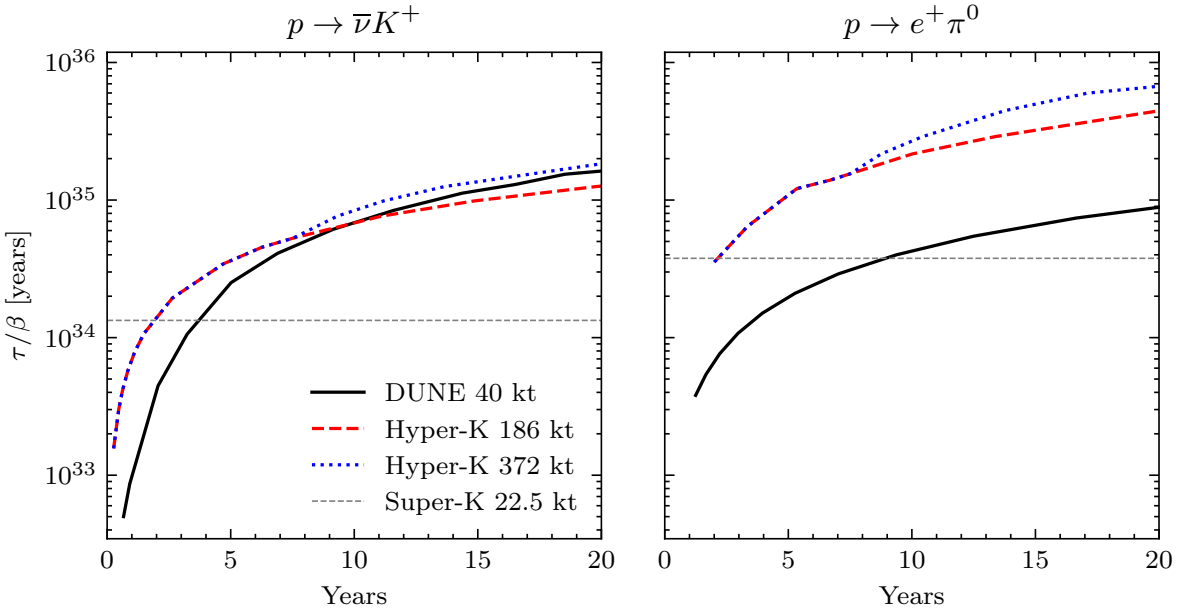

Figure 2.20: 3σ limit proton lifetimes for the $p \to K^+\bar{\nu}$ (left) and $p \to e^+\pi^0$ (right) channels. Note that Hyper-Kamiokande and DUNE are more or less tied when it comes to the kaon channel, despite DUNE's higher efficiency. This is due to Hyper-Kamiokande's much larger effective mass. Data taken from [35].

Chapter 3

ProtoDUNE

An experiment of DUNE's scale is not built without extensive testing. The planned DUNE far detector is planned to hold nearly 70 kt of liquid argon, making it orders of magnitude more massive than anything built so far. Detectors of the same principle have come before it, such as MicroBooNE [34] and ICARUS [59], but are of a much smaller scale at 170 t and 760 t respectively. The far detector had its own prototype in the 35t detector [60]. This relatively small detector served as a test for new technology such as the membrane wall of the cryostat and succeeded in proving the soundness of the concept.

The fact that the DUNE far detector needs to be modular for its transport into the Homestake mine has the advantage that its components can be tested on a much smaller scale. ProtoDUNE aimed to do just this: take a section of the envisioned DUNE far detector and put it into operation. At CERN, the neutrino platform initiative provided space and a charged-particle beam to build two prototypes: one single-phase version (ProtoDUNE-SP) which holds a volume of entirely liquid argon and one dual-phase version (ProtoDUNE-DP) which makes use of both liquid and gaseous argon. This thesis has its focus exclusively on ProtoDUNE-SP and will refer to that detector simply as ProtoDUNE.

ProtoDUNE's host, CERN (derived from Conseil Européen pour la Recherche Nucleaire), is the European Organisation for Nuclear Research. Seated on the Franco-Swiss border near Geneva, it hosts the largest particle physics community in the world. Its research is centred around its accelerator complex crowned by the Large Hadron Collider (LHC) [61], a synchrotron 26.7 km in circumference that accelerates protons to energies of 7 GeV and collides them in large particle detectors, most notably ATLAS [62], CMS [63], LHCb [64] and ALICE [47]. It was built up from its inception in

1954 through the Proton Synchrotron (PS), Super Proton Synchrotron (SPS) and Large Electron-Positron Collider (LEP). CERN's work culminated in the construction and operation of the LHC and subsequent discovery of the Higgs boson in 2012 [65]. The LHC also allows researchers to record processes occurring at higher energies than ever before, having the capability to generate collisions of up to 14 TeV in the centre-of-mass frame. Aside from proton-proton collisions, CERN hosts experiments involving heavy ion collisions, antimatter and neutrino research.

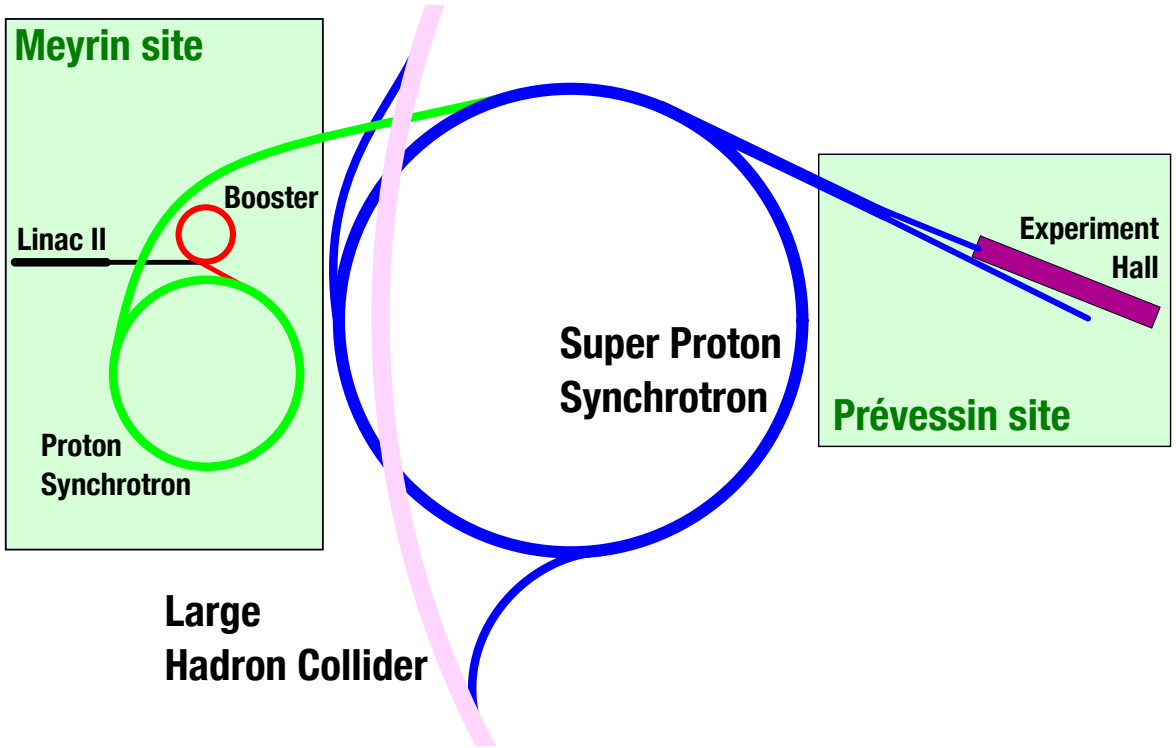

Figure 3.1: Part of the CERN accelerator complex. Protons travel from their initial linear accelerator Linac II via a set of synchrotrons (not to scale) to the LHC where collision experiments take place. Along the way, various paths branch off to lower energy experiments. One branch of the SPS supplies the Prévessin site with protons, where they are used to generate the charged particle beams for ProtoDUNE.

The construction of ProtoDUNE started in 2016 with the excavation of a 12 m-deep trench at CERN's north site. Located at the far end of its test beam hall, the ProtoDUNE area was scheduled to receive a charged particle beam from the Super Proton Synchrotron (SPS) from September until November of 2018. While the starting point of this beam run was given simply by the SPS schedule, the end date was dictated by the Long Shutdown 2 (LS2) of the CERN accelerator complex. Since LS2 was scheduled to last from late 2018 until early 2021, tests performed after this period

would not be in time for decisions taken in relation to the construction of DUNE. This left only two years for the construction and commissioning of the detector. For comparison, the smaller MicroBooNE detector had taken three years to come online. Despite the tight restrictions of the schedule, ProtoDUNE-SP was completed on time and successfully took data during the entirety of its beam run, meeting all of its objectives.

3.1 Experimental goals

Before construction on the four massive DUNE far detector modules can begin, the technology behind it must be subjected to extensive tests. ProtoDUNE uses many of the components the DUNE far detector is planned to use, but instead of 150 APAs, it uses just 6. Instead of featuring four drift volumes, it has two. An overview of ProtoDUNE's Time Projection Chamber (TPC) is shown in figure 3.2. Aside from requiring less space and fewer complex detector components, the smaller size of the detector also reduces the requirements on the data acquisition, the amount of liquid argon, diagnostic subdetectors and many other detector-related systems. The final ProtoDUNE detector is a cube of formidable size still, standing almost 11 m tall.

The most basic test of ProtoDUNE is that of structural integrity. Throughout the filling process and long after the inner volume of the detector has been filled with liquid argon, the cryostat should stay intact. The membrane design of the cryostat finds its origin in the LNG shipping industry and had already been tested with liquid argon in the 35t detector. ProtoDUNE uses this technology on a much larger scale than its predecessor. Aside from the cryostat membrane, the flexing of the outer wall and rest of the structure was monitored closely to stay within tolerances. Besides the detector's main structures, ProtoDUNE is just as much a test for the internal electronics, both those handling beam particle signals and those meant for diagnostics.

One core principle of the planned DUNE far detector is that of a 3.6 m electron drift length, which corresponds to the distance between the cathode and anode planes in the TPC. While MicroBooNE has already successfully operated with a 2.5 m drift length, the longer drift in a larger volume brings extra challenges with it. Argon ions and impurities within the argon are both capable of absorbing electrons on their way to the anode, and these are more disruptive the longer the drift length. ProtoDUNE was able to measure the argon purity over time using a dedicated purity monitor and

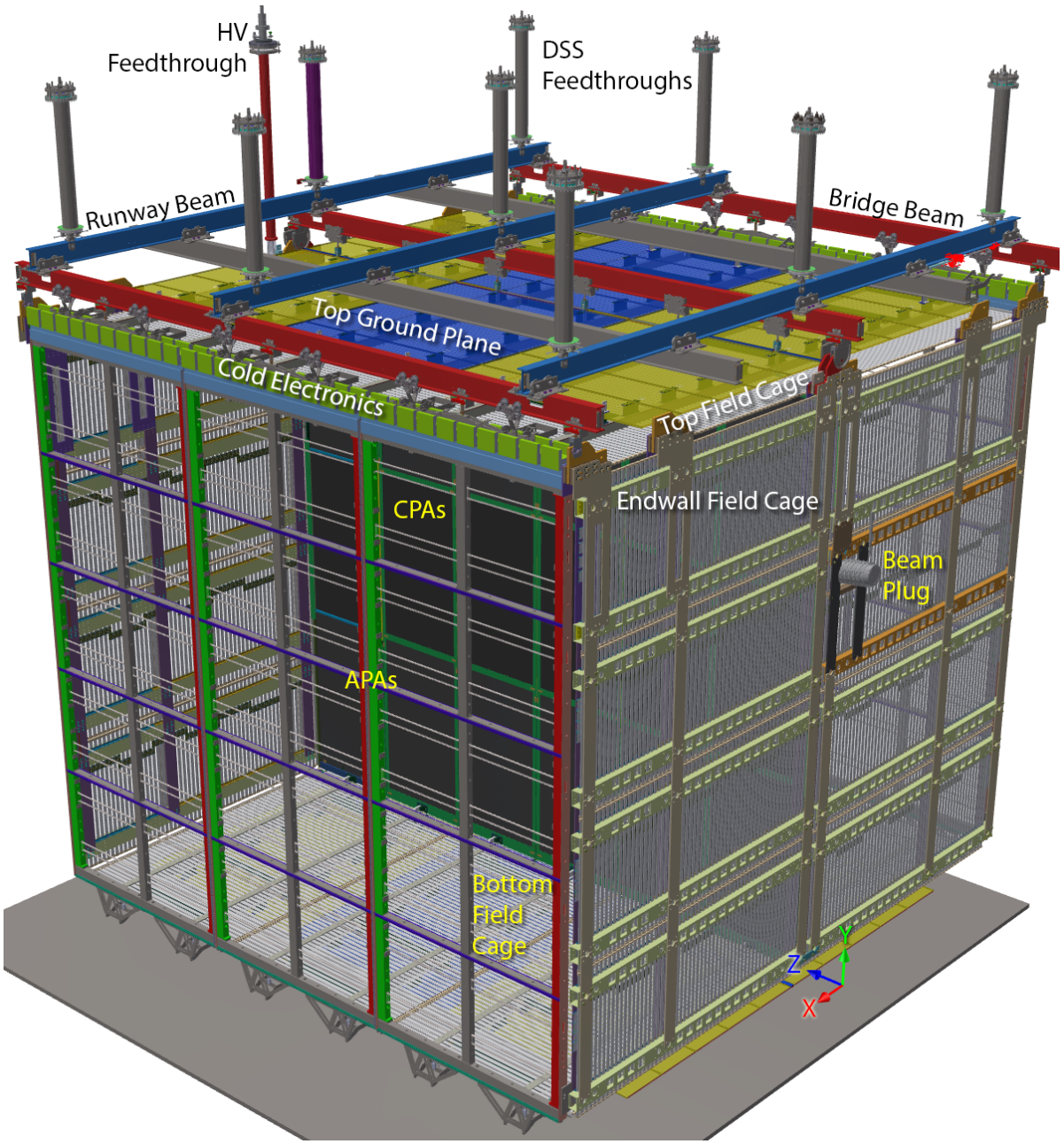

Figure 3.2: An unobstructed view of the ProtoDUNE TPC. The charged particle beam enters from the right through the beam plug. Three APAs are visible in this figure, with three more being positioned on the opposite side of the TPC, making for a TPC with two drift volumes. The entire TPC is submerged in liquid argon during operation. Taken from [55].

TPC-crossing muon tracks. It was also able to investigate reduced visibility in times of low argon purity.

Anode Plane Assemblies are the most important building blocks of the DUNE far detector and are detailed in section 3.2.2. They attract ionisation electrons liberated within the detector and measure their properties, such as total charge, position and time of arrival. Their complex wire planes cause them to be labour-intensive to fabricate and difficult to modify. Additionally, the far detector requires 150 of them per module, or 600 in total if all modules are built according to the same single-phase principle. These structures were therefore tested extensively in ProtoDUNE.

Aside from technical tests, another major goal of ProtoDUNE was that of calibration: using a charged particle beam of known composition and momentum, the response of these particles could be observed in the detector. An array of beam detectors independently measured each particle's type and location so that ProtoDUNE events could be matched accurately with beamline detector information. More information on the ProtoDUNE beam can be found in section 3.2.5.

As the beam run of the DUNE far detector is still years away, certain design choices can be altered based on the performance of ProtoDUNE. One such decision was on the data acquisition system to use. As a baseline, the RCE (Reconfigurable Cluster Elements) DAQ system [66] was used in five of ProtoDUNE's six APAs. The FELIX (Front-End LInk eXchange) system was tested on the single remaining APA. It is in development primarily in the context of the ATLAS detector at CERN [67], but was adapted to be used in ProtoDUNE as well. It is discussed in more detail in section 3.2.4 and chapter 4.

3.2 Experiment overview

DUNE, described in detail in chapter 2, was designed to be modular so that its components, once fabricated on the surface, may be moved and assembled underground. The base components of the inner detector are the anode plane assembly (APA) and cathode plane assembly (CPA). Each of these stands roughly 6 m tall and measures 2.3 m across. The relatively small dimensions of these components make it possible to prototype and characterise their operation at the scale of the eventual DUNE far detector.

3.2.1 Cryostat

The main structure of ProtoDUNE consists of a large $11 \times 11 \times 11\,m^3$ cryostat (inner dimensions $8 \times 7 \times 7\,m^3$). As ProtoDUNE is a homogeneous detector, it holds only liquid argon within its volume. In order to keep the detector structure intact during filling and for months afterwards, several key technologies previously developed in earlier prototypes have been employed. One of these is the membrane cryostat wall. Instead of a smooth surface, the inside layer of ProtoDUNE consists of two-dimensionally corrugated stainless steel that is able to stretch and contract locally without pulling the entire detector geometry askew, as can be seen in figures 3.3 and 3.4. More towards the outside of the detector, multiple layers of thermally insulating material are used to keep the argon cool in a passive manner.

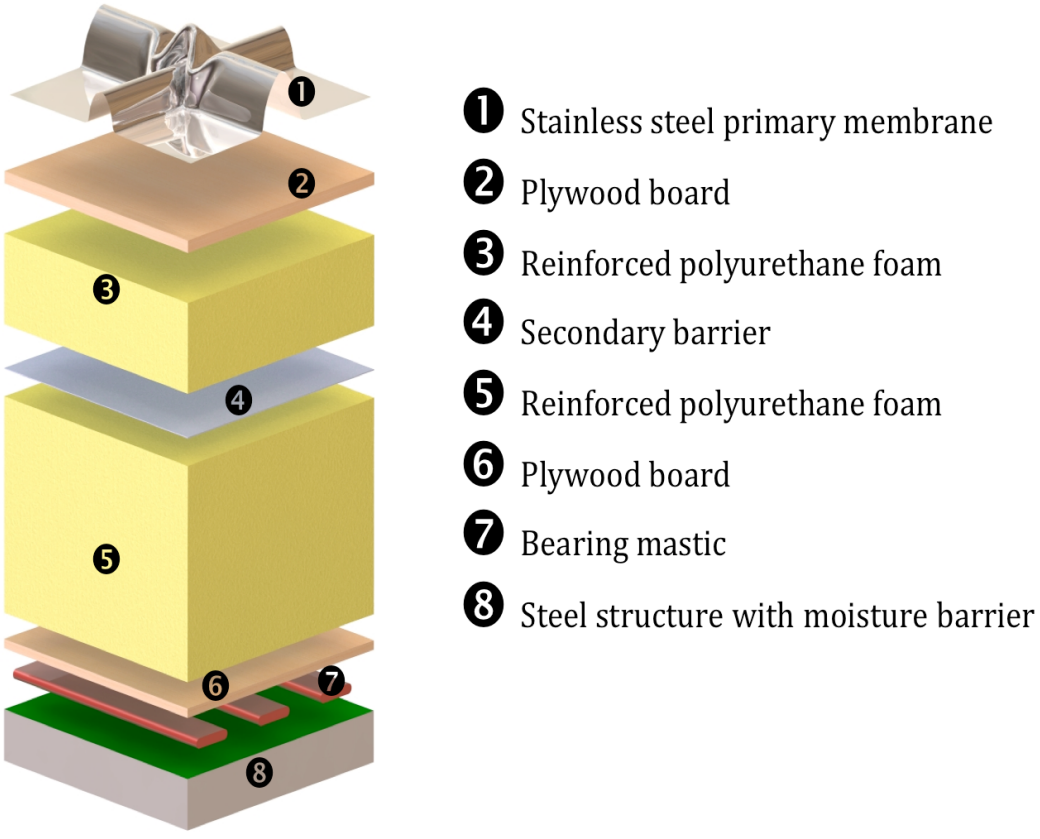

Figure 3.3: The various layers of insulation that ProtoDUNE uses to keep its argon contained and cool. The upper layer, the stainless steel membrane, has the ability to deform and stretch under extreme temperature changes thanks to its corrugated shape. Taken from [68].

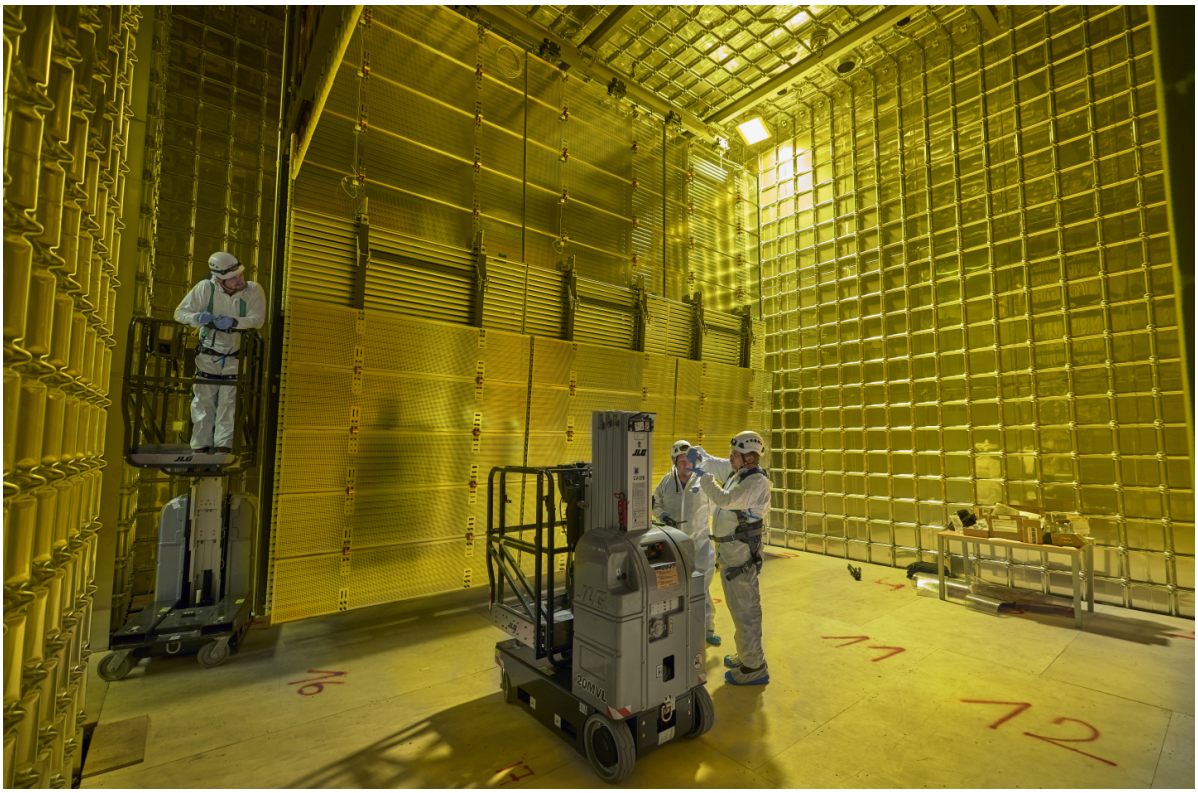

Figure 3.4: Engineers at work in ProtoDUNE before it was filled. The membrane cryostat wall and hanging cathode plane assemblies can be seen under light that has been filtered to preserve photosensitive detector components. Photo: CERN.

ProtoDUNE requires electrons to drift through liquid argon over distances $\sim 3.6\,\mathrm{m}$. However, oxygen and nitrogen contamination capture drifting charges, degrading the signal. Electron lifetime is defined as the drift time after which $1/e$ of the electrons are captured by the detector material. In ProtoDUNE, it must be on the order of milliseconds to retain a good signal-to-noise ratio. The electron lifetime is thus key within this experiment, and as such, the argon inside of the cryostat must be exceedingly pure. Argon is a common atmospheric gas and is used extensively in industry and science. Because of this, the noble gas is relatively cheap and able to be purchased on a large scale. The argon purchased by CERN is of grade 4.6, meaning that it contains 5-10 ppm oxygen equivalent contamination, but it does not meet the goal of < 100 ppt oxygen equivalent. ProtoDUNE therefore incorporates a filtration facility that managed to attain the appropriate argon purity [69].

3.2.2 Time projection chamber

Inside its cryostat, ProtoDUNE features a TPC ($7 \times 6 \times 6\,\mathrm{m}^3$) that measures the ionisation signal of particles passing through it. Consecutive anode and cathode panels (visible in figure 3.2) create an electric field within the liquid argon of 500 V/cm by applying 180 kV to the cathode plane. The ProtoDUNE TPC comprises two drift volumes: a vertical cathode panel divides the TPC volume through the middle and anode panels on either side of the detector complete the drift volumes. The top and bottom of the TPC are closed off with ground planes in order to confine the electric field to the drift volumes and make the electric field as homogeneous as possible.

APAs consist of an electrically insulating frame around which three layers, or planes, of wires are wound. The collection plane collects the electrons produced by particle tracks. Its wires run vertically along the APA frame. Two induction planes envelop the collection layer and feature wires that run at an angle of $\pm 35.7^\circ$ relative to the collection wires, as can be seen in figure 3.5. A final grid wire plane runs vertically around the other layers and helps shape the electric field around and between the various wire planes. The wires in all planes have a spacing of about 0.5 cm, thereby setting the spatial detector resolution. The APA wires consist of a beryllium-copper alloy 152 μm thick and are tensioned to 6 ± 1 N to maintain a consistent spacing [39]. The fragility of these components poses a challenge to the construction, transport and installation of the APAs, especially considering the number of required APAs (150 per single phase module) and the location of the DUNE far detector deep underground.

The front-end electronics connected to the anode plane wires measure their current at a rate of 2 MHz. Clouds of electrons passing by the induction wires produce a bipolar signal and their arrival at the collection wires produces a spike. Due to the manner in which the various wire planes are angled relative to one another, coincidence of signals between all three wire planes can be used to pinpoint the location of the charge deposit on the anode plane. Distance perpendicular to the anode plane, which is the electron drift direction or depth into the detector, is determined from the charge's time of arrival. The electron drift speed must therefore be accurately known. Figure 3.6 shows this detection principle in action. In specific cases, the depth of a charge deposit can be deduced from the TPC signal alone, but in general an older charge deposit created at great depth can seem equivalent to a younger charge deposit created close to the anode plane. An absolute measurement of the time of interaction is needed to determine its precise depth, which can be accomplished via the detection of scintillation photons.

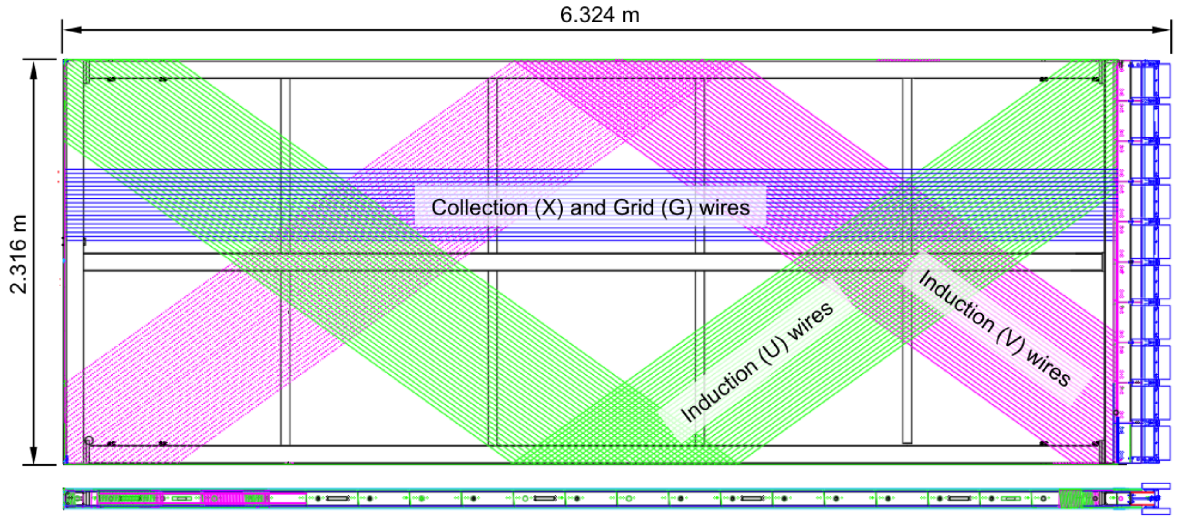

Figure 3.5: A front and side view of an APA. Some wires of each plane are shown in different colours. During normal operation, the APA hangs vertically from what is its right side in this figure. Taken from [39].

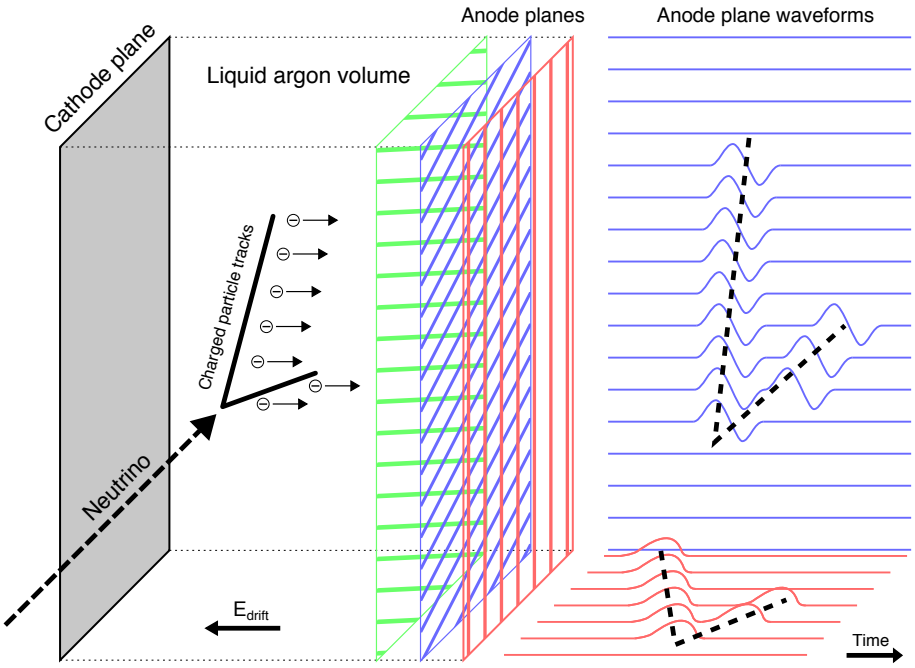

Figure 3.6: The workings of the (Proto)DUNE TPC. Charged particles originating from a neutrino interaction create an ionisation track, the electrons of which are attracted into the APAs. These record the electrons' position and time of arrival for a three-dimensional reconstruction of the event.

3.2.3 Photon detection system

The principle of a LArTPC provides excellent particle track resolution for a large detector volume. However, the absolute position of a detected structure is difficult to determine. This so-called t_0 is ambiguous in a large fraction of events because the arrival time of external particles and the electron travel time to the anode are both unknown. The structure of any connected track structure is therefore known, but its absolute position in the drift direction is not. Another aspect of liquid argon solves this problem: scintillation. Argon is an excellent scintillator and will produce a flash of 128 nm photons as energy is deposited: 20,000 photons are generated per 1 MeV of energy deposition [69]. Compared to the drift electrons, these scintillation photons travel almost instantaneously (photon group velocity $v_g \sim 0.38c$ [52]) and make it possible to fix the time of track creation.

Several photon detector systems are embedded in the detector's APAs, differing in their precise operation but operating under the same form factor. Each APA contains ten horizontally oriented bars distributed uniformly along its entire height. These bars consist for the largest part of waveguide and wavelength-shifting material. Silicon photomultipliers at one extreme of the bars collect the resulting light and send appropriate signals to readout systems outside of the detector. The setup is shown in figure 3.7.

The S-ARAPUCA design is an experimental photon detection system that was installed into the ProtoDUNE APAs alongside the standard systems, shown in figure 3.8. It differs in that it uses 24 photon collection windows per bar with highly reflective internal walls. Photons that enter the windows are trapped by shifting their wavelength and, possibly after many reflections, they are detected by SiPMs installed in the walls of the compartment [69]. Following the ProtoDUNE beam run, the X-ARAPUCA was adopted into the baseline single phase module of the DUNE far detector design. It is an evolution of the S-ARAPUCA system installed in ProtoDUNE [39].

3.2.4 Data acquisition and flow

Three data acquisition systems govern the flow of physics data from the detector front-end to storage servers: FELIX, RCE and SSP, all of which are explained in more detail in chapter 4. TPC data of five of the six ProtoDUNE anode planes are processed by

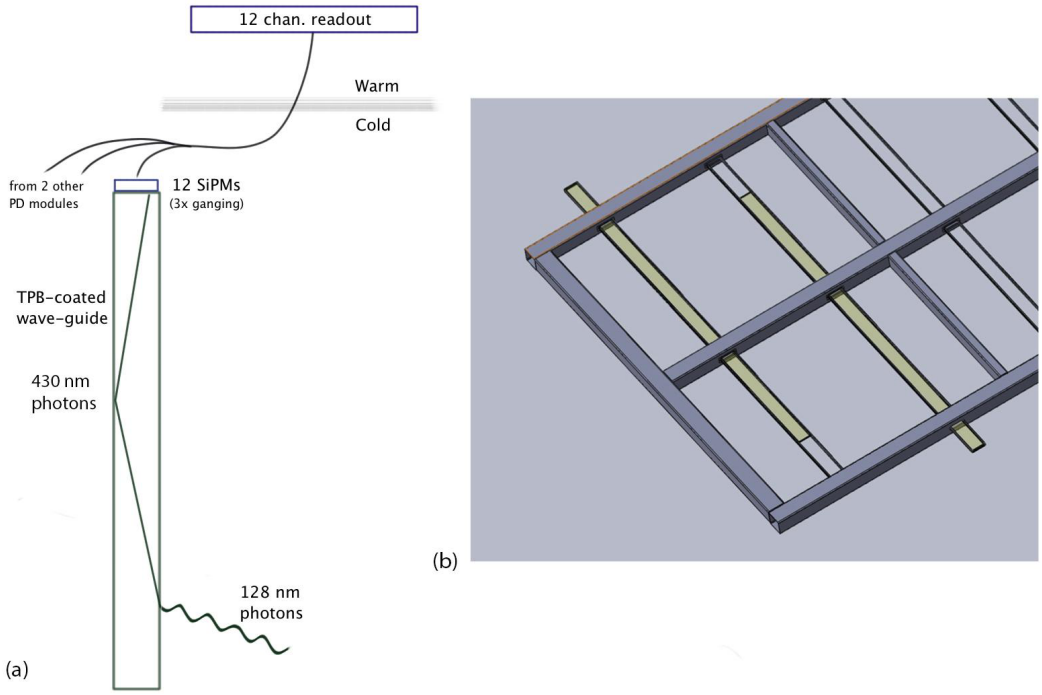

Figure 3.7: A schematic overview of a conventional photon detector bar (a) and the manner in which these fit into an APA (b). The APA is shown face-on without its TPC wires for clarity. The bars in (b) have been displaced to illustrate how they may be installed in the APA; during normal operation they are centred in the frame and are oriented horizontally along the ground. Taken from [69].

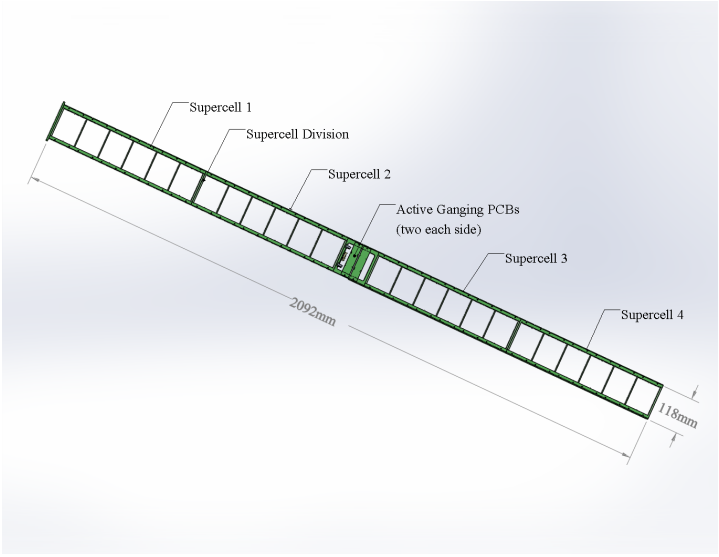

Figure 3.8: A single ARAPUCA module spans the width of an entire APA and consists of 24 cells, divided into four supercells. Taken from [39].

RCE while FELIX processes the remaining APA, being a data acquisition system under development. Each APA holds 2560 wires which are read out at 2 MHz, creating a data flow of $\sim$91 GB/s for the entire ProtoDUNE TPC. The data acquisition's purpose is to record the data from this stream that fall within 3 ms time windows coinciding with test beam particles or cosmic rays. Additional tasks include data compression and facilitating data access through overlays.

3.2.5 Test beam

ProtoDUNE received a charged particle beam from the Super Proton Synchrotron (SPS) between 10 October and 12 November of 2018. The aim of this beam was to simulate as accurately as possible the particles resulting from a neutrino interaction in the eventual DUNE far detector. The recorded data is also used for calibration and reconstruction testing purposes. Because DUNE will receive a broadband neutrino beam and the resulting particles from their interactions will produce an even wider energy spectrum, ProtoDUNE received a wide range of beam momenta itself, from 0.3 to 7 GeV/c.

Depending on the momentum of the beam, its particle composition changes. The beam consists mostly of electrons and positrons at the low end of the momentum range, whereas the higher momentum beams are dominated by pions. This change in composition but also particle rate is shown in figure 3.9 and can be explained by considering the beam creation process. Although a 7 GeV/c momentum beam is needed at most, the SPS delivers 400 GeV/c momentum protons to the test beam hall that ProtoDUNE is situated in. In order to attain the desired beam momentum and composition, the beam is first passed through a primary target (T4) in order to produce a secondary beam (H4). The particle momentum in the secondary beam is selected to be 80 GeV using dipole magnets. This secondary beam is then further attenuated by letting it pass through a secondary target, thereby creating a tertiary beam named H4-VLE, named for its very low energy nature. The secondary target was chosen to be tungsten for momenta of 0-3 GeV/c and copper for 4-7 GeV/c [14]. Dipole magnets are then used to redirect the portion of the beam with the right momentum into the detector's front face. A setup of the beam in the ProtoDUNE section of the hall can be seen in figure 3.10.

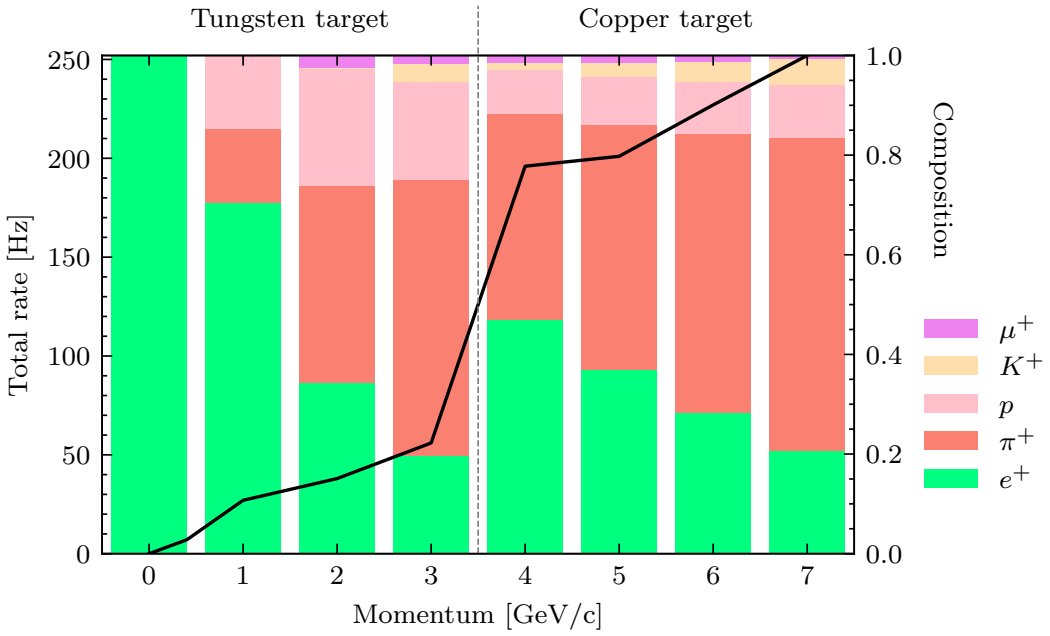

Figure 3.9: The total rate and composition of the H4-VLE beam as a function of selected beam momentum. The beam composition at 0.4 GeV/c is shown at 0 GeV/c for spacing. Data taken from [70].

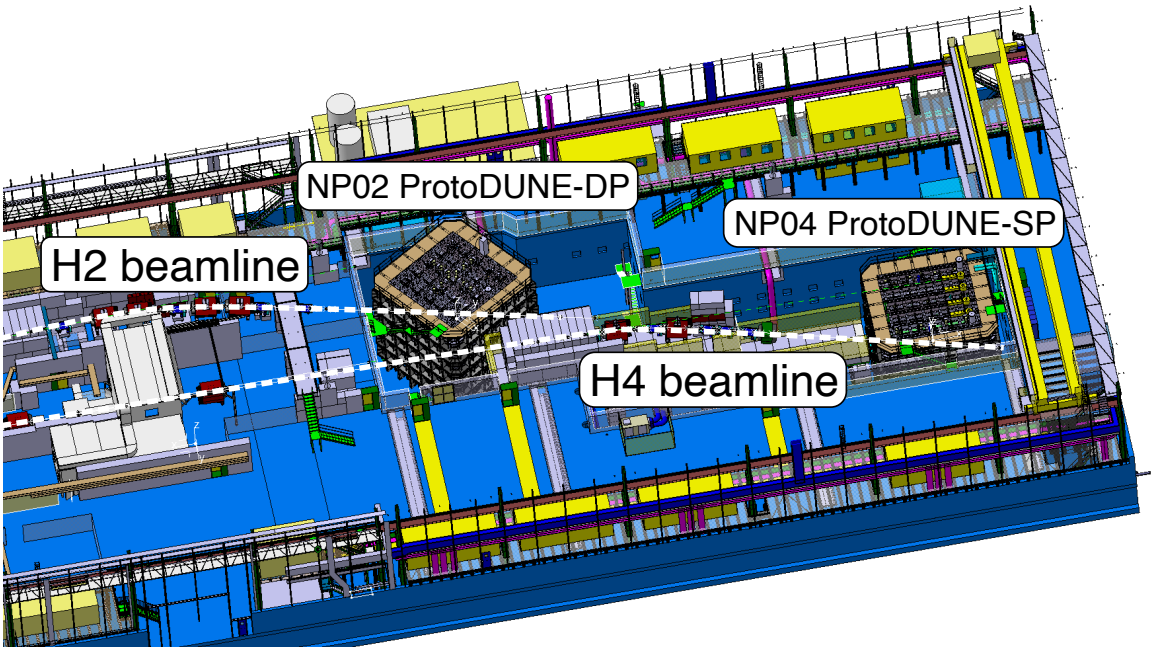

Figure 3.10: A schematic of the ProtoDUNE setup at CERN. The H2-VLE and H4-VLE beam come in from the left and enter ProtoDUNE-DP and ProtoDUNE-SP respectively. Taken from [71].

A close-up of ProtoDUNE-SP and its beam pipe can be seen in figure 3.11, as well as the coordinate system in use. The beamline is curved downward and into the side of the detector so that it enters between the cathode plane and the rightmost anode plane seen from the perspective of the beam. The origin of the coordinate system is located on the centre bottom of the TPC face that the beam enters into. In other words, the x-coordinate signifies the distance from the cathode plane, the y-coordinate the height from the bottom of the TPC and the z-coordinate the depth into the TPC, starting from the beam entrance into the cryostat.

Table 3.1: Coordinates of the active TPC volume within the cryostat taken from [72] and the position and angle of the beam entrance into the cryostat (J. Calcutt, personal communication, 1 September 2020). The beam angles are given in the XZ plane (θ_{xz}) and YZ plane (θ_{yz}) also shown in figure 3.11.

Parameter	Active TPC volume	Beam
x [cm]	-358 – 358	-5.0
y [cm]	0 – 600	448.5
z [cm]	0 – 695	-129.8
θ_{xz} [°]		-10.8
θ_{yz} [°]		11.4

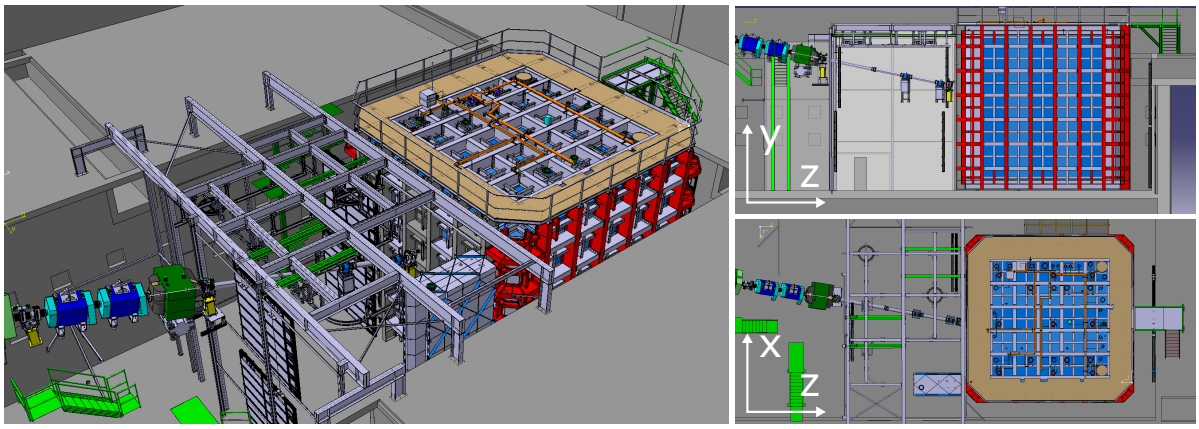

Figure 3.11: A close-up of the beam pipe as it approaches ProtoDUNE-SP, seen from the side (top right), top (bottom right) and in a three-dimensional view (left). Taken from [55].

As one of the main aims of ProtoDUNE is to characterise the detector response to certain charged particles, it is vital to understand the beam that it receives. To this

end, an array of beam instrumentation was set up to measure the beam's properties, pictured in figure 3.12. Aside from focusing quadrupole magnets (Q17-22), the H4-VLE contains two dipoles (B18, B19) that select the beam's momentum and redirect it. A third dipole (B20) was proposed to deflect the beam into different parts of the detector, but was ultimately cancelled. Beam profile monitors (BPROF1-4) measure the two-dimensional beam spread at various points in the beamline and also help accurately determine the beam momentum. Lastly, particle identification is carried out by time-of-flight detectors (TOF0-2) for relatively low velocity particles and Cherenkov detectors (CHER1, CHER2) for particles with a higher velocity. The latter can be modified with various gasses and pressures to change the Cherenkov threshold, adjusting to the beam's current momentum and particle composition [69].

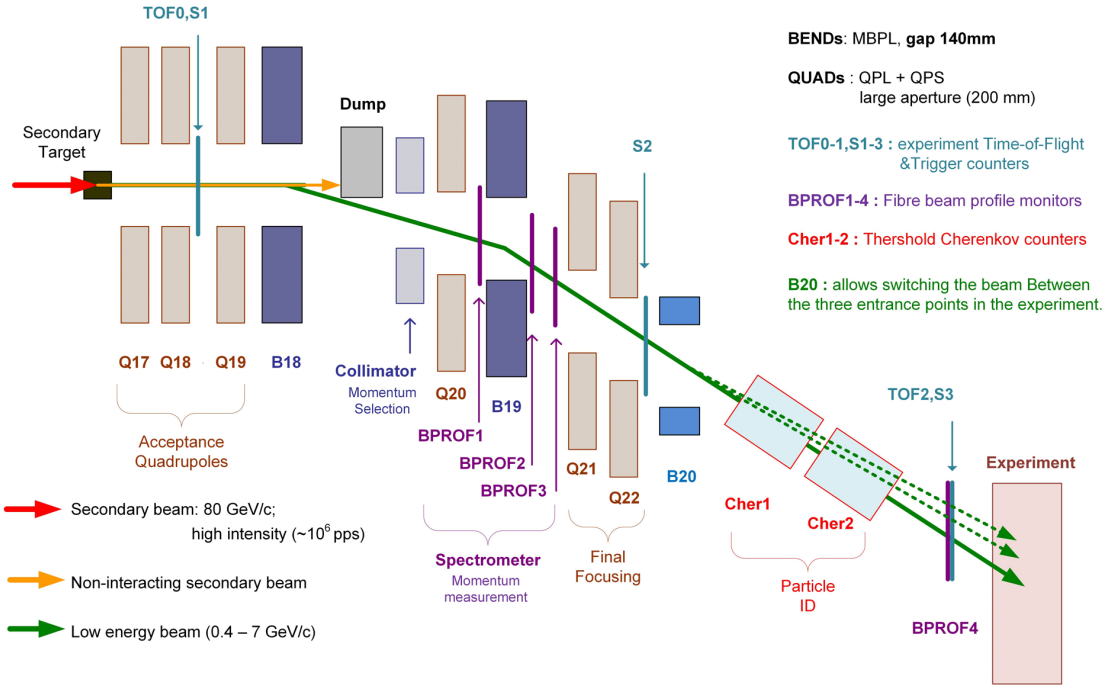

Figure 3.12: A schematic view of the H4-VLE beamline layout. Several magnetic elements and beam instrumentation detectors are visible. Taken from [70].

The Central Trigger Board (CTB) is responsible for sending trigger signals to ProtoDUNE's DAQ systems and does so based on information it receives from systems such as the photon detector, beamline monitors and SPS spill signal. A configurable trigger algorithm is then applied to all input channels for the CTB so that a general trigger can be sent out at the right time. In this manner, interesting physics events without too much background are selected out of a beam spill [70]. Physics runs may include

electron and hadron triggers, whereas runs that focus on detector noise or cosmic muons may employ a simple timed trigger, as these processes occur continuously.

3.3 Beam run

The ProtoDUNE-SP detector structure was completed in July of 2018. In a process that lasted until mid-September, its interior was purged of gaseous contaminants with argon gas and filled with liquid argon. As the liquid argon started circulating through the detector's purification system, the voltage on the cathode reached its nominal value of 180 kV for the first time on the 21st of September. At this time, the concentration of contaminants in the liquid argon made it impossible to detect beam particles, but the first track of a cosmic ray, passing close to one of the anode planes, was spotted that same day. Thanks to the steadily increasing electron lifetime, which is discussed in more detail in chapter 5, the first beam track was observed on the 2nd of October.

A summary of ProtoDUNE's beam run can be seen in figure 3.13. The three main requirements for stable beam data taking during this time were (1) an active charged particle beam, (2) sufficiently high liquid argon purity and (3) a homogeneous electric field at the nominal strength. The first was met simply through the SPS and test facility schedules. ProtoDUNE received three rounds of tertiary SPS beam, each lasting between ten and thirteen days. ProtoDUNE's test beam logistics are described in detail in section 3.2.5.

To attain a sufficiently high liquid argon purity, the material was circulated through the experiment's purification system which filtered its contaminants out over time. This requirement therefore required time above all else. The dip in purity seen in figure 3.13 was caused by a coincidence in purity monitor maintenance and an unscheduled shutdown of the liquid argon pumps. This effectively stopped the purification process and allowed a buildup of contaminants that went unnoticed until the purity monitors were in operation once more [74]. Fortunately, sufficient purity was regained by the time the test beam returned.

Attaining and maintaining an electric field of 500 V/cm proved to be a challenge throughout the ProtoDUNE beam run. The initial 300 kV high voltage power supply proved unreliable, as it caused large current and voltage fluctuations over time. After

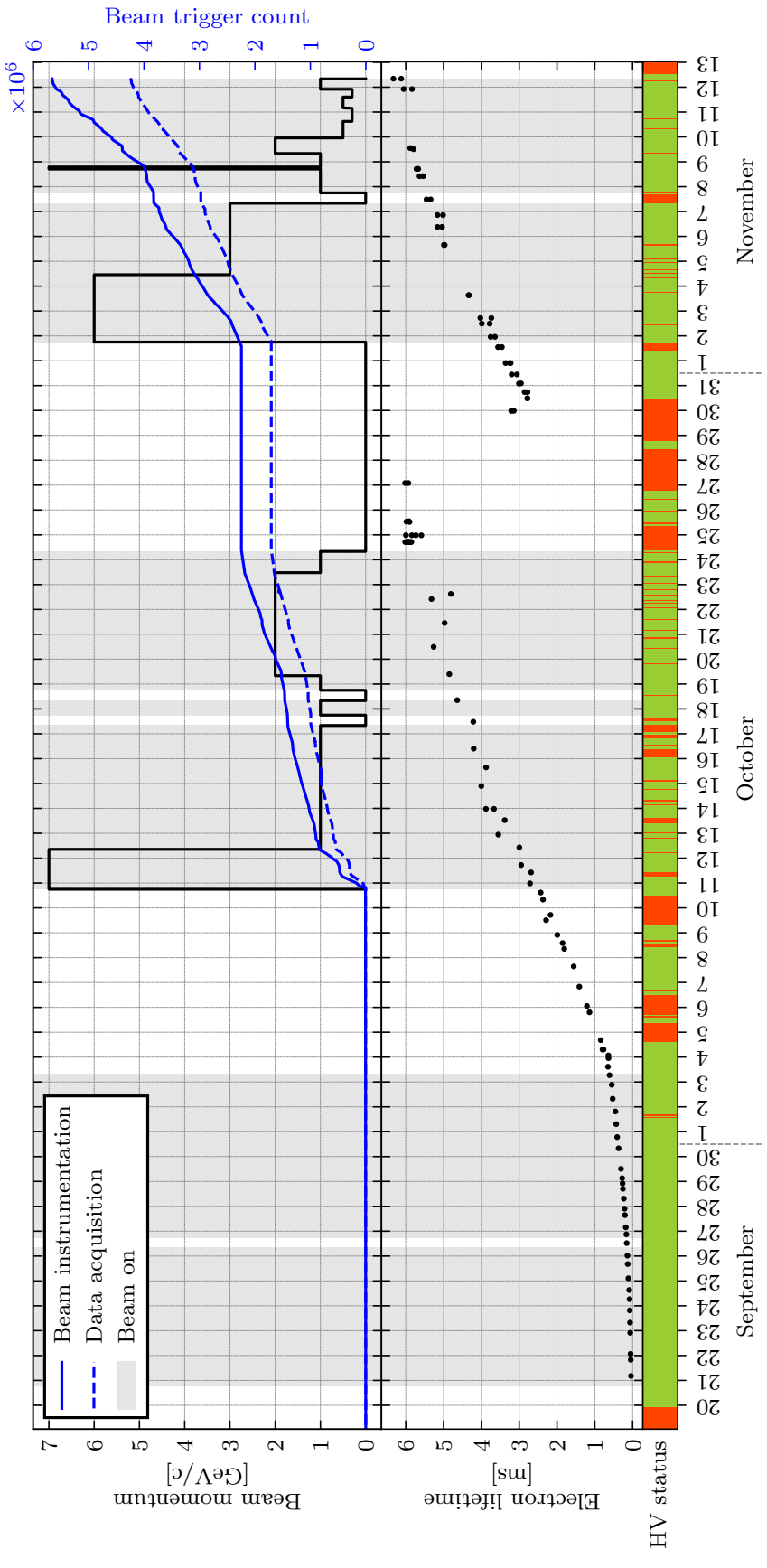

Figure 3.13: A summary of the ProtoDUNE-SP beam run in late 2018. Shown are the nominal beam momentum (top, black), beam trigger count (top, blue), electron lifetime (middle) and high voltage status (bottom) over time. For clarity, only high voltage disruptions exceeding 10 minutes are shown. Adapted from [73].

replacing it with another 200 kV power supply, these issues largely disappeared. Other challenges arose in the context of electronics noise, when high voltage noise filters proved to be susceptible to leaking and short-circuiting. An alternative solution was installed to resolve problems in this area. Within the TPC, the strong electric field created irregularities in the high voltage that, thus far, do not have a conclusive explanation. It is theorised that charge builds up within the detector and eventually creates a path for the high voltage to flow to ground. This causes a drop in the cathode voltage and an elevation in current, visualised in figure 3.14. These long-term anomalies were called streamers and were generally solved by lowering the high voltage for around a minute and then ramping it up again to its nominal value. This was first done manually and later by an automated script that could detect streamers, pause data taking, resolve the instability and resume recording.

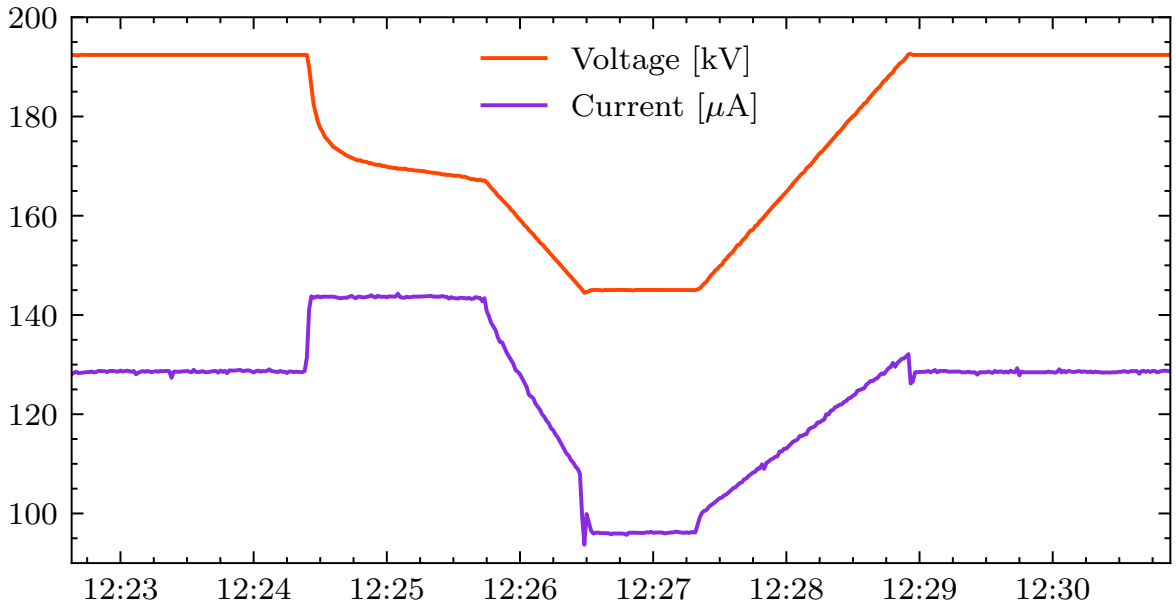

Figure 3.14: The effects of a so-called streamer on the ProtoDUNE cathode high voltage and current, recorded on the 8th of November 2018. Shortly after the streamer starts, the voltage on the cathode is dropped and then ramped up again to its nominal value. Note that the power supply voltage is shown here, being slightly larger in magnitude than the effective voltage on the cathode. Data taken from [74] and averaged per second.

Data taking began in earnest on the 10th of October and lasted until the 12th of November, when the SPS was shut down for a period of upgrades to the LHC accelerator complex. In total, over 4 million test beam events were recorded. Aside from this, cosmic ray events were recorded outside of test beam periods. The estimated

number of test beam events split by particle type can be seen in table 3.2. In some momentum ranges, particularly below 1 GeV/c, data volumes exceeded expectations [75].

Table 3.2: Estimated number of test beam triggers by particle type, obtained from the total trigger count per beam momentum setting and the accompanying beam simulation. Data taken from [75].

Momentum	Number of recorded events $\cdot 10^{-3}$				
[GeV/c]	e^+	π^+	p	K^+	Total
0.3	242	0	0	0	269
0.5	296	1.5	1.5	0	340
1	262	382	420	0	1089
2	173	333	128	5	728
3	113	284	107	15	568
6	197	394	70	28	702
7	98	299	51	24	477
Total	1384	1694	779	73	4175

Chapter 4

Data access and compression in the ProtoDUNE DAQ

In 2015, the ATLAS detector at the LHC detected an unexpected signal with a significance of 2.5 σ [76], leading physicists to propose exciting new physics models in order to explain the excess. At the time, the data pointed to the existence of an unknown particle with a mass around 2 TeV. However, after more data was collected and processed, the signal disappeared, ultimately having been nothing more than a statistical fluctuation. The situation is a reminder that accurate measurements require careful consideration and a large enough data set. Even the enormous data flow of the ATLAS detector can produce significant statistical fluctuations leading to incorrect conclusions. In order to obtain a large enough data sample to draw accurate conclusions, detectors in particle physics must either operate for an extended time or have a large data flow. Since the operation of large-scale experiments tends to be costly, experiments where possible opt for a setup that allows them a high rate of physics interactions. These setups necessarily require sophisticated data acquisition (DAQ) systems.

In a general DAQ system, the front-end of a detector collects the raw data from physical signals. These are the jumps in current that occur when a straw tube is triggered, for example. The data may be digitised immediately and are generally aggregated before triggers are applied. Because modern experiments record vast numbers of processes, triggers are needed to select those that are of particular interest. These triggers may, for example, apply a certain time window in which a collision is known to occur and reject all events outside this time window, or it may record all events in which certain particle signatures are produced. It may also apply more complex patterns that are required for an event to be recorded. In the end, experi-

ments such as ATLAS only keep a fraction of their collected data. After triggers have been applied and potential other data manipulation such as zero suppression and compression has been performed, the various sub-detectors observing an event must have their data pooled in so-called event builders, which collect all detector data and store it on a by-event basis.

Data flows in the ProtoDUNE-SP detector are not expected to be of the same volume as other large CERN experiments, but are nonetheless nontrivial to process. A compression step is therefore included in the DAQ chain to conserve storage space. As an extra challenge, the DUNE collaboration has opted to save all detector data unaltered, decreasing the effectiveness of compression algorithms greatly. The compression ratio is of great importance for the eventual DUNE data infrastructure and storage, and must therefore be made as high as possible. At the same time, compression must be performed online and so compression algorithms should not take so long that they fall behind during data taking.

4.1 Overview

As introduced in chapter 3, the ProtoDUNE TPC contains six APAs, five of which are read out by the RCE system [66] and the sixth is read out by FELIX [67], being a prototype DAQ system. The 20 front-end motherboards (FEMBs) of a single APA process and output 2560 channels of data with a frequency of 2 MHz, where each channel corresponds to an APA wire. Four FEMBs then feed into one warm interface board (WIB), so called because it is the first component outside the experiment's cryostat to receive ProtoDUNE data. It multiplexes the data into two output fibres in the case of FELIX and four fibres in the case of RCE. Because five WIBs process the data of one APA, FELIX receives ten of these 9.6 Gbps fibres in total, each of which contains data corresponding to 256 channels. This entire chain reads and processes the data of an entire APA at a rate of 2 MHz. In total, a single FELIX card processes 2560 channels, corresponding to roughly 96 Gbps [69]. The acronyms that are most commonly used throughout this chapter are listed in table 4.1 for convenience. A summary of the entire ProtoDUNE data flow can be seen in figure 4.1.

As the DAQ systems of RCE and FELIX receive the 2 MHz stream of time slices, they wait for a trigger signal to arrive. As it does, they select a 5 ms-long stretch of data

Table 4.1: A selection of acronyms that describe components around the ProtoDUNE DAQ.

Acronym	Definition
TPC	Time Projection Chamber
APA	Anode Plane Assembly
ADC	Analog-to-Digital Converter
ASIC	Application-Specific Integrated Circuit
COLDATA	COLd DAta Transmission ASIC
FEMB	Front-End MotherBoard
WIB	Warm Interface Board
RCE	Reconfigurable Cluster Element
COB	Cluster On Board
FELIX	Front-End LInk eXchange
BR	Board Reader

from the stream and pass that on to Board Reader (BR) processes. The 5 ms in which data is saved and passed on is called the trigger window, and it corresponds to 10,000 time slices. Because the WIB groups the data into so-called frames of 256 channel values before sending them to RCE and FELIX, these DAQ systems output so-called fragments of 10,000 time slices containing the data of 256 channels. Since an APA contains 2560 wires, 10 fragments are generated per APA per trigger. After optionally applying reformatting and compression operations to the data, BR processes send it to event builders, which combine data from the various ProtoDUNE subsystems into so-called events. Table 4.2 summarises the details of the data stream through the front-end and DAQ's various elements.

The studies shown in this chapter were performed before the operation of the ProtoDUNE detector in late 2018. Initially, a 5 ms readout window of 10,000 time-slices was considered. However, it was later decided that a 3 ms readout window would be sufficient to contain all relevant physics in a beam interaction, greatly reducing the data volume in the process. This length of time corresponds to 6000 time-slices instead. Although the basic processes remain identical between the two cases, the performance of these processes, especially compression, is likely to differ.

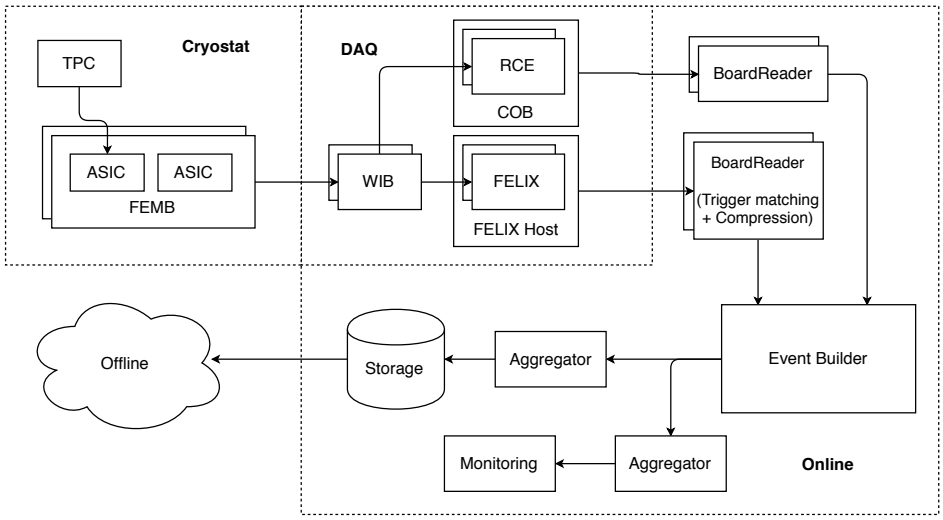

Figure 4.1: A simplified version of the ProtoDUNE DAQ chain.

4.1.1 FELIX

FELIX is based on the concept of using as little custom hardware as possible, focusing its efforts on a single PCIe card and relying on commercial servers and networking solutions in the host machine. Many operations such as trigger matching and compression are therefore performed in software or third party hardware. An advantage of such an approach is that the system can rely on rapid technological advances in commercial computing and networking, making for a system that is highly flexible and upgradeable. The system was developed primarily for use in ATLAS at CERN and was adapted for use in ProtoDUNE [67].

This chapter expands on the ProtoDUNE-SP FELIX team's efforts to meet the stringent compression requirements. This includes the development of software to read the WIB data format, as well as its manipulation and various compression schemes. Finally, the performance results of various compression schemes are shown.

4.2 Frame generation and access

At the time of development, the ProtoDUNE-SP detector was not yet operational and thus provided no data on which compression studies could be performed. However,

Table 4.2: Details of the FELIX data flows.

Element	Quantity	Channels
APA	1	2560
FEMB	20	128
WIB	5	512 (4 · FEMB)
Link	10	256 (1/2 · WIB)
BR process	10	256
Frame	2 MHz	256
Fragment	25 Hz	256 (· 10,000 time slices)

data was available from physics and detector simulations. These are based on models for liquid argon TPC operation and were refined by experience from previously and currently operating experiments such as MicroBooNE [34]. Because the LArSoft software [77] that runs these simulations outputs data in a different format from the ProtoDUNE DAQ chain, some conversion is necessary to accurately simulate incoming frames in a running experiment. For this reason, a minimal C++-based framework was written: one of its functions is to convert simulated detector values into the exact format that the warm interface board (WIB) would send to the ProtoDUNE DAQ. More importantly, it features high-performance data access and manipulation capabilities, and is for this reason in use in many components of the DAQ chain.

4.2.1 Frame and fragment format

The front-end of the ProtoDUNE detector records time slices of the current on its wires at a rate of 2 MHz. 256 of these current values are grouped together by the WIB and sent to the FELIX system in a so-called frame. The frame format, summarised in figure 4.2, contains some notable features. A word length of 32 bits is maintained, and a frame is 120 words long, corresponding to 480 bytes. In practice, the FELIX firmware trims words 0 and 117-119 off the frame before passing it on to the host machine, leaving a 116 word (464 byte) frame exposed to the receiving software. Important identifiers can be found in the frame header. The slot, crate and fibre number indicate the WIB, APA and WIB-FELIX fibre respectively, and a timestamp anchors the frame's

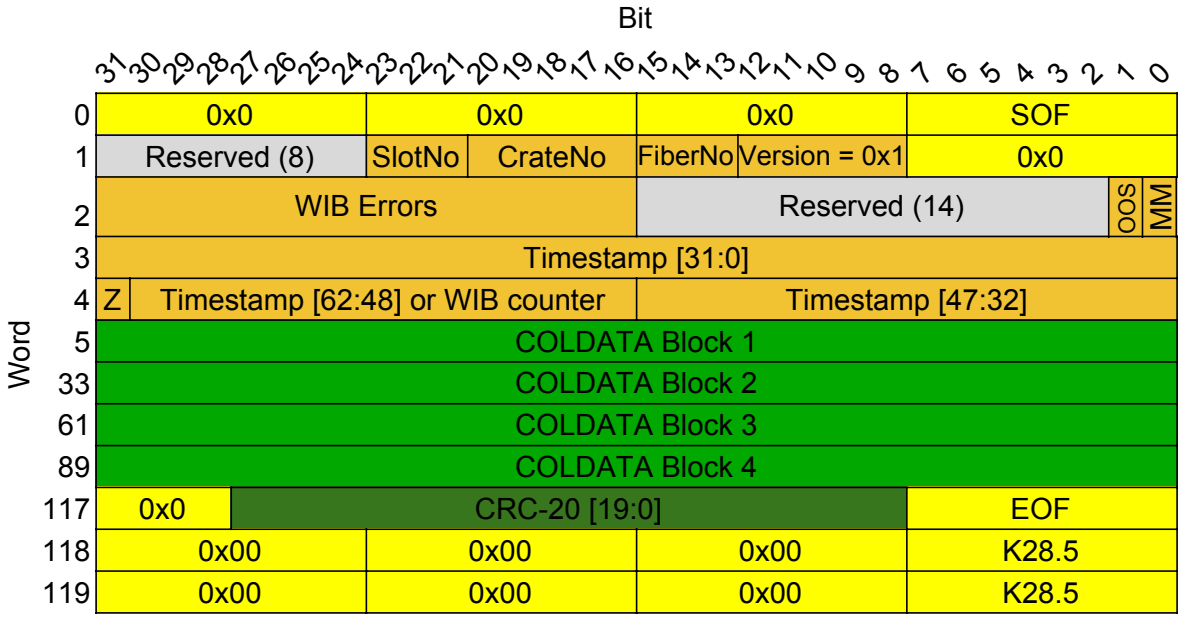

Figure 4.2: The data format going from the WIBs to FELIX. The WIB frame header contains a format version number, identifier fields (fibre, crate and slot numbers) to pinpoint the source of the data, error fields (mismatch, out of scope) and a timestamp. A Z-bit enables switching between timestamp mode and WIB counter mode. Each WIB frame contains four COLDATA blocks that hold the ADC values and a CRC (Cyclic Redundancy Check) in the trailer. Format taken from [78].

data in time. This timestamp is either 48 or 63 bits long depending on the frame's settings and was set to 63 bits for all data recorded in ProtoDUNE.

A frame's wire data are contained in four 28-word COLDATA blocks, the format of which is shown in figure 4.3. It can be seen that each block features its own header to communicate checksums and errors. The data contain 12-bit channel values sorted by their corresponding ADC. These 12-bit fields are not consecutively stored in the frame: successive multiplexing in the FEMB and WIB cause a pattern that requires multiple operations to read. In total, each COLDATA block contains 64 fields that hold raw current values from 64 of the detector's wires in one time slice.

4.3 Data compression

When deciding on the degree of compression for an experiment, a balance must be found between performance and storage requirements. On the one hand, high-

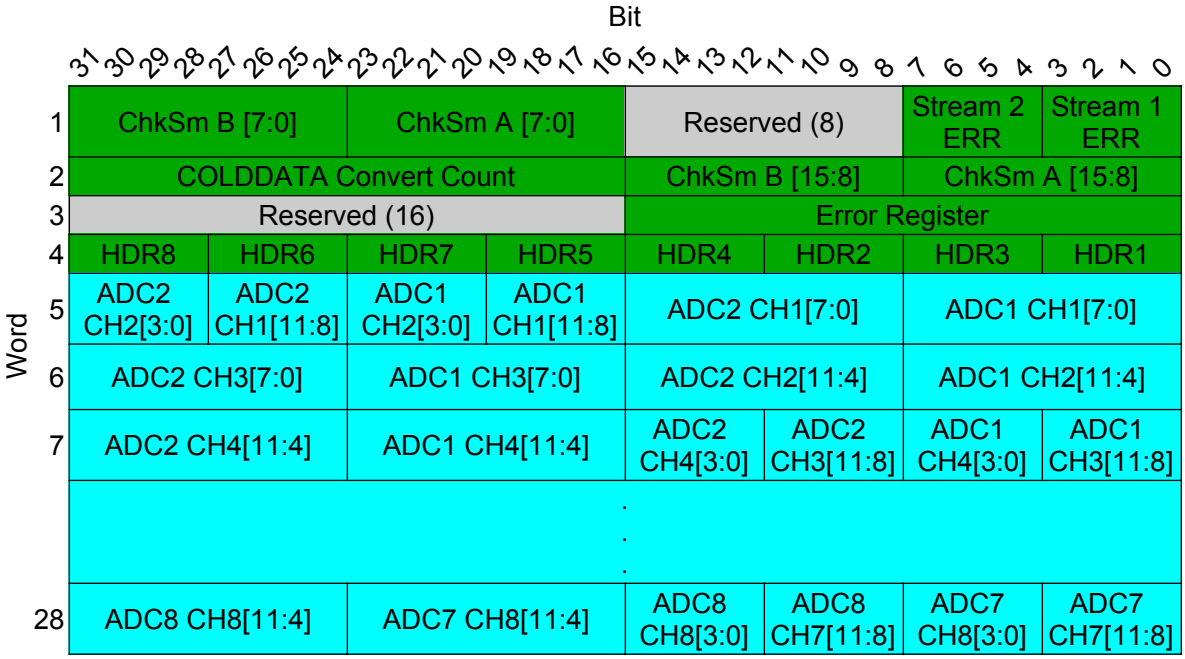

Figure 4.3: A COLDATA block as it is found in FELIX and RCE frames. Its header contains error fields and checksums. The pattern in the first three words of the ADC channel values (words 5, 6 and 7) repeats through the rest of the block. Format taken from [78].

energy physics experiments generate enormous amounts of data. A large compression ratio relaxes requirements on the experiment's networking and storage capabilities. However, it also requires sophisticated compression algorithms to be part of the DAQ chain, placing pressure on components vital to data-taking. Since ProtoDUNE data is expected to be sparse, a compression factor of 4 was estimated. With an expected trigger rate of 25 Hz, the maximal compression time is 40 ms, although in practice it needs to be much shorter to ensure that outliers do not disrupt the rest of the DAQ [69]. The fragment, containing a group of 256 channel values for 10,000 time slices, is the target of compression in this chapter.

The essence of lossless compression is to remove as much redundant information as possible. In this context, an encoder works to reduce the data size and a decoder reverts the compressed data to its original format. There are multiple ways of encoding data in order to decrease its size. If a certain value is repeated many times, for instance, it can be represented by a shorter bit string that the decoder can translate back into the original value. The Huffman compression algorithm uses this principle. Another method replaces repeated bit patterns with a pointer to the first occurrence. Because

this type of algorithm finds repeated bit strings by comparing patterns in a window that moves with the search pointer, the method is called sliding window compression. An algorithm that makes use of this method is LZ77. DEFLATE is one of the most popular compression methods today and it uses Huffman coding in conjunction with sliding window compression.

4.3.1 Software compression studies

Both the reading and compression of the frame format shown in figures 4.2 and 4.3 require the manipulation of individual bits. Two methods were attempted for this purpose: the bit shift method and the bit field method. The bit shift method works by accessing the byte or word on which the desired variable is located, bit-shifting the data container so that the desired value is positioned at the least significant bit and finally applying a mask so that the remaining bits in the container are set to zero. Random bit range access was implemented in this manner.

Random bit range access may also be implemented by making use of bit fields native to C. A struct or class may contain variables that occupy a specifically designated number of bits in memory. This allows the counter-intuitive frame format to be implemented in an object in a straightforward manner. Casting raw data into this object immediately grants access to all relevant fields. Because these structures are 'overlaid' on data in memory, they are often called overlays. Aside from the ease of use, bit fields in this case are also significantly faster than the previously described bit shift method. Extensive tests were performed in which all fields in tens of thousands of frames were accessed. The bit field method performed roughly twice as fast as the bit shift method, making it the method of choice. This dramatic difference in speed is most likely due to bit fields providing more direct access to data: bit shifting requires copying, altering and combining values before the same result is reached.

The main FELIX format structures are contained within the FNAL dune-raw-data repository [79] in FelixFormat.hh as part of the overlay implementation. A dune::FelixFrame class contains a description of the frame format received from the WIB in C bit field style. The overlay for FELIX consists of an object called a dune::FelixFragment (named after the artdaq::Fragment it reads) that casts a raw data buffer into the FelixFrame structure and provides read-only access to its data fields. The data buffer is expected to contain 10,000 frames in normal operation, but this

number is fully configurable according to information received from the Fragment metadata. The metadata also specifies the format of the data, as well as whether it has been compressed. The FelixFragment evaluates this metadata and dynamically selects the appropriate methods to perform all required operations. The end result is a system that is highly configurable and requires no extra effort on the side of the data reader.

Some operations are inefficient when applied to the default frame format shown in figure 4.2. The project therefore includes several data reordering classes, which cast the data into a more accessible format. This format is most useful as input for general-purpose compression algorithms and is expanded upon in section 4.3.2.

The FELIX overlay project in dune-raw-data contains the following classes and structs:

- A dune::FelixFrame class that describes the structure of a single frame. It consists of a WIB header and four COLDATA blocks, which are described by the following structs:
  - A WIBHeader struct that contains the raw WIB header data. As most fields in the WIB header are undivided, they can be accessed for reading and writing directly. The only exception is the timestamp, which spans multiple words.
  - A ColdataHeader struct structured in much the same way as the WIB header.
  - Eight ColdataSegment structs are contained within one ColdataBlock struct. The 64 ADC values in a block have a complex but repeating pattern consisting of eight ADC values corresponding to two halves of two streams. This repeated structure is described by a COLDATA segment. It implements data read and write access through a combination of bit fields and minimal binary operations.
  - Four ColdataBlock structs are contained within one FelixFrame. Every such struct contains its own ColdataHeader and eight ColdataSegments. As all basic data access functionality is contained within its substructures, a ColdataBlock merely passes this functionality on and incorporates some utility functions for printing data.
- A dune::FelixReorderer class to reorder a set of frames and produce a new artdaq::Fragment from their internal data. The two main functions of this class are to place the ADC values in 16-bit containers instead of 12-bit ones and to

order the ADC data such that every channel's waveform is stored contiguously. These facilitate ease of access and more effective compression. The function dune::FelixReorder takes a pointer to a set of frames and uses it to produce an artdaq::Fragment with the reordered structure. Internally, it makes use of the dune::FelixReorderer class.

- A dune::ReorderedFelixFrames class to hold an artdaq::Fragment's worth of reordered WIB frames. It contains an array of all of the original Fragment's ADC values in a 16 bit format. This is followed by the headers of the first frame, which serve as a template for all other headers. If unexpected header values in subsequent frames appear, a bit is set and the unexpected headers are recorded as well.
- A dune::FelixFragment class that takes an artdaq::Fragment as a constant reference and provides read-only access to it. The sole goal of this class is to provide data access that is as complete and fast as possible. Internally, it inherits from dune::FelixFragmentBase and contains a member class that is either dune::FelixFragmentUnordered for a set of unchanged WIB frames or dune::FelixFragmentReordered for a buffer containing reordered data.

ADC value reduction

The bulk of meaningful data that flows into FELIX is comprised of ADC values, making it the most attractive part of the data to compress. They take up 384 bytes per frame, which is over 80% of the entire frame volume, and are most likely to contain values that change from one time slice to the next. In addition, ADC values are only expected to change very slightly from one time slice to the next, especially since most of the data is expected to be noise. In spite of this, all values are recorded in 12-bit fields no matter the variation. There is thus much to gain. Before compression is applied, two so-called reduction methods may be applied to the ADC data.

Pedestal subtraction is one of the simplest and most obvious methods to reduce ADC values to a more compact form. In the ideal case, a channel has a fixed pedestal value that it varies around randomly over time. Most of the time, this variation is expected to be small: large signals from passing particles and high noise amplitudes should be a relative rarity. The 12 bits of space assigned to each value are thus not all

necessary the vast majority of the time. Exactly how much space would be needed instead to account for noise recordings depends heavily on the noise amplitude of the detector, but the general tactic involves determining the pedestal value for a certain channel and subtracting it from any values that enter on that channel. The result is a relatively small range of values centred around zero, with both positive and negative values present. High-value outliers must be collected separately as they would exceed the new range, which has shifted and shrunk.

The main advantages of pedestal subtraction are that it is easy to implement and effective for simple data sets. All that is required is a simple subtraction for every recorded value and as a result large value fields can be decreased in size considerably. Even with just this tactic, 12-bit fields ranging in value from 0 to 4096 can be reduced to 6-bit fields ranging in value from -32 to 31, for example. Aside from the reduced channel values, the pedestal for every channel must be recorded, as well as the outliers and their locations in the data. Because the space requirements for these outliers are naturally higher than that of the original values, the new field size must be chosen such that outliers are few.

Despite its simplicity, the disadvantages of pedestal subtraction are substantial. Firstly, it depends on a constant or at least well-known pedestal. Even if the determined pedestal is effective for a channel initially, there is a chance that the actual pedestal will start to drift, causing the number of outliers to rise, greatly reducing the effectiveness of the method. Additionally, the method requires fine-tuning when actual detector data arrives, as its optimal field size depends heavily on the noise amplitude of the detector. Finally, the method does more harm than good in combination with certain compression methods such as Huffman compression. It makes reading channel values more complex while Huffman compression is not benefitted by it.

Waveform continuity or previous value subtraction (PVS) is a data reduction method that is similar to pedestal subtraction but does away with many of its inflexibilities. As the name suggests, it makes use of waveform continuity, or the principle that a channel does not change very much from one time slice to the next. It relies on recording the difference between one channel value and the next instead of the difference between channel values and a static pedestal. This is fundamentally different: pedestal subtraction is merely a shift of channel values whereas this method is more akin to a time derivative. The reduced data can again be stored in smaller fields complete with outliers, or the values can be used in subsequent compression algorithms as they are.

The principle of waveform continuity does not depend on a pedestal, static or otherwise, and thereby bypasses the problems of determining this value and the possibility of drifting away from it over time. As can be seen in figure 4.4, the reduced data also tends much more towards zero than the original values or a general pedestal subtraction and is therefore better suited for compression. However, reading out a channel value at a random time slice is very time costly, as the decoder needs to fetch the starting value of the channel and then go through the array of values over time while processing all of the recorded differences until the requested value is reached. It is not expected that random access is a requirement, however, and outputting entire waveforms (all time slices for a single channel) is approximately as fast as in the case of pedestal subtraction.

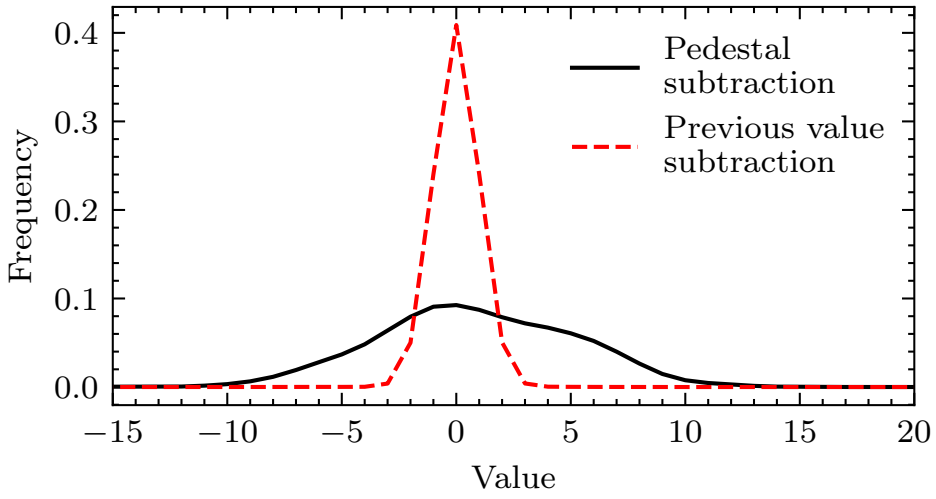

Figure 4.4: A comparison between pedestal subtracted ADC values and previous value subtracted values. The two figures were generated from the same data set using 9600 MicroBooNE frames containing 2,457,600 individual channel values.

ADC value compression

The actual compression of ADC values involves making use of repeated patterns within the data. If a single value or a certain sequence occurs more than once, it can be replaced with a shorter key or reference to a previous occurrence. In a perfectly compressed data sample, all data is unique. The following two compression methods deserve special attention.

Huffman compression [80] is one of the oldest forms of data compression and works well in data sets that are biased toward a small set of values. It relies on the concept that the most frequent values can be represented by a bit string, or code, shorter than their original length whereas uncommon values are assigned a code longer than their original length. Data sets that have a non-uniform value distribution are thereby reduced in size.

Huffman compression determines the code it assigns to each value based on the value's frequency in the data set. This can be determined exactly by counting value occurrences throughout the entire sample or in a representative sample. The resulting frequency table can then be made into a so-called Huffman tree, of which an example is shown in figure 4.5.

In order to assign new codes to values, the frequency of all values first needs to be determined, either over all of the data or a representative sample. The resulting frequency table can then be made into a so-called Huffman tree. This tree structures values in such a way that the higher frequency data is closer to the root than lower frequency values. The construction of a Huffman tree relies on a single simple principle: group together the two lowest-frequency letters or nodes and repeat until no unpaired letters or nodes remain. Ultimately, every node is a parent to two elements, both of which can be a value or a node of their own. Starting at the root and going down the branches, taking a right turn adds a '1' bit to a code whereas a left turn adds a '0' bit. Arriving at a value, the accumulated code is assigned to it and stored in a table. For example, the value 'F' in the Huffman tree of figure 4.5 has an associated code of '10110,' whereas the value 'I' is associated with the much shorter code '01,' owing to its higher frequency in the data. The resulting value-to-code map can then be used to encode a data buffer into a compressed form. Defining 'right' as '1' and 'left' as '0' is an arbitrary choice, as any two child elements may be swapped freely. Additionally, a free choice exists in clustering when more than two elements have identical frequency. The Huffman tree is thus not uniquely defined by its frequency table.

Whereas the encoder uses the Huffman tree to create an encoding map, the decoder instead makes direct use of the tree: as it reads through the bits of the compressed data buffer, it walks down the tree from its root. When it encounters a value, it is recorded and the decoder reverts back to the root. Because values have no child elements, no value's code starts with another value's code. No fixed length codes are thus needed, nor are separation characters necessary. One downside of this method is its inability to handle corruption of the compressed data. Be it mutation, addition or omission of a

single bit, the decoder is entirely unable to determine where the error occurred, nor is it always immediately clear that it occurred in the first place.

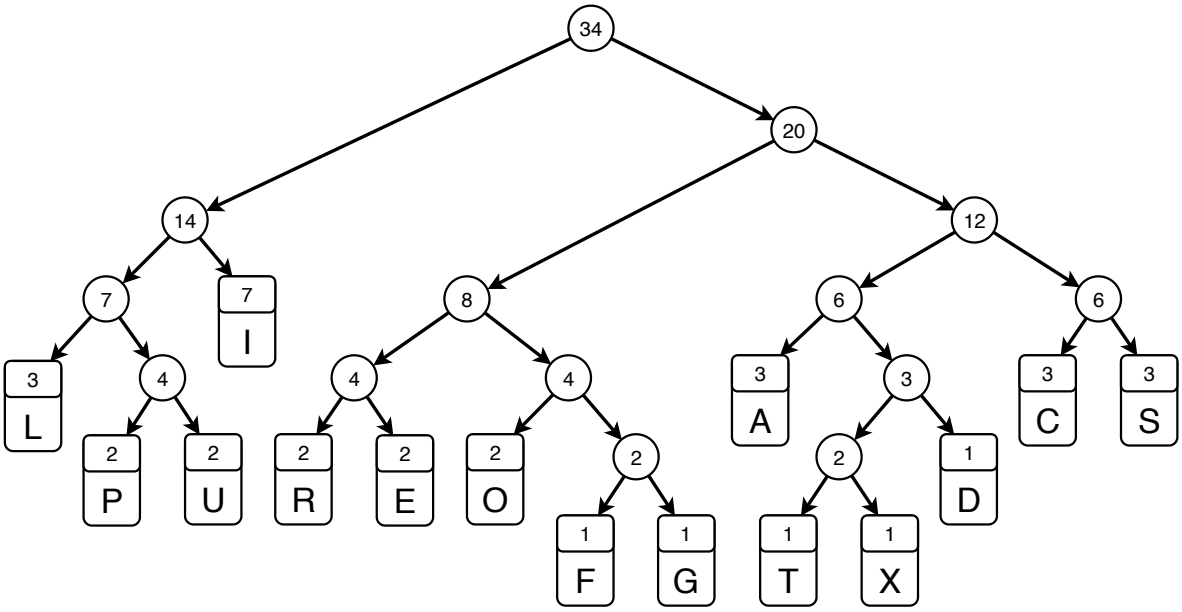

Figure 4.5: An example of a Huffman tree constructed out of the word 'supercalifragilisticexpialidocious', with the root at the top of the figure. The individual letters are shown together with their frequency in boxes. Nodes are denoted as circles and contain the sum of all letter frequencies below it. The tree is not unique as any two letters or nodes with the same frequency can be interchanged freely.

Huffman encoding algorithms can use a static tree which is shared between the encoder and decoder only once or a dynamic tree which is sent along with every encoded packet of data. Static trees allow for simplicity of encoding and decoding, but have the danger of becoming outdated in the case that value ranges shift or the frequency distribution changes. Dynamic trees grant great flexibility by constructing Huffman trees based on the exact data it is applied to, but its necessary inclusion with the data introduces additional overhead in computing time and compressed package size. The choice between dynamic and static trees depends entirely on the nature of the data. In the case of ProtoDUNE, the noise profile and event rate tend to change over time and a consistent compression factor is required, making dynamic Huffman trees a better choice thanks to their flexibility.

Figure 4.4 shows why general pedestal subtraction (or indeed any static offset) is useless in combination with Huffman compression. The actual values do not matter: it is the frequency distribution that determines the encoding. All values in the table could

have a constant pedestal added to them and nothing would change in the compressed data buffer. Waveform continuity, however, could complement this compression method effectively if the pedestal drifts or value changes remain small.

Huffman coding provides the optimal codes for any group of values. However, that is not to say that it is the most effective compression algorithm. One of the bigger limitations of the method is that it requires each code assigned to a value to have an integer number of bits. It is only in the case of an ideal frequency distribution that the overhead from this limitation is zero. Any deviation from this distribution results in an overhead up to 1 bit per code [80]. This overhead becomes vanishingly small if all codes are large, but in the current application it is expected that a small group of values around the noise pedestal are in the majority and could therefore obtain very short codes.

Sliding window methods [81] rely on repeated patterns rather than repeated values. As new data come into view, they are compared to the patterns already recorded. Matches are then replaced with a reference to their first occurrence.

There are many advantages to this compression method. For instance, compression can be performed in one pass: no prior indexing of values is required before compression can take place. Methods like these are universal, meaning that no prior knowledge of the data structure is needed, as is the case with Huffman compression. The data is instead considered as a bit stream featuring repeated patterns. In this manner, even data previously compressed with Huffman compression can be further compressed by sliding window algorithms. This is exactly what the DEFLATE algorithm does.

Header reduction

Omission of fields, difference transmission. Headers take up substantially less space in the data than the ADC values. Regardless, every 120-word frame has its four-word frame header and four four-word COLDATA headers. Fortunately, the values in header fields, insofar as they contain meaningful data at all, are often predictable if not entirely constant. The implementation of header reduction is expanded upon in section 4.3.2.

Every frame header contains a format version number, identifier fields, error fields and a timestamp. The version number and identifier fields are expected to remain

constant within a fragment and therefore only need to be recorded once. As the fragment contains 10,000 time slices of 256 channels, 10,000 different timestamps are expected. Because these increase at a constant rate from one frame to the next, only the very first timestamp needs to be recorded. In case of nonzero error fields or any deviation from expected patterns in the other header fields, the entire header is saved and the frame number is noted.

COLDATA block headers similarly contain many fields that are either empty or predictable. There are many error fields that need to be handled but do not provide information that is useful beyond FELIX. Additionally, the individual stream checksums recorded in these headers no can no longer be used, as some of the data used to construct it is no longer there. As with the WIB header, there is a field that increases monotonically over time. This is called the COLDATA convert count and normally increases by 1 from one frame to the next. As with the frame header's timestamp, if no errors occur, only the first count needs to be recorded.

4.3.2 Hardware-accelerated compression

Intel® QuickAssist Technology (QAT) was developed to accommodate the need for ever-growing cryptography and compression demands in the networking industry. It allows users to offload these processes onto dedicated hardware so that the CPU can be freed up for other tasks. Not only is this a very fast method of compression, it is also the perfect showcase of FELIX's ability to work with off-the-shelf hardware. As far as the compression method goes, it can make use of LZS compression, Huffman coding and DEFLATE, the latter of which is essentially a combination of the first two algorithms [82].

The QAT compression API allows for a high level of customisation. The compression level and type can be specified, as well as details of the algorithms in use, such as the use of dynamic or static Huffman trees [82]. In addition, the QATZip library provides access to the compression and decompression services of QAT hardware. It produces and reads the gzip format, which in turn makes use of the DEFLATE algorithm [83].

Data reordering

The frame format, shown in figures 4.2 and 4.3, packs its 12-bit channel values tightly together. This aids in the high throughput required of the ProtoDUNE DAQ, but impedes reading and compression. Conventional compression methods rely on recognising common values and repeated patterns in the data stream. This is hindered by the fact that the channel values are cut up and distributed throughout the frame in a nontrivial manner. Additionally, a fragment contains 10,000 frames, each of which contains 256 channel values for a single time-slice. It is expected that channel values vary little from one time-slice to the next, whereas they are expected to vary more from one channel to another within a time-slice. With the goal of compression in mind, it is therefore advantageous to transpose the data, storing the time-evolution of each channel contiguously. This has the added advantage of allowing for faster reading of full channel waveforms.

Table 4.3 shows the implemented reordered fragment format. The channel values have been expanded to occupy two bytes and are stored time-consecutively. The reordering algorithm checks all headers of the original frame and saves the first frame's headers entirely. Any anomalous values in subsequent frames trigger it to set the corresponding bit in the faulty header list and record the anomalous frame's set of headers.

4.3.3 Results and discussion

Tests were performed using both software and QAT hardware compression. Using a custom compression program optimised for compressing FELIX frames, a compression factor around 6 was reached, well surpassing the goal of a compression factor of 4. This was done using Huffman compression on MicroBooNE data converted to the FELIX frame format. The data was reduced by utilising waveform continuity. The compression lacked in speed, however, and was thereby unsuitable for online operation. Since data is extracted from the continuous 2 MHz stream at a rate of 25 Hz, only 40 ms is available to perform every operation on the data. The compression software, on the other hand, took approximately three times that amount of time, taking around 120 ms for the compression of a fragment.

Table 4.3: The reordered fragment format, assuming a content of 10,000 frames.

Size (bytes)	Structure
2	Channel 0 [0]
2	Channel 0 [1]
	⋮
2	Channel 0 [9999]
2	Channel 1 [0]
	⋮
	⋮
2	Channel 255 [0]
	⋮
2	Channel 255 [9999]
1250	Bit list of N faulty headers
16	WIB Header [0]
64	COLDATA Header 1-4 [0]
16	Faulty WIB Header [0]
64	Faulty COLDATA Header 1-4 [0]
	⋮
16	Faulty WIB Header [N-1]
64	Faulty COLDATA Header 1-4 [N-1]

MicroBooNE events

As part of the Memorandum of Understanding between the MicroBooNE collaboration and the ProtoDUNE-SP FELIX team, ten Monte Carlo events and ten real data events were released for use in compression studies. In the Monte Carlo case, each event contains 9600 time-slices of 8256 channel values, which are divided into three planes. The real data events contain 9595 time-slices of the same number of channels. In order to form complete fragments, the number of wires extracted per plane was rounded down to the nearest multiple of 256 and grouped into fragments containing 256 channels. This resulted in 31 fragments per event in both simulated and real data. Whereas ProtoDUNE and DUNE frames were expected to contain 10,000 frames, the difference was not believed to have significant effects for the final compression

results. The difference with ProtoDUNE's eventual 6000 frames per fragment is more substantial.

The noise RMS (root mean square) is a measure of the average noise amplitude in a data sample. It is first determined by channel and then averaged over all 256 channels in the fragment to produce a per-fragment average value. Assuming no significant pedestal drift within a fragment, the per-channel noise RMS is determined from the difference of values with the channel's average value. In the case of PVS values, the channel average is close to 0. The results can be seen in figure 4.6. It can be seen that PVS RMS values are much lower and more concentrated than the unaltered ones.

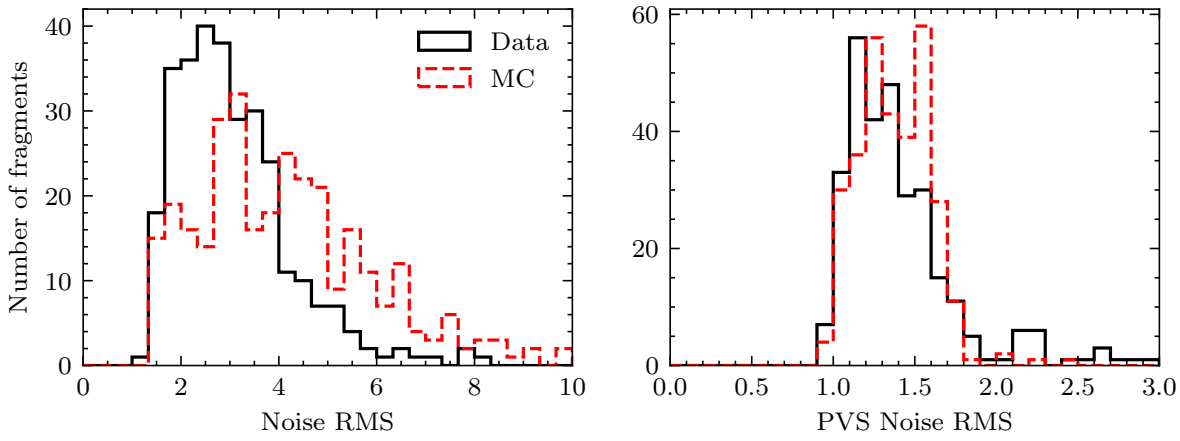

Figure 4.6: Fragment noise RMS distributions in simulated and real data. Both unaltered (left) and PVS (right) values are shown.

Software compression

The used software compression algorithm is applied on a per-fragment basis and consists of three steps. First, waveform continuity of the channel values is exploited, resulting in a set of PVS values. Second, Huffman compression is applied to the PVS channel values and the result is stored along with the Huffman table, the latter of which is many orders of magnitude smaller than the compressed data. Finally, the headers are reduced to their essential information. In order to determine the performance of the algorithm, the process is timed on a system representative of one used in the ProtoDUNE DAQ and the compressed size is compared to the original size.

Figure 4.7 shows cumulative distribution diagrams that indicate the compression ration and time for the data. Distinctions are made between real and simulated data, as well as between unaltered and PVS channel values. The most notable feature is the compression ratio, which increases twofold with the application of waveform continuity. In terms of compression time, it can be seen that performing PVS logically takes some time, lengthening the compression process slightly. It can also be seen that the compression time is unfortunately well above the available time of 40 ms.

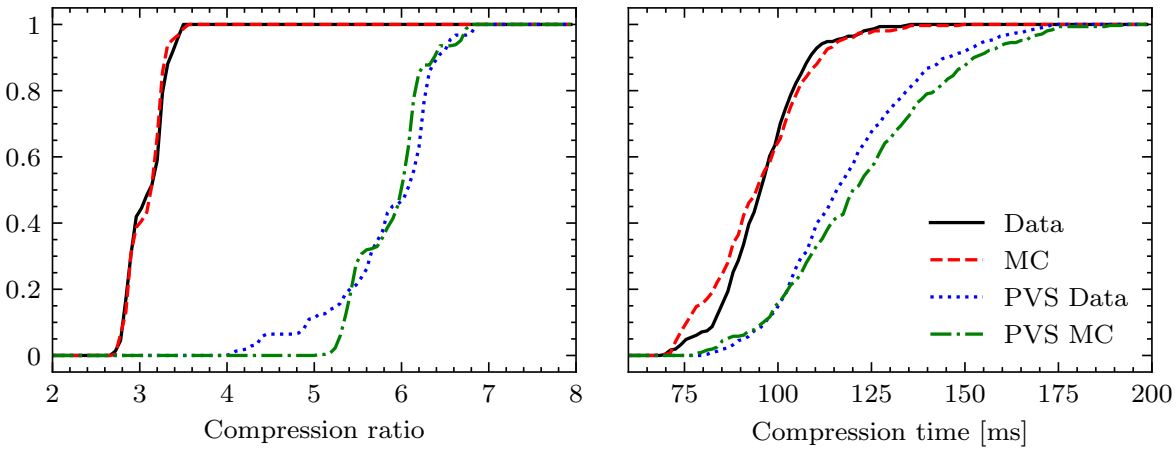

Figure 4.7: Cumulative distribution diagrams for the software compression ratio (left) and operating time (right). All measurements were made on a per-fragment basis, which contain 9600 and 9595 frames in the simulated and real data sample respectively.

It is expected that the attainable compression ratio is dependent on the nature of the data. Unchanging channel values would compress extremely well and completely random values would barely compress at all. To check this prediction, the fragment compression ratio can be compared with its noise levels. Figure 4.8 shows the results of this comparison, again making the distinction between simulated and real data, as well as between unaltered and PVS channel values. It can be seen that the relation between compression ratio and noise RMS is weak in cases where waveform continuity was not applied. Fragments with PVS channel values, on the other hand, show a clear relation between the two variables, indicating a steep drop in the compression ratio for fragments with more noise. By contrast, no relation was found between a fragment's average noise RMS and the time it takes to compress it. This is expected, as the operations performed in compression are largely independent of the nature of the data.

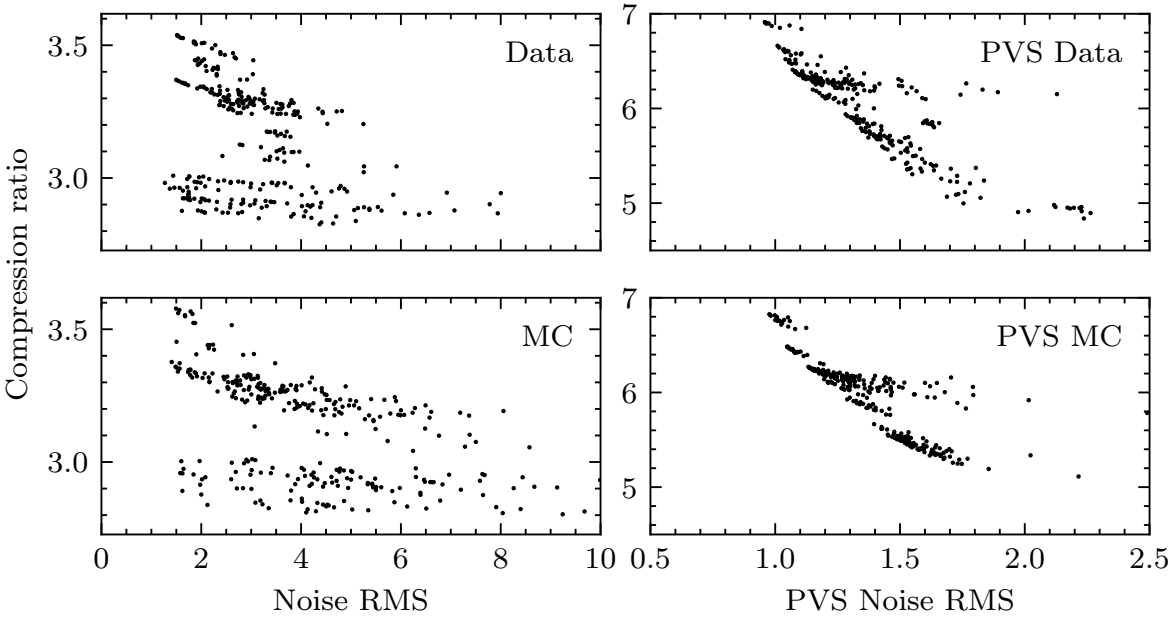

Figure 4.8: The relation between noise RMS and compression factor, shown for real (top) and simulated (bottom) data. A distinction is made between unaltered channel values (left) and PVS values (right).

Intel® QAT compression

QAT compression at the highest level is governed by a compression level, which takes on an integer value between 1 and 9. Levels 1-4 are similar to zlib levels 1-4, levels 5-8 are mapped to level 4 and level 9 is similar to zlib level 9 [83]. It was found that compression at level 9 takes hundreds of times longer than compression at level 4. For this reason, only levels 1-4 were considered. Figure 4.9 shows the per-fragment performance of QAT compression over real MicroBooNE data. It can be seen that the compression time always stays within acceptable bounds. The compression ratio can be seen to improve greatly, going from unaltered to reordered channel values. Interestingly, the compression performs worse in this metric over PVS reordered data. In all cases, the compression ratio increases with the compression level.

Since it features acceptable compression time and the highest compression ratio, level 4 QAT compression performed on reordered fragments was chosen as the method of choice. It can be seen in figure 4.10 that there is only a slight relation between a fragment's compression ratio and its average noise RMS. This is consistent with unaltered fragment software-based compression in figure 4.8.

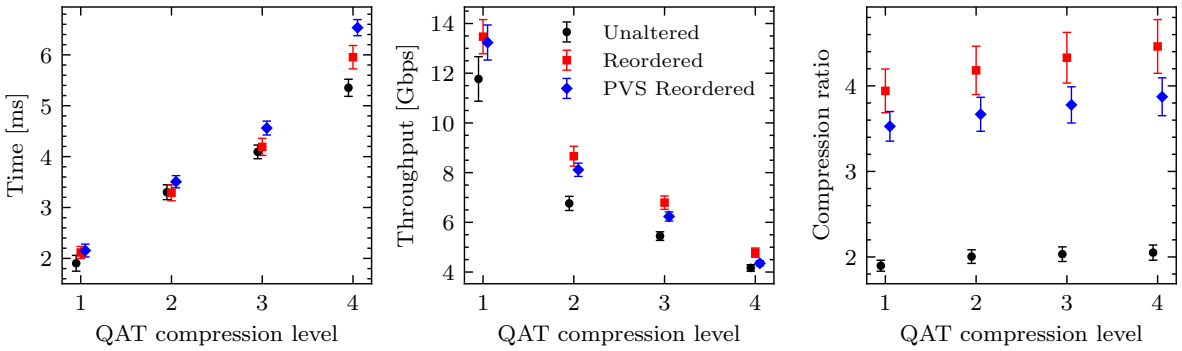

Figure 4.9: Various compression metrics for a range of compression level settings. Shown are the average per-fragment compression time (left), throughput (middle) and ratio (right) for real MicroBooNE data. The error bars denote a 1σ spread within individual data points. Results are shown by fragment format: unaltered, reordered and PVS reordered.

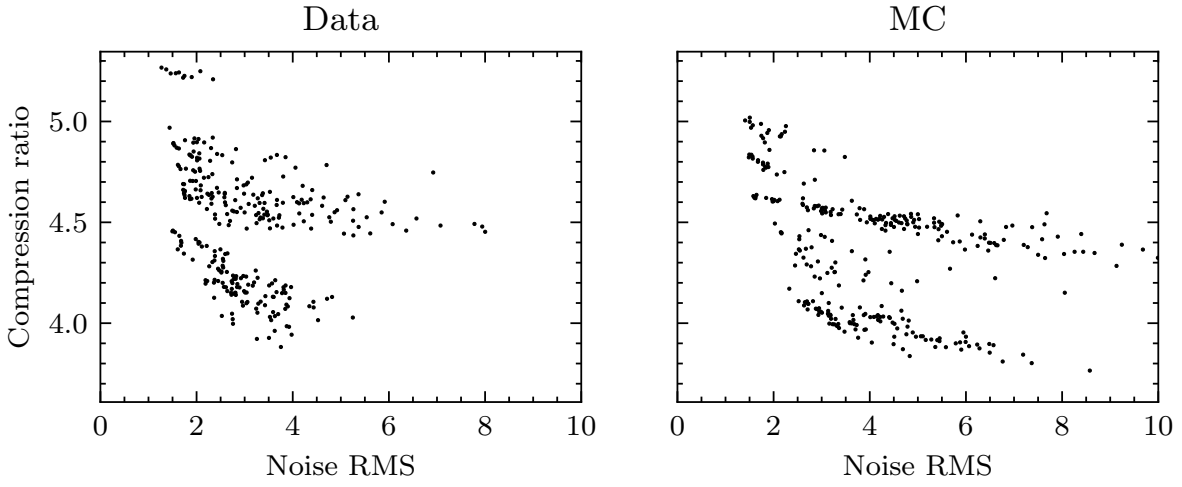

Figure 4.10: The per-fragment compression ratio compared to the fragment's average noise RMS. Both real (left) and simulated (right) MicroBooNE data are shown.

Conclusions

Comparing software to hardware compression, it can be seen that the compression ratio is heavily advantaged by an algorithm that is tailored to the structure of the data. Compression in software, being the more flexible option, is thus better suited to produce the highest compression ratios. In terms of speed, however, software underperforms when compared to hardware-accelerated compression algorithms, making it unsuitable for online operation in this context.

Hardware-accelerated QAT compression is ineffective on the raw data sample, but exceeds the target compression ratio of 4 when the data are reordered. At compression level 4, an average compression ratio of 4.5 ± 0.3 was achieved. At an average of 6.0 ± 0.2 ms, the compression time falls well within the acceptable limits of the 25 Hz trigger rate of ProtoDUNE. Both compression methods show anti-correlation between the compression ratio and the noise RMS of the data. The noise level in ProtoDUNE is therefore expected to play a significant role in determining the compression ratio of the data. As such, this chapter's studies performed on MicroBooNE data samples only provide indicative numbers.

Chapter 5

Reconstruction

In experimental physics, reconstruction refers to the recovery of physics processes from detector signals. In LArTPC experiments such as DUNE and ProtoDUNE (chapters 2 and 3), reconstruction starts with electrical signals from anode plane wires. Each individual wire produces a current signal over time called a waveform. After clearing the waveform of noise as much as possible, it is discretised into individual Gaussian hits. By using coincidence between the three separate anode planes, a three-dimensional position can be assigned to each hit and higher level structures become clear within the detector by clustering hits that are close together. Further classification is made between long line-like clusters called tracks and more cloudy clusters called showers. Lastly, a hierarchy is established on each group of clusters consisting of mother-daughter relationships. A primary track may, for example, have two daughter showers and one daughter track, the latter of which may in turn possess a second-order daughter shower [84]. After this reconstruction of particle traces has taken place, particle identification and characterisation may be performed in physics analysis. However, for the purposes of this chapter it is not included in the reconstruction process.

In the DUNE far detector, detector signals may be traced back to the neutrino that caused a specific event. Figure 5.1 shows an example of such a neutrino interaction. In ProtoDUNE, the momentum and nature of the particle entering the detector is known from beam monitors so that the performance of reconstruction algorithms may be tested. The primary signals in both detectors come from the TPC and photon detection system (PDS), both of which are explained in more detail in chapter 3. In this study, only the information gained through the TPC is used. The PDS is intended to provide information vital to accurate positioning of signatures in the detector, but is at the time of writing not yet incorporated into the main reconstruction chain. The reconstruction

chain relevant to this thesis is therefore contained in the scheme presented in figure 5.2. Its main components will be explained in more detail in the rest of this section.

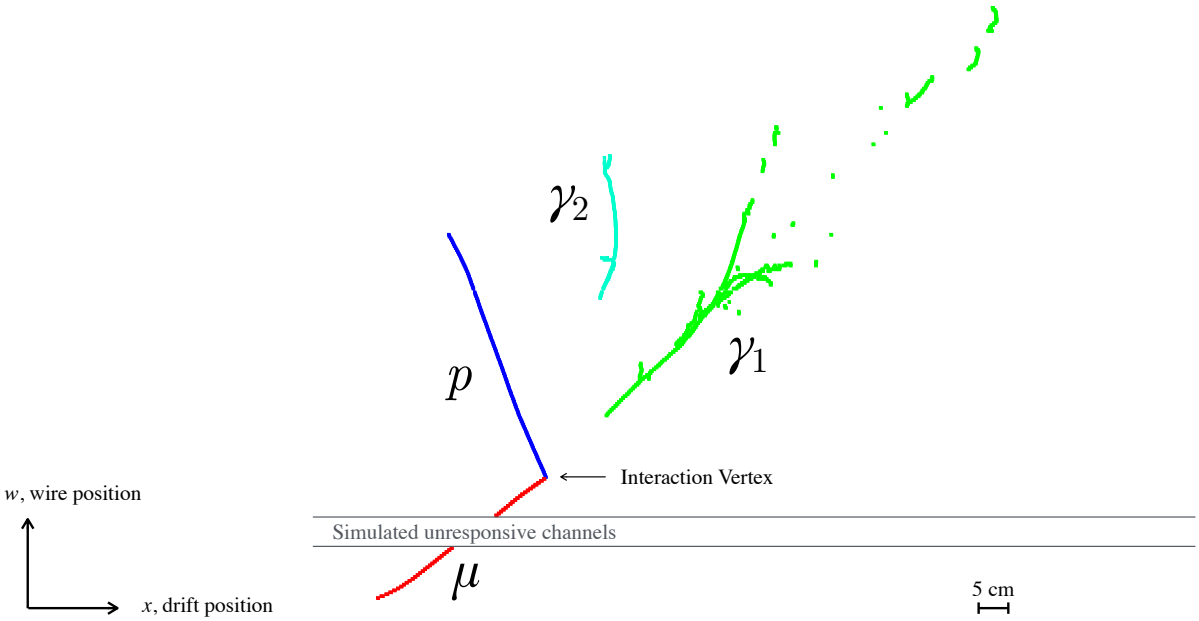

Figure 5.1: An example of a reconstructed event in simulation. An incoming invisible neutrino interacts with the detector material and produces a number of charged particles visible in form of tracks and showers. A proton and muon are responsible for the creation of two tracks, while two photons, likely from π^0-decay origin, produce two showers. Hits have been grouped into several clusters denoted by various colours. Taken from [84].

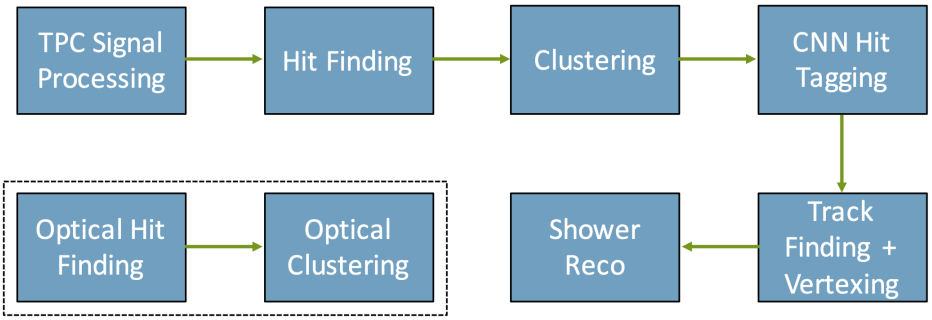

Figure 5.2: The main stages in the TPC and PDS reconstruction chains.

5.1 Signal processing and hit reconstruction

Reconstruction begins as soon as the ionisation electrons are detected and absorbed by the wires in ProtoDUNE's anode planes, described in chapter 3 and visualised in figure 3.6. Each wire in each plane of the detector records a waveform in the units ADC/tick, where ADC represents the number of Analog to Digital Converter counts, or amplitude of the signal, and a tick amounts to 0.5 μs in the ProtoDUNE TPC timing.

Electrons liberated from their argon atoms that have resisted the effects of recombination start to drift along the electric field in the detector. Detector simulations show that the median recombination factor, or the fraction of all charge that makes it to the anode plane, amounts to 0.64 for positron showers [85]. Diffusion causes electrons to spread apart slightly, effectively making the image fuzzier the longer they travel. Modelling has shown that charge created at the ProtoDUNE or DUNE cathode spreads apart 1.8 mm longitudinally and 2.5 mm transversally to the travel direction [86][69]. The field, although it is nearly homogeneous, distorts the shape of the electron cloud further, particularly when it approaches the anode wire planes and the field lines start to deviate substantially. Once the electrons do arrive at the detection wires, the electronics response is not perfectly proportional to incoming charge and depends on a number of factors. The resulting signal is a convolution of all detector effects and must be processed before it can be used [87]. The process of correcting for detector effects on the signal is known as deconvolution and is essential for a proper view of physics processes within the detector. Noise filtering and the removal of detector effects are explained in the following sections. An example of their effects on a single induction channel can be seen in figure 5.3.

5.1.1 Noise filtering

Noise filtering is an important step in the deconvolution process. TPC noise most commonly originates from the readout electronics and the high voltage supply, and because these sources can be characterised in isolation, it is possible to effectively filter out their noise [87]. This per-wire deconvolution is called one-dimensional because it deconvolves only over the dimension of time. A 2D deconvolution algorithm was developed for MicroBooNE and adapted for ProtoDUNE that also takes spatially correlated noise effects into account, where electron clouds induce long-range effects on multiple wires [88][55]. These effects are modelled and compensated for with

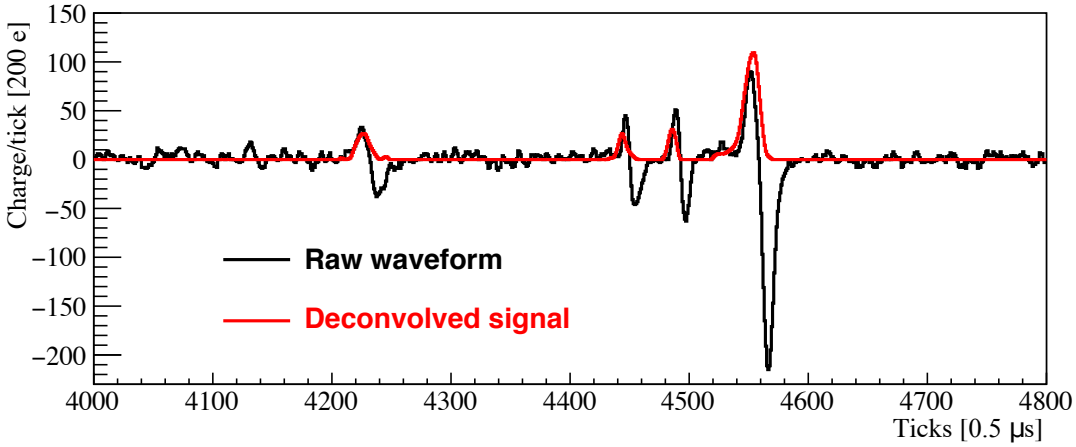

Figure 5.3: An example of a raw (black) and deconvolved (red) waveform. Bipolar signals have been made unipolar and the noise has been removed. The deconvolved waveform has different units than its raw counterpart and has been scaled to a similar amplitude as the original signal. The raw data was extracted from an induction wire in ProtoDUNE. Plot taken from [36].

Garfield [89], a drift-chamber simulation program. An example of the simulated long-range effects is shown in figure 5.4 for one induction wire in particular.

One source of noise is so-called coherent noise, which finds its origin in the low voltage regulator that supplies the cold readout electronics with power. It derives its name from its characteristic to cause correlated fluctuations between channels on the same FEMB. A visual representation of this is presented in figure 5.5, where dark vertical streaks show noise fluctuations across several channels at the same time. Because the FEMB is the origin of this noise, all 40-48 channels on this FEMB can be combined into a median waveform. In this waveform, the incoherent noise cancels out and the coherent noise remains, making it suitable for subtracting this type of noise from the relevant channels [90].

Another major type of noise found in ProtoDUNE is harmonic noise originating from the high voltage power supply. Fluctuations in the supplied high voltage can produce small swings in the electric field amplitude that are nevertheless large enough to be picked up by the sensitive anode plane wires. The noise largely disappeared when a different power supply was installed, but the noise did make its way into the beam run data of 2018. Fortunately, as the name suggests, it is highly harmonic. By applying a Fourier transform to the signal and iteratively removing the spikes at multiples of 50 kHz, a large portion of the noise is mitigated [87].

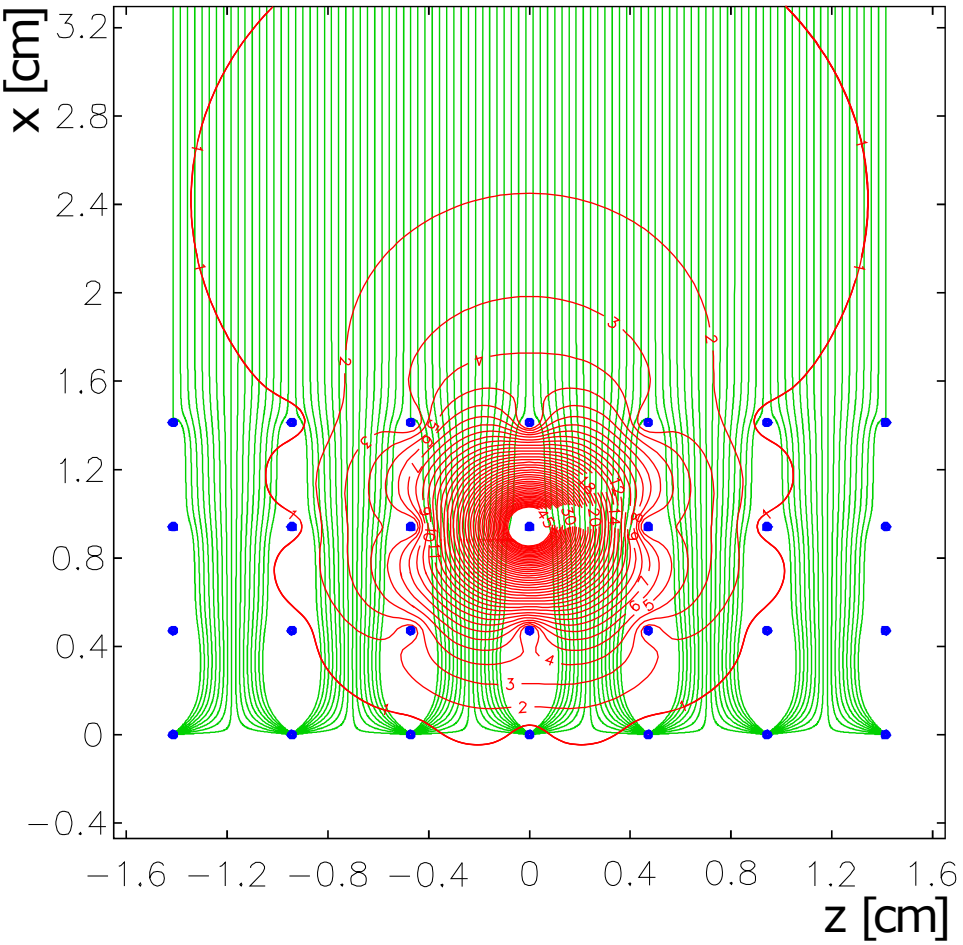

Figure 5.4: A top view of long-range effects as simulated by Garfield [89] seen from the top. From top to bottom, the rows of blue dots in the figure represent the grid wires, induction wires (x2) and collection wires. Electric field lines are shown in green and weighting potential lines are shown in red. The potential line values indicate what percentage of the central wire's signal is transferred to the surrounding wires. Taken from [55] and edited for clarity.

Other sources of electronic noise exist in the detector, such as sticky channels. These are readout channels that prefer producing certain values over others, seeming to 'stick' at these values. The glitch originates in the analog-to-digital converter of a channel, where the transistor responsible for the conversion fails in producing the six least significant bits of the channel's value. This results in the channel only producing the number 0 or 63 plus a multiple of 64. Because sticky and otherwise bad channels are easily spotted, identification of faulty channels in ProtoDUNE was done by eye. Mitigation consists of disregarding the sticky channel's value and using a

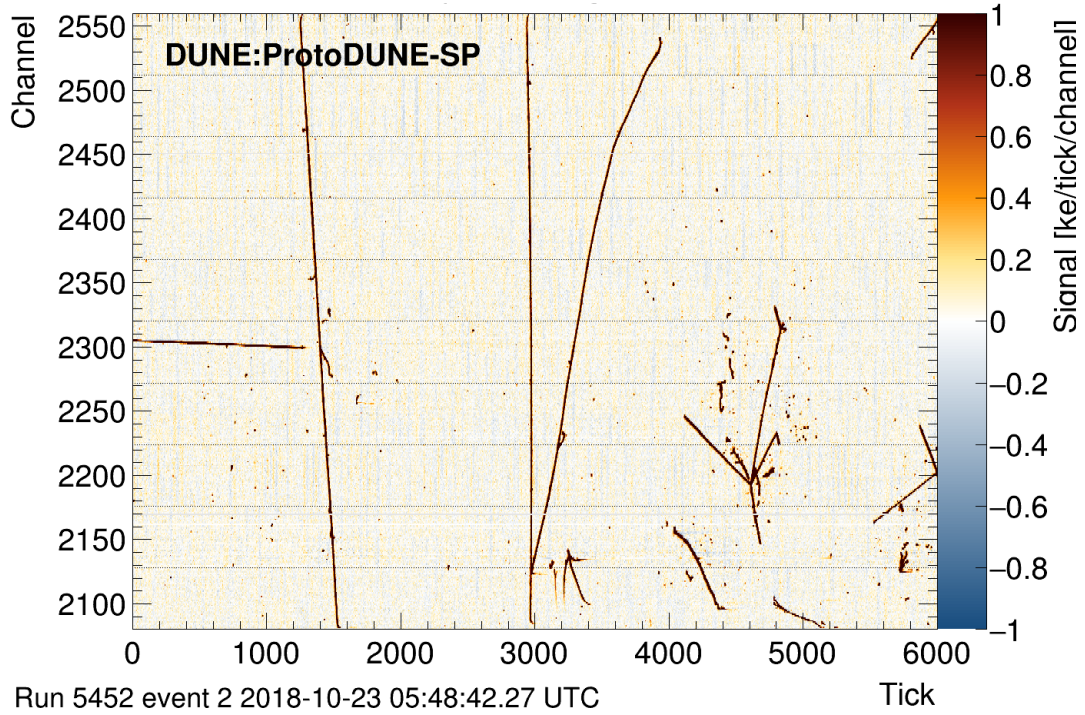

Figure 5.5: An event in ProtoDUNE as seen by the collection plane of the APA closest to the beam entrance into the TPC, shown before coherent noise removal. The horizontal axis indicates time in ticks of 0.5 μs and the vertical axis denotes the channel number in the wire plane, with FEMBs separated by horizontal dotted lines. Consisting of coincident signals between channels of the same FEMB, coherent noise manifests itself as alternating light and dark vertical bands. Taken from [55].

value interpolated from the neighbouring channels instead. 312 channels out of 15360 were identified as sticky in ProtoDUNE [55].

5.1.2 Detector effects

In order to form an accurate picture of ionisation tracks in the detector, the arrival time of all charges must be matched to their position at the time of creation. In other words, the electron travel time must be known to a high degree of precision. A perfectly homogeneous electric field would provide a perfectly linear relation between the arrival time and distance travelled of the collected charges, but in reality there are two main sources of field distortion in the detector. The first is caused by the physical shape of the detector and its anode plane wires, but is small compared to the second cause of distortion: the space charge effect (SCE) [55]. This effect describes the buildup of positively charged ions in the detector's active volume due to the

constant bombardment of cosmic muons and charged beam particles. The particles leave ionisation electrons in their wake, but also produce positively charged ions that move much more slowly along the electric field. The resulting cloud of ions distorts the electric field in a noncontinuous manner: over time, it changes depending on the length of the run and nominal field strength of the detector. As a result, electrons no longer drift perpendicular to the anode planes, causing an apparent displacement in their origin. An example of this effect is shown in figure 5.6. This figure shows clear asymmetry between ProtoDUNE's drift volumes that has not been traced to a certain cause. Possible explanations include slight differences in drift length, fluid flow and field cage resistance.

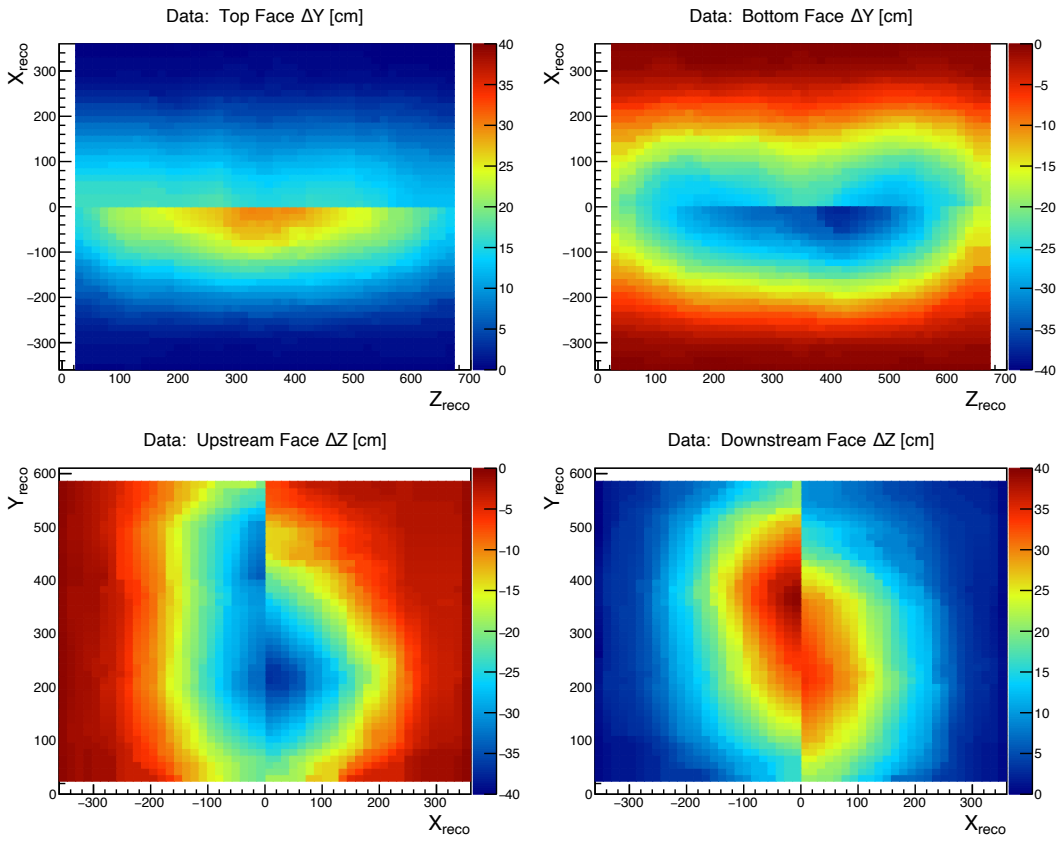

Figure 5.6: A visualisation of the space charge effect at four faces of the ProtoDUNE detector. The figures show the additive transverse correction in cm that must be applied to hit positions at the horizontal faces of the detector (top), as well as its front and back face when seen from the beam position (bottom). The SCE in reality causes three-dimensional apparent displacement of hits throughout the volume of the detector. The visible discontinuity is caused by the cathode plane, which effectively divides the detector into two separate drift volumes. Taken from [55].

Aside from being deflected by positive ion clouds, drift electrons may also be captured by electronegative atoms in the detector volume, oxygen in particular. The result of excessive impurities in the liquid argon is a decrease in collected charge that is seen as a fading of signals. Because electron capture occurs proportional to distance travelled, it is especially prevalent in signals coming from deep within the detector. The capture of electrons and its relation to the reduction of measured charge can be described as follows:

$$Q(t) = Q_0\, e^{-t/\tau}, \tag{5.1}$$

where $Q(t)$ describes the measured charge after an electron travel time of t, Q_0 is the initial charge at the time of ionisation and τ is the electron lifetime. This electron lifetime characterises how long electrons may travel on average before they are captured by impurities. ProtoDUNE employs filters to purify its argon, increasing the electron lifetime drastically, as can be seen in figure 5.7.

Purity measurements were performed using dedicated purity monitors within the ProtoDUNE cryostat during the beam run. Afterwards, however, it was found that the argon purity could accurately be determined from long tracks that cross both the cathode and anode panels. Cosmic muons are abundant and relatively uniform over the distance between the cathode and anode, making their tracks ideal for measuring the electron lifetime within the TPC. By comparing their charge deposition near the anode (Q_a) to that near the cathode (Q_c), it can be determined what fraction of the ionisation electrons survives its journey across the drift volume [55]. Analogously to equation 5.1:

$$Q_c/Q_a = e^{-t_c/\tau}, \tag{5.2}$$

where t_c is the fixed time that electrons take to drift from the cathode to the anode. As the electron lifetime τ approaches infinity, the surviving fraction of electrons Q_c/Q_a approaches 1.

To put the electron lifetime measurements into context, the drift velocity of electrons in the ProtoDUNE electric field is around 1.6 m/ms and the detector's drift length amounts to 3.6 m [52][55], making the maximum drift time around 2.25 ms. An electron lifetime of 3 ms was therefore considered the lower bound of the acceptable lifetime range. This value was far surpassed during and after ProtoDUNE's beam run, which can also be seen in figure 3.13.

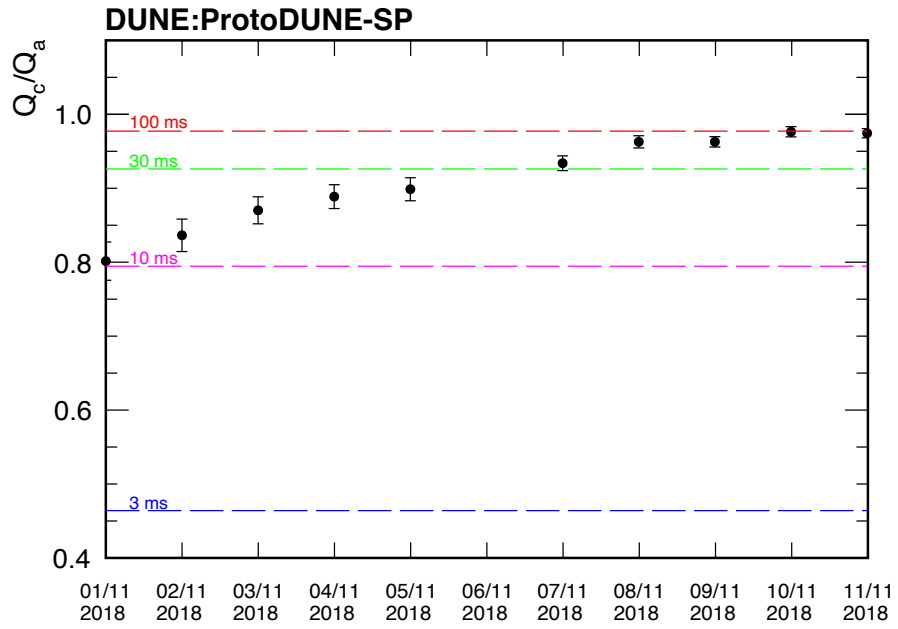

Figure 5.7: The argon purity of ProtoDUNE in terms of charge measured at the cathode (Q_c) over the charge measured at the anode (Q_a) in November of 2018. These measurements were made using tracks within the TPC, calibrated with cosmic ray tagger (CRT) data. Taken from [55].

Another detector feature that is rectified is the bipolar nature of induction wire signals. Ionisation electrons pass between the induction wires of the anode plane on their way to the collection wires, generating a positive signal while travelling towards the induction plane and a negative signal once they have passed it. The bipolar signal is converted into a unipolar signal akin to that of the collection plane at the end of the deconvolution process [91].

5.1.3 Hit finding

The waveforms left over after deconvolution are considered to have a high signal-to-noise ratio, but still contain too much information to practically work with. For this reason, an algorithm called gaushit [92] captures the waveform's shape into a series of Hit objects, which are Gaussian peaks of a certain height and width. Figure 5.8 shows an example of the result of hit finding, where Gaussian hits have been fitted to a waveform from a single wire. A Hit object thus contains at the very least a wire ID, and peak time, amplitude and width. The integral of the Gaussian hit is then proportional

to the amount of charge that created it. An important assumption in using these hits is that the waveform's shape can be described accurately with (overlapping) Gaussian peaks.

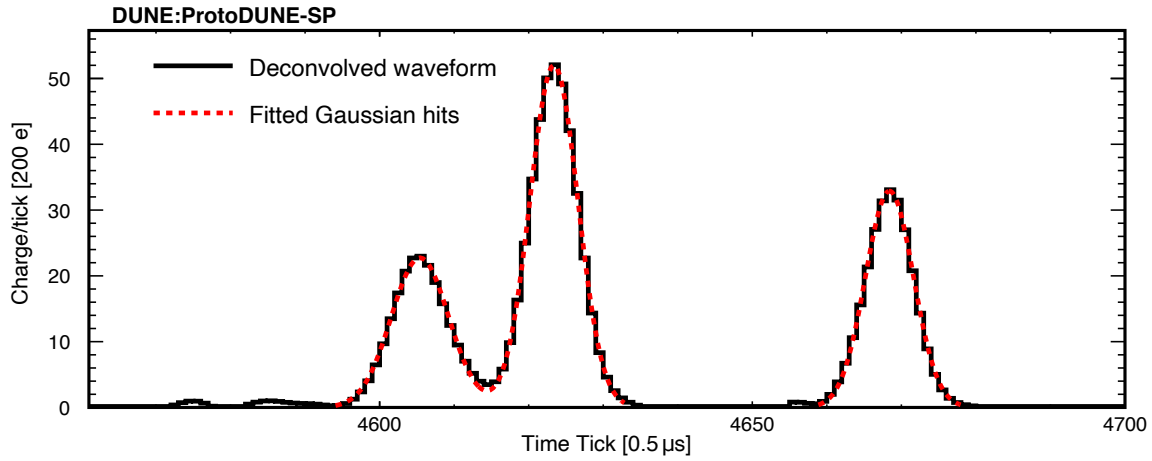

Figure 5.8: The gaushit hit finder in action, fitting Gaussian Hit objects to the channel's waveform where it exceeds a certain threshold. Taken from [55].

Signal processing and hit finding transform the raw detector output into three clear views: two from induction planes and one from the anode's collection plane. In order to reconstruct a three-dimensional image of the event, matches between hits of different planes need to be found. This is done through coincidence in time and distance: three hits from different planes less than 2 mm apart in the drift direction and less than 3.55 mm apart spatially are considered a triplet. Ambiguities still arise when, for example, two tracks overlap in one of the views. These can often be resolved by comparing the charge between hits in triplets [36]. If the deviation is too large, the coincidence was likely coincidental. The result of this process is a set of so-called space points: points in 3D space that have associated 2D induction and collection plane hits.

5.2 Pandora

Although new particle identification frameworks are constantly being developed for LArTPCs, the Pandora Software Development Kit [84] (henceforth called Pandora) has been a consistent baseline. The analyses in this thesis have made use of the Pandora reconstruction chain output. It features a multi-algorithm approach, meaning that separate tasks in reconstruction are handled by independent algorithms. The Pandora

reconstruction chain takes the 2D hits produced by the hit-finding algorithm as input and produces various reconstructed objects that can be used in analyses. Examples include 3D hits, tracks and showers. Note that Pandora chooses to ignore previously created space points and instead creates its own 3D hits based on cluster topology. In addition, Pandora constructs a particle hierarchy from these objects that allows the analyser to identify parent-daughter relations of each object [84].

5.2.1 Reconstruction chain

Developed for MicroBooNE, the Pandora LArTPC reconstruction chain features two main reconstruction paths: PandoraCosmic for the identification of cosmic rays and PandoraNu for the reconstruction of possible neutrino events. For detectors on the surface, tagging the many cosmic rays passing through the active volume is vital. ProtoDUNE, with its charged particle beam, uses yet another reconstruction chain focused around the beam window. This is the region in the active TPC volume directly around and extended in the direction of the particle beam as it enters the detector. The algorithm in use for test beam particles is appropriately named PandoraTestBeam.

As can be seen in figure 5.9, the ProtoDUNE Pandora reconstruction chain first tags unambiguous cosmic rays and passes the remaining hits on to a so-called slicing algorithm. This cuts 3D space up into regions that contain hits coming primarily from one particle. The individual particles are passed to both the Pandora Test Beam and Pandora Cosmic algorithms and are subsequently classified as either a beam particle or a remaining cosmic ray based on their performance in each algorithm and an associated boosted decision tree score [93].

The two main particle reconstruction algorithms approach hit clustering slightly differently. PandoraCosmic is primarily concerned with reconstructing cosmic muons and is therefore strongly track-oriented. Any electromagnetic activity in this pass is assumed to be delta ray activity. Because this algorithm is only used implicitly in this analysis, no deeper explanation of it will be given. The PandoraNu and PandoraTestBeam algorithms focus more on the primary interaction vertex of the neutrino or beam particle in question, as well as the daughter particles originating from this interaction. In a first pass, the input 2D hits are cautiously clustered and merged or split where necessary. Afterwards, possible interaction vertices are generated in 3D from 2D cluster pairs. In order to transform the two-dimensional clusters to 3D objects,

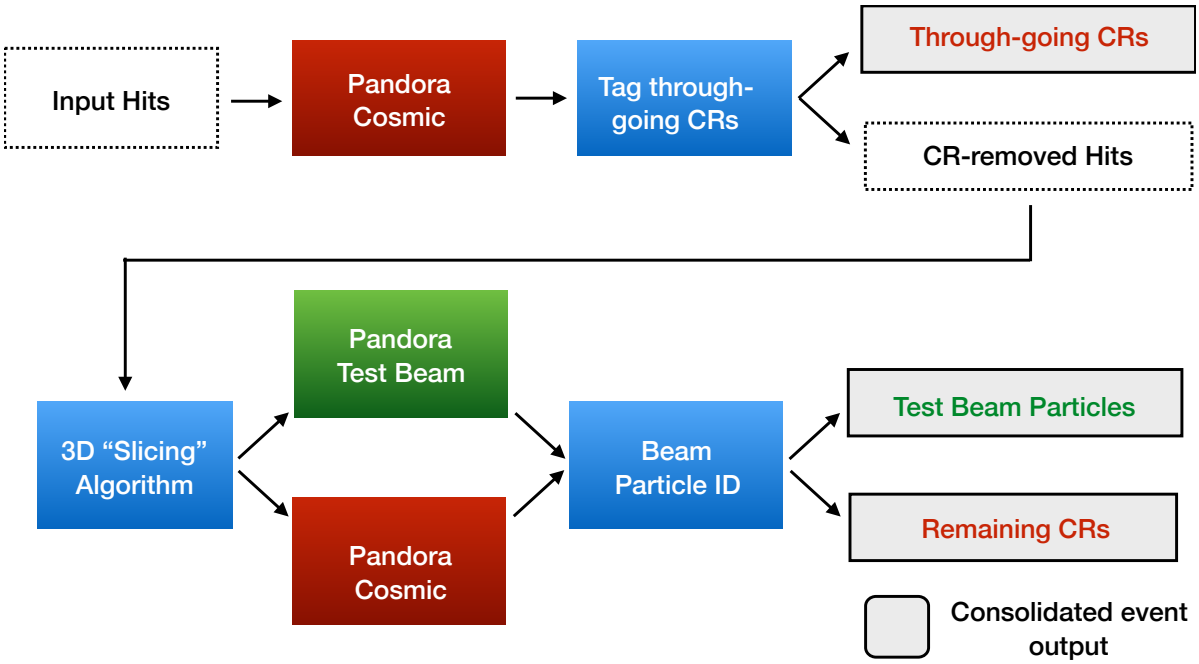

Figure 5.9: A global and greatly simplified overview of the Pandora reconstruction chain. Taken from [93].

three clusters need to be matched, one from each plane. This is done by taking two clusters on different planes and predicting the possible 3D cluster projection on the third plane. The projection is then compared with nearby reconstructed 2D clusters on the plane to create a χ^2 value. A 3D cluster candidate obtains its total score by cycling through all three possible 2D cluster pairs in the triplet and combining the χ^2 values from the resulting projections [94].

Most well-formed clusters, regardless of whether they have been tagged as cosmic, are formed into PFParticle objects. The PFParticle object does not hold much information on its own, but it does have numerous associations with related objects. As can be seen in figure 5.10, it is connected to a vertex, the original 2D clusters and a set of 3D spacepoints corresponding to the clusters' hits. Importantly, each PFParticle is associated with one Track or one Shower object. Distinguishing between tracks and showers is a nontrivial matter and is based on a set of cluster features. It involves object length, variations in sliding-fit direction along the length of the cluster, an assessment of the extent of the cluster transverse to its linear-fit direction, and the closest approach to the projected neutrino or beam interaction vertex [95].

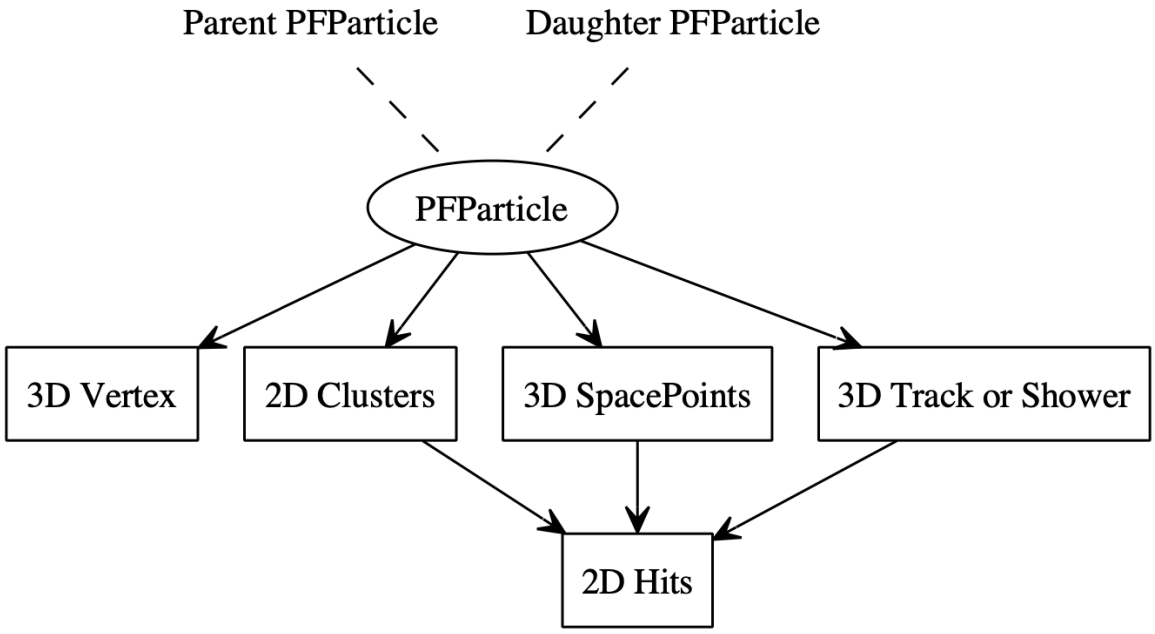

Figure 5.10: An overview of some of the links between a PFParticle and other objects, including substructures such as Clusters, but also other PFParticles in the same hierarchy. Taken from [54].

Oftentimes, analyses are concerned with more than one particle at a time. In the case of a specific process in physics, both the incoming and outgoing particles contain vital information. Pandora supplies PFParticles with parent and daughter particle relations wherever applicable, as shown in figure 5.10. It performs this process starting at the perceived primary particle, which in the case of the DUNE far detector is a neutrino and in ProtoDUNE is a charged beam particle. Having associated an interaction vertex with the primary particle, other reconstructed PFParticles can be tagged as daughter particles based on proximity and angle with respect to this vertex.

5.3 Usage in the LArSoft framework

LArSoft [77] is a software toolkit developed for simulation and reconstruction in LArTPC detectors, particularly in the context of neutrino interactions. Such a detector-agnostic approach is possible because of the relative similarity within this class of detectors, consisting of a volume of argon and detecting ionisation tracks by means of an electric field and a set of wire planes. Built on the *art* [96] event-processing network, data manipulation is performed through so-called modules. These are written in C++

according to a rigid class structure that loops through all specified events. The general configuration of the module is stored in a so-called FHiCL (Fermilab Hierarchical Configuration Language [97]) file. This configuration file can hold simple variables and also specifies the names of modules that produced objects used in the module. For instance, both the Pandora and SpacePointSolver algorithms produce space points that are saved as part of each event. One algorithm's results may be selected over the other by specifying the name of the producer.

From data taking to particle simulation to reconstruction, all information is recorded on a per-event basis in the CERN ROOT [98] format. Although it is technically possible to access the information via ROOT macros, analyses are in practice performed using LArSoft-based modules. LArSoft-style objects such as PFParticles, Tracks and Showers are contained within C++ STL vectors and are easily accessible from an Event object by specifying the type of object required, as well as the name of the module that produced the object. Each such Event corresponds to the data collected during a single global trigger as specified in chapter 4. In addition to object information, links between objects are also stored in each Event. The links in this context are called Associations and are used by supplying the Event with the object to be linked to, as well as the module that produced the linked object. The result is a vector of objects that are associated with the original object. For example, a Cluster object is associated with a vector of Hit objects, and every PFParticle is associated with either a Track or Shower object.

5.3.1 Links between simulation and reconstruction

A useful feature that is not directly built into the reconstruction chain is the link between simulated and reconstructed particles when looking at MC events. A separate set of algorithms in LArSoft that goes by the name of BackTracker [99] finds the origin of each reconstructed Hit object by finding coincidences with simulated charge depositions. The simulated charge positions have links to potentially multiple MC particle objects, but the ambiguity only starts there. A reconstructed object is in essence a collection of hits which may have come from multiple sources, thus a reconstructed object may in reality be a conglomeration of particle traces. Conversely, hits from a single MC particle may lie so far apart that they are split between several reconstructed objects. For the purposes of the analyses described in the following chapters, every reconstructed object is linked to a single MC particle, which is chosen based on the

total number of hits contributed to the object. This has the unfortunate side-effect that a MC particle will only have an associated reconstructed object if it is its primary contributor in terms of hits. Otherwise it is considered to not have been reconstructed at all.

Once a link is established between a reconstructed object and a MC particle, some quantities can be determined that give a general idea of the object's reconstruction quality. The first is the object's completeness, defined as the fraction of hits created by the MC particle that is captured in the reconstructed object. An object has a completeness of 1 if it captures all of the MC particle's hits. The second quantity is the object's purity, which is defined as the fraction of the object's hits that was created by the linked MC particle. An object's purity is equal to 1 if all of its hits were generated by the same MC particle. A simplified example of completeness and purity calculations is shown in figure 5.11 and their definitions are described in equation 5.3. Figure 5.12 and table 5.1 shows the performance of the Pandora clustering in terms of completeness and purity. It is worth noting that alternative definitions of these quantities exist. Most notably, the completeness and purity of an object may be described in terms of combined hit energy rather than the number of hits.

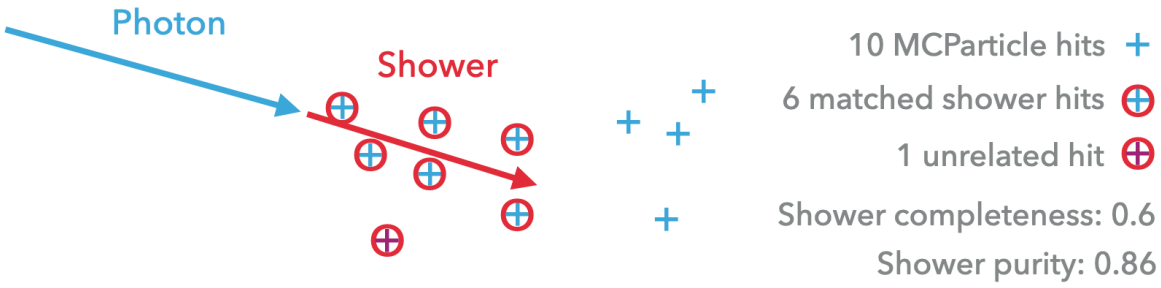

Figure 5.11: A schematic example of an incoming invisible photon (blue arrow) causing charge depositions in the detector (blue crosses). Pandora creates a Shower object (red arrow) out of selected hits (red circles). Only 6 out of 10 of the original photon's hits are captured in the reconstructed object, making the completeness $6/10 = 0.6$. An unrelated hit from another particle is inadvertently included in the Shower object, lowering the purity to $6/7$, or approximately 0.86.

$$\begin{aligned} \text{Completeness:} &\quad \frac{\text{Number of shared hits between reconstructed and MC objects}}{\text{Number of hits in MC object}} \\ \text{Purity:} &\quad \frac{\text{Number of shared hits between reconstructed and MC objects}}{\text{Number of hits in reconstructed object}} \end{aligned} \tag{5.3}$$

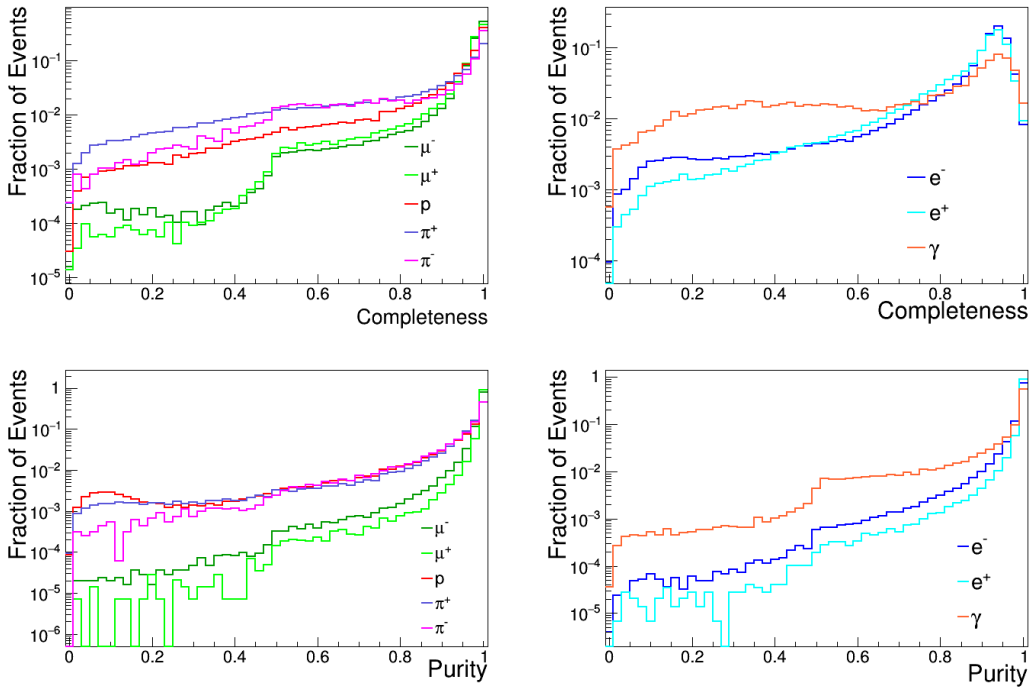

Figure 5.12: Completeness (top) and purity (bottom) of simulated track-like (left) and shower-like (right) particles originating from simulated accelerator neutrino interactions in the DUNE far detector, processed by the Pandora reconstruction chain. Taken from [36].

Table 5.1: The mean completeness and purity of Pandora-reconstructed particles. Data taken from figure 5.12.

	Track-like					Shower-like		
	π^+	π^-	p	μ^+	μ^-	γ	e^+	e^-
Completeness	82%	79%	89%	95%	96%	67%	84%	84%
Purity	84%	88%	81%	98%	96%	85%	96%	94%

5.3.2 Analysis design

Simple analyses can be run directly within LArSoft modules, but complications arise when more complex or changeable calculations have to be performed. The data sets used in the analysis are large, containing half a million events which each hold tens of thousands of MC particles and hundreds of thousands of recorded hits. This makes the analysis a lengthy process, running on Fermilab's FermiGrid [100] distributed computing resources over the course of 5-8 hours. Data analysis, especially in its exploratory phase, benefits from rapid prototyping and visualisation. Being perhaps the largest time sink, debugging code is also more effective with shorter turnaround times for the code to run. For these reasons, the LArSoft module stage is used exclusively for the simple extraction of relevant data from the main reconstruction files. It writes the relevant values for each event to a flat ROOT tree for easier access in later stages. Most of these values can be copied without alteration, such as the position and direction of relevant particles, as well as basic information on their hits. Some quantities require simple calculation to avoid having to store all MC hit information, such as the completeness and purity of reconstructed objects.

Having extracted the relevant information from the reconstruction data set, the total data size is reduced to the order of tens of gigabytes, which is manageable enough to analyse using local scripts instead of grid jobs. A ROOT macro first processes the data to determine quantities that depend on changeable or complex calculations. Two prominent examples of this are the energy and $\mathrm{d}E/\mathrm{d}x$ of reconstructed objects, which both depend on calorimetric constants and spatial correction maps. Both of these quantities are explained in more detail in chapters 7 and 6 respectively. Since not all extracted event parameters are immediately useful, some data selection takes place as well. In addition, the macro cycles over and combines the ROOT files from all runs, recording the run number and associated nominal beam momentum for each file. The results are written to one output file, which is read by various Python scripts for data exploration and visualisation. Heavy use was made of the uproot [101] and pandas [102] frameworks. A visual representation of the analysis structure is shown in figure 5.13.

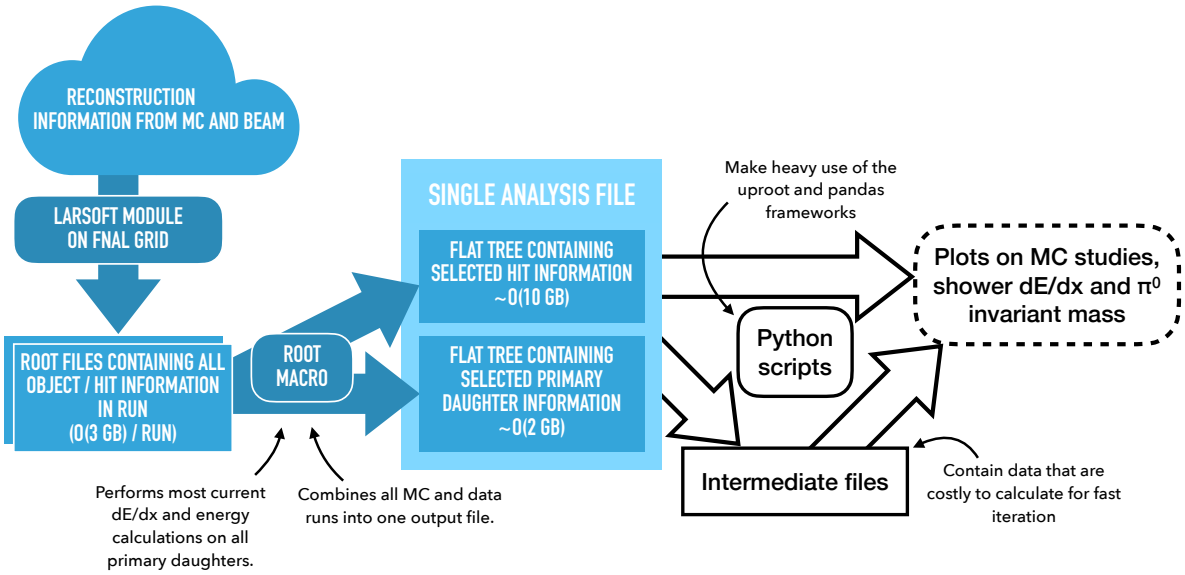

Figure 5.13: The analysis structure in use in this thesis.

5.4 Data used

This section describes the data used in the following chapters. It pertains a mix of Monte Carlo simulations and data recorded during the ProtoDUNE beam run of 2018 described in section 3.3. An overview of the data is shown in table 5.2. All used data was produced as part of the ProtoDUNE Production 2 campaign, wherein simulated beam events were newly created and reconstructed, and existing raw real events from the listed runs were reconstructed using the same algorithms.

Most real beam events in the used data set were recorded using the "Hadrons" trigger, using the beam monitors described in section 3.2.5. As the name suggests, it selects beam particles that have track-like signatures in the ProtoDUNE TPC, the vast majority of which being protons, pions and kaons. Since the analyses in the following chapters rely on hadronic interactions and their secondary particles, it is useful to select these particles and to leave out the many positrons that occur particularly at lower beam momenta. Simulated data sets were not divided by particle type through a similar trigger, instead containing the full range of particles expected from the beam. Since beam monitor information was not available to accompany these simulated events, a cut was instead introduced on the truth information of the beam particle, requiring it to be a proton, pion or kaon. This cut therefore assumes that the Hadrons

Table 5.2: A summary of the data sets used, sorted by nominal beam momentum in GeV/c. Both recorded and simulated ProtoDUNE data was used. Simulated data always use the "All" trigger and have no meaningful run number.

Beam	Data			Simulated
[GeV/c]	Run	Trigger	Total events	Total events
1	5387	Hadrons	126381	95690
2	5432	Hadrons	78441	2600
3	5786	Hadrons	215618	271550
6	5770	All	186625	287430
7	5204	Hadrons	57535	15230

beam trigger operates near 100% efficiency in data. In order to approximate this perfect efficiency further, both simulated and real events are required to contain a track-like reconstructed beam particle.

5.4.1 Event description

An event in the ProtoDUNE context refers to a 3 ms recording of data from all detector systems, centred around a trigger. The three detector systems involved are the TPC, which measures drift electrons on its wire planes, the PDS, which records flashes of scintillation light in the liquid argon, and the beam monitor, which registers and characterises particles as they travel along the beamline into the ProtoDUNE cryostat. All of these are described in more detail in chapter 3. In the case of the data sets listed in table 5.2, the beamline monitors supplied the information to trigger event recording. Depending on the characteristics of the passing beam particle, it was classified as either electromagnetic or hadronic, allowing for data sets that contain either type of particle almost exclusively. Muons are an exception in this system, as they have the characteristics of pions to the beamline instrumentation. These two particle types are instead distinguished by their track length in the detector. Each event that was produced as the result of a beamline trigger is meant to contain exactly one beam particle, although it is possible for there to be none or more than one in rare cases.

Figure 5.14 shows two examples of beam particles interacting in the ProtoDUNE TPC. Having been recorded by the collection plane, the events are shown in a top-

down view, with the beam entering from the left. The beam tracks seem to appear out of nowhere because of the space charge effect, which pushes ionisation electrons inwards near the edges of the TPC. Great detail is visible in both events, with hadronic particles visible as clear tracks and electromagnetic showers even showing some internal structure.

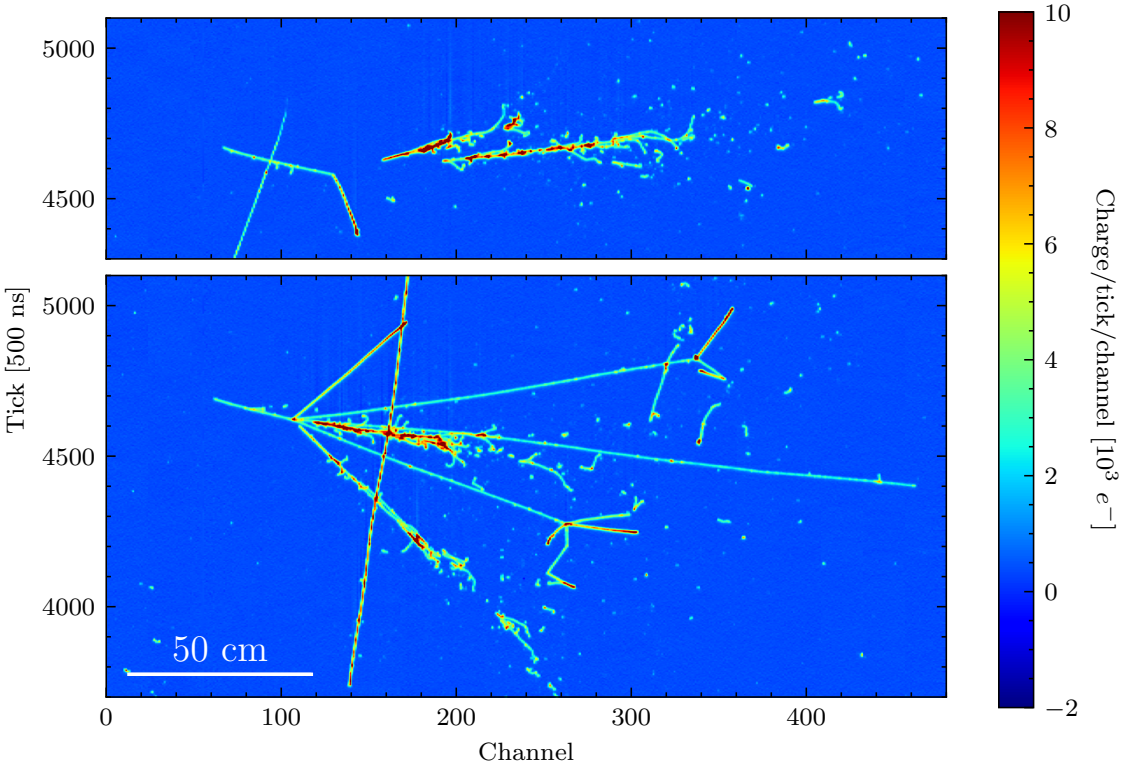

Figure 5.14: Two examples of candidate π^+ events, shown as charge recorded by the collection plane over time. The horizontal axis denotes the channel or wire number in the Z direction with a wire spacing of 4.79 mm [55]. The vertical axis represents time, or the negative X direction. In both events, the axes have been adjusted using the electron drift speed to represent the same scale. The beam particle enters the view around channel 60, tick 4700, and travels rightwards, in both cases interacting with the liquid argon before channel 120. The top event likely shows a 2 GeV/c π^+ particle undergoing a charge exchange interaction, resulting in two clear photon showers. The bottom image shows a more energetic 6 GeV/c π^+ candidate, creating multiple daughter tracks and showers. Event images taken from [55].

5.4.2 Beam particle

Using the Pandora reconstruction output as described in section 5.2, the beam particle can be identified. Figure 5.15 shows a summary of the beam track start position and direction within the TPC. Significant discrepancy exists between data and MC due to mismodelling of the beam in simulation. Overall, the simulated beam is more uniform, more focused and contains fewer imperfections. In particular, the data in figure 5.15 shows particles around the position of the cathode plane at $x = 0\,\text{cm}$ that have been wrongly tagged as beam particles. It is likely that these are cosmic rays that had their ionisation tracks cut in two by passing through the cathode. Another feature in the data that lacks in simulated events is a bump around $y = 450\,\text{cm}$. It is unclear what the exact origin of these particles is, but one possibility is a small leak in the beamline around one of the steering dipole magnets.

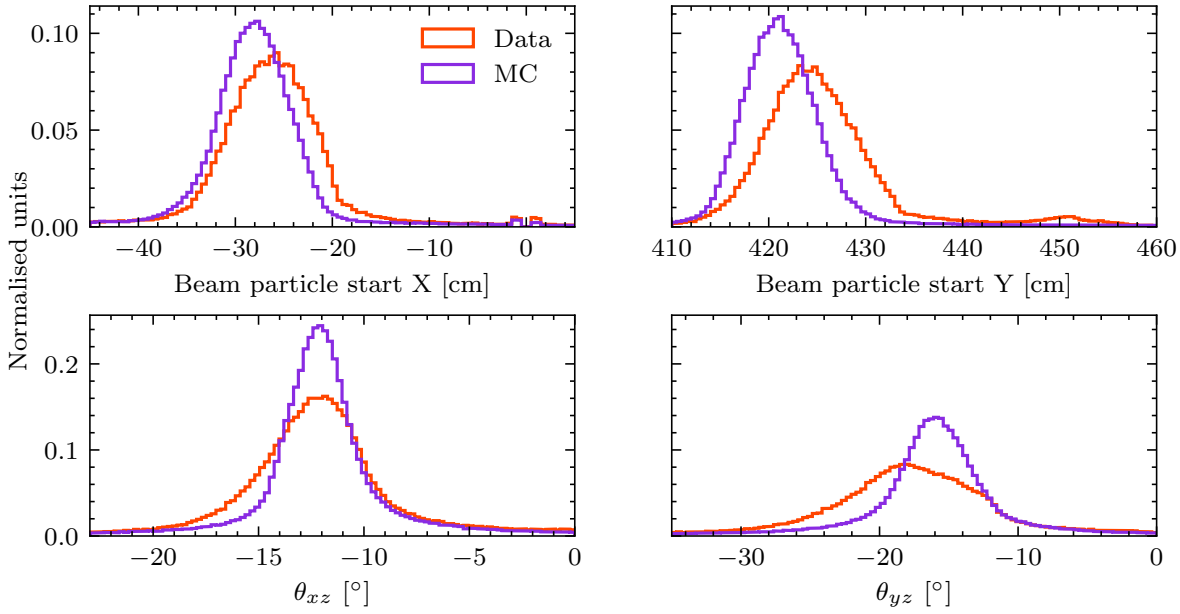

Figure 5.15: The beam particle start position in the XY plane (top), as well as its direction in the XZ and YZ planes (bottom). The coordinate system is the same as in figure 3.11.

Excluding particles leaving the detector volume, the uniform nature of the detector makes it practically irrelevant where the beam particle originates and where it interacts. Aside from possible slight biases related to wire orientation and space charge effects, the ProtoDUNE TPC records particle traces equally well throughout its active volume.

For this reason, slight differences in beam particle positioning between simulated and real events are almost never cause for concern in physics analyses.

5.4.3 Daughter particles

Through the particle hierarchy put in place by Pandora, daughter particles of the beam track can be identified. These particles are divided into two groups corresponding to tracks and showers. Figure 5.16 shows the number of tracks and showers per beam track, highlighting a difference between data and MC. On average, simulated beam tracks seem to generate more daughter particles than their data counterparts.

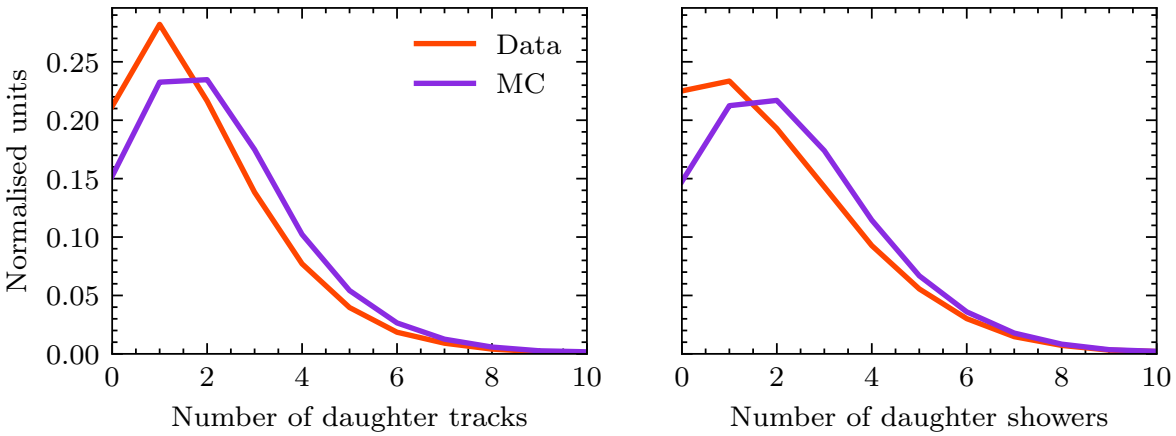

Figure 5.16: The number of daughter particles per beam particle track, divided into tracks (left) and showers (right).

To get a sense of scale of the particles' signatures in the TPC, figure 5.17 shows the length of the beam particle track as well as that of its direct daughters. For comparison, the TPC length in the beam direction is 695 cm. The distribution of the beam particle track lengths shows spikes at 200 cm, 450 cm and 680 cm. The former two come from anode plane transitions: the gap between anode planes causes tracks to seem broken, hence many tracks seem to end at the anode boundaries. The last spike comes from the end of the TPC, where all recorded tracks must necessarily end. Space charge effects cause tracks to seem shorter than the physical length of the TPC. Daughter tracks, being less energetic and produced further into the TPC, are on average shorter than their beam parents. Showers produce a cloud of ionisation much wider than a single track and thus result in much shorter objects.

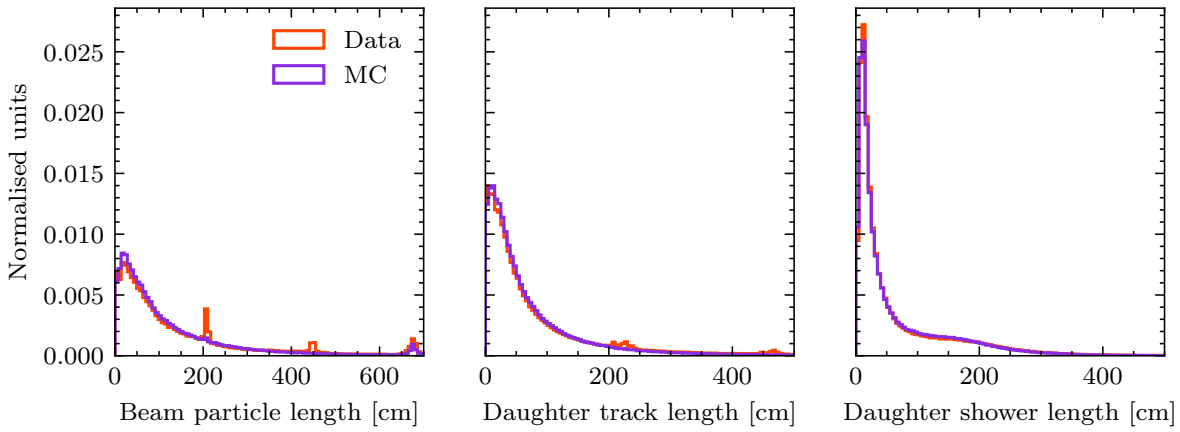

Figure 5.17: The length of beam particle tracks (left), beam daughter tracks (middle) and beam daughter showers (right) in cm. The length of tracks was calculated linearly from their start and end positions. Pandora supplies a length measure for its shower objects.

The first-generation daughter particles of the beam particle already paint a picture too complex to generalise in one dimension. Still, some quantities give an indication of the general structure of an event. Two such quantities are the ionisation energy of each daughter particle and its angle relative to the beam particle. Figure 5.18 shows these measures separately for tracks and showers. It shows that MC and data correspond very well overall, with the exception of some low-angle track daughters that are not simulated. The explanation for this discrepancy once again comes from the transition between anode planes: as a beam track is cut up by the gap between two anode planes, the second part of the beam track is marked as a daughter of the first part with only a small angle between the two track segments.

Shower daughters are emitted more uniformly than tracks and peak at a much lower energy. Both observations indicate that shower-creating particles are often second-generation daughters to the beam particle. For instance, an incoming π^+ beam particle may produce a π^0 particle which in turn decays into two photons, each of which may produce a shower object. The energy per shower is reduced by dividing the first-generation daughter energy between multiple showers. The angular distribution between shower and beam particle is also broadened due to the intermediate particle.

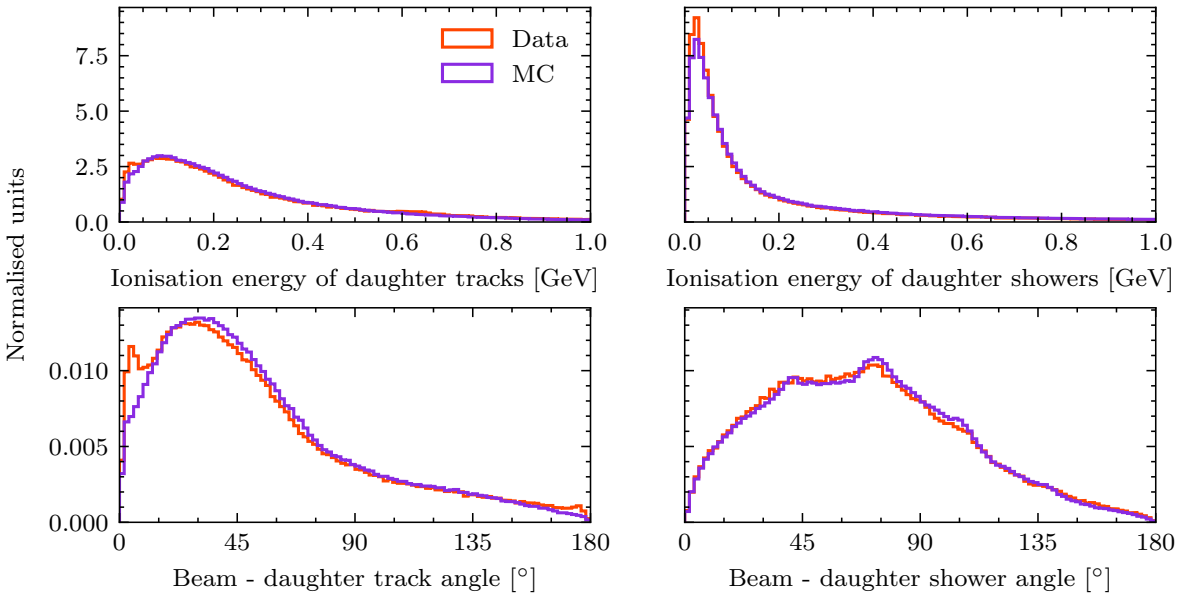

Figure 5.18: The reconstructed ionisation energy (top) and angle relative to the parent (bottom) of track (left) and shower (right) daughters to the beam particle.

Chapter 6

Photon shower identification

Particle identification is a necessity for any high energy particle physics experiment. Neutrinos themselves are not electrically charged and thus leave no ionisation tracks, but their reaction products often do, and it is by these products that the original neutrino may be identified.

The π^0-particle is of special interest to neutrino experiments due to its ability to interfere with neutrino identification. An electron neutrino ν_e undergoing a charged-current interaction is primarily recognised by the resulting electron causing an electromagnetic shower in the detector. In DUNE, the estimated neutral-current background to the ν_e ($\bar{\nu}_e$) appearance signal is 7.3% (12.7%), being the largest background after beam contamination [36]. A π^0-particle produced in any neutral-current neutrino interaction may produce a similar signature in a number of cases, including the following:

1. The photon showers from π^0-decay overlap, making them appear as a single shower.
2. One of the photon showers is far from the decay vertex, or even outside of the detector.
3. One of the photon showers is too small to be reconstructed, track-like or otherwise not recognised as a shower.

In all of these instances, the two π^0-decay showers are seen as a single shower, more closely matching the signature of an electron neutrino even though any neutrino could have caused the interaction. It is therefore important to be able to distinguish electron and photon showers to a high degree. One distinguishing feature between the two

types of showers is their differing initial energy deposition, or $\mathrm{d}E/\mathrm{d}x$, which is the focus of the analysis presented in this chapter.

Charged particles lose energy over time when travelling through a medium. In liquid argon, this energy goes into the ionisation as well as the excitation of argon atoms. Although excitation is visible through scintillation light, this chapter estimates particles' energy loss over distance travelled from their ionisation charge. The energy loss of a moderately energetic particle in matter is described accurately by the Bethe equation and corrections to it, shown in figure 6.1 in the momentum range between $\beta\gamma = 0.1$ and 800. Although the figure describes the energy loss of muons in a copper target, the Bethe equation is independent of the incoming particle and is qualitatively identical between target materials. Of particular note is the minimum-ionisation point, since most relativistic particles in detectors spend most of their time close to this minimum-ionisation value [8]. In liquid argon, this value is 2.12 MeV/cm [52].

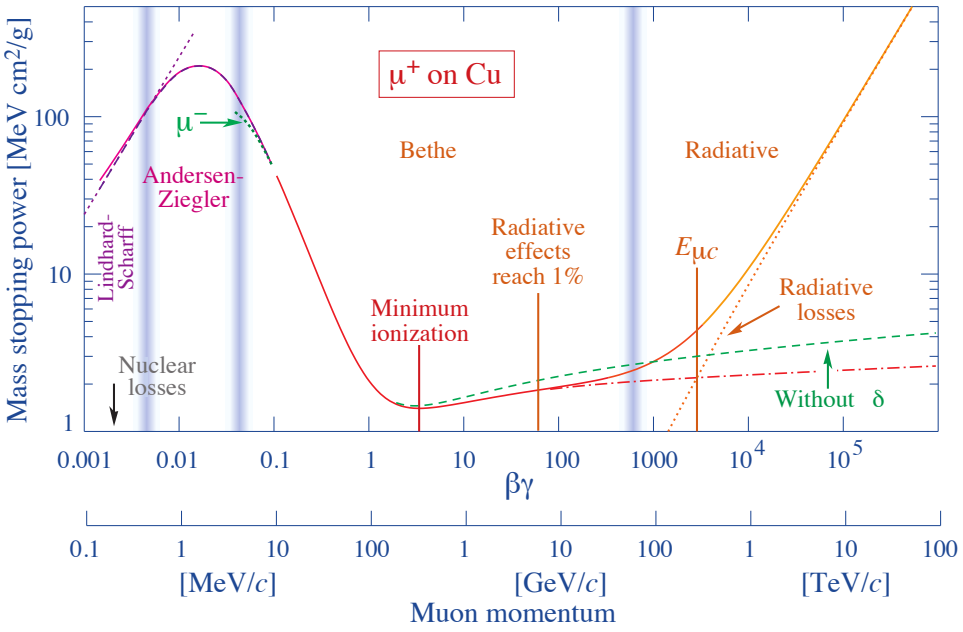

Figure 6.1: The energy loss rate of muons in copper as a function of their momentum. Taken from [8].

In a track-like particle such as a muon or pion, the hits that make up the track are lined up, making the distance between one hit and the next trivial to determine. The hit-to-hit distance can in this case be described by $\mathrm{d}x = p/\cos\theta$. Here, p is the horizontal distance between vertically oriented collection wires, or wire pitch, and θ is the angle between the particle direction and the horizontal-forward z-direction. It should be noted that the choice of wire plane is arbitrary, since only those hits are

used that have coincident hits between all three planes. In this analysis, dx simply describes the distance between consecutive hits and dQ the deconvolved hit integral, which describes the charge contained in a hit. The dE/dx of the hit can be determined from its dQ/dx value through various models, among which are the Jaffe-Birks model and Box model, described in section 6.1.

Determining the dE/dx of shower hits is less straightforward than doing so for track hits, as showers are made up of many minimum-ionising particles: mainly electrons, positrons and invisible photons. Because the radiation length in liquid argon, at 14 cm, is much larger than the spatial detector resolution of ~ 0.5 cm, the track-like start of showers can in most cases be observed in some detail. Their cloud-like main body, on the other hand, contains particles with no clearly defined direction. As such, only the very initial portions of showers are considered. Herein lies also the difference between an electron and photon shower: electrons start their shower as a solitary particle before creating other particles through interaction with the detector material. Photon showers, on the other hand, only become visible once the photon undergoes pair production, producing an e^+e^- pair. The electron and positron are created close together and are heavily boosted in the forward direction, each creating a shower of its own. Thus, in the initial part of an electromagnetic shower, the charge deposition of a photon shower is expected to be twice that of a single electron or positron shower.

The result of a prior study involving positron showers can be seen in figure 6.2. The dE/dx values of positron showers peak strongly towards the 2 MeV/cm minimum-ionising peak, containing only a single positron to create charge depositions near the shower start. Photon showers with sufficiently overlapping e^+e^- pairs are thus expected to peak around 4 MeV/cm. As can be seen in figure 3.9, positrons are common beam particles in ProtoDUNE and are easily selected for, as they are the only shower-like beam particle. The positron sample is therefore much larger, more pure and higher in energy than can be expected for the photon sample presented in this chapter's analysis.

6.1 Calorimetry

The energy deposition reconstruction process contains two distinct steps. First, the wire signal is corrected and converted so that it accurately represents the original

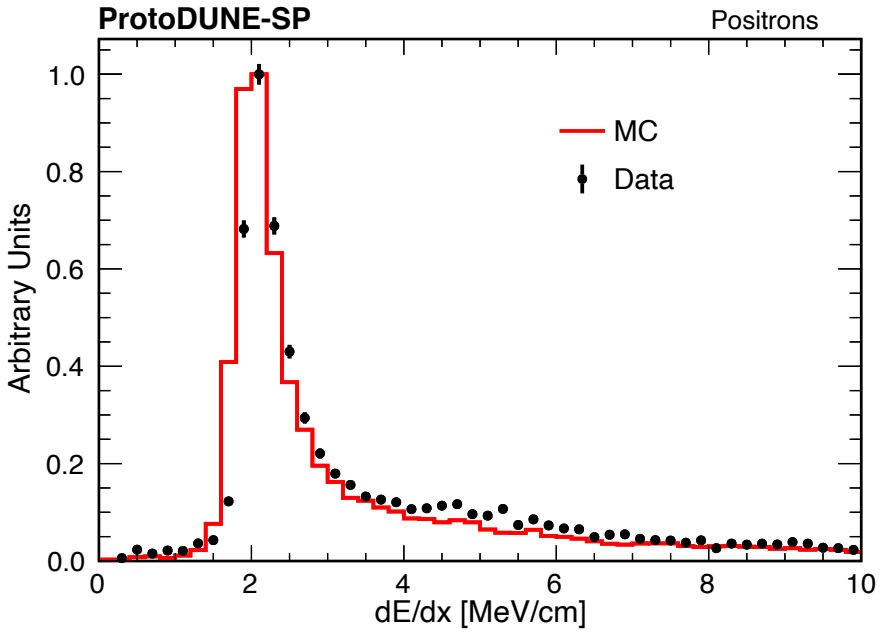

Figure 6.2: The median $\mathrm{d}E/\mathrm{d}x$ of hits captured in a cylinder near the shower start for a sample of positron objects. Simulation and data distributions were normalised relative to their maximum bin values. Taken from [85].

charge deposition $\mathrm{d}Q/\mathrm{d}x$. The detector environment is complex to a degree that requires several correction factors, some of which are position dependent and differ from one run to another. Secondly, the obtained charge deposition is converted into the corresponding energy deposition $\mathrm{d}E/\mathrm{d}x$, accounting for nonlinear electron-ion recombination effects.

6.1.1 Correction and conversion of wire signals

The detector signal consists of current measurements measured in ADC counts. Between charge deposition and detection, ionisation electrons must travel from their point of origin to the anode plane wires. Although their mitigation is briefly discussed here, section 5.1 goes into detail on the specific effects that take place on this journey.

The first of these effects concerns space charge, which denotes a buildup of positively charged ions within the detector volume. The resulting distortion of the electric field alters the electron recombination, but also distorts the shape of particle tracks. In practical terms, $\mathrm{d}Q$, $\mathrm{d}x$ and the electric field $\mathcal{E}$ are all affected. Using cosmic ray tracks, a three-dimensional space charge map was constructed to correct for the effect.

Two-dimensional slices of this map can be seen in figure 5.6. Other detector irregularities may cause further attenuation of the charge collected at the anode wire planes. These include electronegative impurities and imperfections in the wire planes, and are generally position-dependent. Cathode-crossing cosmic rays are used to probe these effects, resulting in a multiplicative correction factor $\sigma(\vec{x})$ [55].

In order to create consistency between various runs, the charge deposition is normalised at this point. This is done by way of a normalisation constant n, generated by comparing the median cosmic ray $\mathrm{d}Q/\mathrm{d}x$ between runs. The normalisation factor of the run with the highest argon purity is set to 1 and all other runs are scaled linearly according to their median charge deposition. Finally, the units of the charge deposition can be converted from ADC/cm to e/cm. The calibration constant f_c is responsible for this conversion and captures the electronics response and any other constant effects that have not explicitly been accounted for. Taking the previously set normalisation into account, it is determined through stopping muon tracks within the detector volume [55].

The correction, normalisation and calibration factors for use in $\mathrm{d}E/\mathrm{d}x$ calculation can be determined from any files containing larsoft reconstruction data. For this analysis, all available reconstruction data were used for each run listed in table 6.1. Using the larsoft and dunetpc projects at version v08_45_00, utility scripts were used to produce the relevant factors [103]. Although the normalisation factor n is expected to change from run to run due to time-dependent impurity levels, the calibration factor f_c is largely dependent on the electronics response and is expected to remain constant. It can be seen that the obtained values generally fit this premise. In addition, the run with the highest recorded argon purity is not among those listed, resulting in all normalisation factors deviating from 1.

Table 6.1: Normalisation (n) and calibration factors (f_c) for all used runs.

Run number	5387	5432	5786	5770	5204	MC
p_{beam} [GeV/c]	1	2	3	6	7	all
n	1.02647	1.00847	1.02747	1.07119	1.14753	0.9947
f_c [10^{-3} ADC/e]	5.3956	5.3954	5.3982	5.3948	5.3601	4.81

Combining all factors, the $\mathrm{d}Q/\mathrm{d}x$ is obtained from the space charge-corrected raw signal in the following manner:

$$\mathrm{d}Q/\mathrm{d}x = (\mathrm{d}Q/\mathrm{d}x)_{\mathrm{raw}} \cdot \frac{n\sigma(\vec{x})}{f_c}. \tag{6.1}$$

6.1.2 Recombination

High-energy particles travelling through a medium leave a cloud of ions and liberated electrons in their wake. In a TPC, the electric field rapidly moves these ionisation electrons towards the anode plane, ultimately forming the signal used for reconstruction. Before reaching the anode plane, however, ionisation electrons have a chance to disappear from view by recombining either with their parent atom or another ion in the cloud, aided by the Coulomb force. The rate of electron recombination is dependent on the density of ions, which is in turn dependent on the energy deposition, or $\mathrm{d}E/\mathrm{d}x$, of the initial ionising particle. The fraction of electrons that does not recombine is called the recombination factor, often denoted by $\mathcal{R}$. Practically, it is used to relate the observed charge to the total charge created in ionisation.

Several models exist to describe electron recombination, such as the Jaffe-Birks model and the Box model. Using data of highly ionising stopping protons and deuterons in the ArgoNeuT detector [104], parameters that govern these models were fitted to observations [105]. In the Jaffe-Birks model, the fraction of electrons that survive recombination can be described by

$$\mathcal{R}_{\mathrm{J}} = \frac{A_B}{1 + k_c(\mathrm{d}E/\mathrm{d}x)/(\rho\mathcal{E})}. \tag{6.2}$$

Here, A_B is a fit parameter, $\mathcal{E}$ is the electric field strength, k_c is a detector material-specific constant and ρ is the density of the detector material.

In the Box model, the fraction of surviving electrons is given by

$$\mathcal{R}_{\mathrm{Box}} = \frac{1}{\xi}\ln(\alpha + \xi), \text{ where } \xi = \frac{\beta(\mathrm{d}E/\mathrm{d}x)}{\rho\mathcal{E}}, \tag{6.3}$$

where α and β are fit parameters. The Box model describes ξ as being dependent on the charge mobility, material constants and local charge density. However, these parameters can be absorbed into β, ρ and $\mathrm{d}E/\mathrm{d}x$.

The energy deposition can be determined from the charge deposition as follows:

$$\mathrm{d}E/\mathrm{d}x = \frac{(\mathrm{d}Q/\mathrm{d}x)W_{\mathrm{ion}}}{\mathcal{R}}, \tag{6.4}$$

where W_{ion} is the energy required for the ionisation of a single atom. For argon, this amounts to 23.6 eV/e [52]. Entering the Jaffe-Birks survival rate of equation 6.2 into equation 6.4 and solving for $\mathrm{d}E/\mathrm{d}x$ then yields

$$\mathrm{d}E/\mathrm{d}x = \frac{\mathrm{d}Q/\mathrm{d}x}{A_B/W_{\mathrm{ion}} - k_c(\mathrm{d}Q/\mathrm{d}x)/(\rho\mathcal{E})}\,, \tag{6.5}$$

which contains a singularity at $\mathrm{d}Q/\mathrm{d}x = A_B\rho\mathcal{E}/(W_{\mathrm{ion}}k_c)$, corresponding to roughly $5 \cdot 10^5\, e$/cm with ArgoNeuT parameters. Although the singularity only occurs at this unusually large charge deposition, it is an undesired feature.

Entering the Box model electron survival rate of equation 6.3 into equation 6.4 yields

$$\mathrm{d}E/\mathrm{d}x = \frac{\exp[\beta W_{\mathrm{ion}}(\mathrm{d}Q/\mathrm{d}x)/(\rho\mathcal{E})] - \alpha}{\beta/(\rho\mathcal{E})}, \tag{6.6}$$

where the canonical version of the model has $\alpha = 1$. Although this equation does not feature the problematic singularity of the Jaffe-Birks model, it was found to perform poorly at low $\mathrm{d}E/\mathrm{d}x$. This was resolved by allowing $\alpha < 1$, resulting in the modified Box model.

Figure 6.3 shows the expected recombination factor and resulting charge deposition as a function of the deposited energy in the detector material. At low values of $\mathrm{d}E/\mathrm{d}x$, the canonical Box model deviates significantly from the Jaffe-Birks model, a discrepancy which is resolved in the modified Box model. It should be noted that very low energy depositions under 1 MeV/cm return anomalous values in the modified Box model. The valid energy deposit range for both models is described to be 2–24 MeV/cm and the modified Box model accurately describes higher energy deposits up to 35 MeV/cm [105]. All fit parameters and used constants are shown in table 6.2.

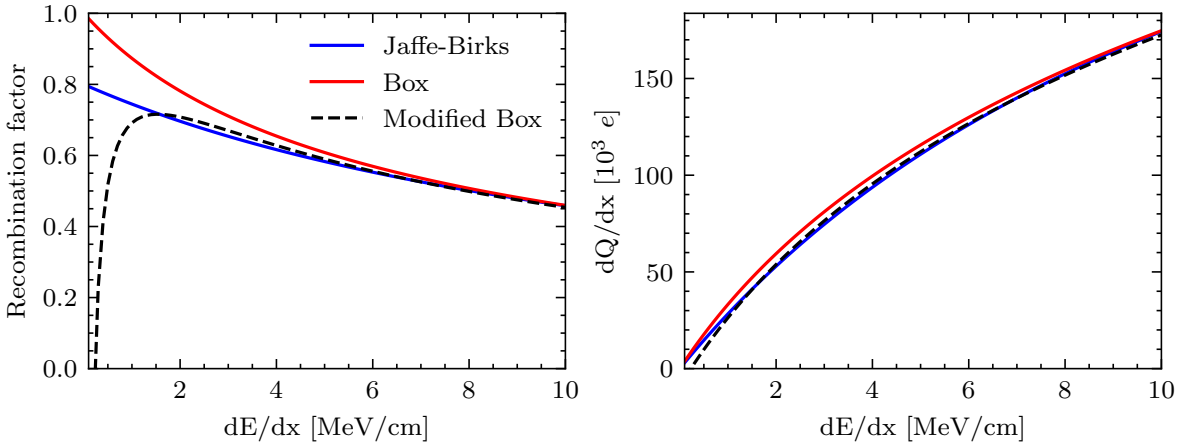

Figure 6.3: The expected recombination factor (left) and charge deposition (right) as a function of the energy deposition in liquid argon. The various models make use of the parameters listed in table 6.2.

Table 6.2: Values used in the determination of a hit's $\mathrm{d}E/\mathrm{d}x$. Various fit parameters taken from [105].

Parameter	Value
A_B	0.8
k_c	$0.052\,(\mathrm{kV/cm})(\mathrm{g/cm^2})/\mathrm{MeV}$
α	0.93
β	$0.212\,(\mathrm{kV/cm})(\mathrm{g/cm^2})/\mathrm{MeV}$
ρ	$1.396\,\mathrm{g/cm^3}$
W_{ion}	$23.6\,\mathrm{eV}/e$
$\mathcal{E}$	Variable ($\sim 0.5\,\mathrm{kV/cm}$)

6.2 Shower sample description

A shower in the Pandora reconstruction framework is an object that has some properties of its own and is associated with lower level objects such as hits, as well as one higher level object: the PFParticle. As described in section 5.2, PFParticles are the links that form the particle hierarchy in an event through parent-daughter associations. Although the Pandora shower object contains energy and $\mathrm{d}E/\mathrm{d}x$ getter functions for future use, both of these values must at present be determined from hit information. By contrast, the shower object does supply a start position, direction and length. The hit objects associated with tracks and showers contain information about the amount of charge they hold and have associated spacepoints that indicate the hits' positions in three-dimensional space.

6.2.1 General sample characterisation

In order to select photon showers from the sample of beam daughters described in section 5.4, it is useful to compare reconstructed metrics to truth information. Table 6.3 shows the Pandora track/shower classification for beam daughters by true particle type. This was done through the backtracker service, described in section 5.3.1. It can be seen that considerable confusion exists for the track-like particles, especially the muons and nucleons. The shower-like particles, on the other hand, are fairly well reconstructed as showers by Pandora.

Table 6.3: Pandora track/shower classification of beam daughter particles compared to their true particle type.

	Track-like					Shower-like	
	$\pi^{\pm}$	p	$\mu^{\pm}$	$K^{\pm}$	Nuc.	γ	$e^{\pm}$
Track class.	70%	62%	57%	74%	33%	12%	13%
Shower class.	30%	38%	43%	26%	67%	88%	87%
Total $\cdot 10^{-3}$	367	585	102	12	28	433	44

In addition to Pandora's classification of reconstructed particles into tracks and showers, a convolutional neural network (CNN) was developed [106] to determine

whether individual hits are the result of electromagnetic activity. It assigns a track-like and shower-like score to each hit. Averaging over all collection plane hits, a shower-like score for each object can be obtained. The distribution of scores, ranging from 0 to 1, is shown in figure 6.4 per true particle type. It can be seen that photons and electrons are in general very well classified. Most confusion comes from protons, pions and muons obtaining a large shower-like score.

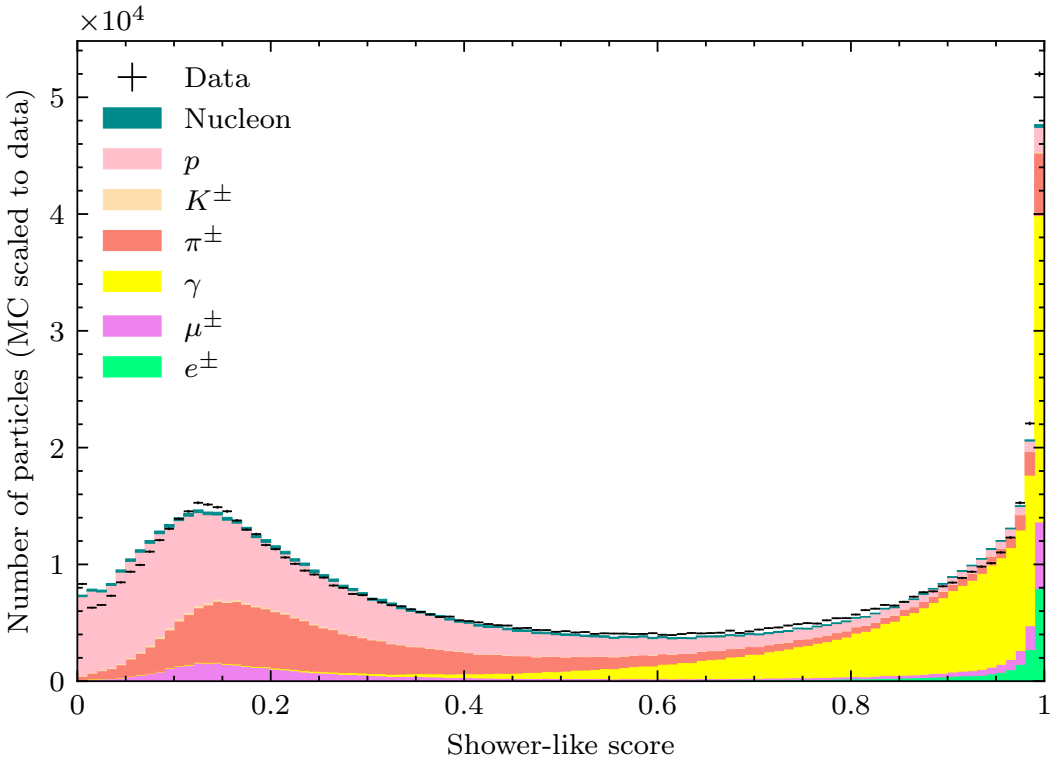

Figure 6.4: The CNN score of reconstructed daughter particles of the beam track.

The number of hits in an object is an indicator of its physical extent within the detector. Additionally, despite not all hits being created equal, objects with more hits are predominantly more energetic. In general, the Pandora reconstruction chain performs better with larger objects than with smaller ones. Two metrics to judge the reconstruction quality of a particle were introduced in section 5.3: completeness and purity. To summarise these concepts: the completeness of an object is the fraction of the original particle's hits that was captured, whereas the purity of an object is the fraction of an object's hits that was produced by the original particle, as opposed to other particles. The desired value of each of these metrics is 1, in which case the reconstructed object contains all hits created by one and only one particle. As can be seen in figure 6.5, the completeness of daughter objects increases rapidly between 0

and 100 collection plane hits. Larger objects have an increasingly larger tendency to overlap, resulting in a steadily decreasing purity.

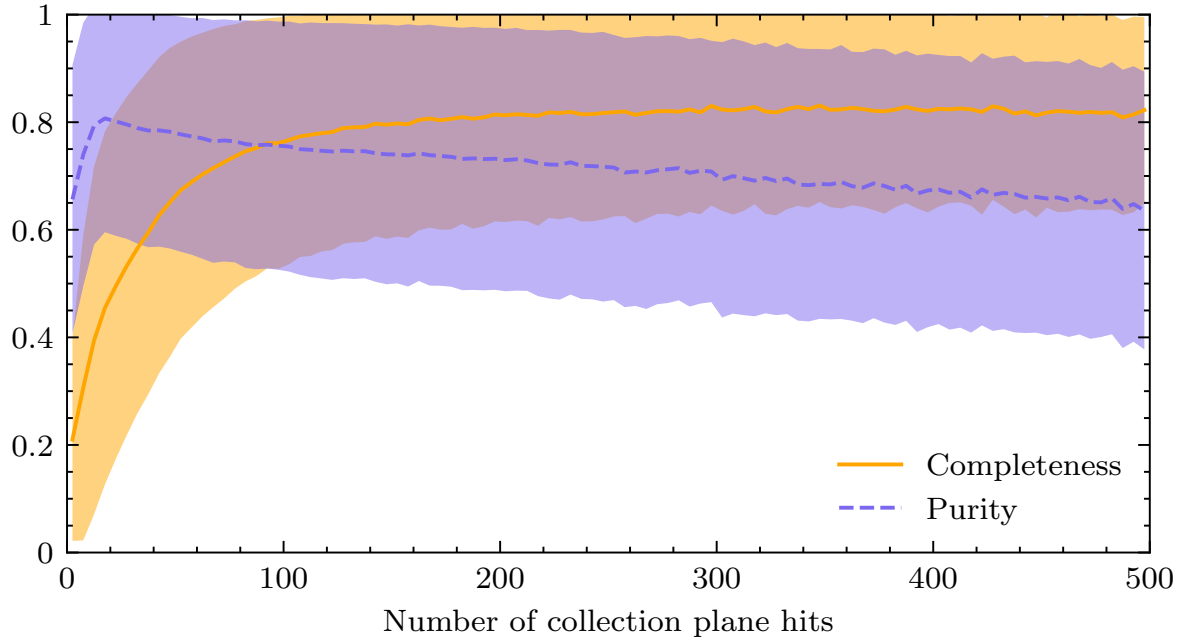

Figure 6.5: The completeness and purity of beam track daughter objects as a function of their number of collection plane hits. The colour bands signify a 1σ spread of the values in each bin.

Different particles produce objects of different sizes. Figure 6.6 shows the number of collection plane hits of objects per true particle type. It can be seen that the majority of beam daughters at a low number of hits is expected to consist of pions, protons and photons. Crucially, the electrons present in the sample are much less abundant than the photons, peaking more strongly towards the low end of the scale. The full distributions show dominance of photon and electron showers at a large number of hits. Discrepancies between data and MC are present around ~ 500, ~ 1000 and ~ 1500 hits. Since collection plane hits in a linear track are $\sim 0.5\,\text{cm}$ apart, these numbers of hits correspond roughly to multiples of the width of an APA. It is therefore likely that this difference is related to tracks being cut up by gaps between APAs, as in figure 5.17.

Further investigation revealed that the large electron showers visible in 6.6 around 2200 hits are beam particles. In some cases, the initial part of an electron shower appears so linear as to be a track of its own. The Pandora reconstruction chain classifies this track-like portion of the electromagnetic shower as an independent track,

with the main body of the shower being classified as a daughter particle. The strictly track-like portion of an electron shower is short, producing a shower very close to the beam entrance.

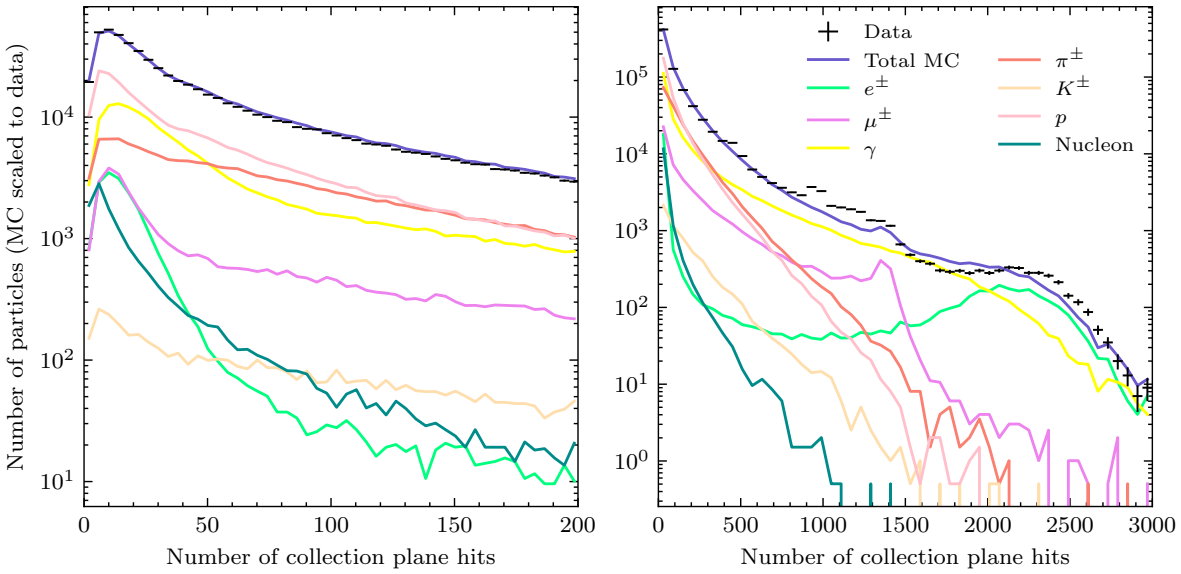

Figure 6.6: The number of collection plane hits of beam track daughter objects by true particle type. A close-up of the low end of the distributions is shown on the left, as well as a more complete range on the right.

6.2.2 Photon shower selection

Based on the sample characteristics discussed in the previous section, it is possible to isolate a pure photon sample from the beam track daughter particles. Table 6.4 shows an overview of the cuts and the resulting cumulative purity and efficiency. Here, purity is defined as the fraction of the sample that consists of true photon daughters and efficiency is the remaining fraction of the original true photon daughters. Although the resulting sample is invariant under a change in cut application order, the intermediate purity and efficiency values are not. This is a result of overlap between cuts. For example, most reconstructed objects with a CNN shower-like score greater than 0.8 are already classified as a shower by the Pandora reconstruction chain. In addition to the cuts' purity and efficiency, table 6.4 shows the ratio between data and MC events in the resulting sample. A ratio of 0.5 signifies the presence of twice as many MC events compared to data. A change in this ratio indicates that cuts affect data and MC differently, pointing to inaccuracies in the physics simulation. Once again, there may

be overlap between cuts causing intermediate cumulative ratios to change under a permutation of the cut order.

The last cut requires a certain number of hits to be contained within a specific region around the shower start. The exact definition of this region and the reasoning behind the cut is expanded upon in section 6.3.

Table 6.4: The applied cuts to the sample of beam track daughter particles. The purity and efficiency of the cuts are shown cumulatively, meaning that values include all cuts up to and including the relevant row. Also shown are cumulative and individual cut data/MC bias fractions, referring to the size of the entire sample. Since the completeness and purity use truth information, these are only shown for MC.

			Data/MC ratio	
Cut	Purity	Efficiency	Cumulative	Individual
	0.28	1.00	0.50	
Pandora shower class.	0.47	0.88	0.50	1.02
CNN shower-like score > 0.8	0.72	0.65	0.50	1.01
Number of coll. hits > 50	0.91	0.35	0.44	1.00
Shower start z > 40 cm	0.92	0.33	0.43	0.99
Shower start hits > 4	0.97	0.05	0.41	1.06

6.3 Method

Pandora assigns all shower objects a three-dimensional start point and direction, as well as a length. Using this, a shower start region can be defined in which the shower's particles are assumed to have a well-defined direction, taken to be the direction of the shower as a whole. One straightforward method of defining the shower start region is to consider the hits in a small cylinder around the shower start point, pointing in the direction of the shower. The dimensions of the cylinder may be optimised for maximal separation between photons and electrons and a cut may be placed on the number of hits captured in the cylinder. Whereas a misplaced shower start point may result in a few captured hits by chance, a correctly placed one will on average capture more.

A previous study that focused on positron shower $\mathrm{d}E/\mathrm{d}x$ determination [85] chose a capture cylinder 5 cm long, pointed in the direction of the shower object and preceding the shower start by 1 cm. Its radius is likewise 1 cm. In order to treat simulated and real showers on an equal footing, it is important to consider the structural differences between the two. Figure 6.7 shows the longitudinal and transverse charge profiles of showers in simulation and data. Whereas the length of shower objects agrees relatively well between the two, there is a clear bias in the shower width, where real showers are substantially wider than simulated ones. A multiplicative correction factor of 1.16 can be applied to the simulated shower width to match the shower width distributions, which is also shown in the figure. The cylinder radius in data was therefore modified to be 1.16 cm instead of 1 cm.

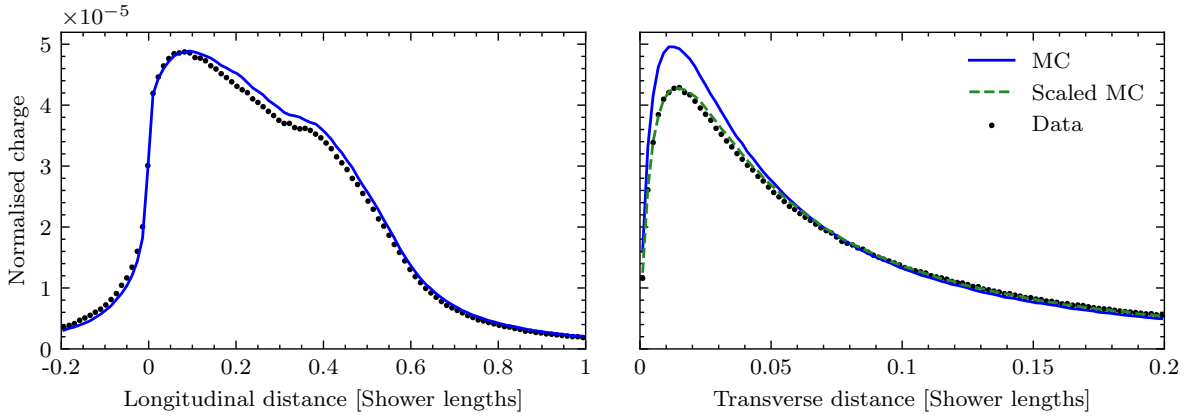

Figure 6.7: The longitudinal (left) and transverse (right) normalised charge distributions for showers in the sample. To make these distributions, the relative longitudinal and transverse distance of hits to the shower start was determined and weighted by their relative charge. Distance measures were scaled by the shower length and charge by the total object charge. Longitudinal and transverse distances are relative to the shower object's direction. The transverse distance is also shown scaled by a factor of 1.16 to better match the distribution in data.

After defining the shower start, the hits included in it may be investigated. As shown in figure 6.8, a large portion of the showers in the sample feature few hits at their start, pointing to a reconstruction failure of either the start point or the direction of the shower. Either of these issues is likely to have its origin in the hit clustering stage. The shower reconstruction process is a complex procedure that is elaborated on in more detail in section 5.2. Although a cut on the number of hits at the start of the shower drastically decreases the sample size, it is necessary to ensure good reconstruction quality. Additionally, most remaining non-photon objects are biased

strongly towards low numbers of captured hits. The resulting sample therefore attains higher purity as a result of any cut on this quantity. Lastly, it can be seen that showers in data capture disproportionately few start hits compared to simulated showers, despite the wider shower start cylinder to compensate for disparities in shower width. Together, these discrepancies point to a more fundamental inaccuracy in the shower structure simulation.

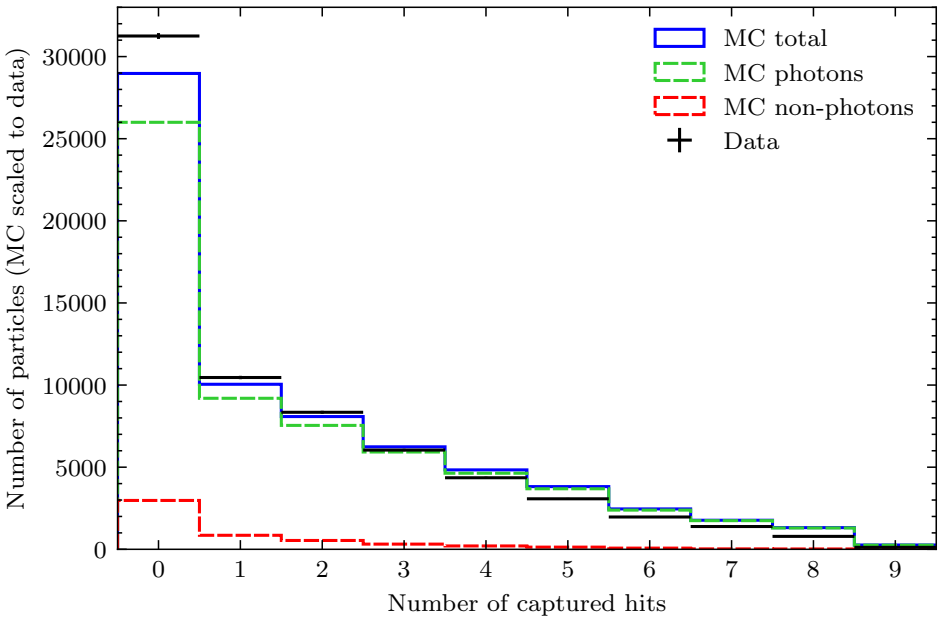

Figure 6.8: The number of hits captured by the shower start cylinder, shown both for simulated and real showers.

Because every hit in the shower start has its own $\mathrm{d}E/\mathrm{d}x$ value, described in section 6.1, one way of combining these would be to compute their mean. However, the distribution of hit $\mathrm{d}E/\mathrm{d}x$ values in a single shower can be highly skewed, especially in cases where the shower points predominantly in the vertical direction and $\mathrm{d}x$ is very large. Additionally, only a handful of hits generally fall within the shower cylinder, making the mean especially susceptible to outliers. A more robust method for combining the hit $\mathrm{d}E/\mathrm{d}x$ values is therefore to take their median.

Figure 6.9 shows median $\mathrm{d}E/\mathrm{d}x$ distributions for various cuts on the minimal number of captured hits. It can be seen that two peaks emerge as the cut becomes more stringent: one around 2 MeV/cm that is characteristic of minimum-ionising particles in liquid argon, and one around double that value, around 4 MeV/cm. The latter peak describes showers in which the e^+e^- pair created by the initial photon

travel close enough together that their hits overlap. As the cut becomes more strict, the sample also becomes substantially smaller, as can be seen in figure 6.8. Placing this cut therefore requires some balance. In the following, showers are required to contain > 4 hits near their start points. This number of hits obtains a pronounced peak around the characteristic photon $\mathrm{d}E/\mathrm{d}x$ value while retaining 14% of the simulated sample and 11% of the data. The requirement of a larger number of hits yields diminishing returns in terms of the peak height and decreases the sample size greatly.

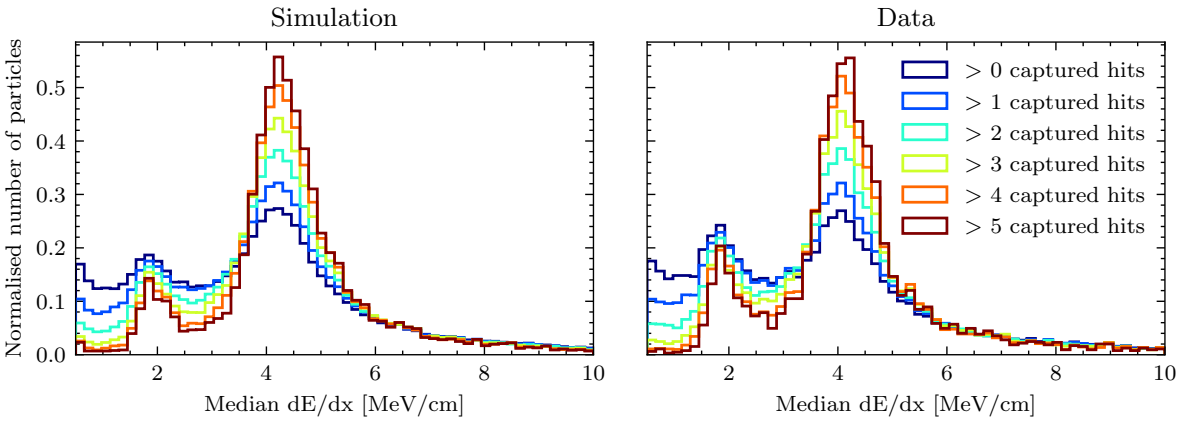

Figure 6.9: Median $\mathrm{d}E/\mathrm{d}x$ distributions for various cuts on the number of hits captured in the shower start in simulation (left) and data (right).

The configuration used for shower $\mathrm{d}E/\mathrm{d}x$ determination is presented graphically in figure 6.10. Based on the distribution of hits relative to the shower object, it can be seen that this particular shower has been reconstructed well. The shower start is placed correctly at the start of its track-like initial portion and the reconstructed shower direction follows the general direction of the hit cloud. A cylinder was placed around the shower start as described in this section, visible here in transparent grey. Compared to the shown shower length, which is representative of electromagnetic showers in ProtoDUNE, the cylinder is quite small. Although the entire shower contains 154 three-dimensional hits, only 4 were captured within the volume of the cylinder. It can be seen that only the hits at the very start of the shower are decidedly photon-like, being close to the characteristic 4 MeV/cm value. After a brief region of particle creation and increased energy deposition, the particles lose energy and spread out, yielding less energy per hit as the shower progresses.

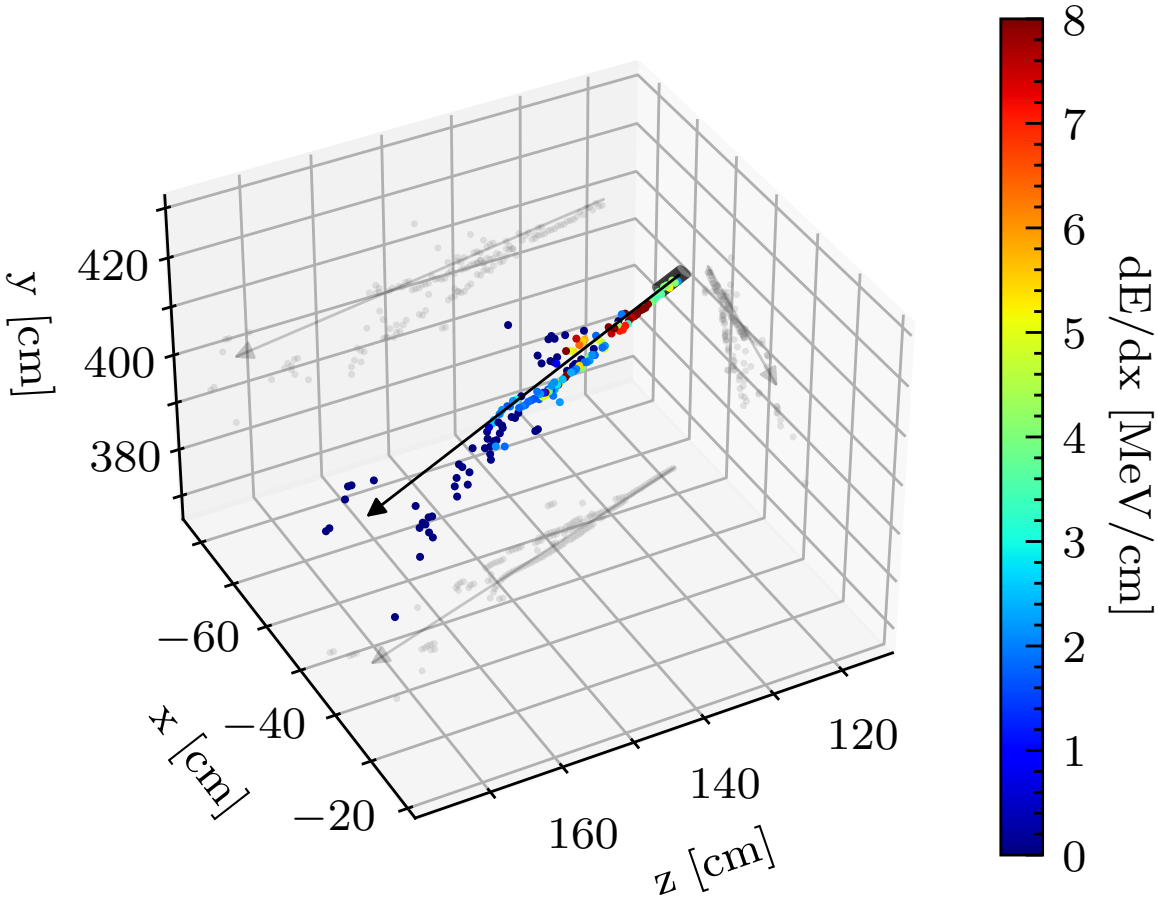

Figure 6.10: A shower in isolation, shown in three dimensions as a cloud of hits (coloured markers) and as a reconstructed object with a start point, direction and length (solid arrow). Two-dimensional projections of the shower are shown as transparent grey markers and arrows. The colour of the markers corresponds to their individual $\mathrm{d}E/\mathrm{d}x$ value. The axes follow the convention of ProtoDUNE, where the z-axis lies horizontally roughly in the beam direction and the y-axis points vertically up. A transparent grey cylinder for $\mathrm{d}E/\mathrm{d}x$ determination is shown around the shower start.

6.4 Results

Figure 6.11 shows the result of the method described in section 6.3, applied to the sample resulting from the cuts described in table 6.4. The doubly minimum-ionising peak characteristic of photon pair production is clearly visible around 4 MeV/cm, but the singly minimum-ionising peak at 2 MeV/cm is substantial as well, particularly in data. Possible reasons for this are listed in section 6.4.1. The purity of the sample is

highlighted by the rarity of non-photon objects, concentrating around and making up only a small fraction of the 2 MeV/cm peak.

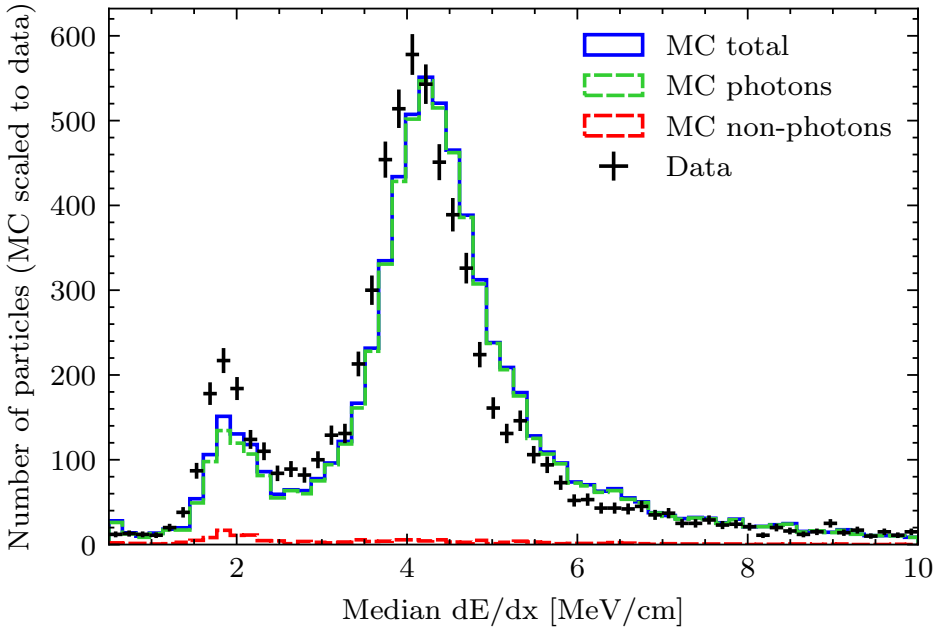

Figure 6.11: The median dE/dx of hits near the start of shower objects in the selected sample. The simulated distribution is split into its photon and non-photon components and all simulated distributions are scaled to the data by area.

It can be noted that the discrepancy between simulation and data in figure 6.2 is opposite that of the current analysis, shown in figure 6.11. This is the result of differing calorimetric constants, described in section 6.1. The positron study was concluded months before the current photon analysis and made use of currently outdated calorimetric methods.

6.4.1 Limitations of the method and conclusion

The method used for this analysis relies on the distribution of single hit dE/dx values. There are some cases in which this method does not produce a shower dE/dx value that differs from that of an electron shower. For instance, when a photon undergoes Compton scattering instead of producing an e^+e^- pair, the resulting shower is caused by a single scattered electron, necessarily resulting in low hit dE/dx values. However, it was found that Compton scattering photons represent a negligible fraction of the sample in simulation.

In another scenario, the electron and positron resulting from a photon's conversion may be produced nearly transversally to the photon's direction, causing them to diverge rapidly. In this situation, the two particles may not produce many or any hits in collaboration. Although the total charge deposited in the shower's cylinder would indicate its source to be a photon, this would not be visible from the individual hits' $\mathrm{d}E/\mathrm{d}x$ values. Figure 6.7 indicates that showers in data are substantially wider than those in simulation. Given the detector's finite resolution, this suggests that charge depositions from e^+e^- pairs may be split into two hits more often in data, resulting in a more substantial 2 MeV/cm $\mathrm{d}E/\mathrm{d}x$ peak.

Lastly, highly asymmetric e^+e^- production by a photon may cause one particle of the pair to deposit only a minimal amount of ionisation charge, again making the shower overall resemble an electron shower. Unfortunately, neither divergent e^+e^- pairs nor asymmetric e^+e^- production can be studied in the current sample, as the truth information of particles within electromagnetic showers was not recorded. This was a decision that arose from data volume considerations: a single shower may create hundreds of individual ionising particles. This information will be available in future large-scale reconstruction campaigns, opening up new analysis opportunities.

This chapter describes an analysis that characterises showers by their initial energy deposition, with a focus on photon showers. A comparison between the current photon study and a previous analysis concerning positron showers shows excellent separation between the two types of particles, providing a promising starting point for future $e^\pm/\gamma$ separation studies.

Chapter 7

Neutral pion reconstruction

Two principles vital to DUNE intersect in the π^0-particle. On one hand, the particle forms a major background to the neutrino appearance signal, requiring careful study in order to suppress its effects. On the other hand, its decay signature allows for the energy calibration of electromagnetic showers. The reconstructed π^0-particle mass is dependent on the overall energy scale of its decay photon's showers. Comparing the reconstructed value with the true particle mass thus yields a scale factor that can be used to correct the energy of showers in a data-driven manner. This is particularly important in DUNE, where the energy of a neutrino is reconstructed from its interaction products, among which are showers. This chapter will demonstrate the reconstruction of the π^0-particle mass in ProtoDUNE charged particle beam events, as well as its use in shower energy calibration.

7.1 The neutral pion

π^0-particles are common products of hadronic interactions. Consisting of $u\bar{u}$ and $d\bar{d}$ quarks (more precisely $\frac{u\bar{u}-d\bar{d}}{\sqrt{2}}$), these mesons are without electric charge or spin and have a mass just under $135\,\mathrm{MeV/c^2}$. Because of their extremely short lifetime ($8.4 \pm 0.6 \times 10^{-17}$ s), π^0-particles do not travel appreciable distances before decaying. Their signature decay is $\pi^0 \rightarrow \gamma\gamma$ with a branching ratio of 0.98823, which in LArTPCs ideally translates to two electromagnetic showers originating from the same vertex. Their next most likely decay is the Dalitz decay: $\pi^0 \rightarrow \gamma + e^- + e^+$ with a branching ratio of 0.01174, which creates up to three showers in the detector. Due to the rarity of

this decay, it will not be considered further in this chapter. The values given in this paragraph, as well as more information on the π^0-particle, can be found in [8].

Neutrino interactions lead to the creation of π^0-particles primarily through resonant and coherent production:

$$\begin{aligned} \nu N^{+/0} \rightarrow \nu \Delta^{+/0} \rightarrow \nu \pi^0 N^{+/0} \quad \text{(resonant)} \\ \nu A \rightarrow \nu A \pi^0 \quad \text{(coherent)}, \end{aligned} \tag{7.1}$$

where $N^{+/0}$ denotes a nucleon (proton or neutron) and A a nucleus. In resonant production, the neutrino interacts directly with a nucleon, exciting it to a Δ^+ or Δ^0 particle depending on the electric charge of the original nucleon. This excited state reverts to its original form by producing a π^0-particle. In coherent π^0-production, the neutrino excites the entire nucleus, which sheds its excess energy in the form of a π^0-particle through coherent scattering of its nucleons. Coherent π^0-production is characterised by the direction of the π^0-particle being very close to that of the interacting neutrino [107].

In ProtoDUNE, π^0-particle production is induced by electrically charged beam particles instead of neutrinos. Of particular note is the charge exchange reaction, in which a charged pion interacts elastically with a nucleon in the detector material as follows:

$$\pi^+ n \rightarrow \pi^0 p \quad , \quad \pi^- p \rightarrow \pi^0 n, \tag{7.2}$$

where interactions involving positively charged pions are expected to be much more common in the ProtoDUNE data, given that the selected beam only contained positively charged particles. Nonetheless, a significant number of negatively charged pions are produced in interactions between beam particles and argon nuclei in the detector. As in the case of neutrino interactions, π^0-particles are also a common product of nuclear excitation.

7.1.1 Kinematics

Figure 7.1 shows the kinematics of a π^0-particle decay into two photons. In the rest frame of the π^0-particle, each of the two decay photons receives exactly half of the π^0-mass as energy. In order to conserve the system's momentum, they are also emitted

in exact opposite directions. Due to the π^0-particle having a spin of 0, in its rest frame the photon emission is otherwise entirely isotropic. In the lab frame, where the π^0-particle has a certain relativistic momentum, the photons resulting from its decay are boosted in the direction of their parent particle. They therefore receive an energy that is dependent on the angle between the forward-facing photon and the π^0-direction θ^*:

$$E_{1,2} = \frac{\gamma m_\pi}{2} \left(1 \pm \beta \cos\theta^*\right), \tag{7.3}$$

where E_1 and E_2 correspond to the most and least energetic photon respectively, β is the velocity of the π^0-particle relative to the speed of light and $\gamma = 1/\sqrt{1-\beta^2}$ is the particle's Lorentz factor. The special cases $\beta = 0$ and $\theta^* = 90^\circ$ both result in $E_1 = E_2 = \gamma m_\pi / 2$.

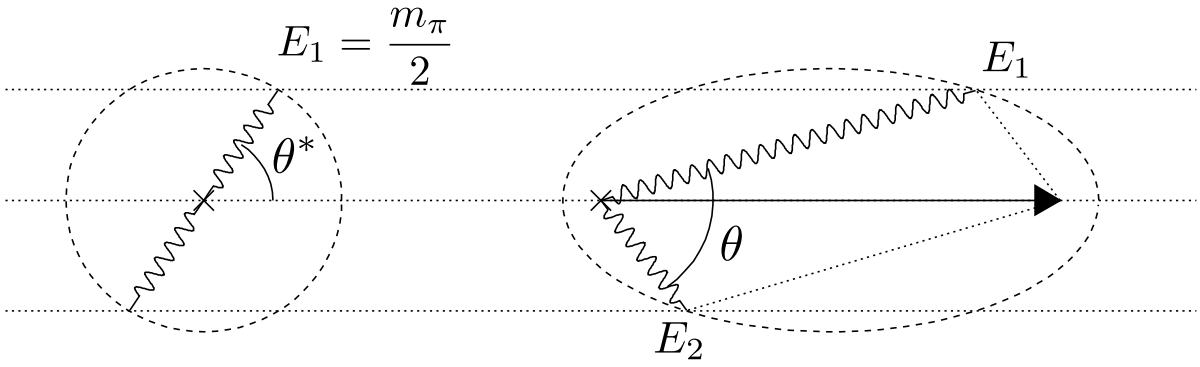

Figure 7.1: A visual representation of π^0-particle decay kinematics. Shown is a $\pi^0 \to \gamma\gamma$ decay process in the π^0-particle rest frame (left) and in the lab frame (right) where the particle has some momentum. Taken from [108] and edited for clarity.

The angle between the two decay photons can be expressed in terms of the original π^0-particle's energy as follows:

$$\cos\theta = 1 - \frac{2m_\pi^2}{E_\pi^2(1-\alpha^2)}$$
$$\text{where} \quad \alpha = \frac{E_1 - E_2}{E_1 + E_2} = \beta \cos\theta^*. \tag{7.4}$$

Here, α is the asymmetry between the two photons, expressed in terms of both photon energy and lab frame angle. As the energy of the initial π^0-particle increases, it can be seen that the opening angle decreases, asymptotically nearing 0.

From the energy and direction of π^0-decay photons, the invariant mass of the π^0-particle may be determined. Since photons are massless, this quantity can be described quite simply:

$$m_\pi^2 = 2E_1E_2\,(1-\cos\theta)\,. \tag{7.5}$$

The π^0-particle mass is independent of the particle's kinetic energy and direction of its decay photons. It therefore offers a data-driven verification and calibration of the reconstructed quantities, in particular the reconstructed energy of the photons.

7.1.2 Neutral pions in ProtoDUNE

The π^0-particle is electrically neutral and extremely short-lived. Decaying almost always into two photons, the showers that these photons produce are the only trace that π^0-particless leave in the detector. The general goal of this reconstruction study is therefore to detect both showers originating from a π^0-particle.

Two aspects of the π^0-decay complicate matters. Firstly, showers have a chaotic hit distribution compared to tracks. Since they are made up of tens or even hundreds of individual particles that all create charge deposits, showers are more cloud-like than track-like. Photons produced within the shower also have a chance to travel undetected for substantial distances before creating a sub-shower of their own. The disconnected cloud-like nature of showers makes it a challenge to properly catch all hits within the same object and to assign a clear start and direction to the shower. This has implications for the energy reconstruction of individual showers, but also the invariant mass reconstruction of the original π^0-particle, which relies on the angle between the two showers as well as their energy.

Secondly, photons originating from a π^0-decay travel undetected until they interact with the detector material. This separates π^0-decay showers from their photons' origin as well as from each other, making it more difficult to recognise that showers have come from a certain vertex. Photons that travel very far may create showers that Pandora fails to identify as daughters of their true origin.

In ProtoDUNE, π^0-particles are produced in a variety of interactions, including those from cosmic rays, but they are most consistently produced at detectable energies in beam particle interactions. Selecting π^0-particles that have their origin in beam

interactions also ensures that their decay showers are likely to stay within the detector's active volume. For these reasons, this π^0-particle study starts at the first interaction of the primary beam particle.

7.2 Shower energy reconstruction

Chapter 6 examined only the very start of the showers in the sample. In order to reconstruct the energy of a photon from the shower it creates, however, the entire shower must be taken into account. Previously, the energy deposition of a single hit was determined by an exponential relation described in equation 6.4. Using this method, it is implicitly assumed that the particle creating the charge deposit is travelling in a well-known direction, namely the overall shower direction. This approximation holds well for the initial part of the shower, which is track-like. It is not accurate for the chaotic main body of the shower, however, where particles may travel in different directions. For this reason, the method for shower energy reconstruction used in this analysis is a simpler linear relation with the deposited charge:

$$E = \frac{QW_{\text{ion}}}{f_c R} = \frac{W_{\text{ion}}}{f_c R} \sum_{\text{hits}} \sigma(\vec{x}_h) Q_h, \tag{7.6}$$

where Q and Q_h denote the charge of the shower as a whole and that of a single hit respectively. f_c is the same run-dependent calibration factor used in chapter 6 and $\sigma(\vec{x}_h)$ is the same run and position-dependent correction factor. Section 6.1.1 gives a more detailed description of these parameters. W_{ion} is once again the energy required to ionise a single atom. Any charge losses are captured in R. This factor includes recombination effects discussed in section 6.1, but also any inefficiencies of the clustering algorithm.

Figure 7.2 shows the distribution of the factor R for all simulated showers in the sample after applying the cuts shown in section 7.4. In this figure, R was determined using equation 7.6, taking E to be the particle's true energy and Q to be the total charge as seen by the reconstruction. The distribution is strongly peaked around one central value, but a large portion of the sample deviates from this value. It is unlikely that different showers in the same sample have wildly varying recombination factors. Rather, this highlights the inherent difficulties of shower hit clustering, creating objects

that are incomplete or composed of multiple showers. In this analysis, $R = 0.575$ was taken as a constant for all MC showers.

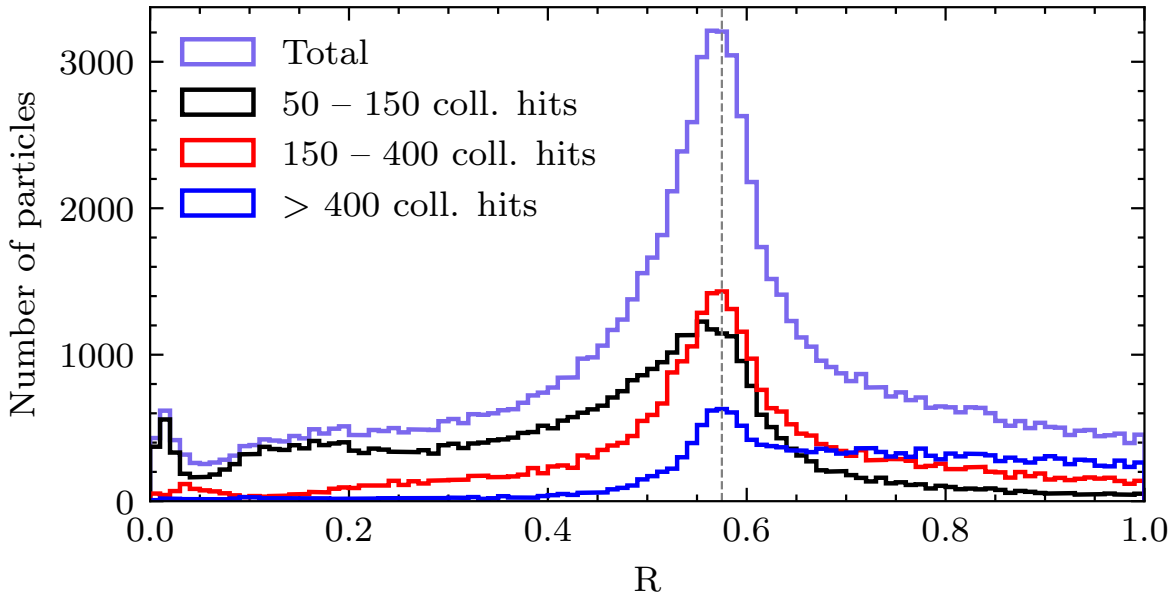

Figure 7.2: The R factor of simulated reconstructed objects in the sample. The vertical grey line denotes the mode of the distribution, with a value of $R = 0.575$.

Since the distribution of figure 7.2 is based on truth information, it cannot accurately be recreated for showers in data. Indeed, finding an accurate link between the deposited charge in a reconstructed shower object and the true energy of its parent particle is one of the goals of this analysis.

7.3 Simulation studies

Simulated ProtoDUNE events can give an indication of the π^0-particle sample in data. In this section, various parameters will be investigated that can only be observed by considering the truth information of simulated events. Only π^0-particles produced within the direct vicinity of the beam particle end (< 1 cm) are considered to make sure that their photon showers should be among the direct reconstructed beam daughters.

Firstly, the number of π^0-particles per beam particle can be considered. Figure 7.3 shows the distributions corresponding to this metric, separated by beam momentum. In general, the number of π^0-particles goes up with the available energy from the

beam particle. This might seem like a positive effect, increasing the sample size of π^0-decay shower candidates, but in general the presence of multiple π^0-particles in a single event greatly hinders individual π^0-particle reconstruction. Confusion between shower objects from different π^0-particles will prove to be one of the main backgrounds in the analysis in the following sections.

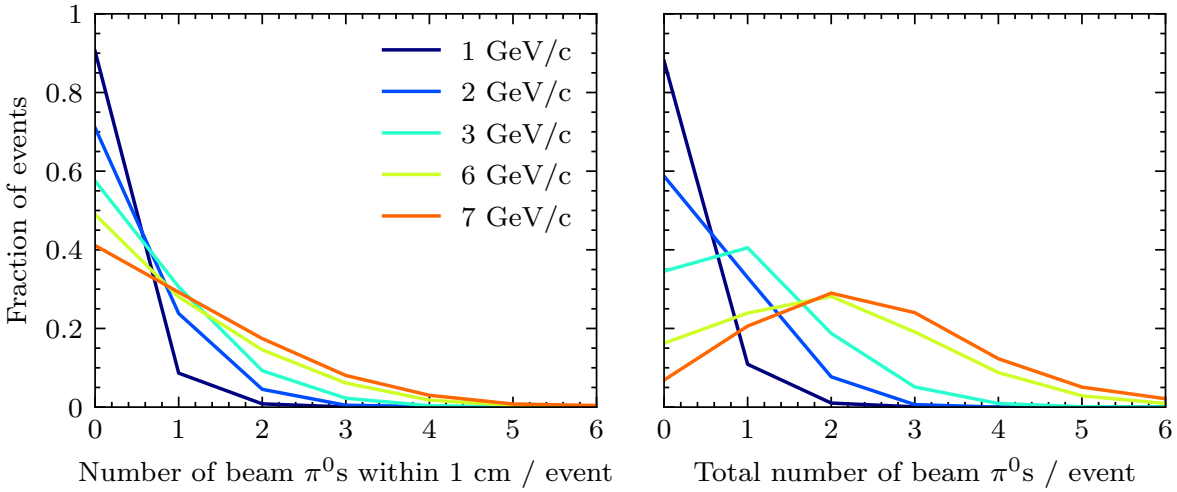

Figure 7.3: The number of π^0-particles resulting from a beam interaction within 1 cm from the beam track end (left) and in total (right). Distributions are shown by beam momentum.

The energy of the π^0-particles and their photon daughters is shown in figure 7.4. Since the π^0-particle has a rest mass of 135 MeV/c^2, its energy spectrum starts at this value. It can be seen that the energy distributions change only minimally between the various beam momenta, invariably peaking below 250 MeV. For the reconstruction of a π^0-particle, both of its photon daughter showers must be reconstructed. However, its energy is distributed asymmetrically over its two daughters: the subleading photon energy distribution peaks at just 50 MeV, which makes these showers a great challenge to reconstruct. Due to the nebulous nature of electromagnetic showers in liquid argon, smaller shower objects may not be recognised by the clustering algorithm as being separate from the noise that accompanies the energetic main event. This effect will be shown in later figures as well.

In order to analyse π^0-particles in data, their photon daughters must first have been reconstructed as shower objects. Here, a particle is said to "be reconstructed" if it is the primary contributor to any reconstructed object, described in more detail in section 5.3.1. The reconstruction efficiency, in turn, describes the fraction of all particles

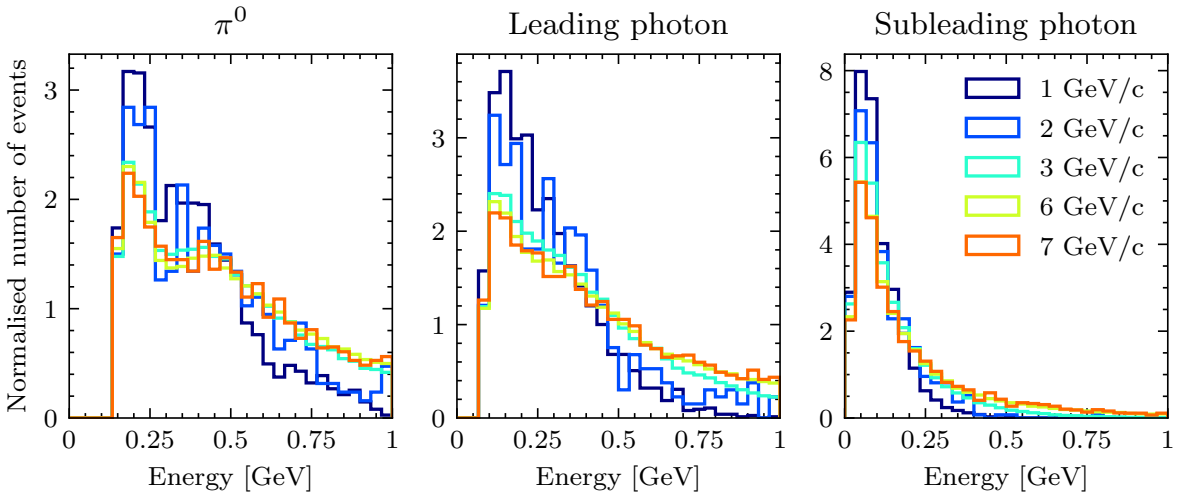

Figure 7.4: Energy distributions for π^0-particles resulting from beam interactions (left) and their decay photons, split between the leading photon (middle) and subleading photon (right). Distributions are shown by beam momentum.

in a sample that is reconstructed, or how likely any given particle in the sample is to be reconstructed. Figure 7.5 shows the reconstruction efficiencies of π^0-particles and their decay photons in the sample as a function of energy. Since the π^0-particle itself is invisible, it is said to be reconstructed if both of its daughter photons are. It can be seen that the leading photon follows a predictable rise in reconstruction efficiency as its energy grows. The subleading photon follows suit until approximately 200 MeV, after which it suddenly decreases in efficiency again. This is due to it starting to overlap with the leading shower: if the efficiency is examined as a function of the photon opening angle, also shown in figure 7.5, it can be seen that the subleading photon reconstruction efficiency decreases drastically for lower angles.

Beyond the question of whether a particle is reconstructed at all, it is useful to determine how well it is reconstructed. This is done by the completeness and purity metrics defined in section 5.3.1. Figure 7.6 shows the distributions of these values for π^0-photons in the sample that have been reconstructed. It can be seen that the purity of these reconstructed objects is good, with a large peak at a purity of 1. Completeness plays a direct role in the energy reconstruction, and as such it is important to create a sample in which the completeness varies little and is preferably at a high value. The complete sample features a dual peak structure for the leading photon and a very broad near-constant distribution for the subleading photon. Both are sub-optimal for the determination of the original photon's energy.

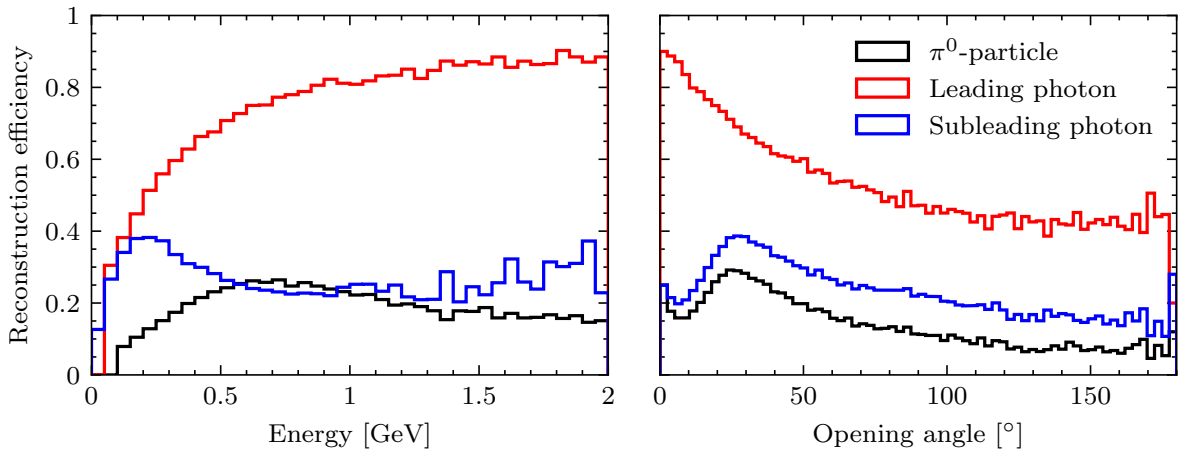

Figure 7.5: The reconstruction efficiency of π^0-particles and their daughter photons in the sample as a function of their energy (left) and the photon opening angle (right).

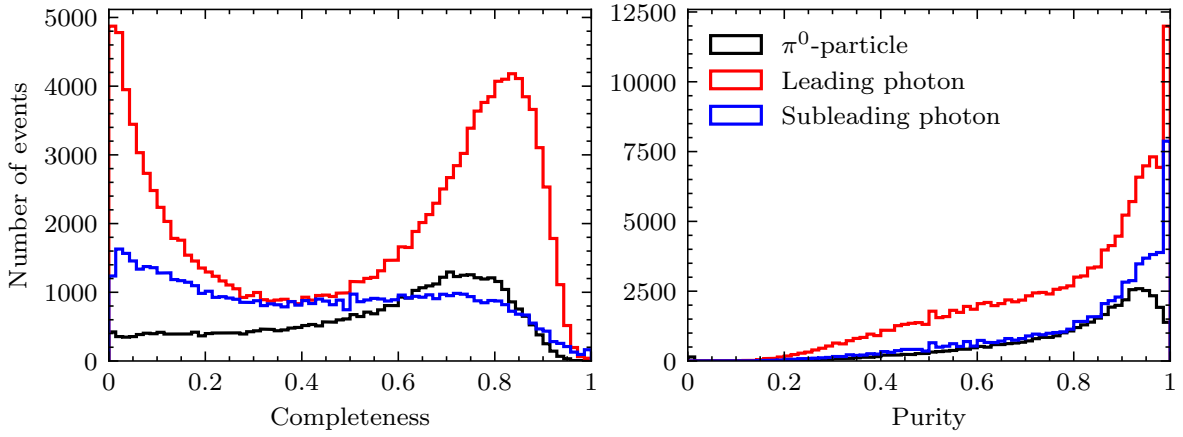

Figure 7.6: Completeness (left) and purity (right) distributions for reconstructed photons coming from π^0-decay. These metrics were also computed for π^0-particles of which both photons were reconstructed by weighting the values of their two daughter photons by their number of hits and averaging them.

Both the completeness and purity are highly dependent on the size of the reconstructed shower. Small showers are likely to have low completeness since their shape tends to be less well-defined, which poses a challenge to the hit clustering algorithm. In contrast, as electromagnetic showers become larger, they start overlapping with other objects, pushing their purity down. Figure 7.7 shows these effects graphically. It is useful to link truth and reconstructed information in this manner, so that well-reconstructed objects may be selected with high certainty even in data, where truth

information is not available. In this case, more complete photon showers may be selected by requiring a minimum number of hits.

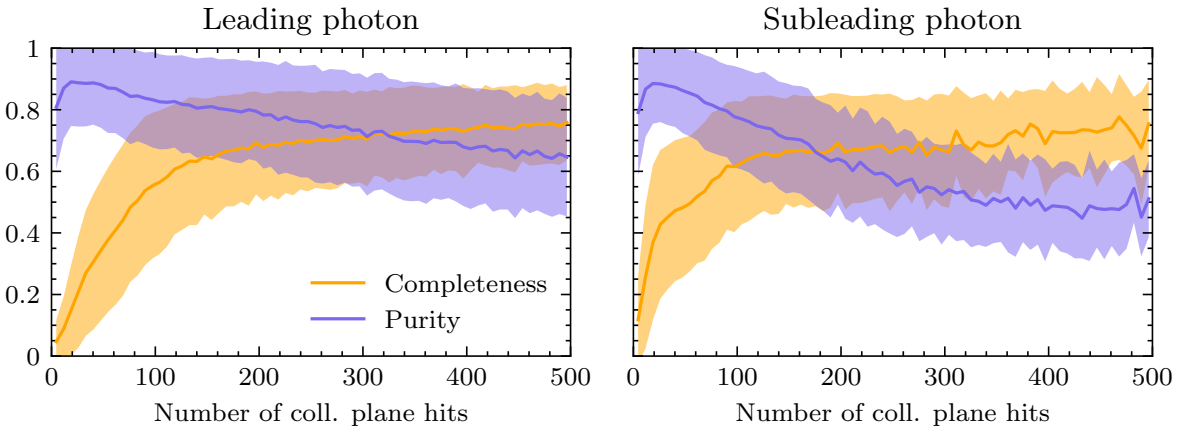

Figure 7.7: The completeness and purity of leading (left) and subleading (right) π^0-decay photons as a function of the number of hits in the reconstructed object. The colour bands signify a 1σ spread of the values in each bin.

In order to investigate the quality of the photon energy reconstruction, a useful metric is the relative difference between reconstructed and true photon energy $(E_{\text{reco}} - E_{\text{MC}})/E_{\text{MC}}$, where the reconstructed photon energy is derived from the process described in section 7.2. Once again, the metric may be considered as a function of the shower size, shown in figure 7.8. It can be seen that small objects feature energies that fall well below the original photon's energy due to their low completeness. Larger shower objects exhibit a surplus in energy that can be explained by the decreasing purity. As shower objects are supplemented with hits from other particles, their reconstructed energy rises to values above that of the original photon. Despite this, a clear peak is visible around a relative energy difference of 0, especially for showers containing 50 or more collection plane hits.

The opening angle between the two π^0-daughter photons provides the final ingredient for the reconstruction of the π^0-particle mass. To obtain an accurate opening angle, the direction of each shower object must represent that of its photon parent well. Similarly to figure 7.8, figure 7.9 shows the angle between π^0-daughter photons and their associated reconstructed objects as a function of the number of collection plane hits contained within the shower object. It shows remarkable similarity between the leading and subleading π^0-daughter photons, apart from the lacking large subleading showers. Both distributions are relatively well-peaked towards small angles, although the mean of the distribution never drops below 15°. In general, few showers are recon-

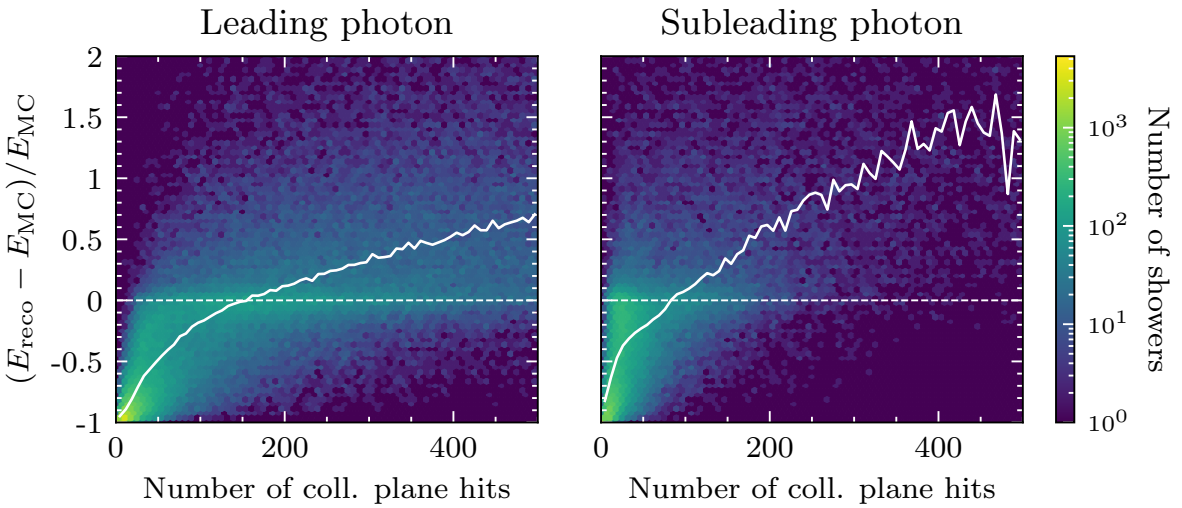

Figure 7.8: Distributions containing reconstructed leading (left) and subleading (right) π^0-decay photons. The relative difference between reconstructed and true energy $(E_{reco} - E_{MC})/E_{MC}$ is shown as a function of the number of collection plane hits in the reconstructed shower. Solid white lines denote the columnar mean and dashed white lines signify the desired value of 0.

structed backwards, although there is some tendency towards this at lower numbers of collection plane hits.

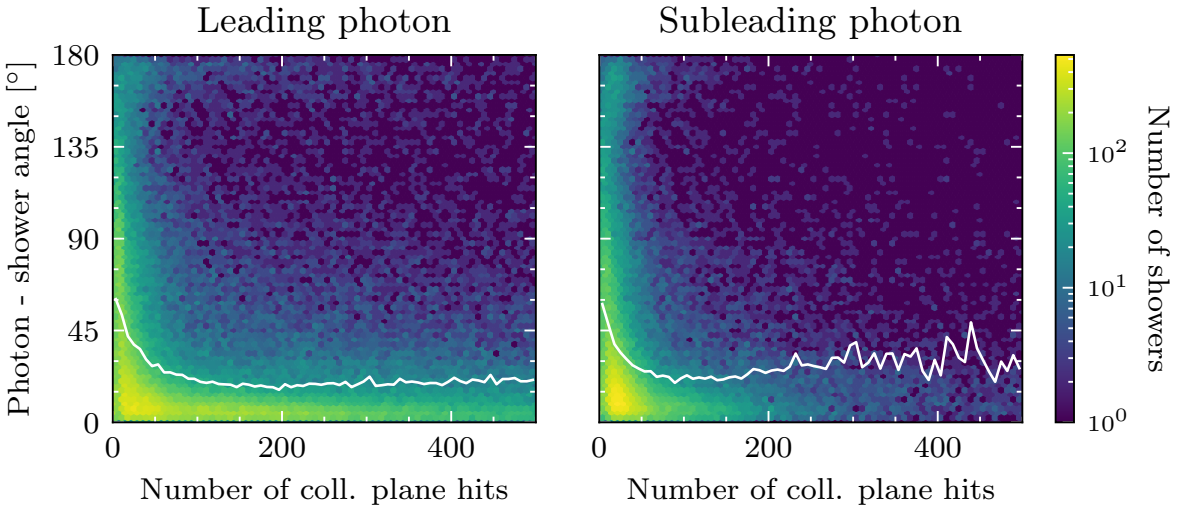

Figure 7.9: The three-dimensional angle between π^0-daughter photons and their associated shower objects, shown for the leading (left) and subleading (right) π^0-decay photons. The solid white line represents the columnar mean.

The shower direction used in figure 7.9 is taken from the shower object created by the Pandora reconstruction algorithm. It is determined from a principal component analysis (PCA) of all the hits contained in the object. Other ways to determine the shower angle exist, two of which are explored in figure 7.10. Both are derived from the interaction vertex of the beam particle, which is defined as the end of the beam track. In the first, the direction of the shower is defined by the vector between the beam interaction vertex and the start of the shower. The second method relies on the vector between the interaction vertex and the shower centre of charge, which is the average position of all hits in the shower weighted by charge. The latter method has the advantage that it bypasses possible misidentified shower start positions. It can be seen that the method involving the interaction vertex and shower start performs poorly in the leading photon sample, whereas the method utilising the shower CoC performs best of all. This suggests that a large fraction of the sample's showers have a start position that is misplaced, or that the interaction vertex is not in the right place. Subleading photons, in contrast, perform best under the method that uses the vertex position and shower start, although the ranking is much less clear in their case. In general and as expected from previous observations, subleading photons are reconstructed less well than their leading counterparts.

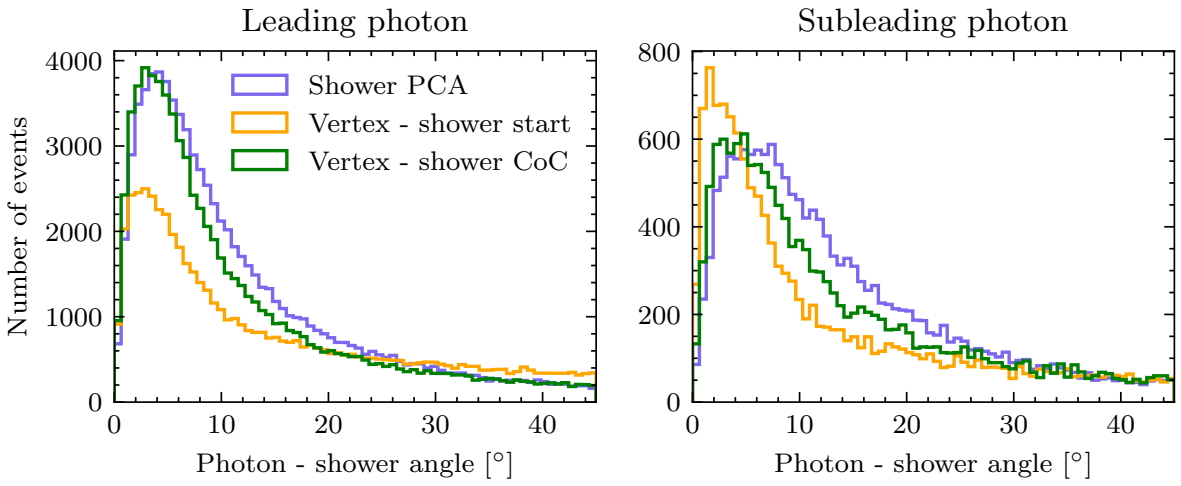

Figure 7.10: The angle between π^0-decay photons and their associated shower objects, shown separately for the leading (left) and subleading (right) photon. Three different methods of determining the shower direction are shown: the default principal component analysis (PCA) value contained in the shower object, the direction of a vector between the beam interaction vertex and shower start, and the direction of a vector between the beam interaction vertex and the shower centre of charge (CoC). Only showers containing > 50 collection plane hits are shown.

There are multiple ways to judge the quality of the reconstructed opening angle θ between two π^0-daughter photon showers. Figure 7.11 shows two of these: an absolute difference between reconstructed and true opening angles as well as a metric that is more applicable to the reconstruction of the π^0-particle invariant mass. C, here defined as $1 - \cos\theta$, appears directly in the invariant mass calculation defined in equation 7.5. It can be seen that both distributions peak around 0, but also that they are both very wide. The relative C difference distributions involving the beam interaction vertex feature pronounced peaks near -1, suggesting that some shower pairs are close together in terms of start position or centre of charge, but come from photons that are angled further apart from one another.

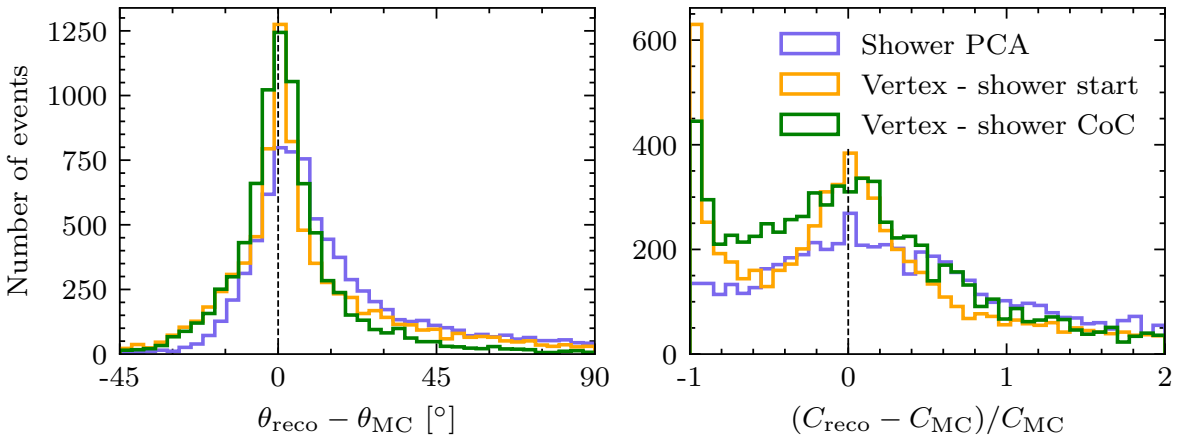

Figure 7.11: The absolute difference between reconstructed and true opening angles (left), and the relative difference between reconstructed and true $C = 1 - \cos\theta$ (right). Reconstructed opening angles were derived from methods described in figure 7.10, whereas the MC opening angle is defined as the true angle between photons. Only showers containing > 50 collection plane hits are shown.

The origin of badly reconstructed opening angles can be explored by examining the relative C difference as a function of the reconstructed opening angle. This is done in figure 7.12 (top). It can be seen that reconstructed opening angles below $\sim 10^\circ$ have a high tendency of being badly reconstructed in all cases. Taking only shower pairs into account with opening angles larger than 10°, figure 7.12 (bottom) revisits the performance metrics shown in figure 7.11. Comparing the two figures, it can be seen that the accuracy of the opening angle improves greatly with the applied cut. Both methods involving the beam interaction vertex now perform better than the method that uses the default PCA values for the shower directions.

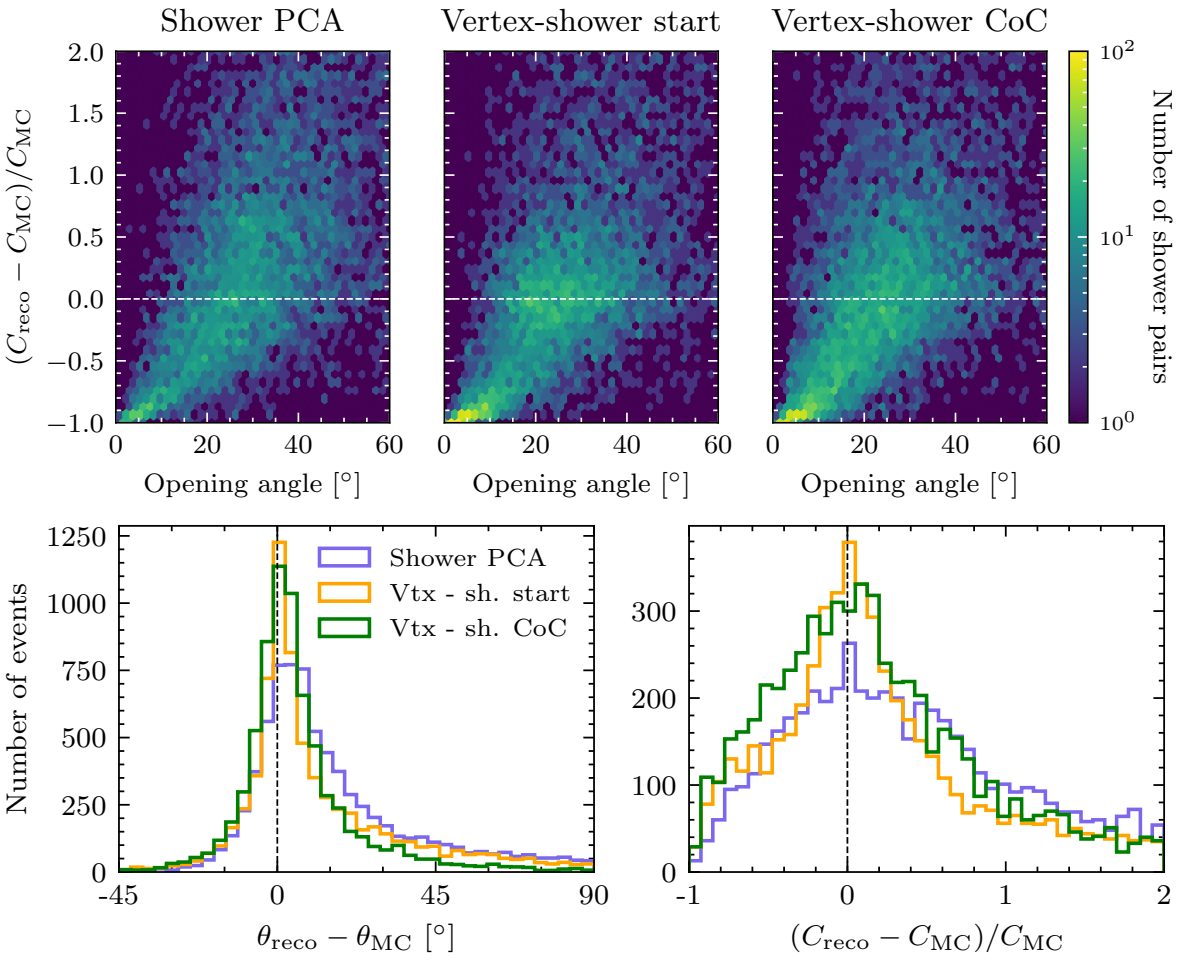

Figure 7.12: Top: the relative difference between reconstructed and true $C = 1 - \cos\theta$ as a function of the opening angle θ_{reco} for various shower direction reconstruction methods. Bottom: the absolute difference between reconstructed and true opening angles (left), and the relative difference between reconstructed and true C. Only shower pairs that contain more than 50 collection plane hits per shower and span an opening angle greater than 10° are shown.

Using the energy and opening angle reconstruction, a best-case scenario can be sketched for the invariant π^0-particle mass. Figure 7.13 shows the distribution for this value according to equation 7.5. The three methods of determining shower direction discussed above are shown in the figure. It can be seen that all centre around the true π^0-particle mass, but the two methods that hinge on the beam interaction vertex perform slightly better. In order to quantify the results for the three methods, a Gaussian fit was performed within the range $135 \pm 50\,\mathrm{MeV/c^2}$. The results of this fit, as well as some general statistics describing the distributions can be seen in table 7.1. It can be seen that the mode of each distribution is on or very close to the true value,

but the Gaussian fit mean reveals that both shower direction determination methods involving the interaction vertex underestimate the invariant mass.

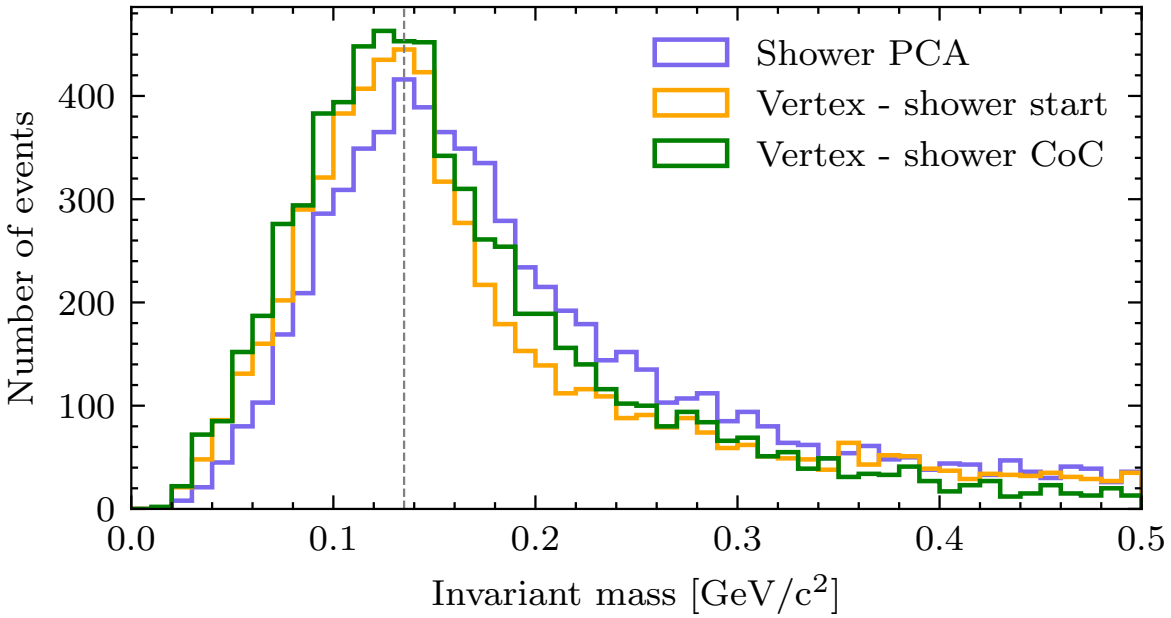

Figure 7.13: The invariant mass determined from shower pairs truth-matched to π^0-daughter photons. The shower pairs shown here are required to have an opening angle greater than 10° and contain at least 50 collection plane hits per shower. The results of three different methods to determine shower direction are shown. The true π^0-particle mass is shown with a vertical dashed line.

Table 7.1: Various statistical parameters for the distributions shown in figure 7.13. Gaussian parameters are shown for a fit applied within the range 135 ± 50 MeV. Also shown are general statistical parameters concerning the distributions.

	Gaussian fit [MeV/c^2]		General stats. [MeV/c^2]		
Method	μ	σ	Mean	Median	Mode
Shower PCA	141 ± 2	56 ± 5	293	183	135
Vertex - shower start	127 ± 1	39 ± 2	260	157	135
Vertex - shower CoC	126 ± 2	45 ± 3	193	147	125

It is important to keep in mind that this section considers reconstructed shower objects that have been linked to π^0-daughter photons using truth information. It shows best-case scenarios, in which it is assumed that all photons that have substan-

tial associated reconstructed objects are correctly identified and matched with their sister π^0-showers. In data, achieving a 100% efficient shower selection is impossible. Additionally, this section only considers π^0-particles created within 1 cm of the beam interaction vertex. Particularly at higher beam momenta, π^0-particles may be created by secondary particles, as shown in figure 7.3. Not only does this introduce more shower objects that cause confusion in pairing π^0-daughter showers, it also casts doubt on any methods relying on the beam interaction vertex. Although the Pandora reconstruction chain attempts to construct a robust particle hierarchy, the invisibility of photons and chaotic nature of showers could link tertiary showers directly to the primary beam interaction vertex.

7.4 Photon pair selection

In order to reconstruct π^0-particles from their photon showers, a pure photon shower sample must first be established. The procedure for this is similar to that described in chapter 6, where a pure photon shower sample was likewise sought. The same data sample was used, described in section 5.4, as well as many of the cuts outlined in section 6.2. Fundamentally different is the present need for shower pairs, where at least two well-reconstructed showers per event are required. As shown in figure 7.5, the reconstruction efficiency for the subleading photon is rather low, and figures 7.8 and 7.9 show that the subleading photon is usually small in size. The cuts outlined in chapter 6 can therefore be used, but must be relaxed to allow for more shower pairs to form. In this section, various selection variables are explored, with comparisons between the real and simulated samples.

Although the number of hits at the start of a shower is a metric useful for $\mathrm{d}E/\mathrm{d}x$ studies, as explored in section 6.3, it has also proved convenient in selecting showers that have higher quality energy and direction reconstruction. As can be seen in figure 7.14, both the relative energy difference and angular difference between reconstructed objects and their true parent particles decreases with more hits at the shower start. On one hand, this can be explained by the correlation with total shower size: larger showers generally contain more hits at the shower start and have a better average reconstruction quality. On the other hand, shower objects with few or no hits near the shower start necessarily have a misplaced start position, misreconstructed direction and/or missing hits. For the purposes of this analysis, requiring more than one hit

near the shower start provides an adequate balance between reconstruction quality and sample size.

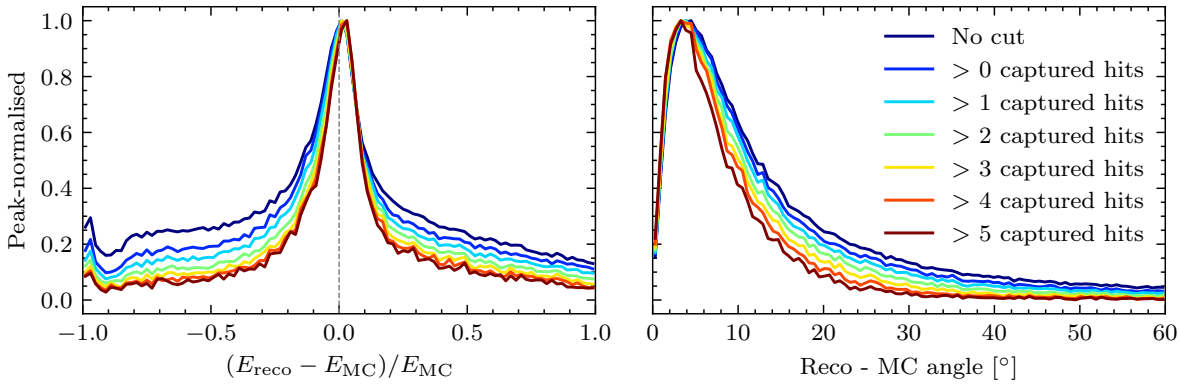

Figure 7.14: The relative energy difference (left) and direction difference (right) between reconstructed objects and their true parent particles. Shown here are reconstructed objects that are classified as shower by the Pandora reconstruction algorithm, have a CNN score greater than 0.6 and contain more than 50 collection plane hits.

The individual reconstructed object cuts of this analysis are summarised in table 7.2. Additional details concerning individual reconstructed object variables in this table are listed in section 6.2. When compared to table 6.4, it can be seen that the purity of photons is slightly lower in this sample, but the efficiency is much higher. This was attained by relaxing the cut on the CNN shower-like score and number of shower start hits, as well as the omission of the requirement for a shower to start some distance into the detector volume. The increased sample size is vital for applying subsequent cuts to the sample of shower pairs.

In order to form a shower pair, an event needs to be shared by at least two adequately reconstructed showers. Applying the cuts listed in table 7.2 to the simulation and data samples yields a number of showers per event, shown in figure 7.15. Unfortunately, the majority of beam events does not contain any showers passing the cuts, especially at lower momenta: low energy beam particles generally produce fewer beam particles that are also smaller in size. The number of 1 and 2 GeV/c beam events containing at least two showers is so low that it renders events with these beam momenta unusable in this analysis. Aside from differences between various beam momenta, there is a clear discrepancy between simulated and real events, where the data in all cases features fewer showers per event on average. This can partly be attributed to the object-level cuts of table 7.2, in which a clear difference arises in favour

Table 7.2: The cuts applied to the sample of beam track daughter particles. The purity and efficiency of the cuts are shown cumulatively, meaning that values include all cuts up to and including the relevant row. Also shown are cumulative and individual cut MC/data bias fractions. Since the purity and efficiency quantities use truth information, only values from simulation are shown.

			Data/MC ratio	
Cut	Purity	Efficiency	Cumulative	Individual
	0.28	1.00	0.50	
Pandora shower class.	0.47	0.88	0.50	1.02
CNN shower-like score > 0.6	0.68	0.82	0.52	1.03
Number of coll. hits > 50	0.85	0.47	0.48	1.00
Shower start hits > 1	0.89	0.21	0.44	0.97

of simulated events. Especially the requirement of two or more hits near the shower start contributes to this discrepancy. Another factor in the data-simulation discrepancy is the initial number of beam-daughter objects per event, shown in figure 5.16. It can be seen that simulated events contain significantly more beam daughter particles than real events, even without applying bias-introducing cuts. A data-simulation difference is therefore inherent to the sample.

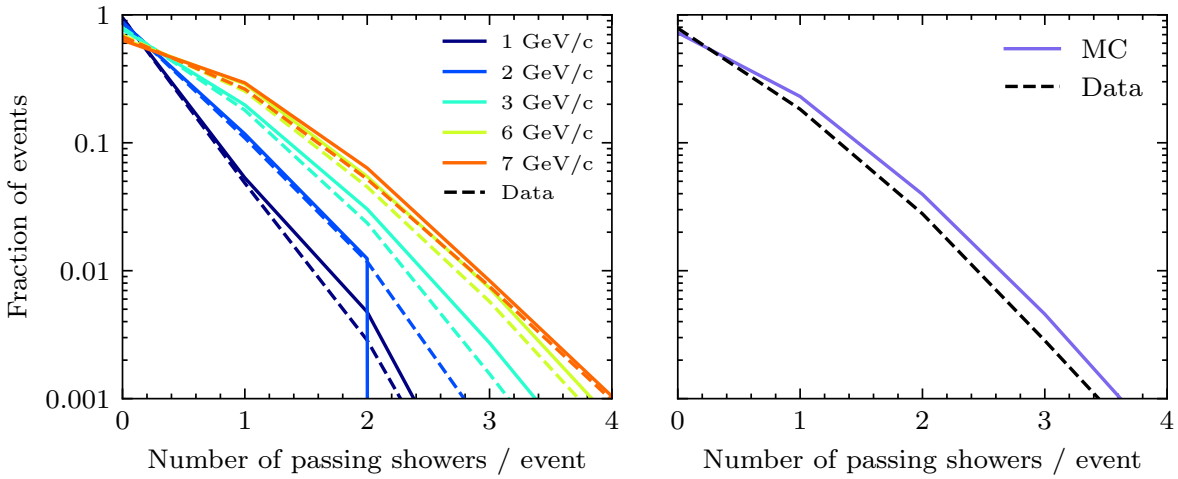

Figure 7.15: The number of showers per event passing the cuts in table 7.2, shown by beam momentum (left) and in total (right).

Events containing more than two showers give rise to an ambiguity in matching showers to form pairs. In this analysis, if an event contains more than two showers, the two most energetic showers are selected and the rest are discarded. This results in minimal loss of information since the number of events containing three or more showers is well below 1% of the total sample size. More sophisticated methods exist to resolve this ambiguity. However, there is little to gain from refining this selection. Since more energetic showers generally have a higher reconstruction quality, selecting the two most energetic showers of an event preserves the most valuable information for this analysis.

Once shower pairs are formed, it is useful to check whether they correspond to π^0-decay photons in simulated events. Table 7.3 shows that for all beam momenta, the majority of shower pairs are associated with two true π^0-decay photons. Having said that, as the beam momentum increases, one can see a significant rise in the fraction of shower pairs associated with at least one true non-photon. It is to be expected that more energetic interactions produce more contamination in the sample. More striking and more difficult to mitigate is the fraction of shower pairs that originate from different true π^0-particles. This fraction increases dramatically with the beam momentum, going from 18% at 1 GeV/c up to 69% at 7 GeV/c. The explanation for this is the average number of π^0-particles produced at each beam momentum, shown in figure 7.3, which naturally rises as the beam momentum increases. An ideal shower pair sample would be made from an enormous 1 GeV/c hadronically triggered beam run, where there is least chance of confusion between multiple π^0-particles. However, the 1 GeV/c run in the present sample is too small to provide a meaningful number of shower pairs.

In order to determine the opening angle of a shower pair, it is useful to consider the shower direction reconstruction quality of the current sample of beam daughter particles after applying table 7.2's cuts. Section 7.3 describes three distinct methods to determine the shower direction, the performance of which when applied to the beam daughter sample is shown in figure 7.16. It can be seen that much of this looks familiar: the vertex-based direction reconstruction methods perform better or worse than the default shower PCA depending on whether the shower is leading or subleading, just as was observed for true π^0-decay photon showers in section 7.3. Of more importance is the C parameter, which shows a previously unobserved peak near -1 for vertex-based direction reconstruction methods. It is likely that these shower pairs were misplaced

Table 7.3: Truth information concerning shower pairs in the sample. Aside from the total pair count in simulation and data, the simulated sample is divided into several fractions. The first denotes whether at least one true particle associated with the shower pair is a non-photon. The second shows the fraction of shower pairs that originate from the same photon. The third indicates the shower pairs that have two associated true π^0-decay photons coming from two different π^0-particles. The last fraction indicates the correctly matched showers, which have associated true photons originating from the same π^0-particle.

Beam mom.	Non-photon	Same photon	Different π^0s	Correctly matched
1 GeV/c	6.56 %	1.64 %	18.0 %	73.8 %
2 GeV/c	14.3 %	0.0 %	35.7 %	50.0 %
3 GeV/c	12.2 %	2.08 %	52.1 %	33.1 %
6 GeV/c	20.2 %	2.57 %	63.4 %	13.3 %
7 GeV/c	17.8 %	2.29 %	69.1 %	10.0 %

in the particle hierarchy, being listed as direct beam track daughter objects, but in reality originating from a different vertex. In order to avoid this effect, shower PCA was selected as the method of choice for determining the opening angle of shower pairs.

The object-level cuts in table 7.2 pose no upper limit on the shower energy, making it possible to obtain shower pairs that create highly unlikely π^0-particle energies. Figure 7.17 shows that correctly matched shower pairs mostly manifest energies within the 0.15 - 1.5 GeV range, in agreement with the true π^0-particle energy spectrum shown in figure 7.4. Photons originating from different π^0-particles, on the other hand, have a much broader range, allowing for a fraction of this background to be cut out. It is important to note that the shower energy in data is the subject of calibration in this analysis. Any cut on this variable must therefore be applied with caution in order to avoid treating simulation and data differently. It can be seen that the total energy spectrum in simulation and data agree quite well, although the energy in data seems to be slightly lower than that in simulation.

As described in section 7.1.1, the opening angle between showers depends on the initial π^0-particle energy. Since the energy spectrum of π^0-particles in the sample is limited, it is also expected that the opening angle of correctly matched shower pairs is not uniform across the entire range. Figure 7.18 shows this effect. Incorrectly

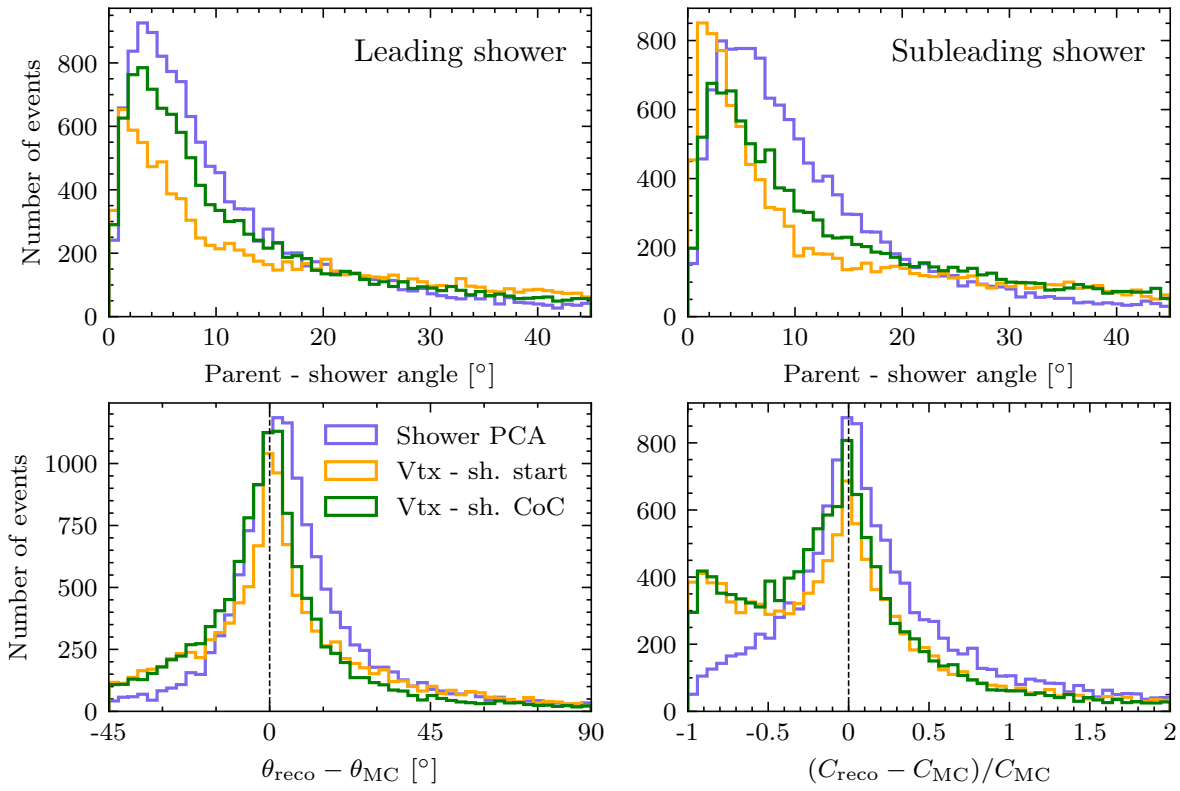

Figure 7.16: A summary of the direction reconstruction quality of beam track daughter object pairs. Shown are the leading (top left) and subleading (top left) shower angles to their true parent particles. Also shown are difference between reconstructed and true opening angle (bottom left), as well as the relative difference between reconstructed and true $C = 1 - \cos\theta$ (bottom right).

matched shower pairs are only slightly biased towards smaller opening angles due to the overall boost imparted to the system by the beam particle. However, photon pairs from π^0-particle decay tend more strongly towards small opening angles, being directly boosted in one particular direction by their parent particle. It can also be seen that the data corresponds very well to the simulated sample, which indicates that the general event structure is well-simulated.

Table 7.4 shows a summary of the cuts applied to the sample of shower pairs constructed in this section. Not shown in the table is the process by which the shower pair sample was formed. While events and individual shower objects were discarded in this process, the definition of what constitutes a shower pair is ambiguous before the completion of this stage. The efficiency of any subsequent cut is therefore shown relative to the first instance of the shower pair sample. By selecting only 3 GeV/c beam particle events, a large data/MC sample size difference is introduced. This is

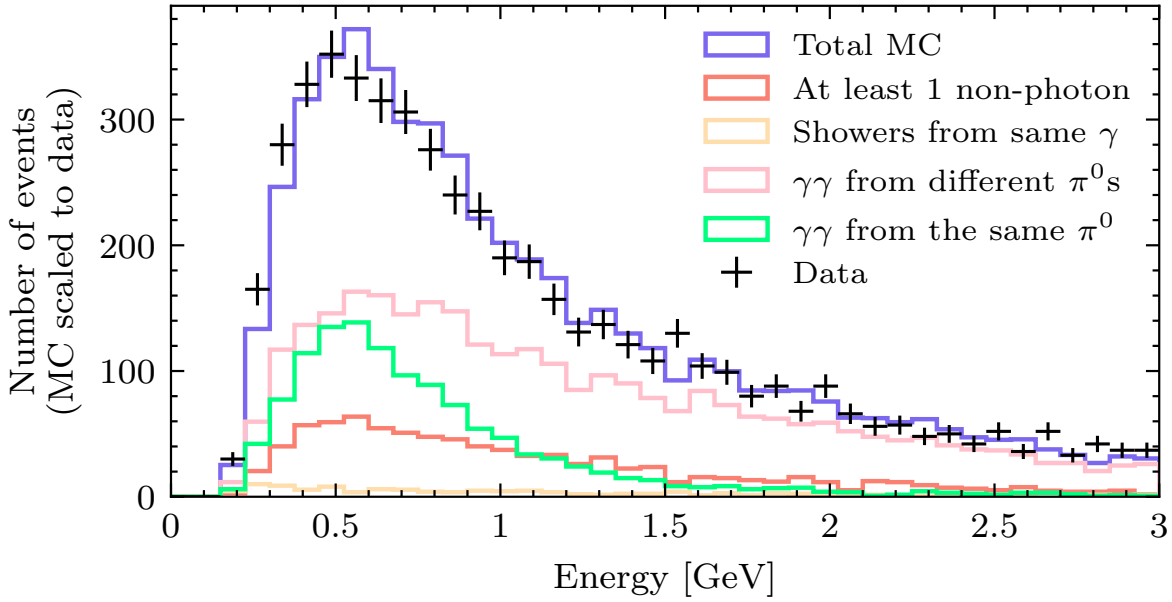

Figure 7.17: The reconstructed π^0-particle energy for all shower pairs in the sample, determined by the sum of the shower energies in each pair. The various categories correspond to those in table 7.3.

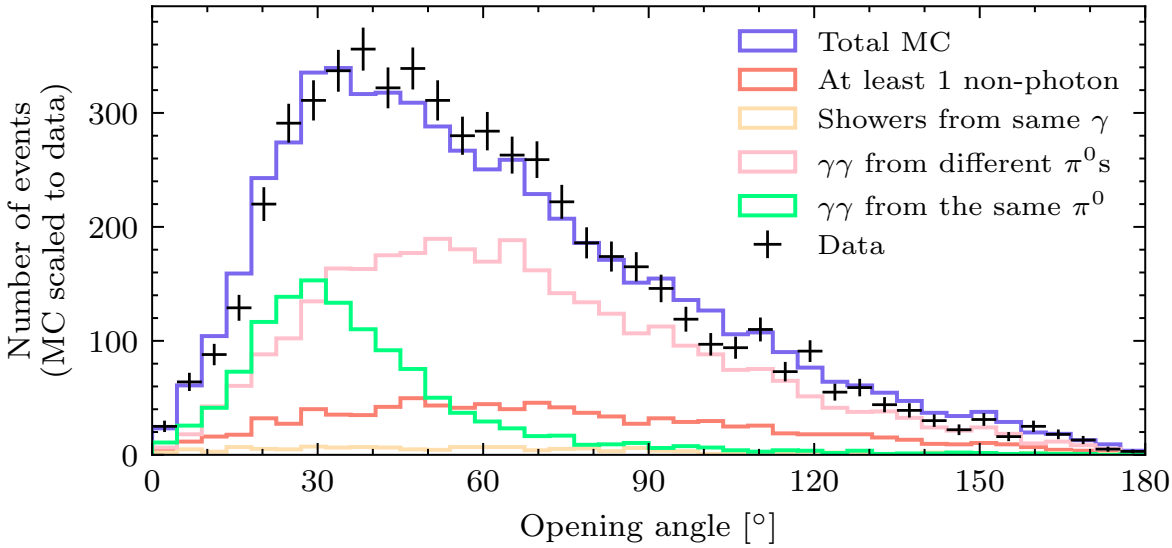

Figure 7.18: The opening angle of shower pairs in the sample. The various categories correspond to those in table 7.3.

expected, as the original simulation and data samples contain differing fractions of 3 GeV/c beam particle events, reflecting the need for a different normalisation when

comparing the two. By similar reasoning, the efficiency in this step has little bearing on data.

Table 7.4: A summary of the cuts applied to shower pairs in the sample. The format is identical to that in table 7.2. The desired events contain two showers originating from different photons that come from the same π^0-particle.

			Data/MC ratio	
Cut	Purity	Efficiency	Cumulative	Individual
	0.24	1.00	0.40	
Reco. π^0-energy < 1.5 GeV	0.28	0.92	0.39	0.98
10° < Opening angle < 60°	0.47	0.79	0.39	1.01
Beam momentum = 3 GeV/c	0.56	0.52	0.32	0.86

From table 7.4, it can be seen that the final purity of the shower pair sample is a mere 0.56. This highlights the difficulty in coupling the right showers together in multiple-π^0 events, incorrectly matched photon showers being the main background to the sample. On the object level, there is nothing to be done: all of these showers indeed originate from π^0-decay photons. On the event level, the options are likewise limited. Two showers originating from the same vertex may originate from the same π^0-particle just as well as two different ones. Nonetheless, the cuts explored in this section aim to mitigate the background as much as possible.

7.5 Results and discussion

Combining the results obtained in the previous section to obtain the π^0-particle invariant mass distribution, figure 7.19 is obtained. Immediately visible is the prominent peak around the true π^0-particle mass. Even though the purity of the sample is relatively low, as listed in table 7.4, the correctly matched shower pairs yield a recognisable signal in the overall sample.

Table 7.5 shows fit parameters concerning the invariant mass of correctly matched shower pairs in simulation. It can be seen that the result of the 3 GeV/c beam momentum sample is good, but is not universally shared with the other available momenta. At 1 GeV/c, where low statistics but high purity are expected, the results are similar

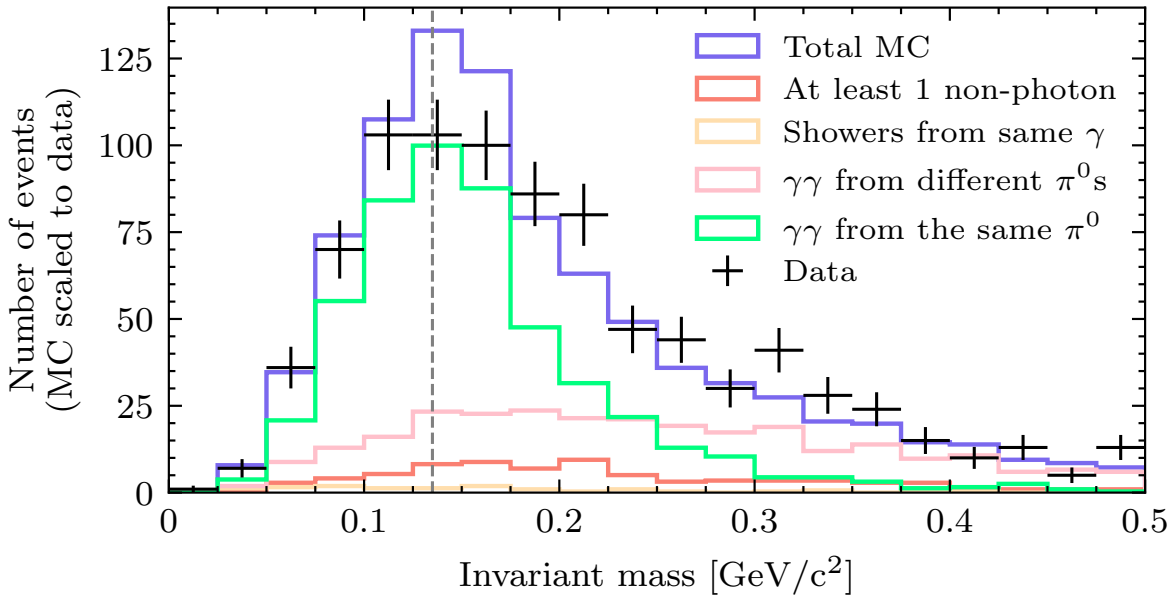

Figure 7.19: The invariant mass computed from the shower pair sample after applying the cuts listed in tables 7.2 and 7.4. Also shown is the true π^0-particle mass at 135 MeV/c^2.

to the 3 GeV/c case. The 2 GeV/c sample contains too few simulated showers to draw any definite conclusions, although the found parameters are mostly within the right range. Momenta above 3 GeV/c skew toward higher masses, peaking around 145 MeV/c^2. This effect is likely caused by diminished purity of larger showers leading to an overestimation of the shower energy, shown in figures 7.7 and 7.8.

With reasonable confidence that the shower pairs in both simulation and data are reconstructed well, the invariant mass peak in simulation can be fitted to that in data. This is done by applying a multiplicative scale factor b_E to the invariant mass calculation in simulation so that $E_1 \to b_E E_1$ and $E_2 \to b_E E_2$ for the leading and subleading photon respectively. In essence, this corrective factor signifies a difference between simulation and reality in the R factor discussed in section 7.2. By computing the χ^2 value between simulation and data invariant mass distributions for a range of b_E, the best fitting value for the parameter can be determined. The result of this procedure is shown in figure 7.20, where a scale factor of 1.05 ± 0.12 was obtained.

The simulated shower energy was previously corrected in section 7.3 using truth information. The obtained scale factor is therefore inverted and applied to the reconstructed invariant mass in data, yielding figure 7.21. As the scale factor of 1.05 is close to 1, the distribution is almost unchanged from figure 7.19. Although the general

Table 7.5: Various statistical parameters for distributions of the reconstructed invariant mass of the π^0-particle, considering only correctly matched shower pairs in simulation. Gaussian parameters are shown for a fit applied within the range of an estimated central value $\pm 50\,\mathrm{MeV/c^2}$. The estimated central value was obtained through a coarse Gaussian fit in the range $135 \pm 100\,\mathrm{MeV/c^2}$. Also shown are general statistical parameters concerning the distributions.

	Gaussian fit [MeV/c^2]		General stats. [MeV/c^2]		
Beam mom.	μ	σ	Mean	Median	Mode
1 GeV/c	139 ± 4	33 ± 5	144	140	130
2 GeV/c	130 ± 6	21 ± 6	145	129	70
3 GeV/c	137 ± 2	47 ± 4	159	146	150
6 GeV/c	145 ± 3	48 ± 5	176	158	150
7 GeV/c	143 ± 46	55 ± 94	169	163	130
All	140 ± 2	44 ± 3	164	149	150

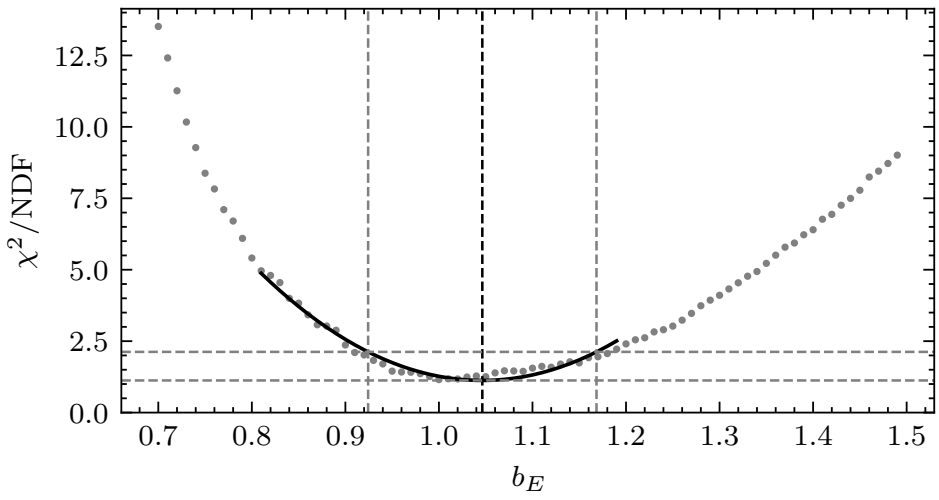

Figure 7.20: χ^2 results as a function of the scale parameter b_E, shown as grey dots. The χ^2-test was applied within the range 0-1 GeV/c^2, with bins 0.025 GeV/c^2 wide. The minimum of the test results was fitted with a parabola (black line) and the obtained minimum value for b_E is 1.05 (dashed black line) with $\chi^2/\mathrm{NDF} = 1.12$. A $\Delta\chi^2$ of 1 from the minimum is shown in horizontal grey dashed lines, from which a 1σ statistical uncertainty of 0.21 on b_E was deduced.

shapes of the distributions in simulation and data distributions are similar, fluctuations are visible in the ratio between the two, particularly near the true π^0-particle mass. It

is possible that these are caused by fundamental discrepancies between simulation and reality, but it cannot be said with certainty, given the limited size of the sample.

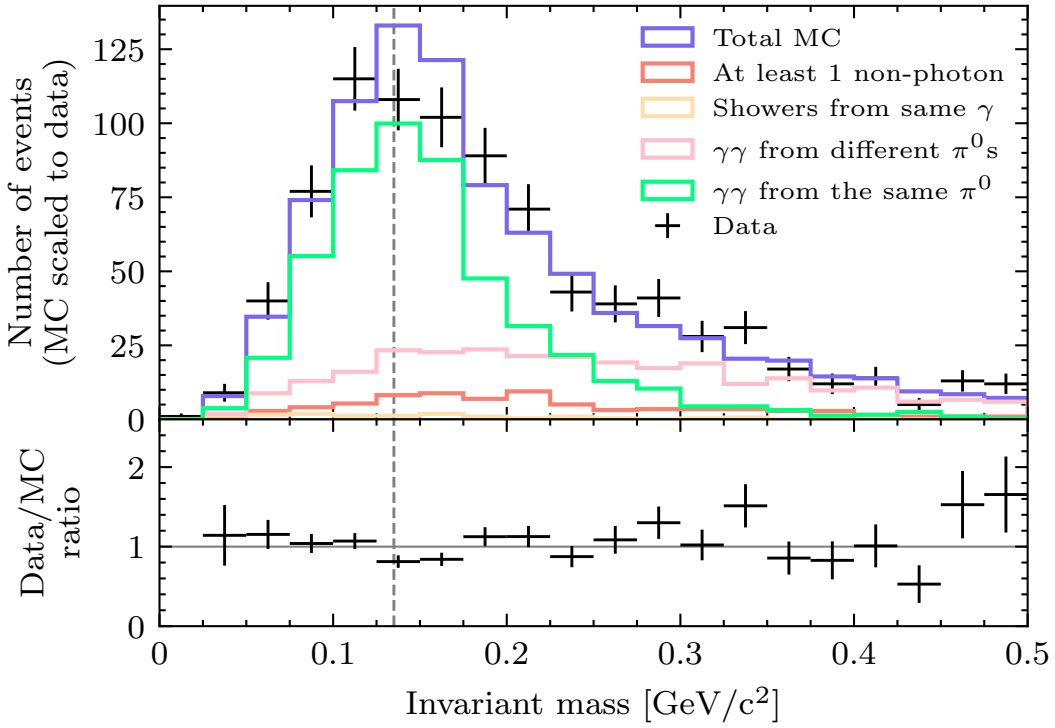

Figure 7.21: The invariant mass reconstructed from shower pairs in the sample. A scale factor of 1/1.05 has been applied to the distribution in data. Shown in the bottom plot is the ratio between the data and total simulated sample.

Since it corresponds to the true rest mass of the π^0-particle, the invariant mass measurement should be independent of its components. Figure 7.22 shows the correlation between the invariant mass and the shower opening angle, leading shower energy and subleading shower energy in both simulation and data. It can immediately be seen that the simulated sample is much larger and thereby better defined than the data sample. Despite this, the overall structures of the two cases are in agreement and show that there is a minor dependence of the invariant mass on the various parameters. Especially the opening angle shows a noticeable relation that is consistent with figure 7.12, in which it was shown that small opening angles are generally reconstructed smaller than they are. Although this feature is undesirable, it is an encouraging fact that it occurs in simulation as much as it does in data.

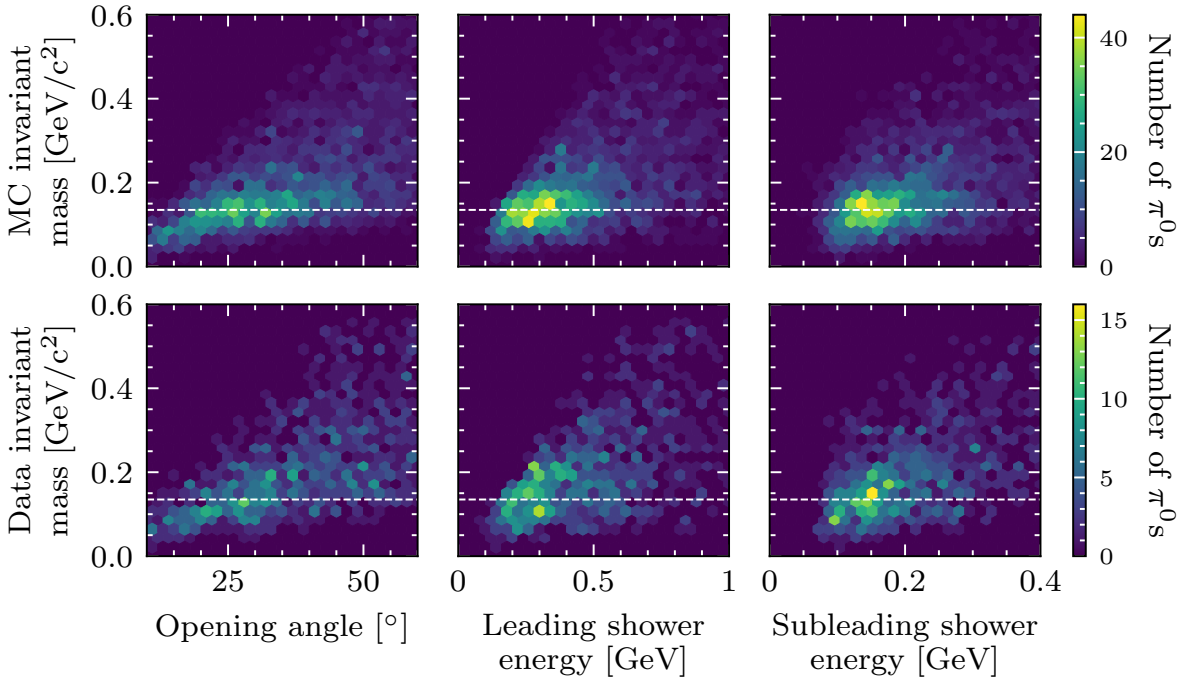

Figure 7.22: The invariant mass in simulation (top) and data (bottom) shown in relation to the opening angle (left), leading shower energy (middle) and subleading shower energy (right). The true π^0-particle mass is shown as a horizontal dashed line.

7.6 Conclusions and next steps

In this chapter, it was shown that π^0-particles can be reconstructed in ProtoDUNE data. A clear invariant mass peak around the true π^0-particle mass was constructed with good agreement between simulation and data. A calibration procedure was demonstrated, verifying that the energy reconstruction of electromagnetic showers is well-simulated, with a scale factor of 1.05 ± 0.12, which is compatible with the nominal value of 1. Also shown was the reconstruction efficiency of π^0-decay photons in simulation, establishing that it is infeasible for π^0-related background suppression to depend on both decay photons being reconstructed and matched.

Several improvements should aid in similar studies moving forward. Firstly, a larger data sample should prove whether the remaining discrepancies between simulation and data are statistical fluctuations or systematic differences. Additionally, the shower energy reconstruction currently lacks sophistication, consisting of a linear relationship between the total observed charge in the shower and its reconstructed energy. In reality, electromagnetic showers in LArTPCs are complex structures, in

which each hit has a different relation between charge and energy depending on its charge density and position within the shower. Lastly, shower clustering and direction reconstruction algorithms improve over time. This analysis used the results of the Pandora reconstruction algorithm, which is a constant work in progress. Improvements in both the energy and direction reconstruction of photon showers should increase the accuracy of the invariant mass measurement, creating a sharper peak that allows for a more accurate energy calibration.

A study similar to the one presented in this chapter can be carried out in the context of DUNE to provide electromagnetic shower energy calibration in its far detector. It is known that π^0-particles are produced by neutrino events, but they are also a product of cosmic rays that interact with the detector material. Several analysis projects are already under way on simulated far detector events in order to isolate these particles and determine their invariant mass. π^0-particles in the near detector are likewise under study, using novel machine learning reconstruction techniques. Between being a major background to the neutrino signal and a valuable calibration method, the neutral pion will undoubtedly be the subject of many more analyses in the future.

Conclusions and outlook

With the construction of DUNE well under way, the field of neutrino physics stands on the precipice of a new era defined by next-generation precision experiments. In the next decade, DUNE promises to solve long-standing questions regarding the neutrino mass ordering and leptonic CP violation. It will also measure a number of oscillation parameters to an unprecedented degree while providing invaluable checks on some of the other parameters. At the same time, the near and far detectors of the experiment will have rich individual physics programmes, ranging from beam neutrino cross-section measurements to the detection of supernova neutrinos. Both the near and far detector of DUNE are expected to start their data taking in 2027 after a period of construction and commissioning [39].

An essential step in the development of DUNE, ProtoDUNE was expanded upon in chapter 3. It is an experiment of unprecedented scale in its own right and provided a wealth of technical knowledge and charged particle data around its beam run in late 2018. Making use of the lessons learned from the first beam run, the hardware, software and firmware of ProtoDUNE will be upgraded to match that of the DUNE far detector for a proper dress rehearsal under the name of ProtoDUNE II. This experiment will be built up during the course of 2021 and operations are expected to start in the first quarter of 2022 [109].

The FELIX system detailed in chapter 4 was implemented in the ProtoDUNE DAQ chain shortly before the beam run. It was initially tested as a prototype, but thanks to its performance in ProtoDUNE, it has since become the DAQ system of choice for the DUNE far detector TPC. The data access and compression algorithms introduced in the same chapter provide an important link between FELIX and any systems downstream that request TPC data, such as online monitors and event builders. The data format is expected to be streamlined for the ProtoDUNE II run [110] and thanks to the flexibility of the developed data access overlays it will be made readable with minimal effort.

The developed data reordering and compression capabilities will likewise continue to persist in altered forms.

Making the distinction between electron and photon showers in DUNE is an essential step in identifying the neutrino that created the shower. Chapter 6 demonstrated a method by which one characteristic difference between these particles can be exploited to great effect, based on the energy deposition in the initial portion of their showers. Locating the exact start and direction of the shower is one of the main challenges of this analysis and was made more difficult by the busy particle environment in ProtoDUNE. Fortunately, single-neutrino interactions in the DUNE far detector are expected to be much cleaner than the cosmic ray-riddled ProtoDUNE beam events. With clear interaction vertices and few overlapping objects, showers produced in neutrino interactions are likely to have well-reconstructed start points and directions, allowing for more sophisticated charge profile analyses. In combination with other shower characteristics, the described method provides an effective complementary handle on the particle type of electromagnetic showers.

In DUNE, determining the type of interacting neutrino is only one part of the measurement. In order to construct an accurate image of the neutrino oscillation parameters, the energy of incoming particles must be determined to a high degree of certainty. Chapter 7 demonstrated a simulation-independent calibration method for the energy reconstruction of electromagnetic showers. Making use of the invariant π^0-particle mass, the method depends on accurate shower direction and energy reconstruction. Whereas π^0-particles in ProtoDUNE are created in hadronic beam particle interactions, the DUNE far detector may obtain these particles from cleaner neutrino interactions, as well as from the few cosmic rays that reach the detector deep underground. Using π^0-decay showers, a robust energy calibration may be constructed along the same lines as the ProtoDUNE study.

Electromagnetic showers are one of the main components that make up particle interactions in DUNE and ProtoDUNE. Their reconstruction and characterisation is an ongoing field of study that will play an invaluable role in the classification of neutrino-induced interactions. Being one crucial piece of the oscillation puzzle, it will ultimately aid DUNE in answering some of the biggest outstanding questions in particle physics.

Appendix A

Neutrino oscillation

The propagator S can be obtained by exponentiation of the Hamiltonian:

$$S = \exp(-iHx) \quad \rightarrow \quad S_{ij} = \langle \nu_j | S | \nu_i \rangle = \delta_{ij} \exp\left(-i\frac{m_i^2}{2E}x\right), \tag{A.1}$$

The probability to find a neutrino in a certain mass eigenstate given a pure initial state (abbreviated to ν_i and ν_j) is given by the following:

$$P(\nu_i \rightarrow \nu_j) = |S_{ij}|^2 = \delta_{ij}, \tag{A.2}$$

Using the mixing matrix U to transform between the flavour and mass bases, the probability to find one flavour eigenstate given an initial pure flavour eigenstate can be described as follows:

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= |S_{\alpha\beta}|^2 = |\langle \nu_\beta | S | \nu_\alpha \rangle|^2 \\ &= \left| \sum_{ij} \langle \nu_j | U_{\beta j} S U^*_{\alpha i} | \nu_i \rangle \right|^2 \\ &= \sum_{ij} U_{\beta i} U^*_{\alpha i} U^*_{\beta j} U_{\alpha j} \exp\left(i\frac{m_j^2 - m_i^2}{2E}x\right) \end{aligned} \tag{A.3}$$

In order to simplify notation, equation A.3 will be written in terms of the squared mass difference Δm^2_{ij} and the Jarlskog invariant $J^{ij}_{\alpha\beta}$:

$$\Delta m^2_{ij} = m^2_i - m^2_j \quad , \quad J^{ij}_{\alpha\beta} = U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j} \tag{A.4}$$

$$\text{where } J^{ji}_{\alpha\beta} = J^{ij}_{\beta\alpha} = \left(J^{ij}_{\alpha\beta}\right)^* \quad , \quad \sum_{ij} J^{ij}_{\alpha\beta} = \delta_{ij}. \tag{A.5}$$

Switching i and j without loss of generality, equation A.3 can then be written and continued as follows:

$$\begin{aligned}
P(\nu_\alpha \to \nu_\beta) =& \sum_{ij} J^{ij}_{\alpha\beta} \exp\left(i\frac{\Delta m^2_{ij}}{2E}x\right) \\
=& \left(\sum_{i<j} + \sum_{i>j}\right) J^{ij}_{\alpha\beta}\left[\exp\left(i\frac{\Delta m^2_{ij}}{2E}x\right) - 1\right] + \sum_{ij} J^{ij}_{\alpha\beta} \\
=& \sum_{i<j} J^{ij}_{\alpha\beta}\left[\exp\left(i\frac{\Delta m^2_{ij}}{2E}x\right) - 1\right] \\
&+ \sum_{i<j}\left(J^{ij}_{\alpha\beta}\right)^*\left[\exp\left(-i\frac{\Delta m^2_{ij}}{2E}x\right) - 1\right] + \delta_{\alpha\beta} \\
=& \delta_{\alpha\beta} + \sum_{i<j}\left[J^{ij}_{\alpha\beta} + \left(J^{ij}_{\alpha\beta}\right)^*\right]\left[\cos\left(\frac{\Delta m^2_{ij}}{2E}x\right) - 1\right] \\
&+ \sum_{i<j}\left[J^{ij}_{\alpha\beta} - \left(J^{ij}_{\alpha\beta}\right)^*\right] i\sin\left(\frac{\Delta m^2_{ij}}{2E}x\right) \\
=& \delta_{\alpha\beta} - 4\sum_{i<j}\mathrm{Re}\left(J^{ij}_{\alpha\beta}\right)\sin^2\left(\frac{\Delta m^2_{ij}}{4E}x\right) - 2\sum_{i<j}\mathrm{Im}\left(J^{ij}_{\alpha\beta}\right)\sin\left(\frac{\Delta m^2_{ij}}{2E}x\right)
\end{aligned} \tag{A.6}$$

Appendix B

Additional DUNE sensitivities

This appendix shows DUNE's sensitivities to the neutrino mass ordering and CP violation as described in chapter 2. They are shown as a function of the CP-violating phase δ_{CP} for various values of other neutrino oscillation parameters. All plots were taken from [36].

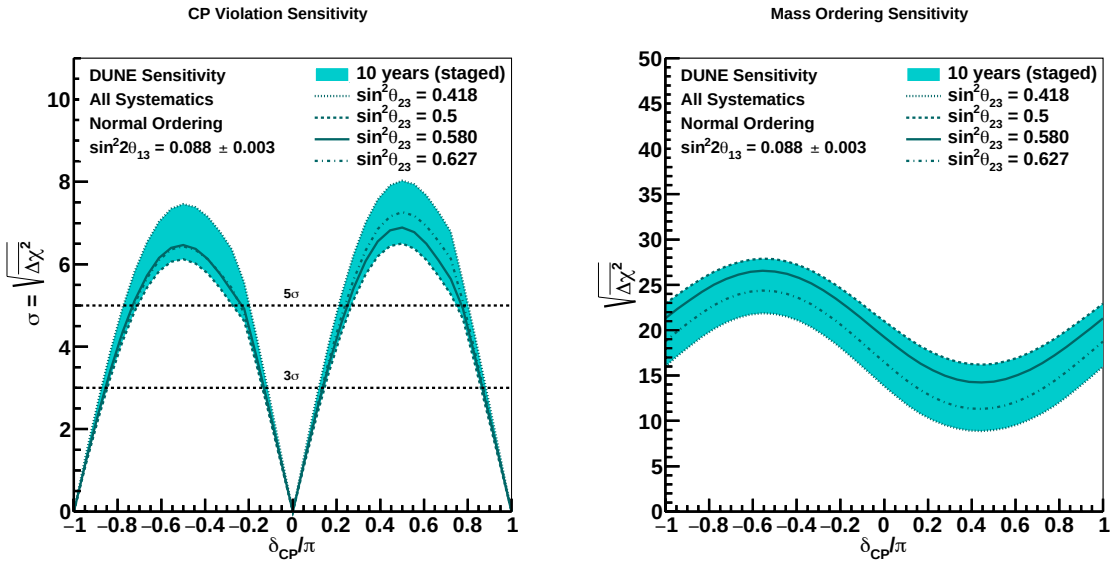

Figure B.1: Sensitivity to CP violation (left) and the mass ordering (right) as a function of δ_{CP} for various values of θ_{23}.

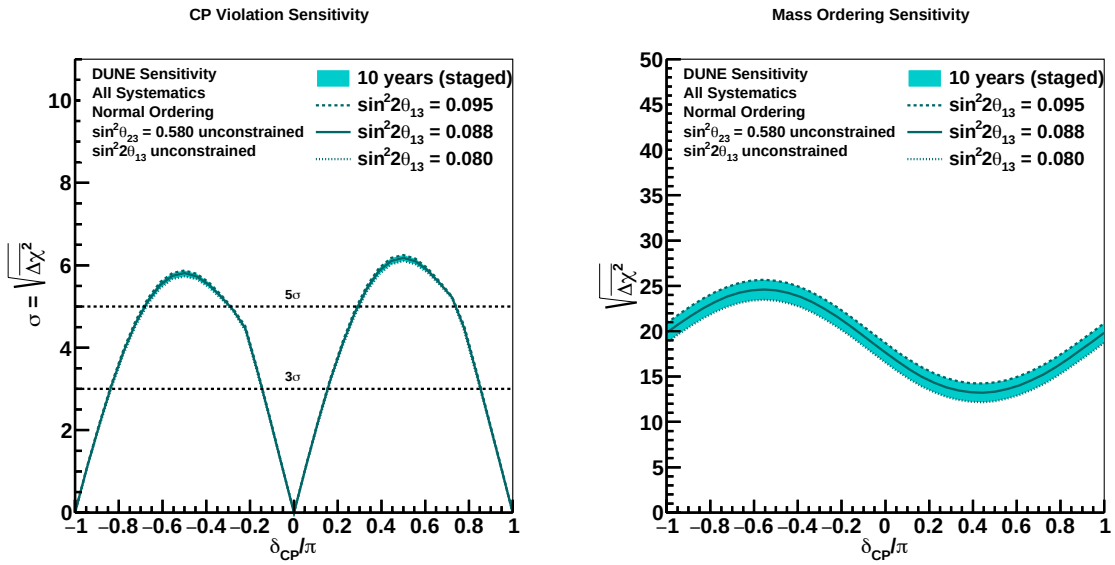

Figure B.2: Sensitivity to CP violation (left) and the mass ordering (right) as a function of δ_{CP} for various values of θ_{13}.

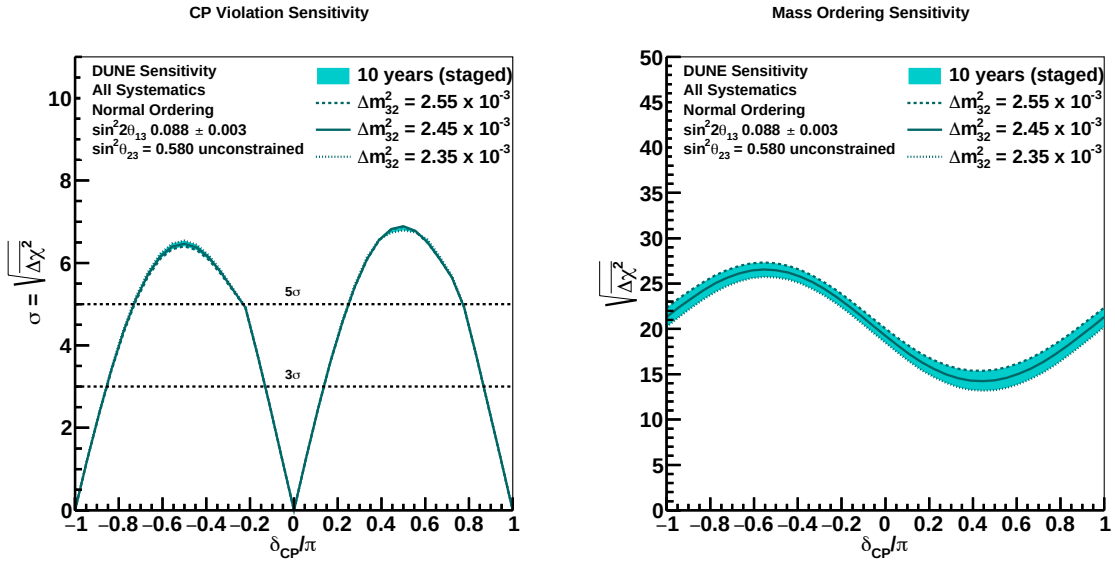

Figure B.3: Sensitivity to CP violation (left) and the mass ordering (right) as a function of δ_{CP} for various values of Δm^2_{32}.

Bibliography

[1] University of Zurich, *Standard Model*, https://www.physik.uzh.ch/en/researcharea/lhcb/outreach/StandardModel.html (visited on 2/7/2021) (cit. on pp. 4, 190, 202).

[2] C. L. Cowan et al., "Detection of the free neutrino: A Confirmation", Science **124**, 103–104 (1956) (cit. on p. 5).

[3] J. N. Bahcall and R. Davis, "Solar Neutrinos - a Scientific Puzzle", Science **191**, 264–267 (1976) (cit. on pp. 6, 25, 39).

[4] J. Boger et al. (SNO), "The Sudbury neutrino observatory", Nucl. Instrum. Meth. A **449**, 172–207 (2000), arXiv:nucl-ex/9910016 (cit. on pp. 6, 25).

[5] M. Thomson, *Modern particle physics* (Cambridge University Press, 2013) (cit. on p. 6).

[6] C. Jarlskog, "Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation", Phys. Rev. Lett. **55**, 1039 (1985) (cit. on p. 9).

[7] I. Esteban et al., "The fate of hints: updated global analysis of three-flavor neutrino oscillations", JHEP **09**, 178 (2020), arXiv:2007.14792 [hep-ph] (cit. on pp. 11, 24, 26).

[8] P. A. Zyla et al. (Particle Data Group), "Review of Particle Physics", PTEP **2020**, 083C01 (2020) (cit. on pp. 12, 16, 24, 27, 126, 146).

[9] C. S. Wu et al., "Experimental Test of Parity Conservation in β Decay", Phys. Rev. **105**, 1413–1414 (1957) (cit. on p. 15).

[10] NagualDesign / Wikimedia Commons CC BY-SA 3.0, *Wu experiment*, https://commons.wikimedia.org/wiki/File:Wu_experiment.jpg (visited on 2/7/2021) (cit. on p. 15).

[11] J. H. Christenson et al., "Evidence for the 2π Decay of the K_2^0 Meson", Phys. Rev. Lett. **13**, 138–140 (1964) (cit. on p. 16).

[12] B. Jacob and G. Guennebaud, *Eigen*, `http://eigen.tuxfamily.org` (visited on 2/7/2021) (cit. on p. 18).

[13] P. Wynn, "On the convergence and stability of the epsilon algorithm", SIAM Journal on Numerical Analysis **3**, 91–122 (1966) (cit. on p. 18).

[14] B. Roe, "Matter density versus distance for the neutrino beam from Fermilab to Lead, South Dakota, and comparison of oscillations with variable and constant density", Phys. Rev. D **95**, 113004 (2017), arXiv:1707.02322 [hep-ex] (cit. on pp. 18, 68).

[15] M. A. Vermeulen, *EigenNeut*, `https://github.com/mivermeu/EigenNeut` (visited on 2/7/2021) (cit. on p. 18).

[16] M. Freund, "Analytic approximations for three neutrino oscillation parameters and probabilities in matter", Phys. Rev. D **64**, 053003 (2001), arXiv:hep-ph/0103300 (cit. on p. 22).

[17] S. J. Parke, P. B. Denton, and H. Minakata, "Analytic Neutrino Oscillation Probabilities in Matter: Revisited", PoS **NuFact2017**, 055 (2018), arXiv:1801.00752 [hep-ph] (cit. on p. 22).

[18] S. K. Agarwalla, Y. Kao, and T. Takeuchi, "Analytical approximation of the neutrino oscillation matter effects at large θ_{13}", JHEP **04**, 047 (2014), arXiv:1302.6773 [hep-ph] (cit. on p. 22).

[19] H. Nunokawa, S. J. Parke, and J. W. F. Valle, "CP Violation and Neutrino Oscillations", Prog. Part. Nucl. Phys. **60**, 338–402 (2008), arXiv:0710.0554 [hep-ph] (cit. on p. 22).

[20] A. Serenelli, "Alive and well: a short review about standard solar models", The European Physical Journal A **52**, 78 (2016) (cit. on p. 25).

[21] Y. Fukuda et al. (Super-Kamiokande), "The Super-Kamiokande detector", Nucl. Instrum. Meth. A **501**, 418–462 (2003) (cit. on p. 25).

[22] K. Eguchi et al. (KamLAND), "First results from KamLAND: Evidence for reactor anti-neutrino disappearance", Phys. Rev. Lett. **90**, 021802 (2003), arXiv:hep-ex/0212021 (cit. on p. 25).

[23] G. Alimonti et al. (Borexino), "The Borexino detector at the Laboratori Nazionali del Gran Sasso", Nucl. Instrum. Meth. A **600**, 568–593 (2009), arXiv:0806.2400 [physics.ins-det] (cit. on p. 25).

[24] M. G. Aartsen et al. (IceCube), "The IceCube Neutrino Observatory: Instrumentation and Online Systems", JINST **12**, P03012 (2017), arXiv:1612.05093 [astro-ph.IM] (cit. on p. 25).

[25] M. Ageron et al. (ANTARES), "ANTARES: the first undersea neutrino telescope", Nucl. Instrum. Meth. A **656**, 11–38 (2011), arXiv:1104.1607 [astro-ph.IM] (cit. on p. 25).

[26] U. F. Katz, "KM3NeT: Towards a km^3 Mediterranean Neutrino Telescope", Nucl. Instrum. Meth. A **567**, 457–461 (2006), arXiv:astro-ph/0606068 (cit. on p. 26).

[27] Y. Abe et al. (Double Chooz), "Indication of Reactor $\bar{\nu}_e$ Disappearance in the Double Chooz Experiment", Phys. Rev. Lett. **108**, 131801 (2012), arXiv:1112.6353 [hep-ex] (cit. on p. 26).

[28] F. P. An et al. (Daya Bay), "Observation of electron-antineutrino disappearance at Daya Bay", Phys. Rev. Lett. **108**, 171803 (2012), arXiv:1203.1669 [hep-ex] (cit. on p. 26).

[29] J. K. Ahn et al. (RENO), "Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment", Phys. Rev. Lett. **108**, 191802 (2012), arXiv:1204.0626 [hep-ex] (cit. on p. 26).

[30] F. An et al. (JUNO), "Neutrino Physics with JUNO", J. Phys. G **43**, 030401 (2016), arXiv:1507.05613 [physics.ins-det] (cit. on p. 26).

[31] K. Abe et al. (T2K), "The T2K Experiment", Nucl. Instrum. Meth. A **659**, 106–135 (2011), arXiv:1106.1238 [physics.ins-det] (cit. on pp. 26, 36).

[32] J. Evans (MINOS), "The MINOS Experiment: Results and Prospects", Adv. High Energy Phys. **2013**, 182537 (2013), arXiv:1307.0721 [hep-ex] (cit. on p. 26).

[33] M. A. Acero et al. (NOvA), "First Measurement of Neutrino Oscillation Parameters using Neutrinos and Antineutrinos by NOvA", Phys. Rev. Lett. **123**, 151803 (2019), arXiv:1906.04907 [hep-ex] (cit. on pp. 26, 36).

[34] R. Acciarri et al. (MicroBooNE), "Design and Construction of the MicroBooNE Detector", JINST **12**, P02017 (2017), arXiv:1612.05824 [physics.ins-det] (cit. on pp. 26, 45, 57, 81).

[35] K. Nakamura, "Hyper-Kamiokande: A next generation water Cherenkov detector", Int. J. Mod. Phys. A **18**, 4053–4063 (2003) (cit. on pp. 26, 55, 56).

[36] B. Abi et al. (DUNE), "Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics", (2020), arXiv:2002.03005 [hep-ex] (cit. on pp. 29–31, 35, 36, 38, 42–46, 48–52, 54, 104, 110, 116, 125, 177).

[37] V. Lebedev (PIP-II), *The PIP-II Reference Design Report*, tech. rep., FERMILAB-DESIGN-2015-01 (June 2015) (cit. on p. 30).

[38] B. Abi et al. (DUNE), "Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I. Introduction to DUNE", JINST **15**, T08008 (2020), arXiv:2002.02967 [physics.ins-det] (cit. on pp. 30, 33, 35, 39).

[39] B. Abi et al. (DUNE), "Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume IV. The DUNE far detector single-phase technology", JINST **15**, T08010 (2020), arXiv:2002.03010 [physics.ins-det] (cit. on pp. 30, 39, 41, 43, 64–67, 173).

[40] R. Acciarri et al. (DUNE), "Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNF", (2015), arXiv:1512.06148 [physics.ins-det] (cit. on pp. 30, 45, 55).

[41] The DUNE Collaboration, *Dunescience*, https://www.dunescience.org (visited on 2/7/2021) (cit. on pp. 30, 39).

[42] V. Papadimitriou (DUNE), "Design of the LBNF Beamline", in 38th International Conference on High Energy Physics (2017) (cit. on pp. 31, 34, 35).

[43] Fermilab, *Fermilab's Accelerator Complex*, (2020) https://www.fnal.gov/pub/science/particle-accelerators/accelerator-complex.html (visited on 2/7/2021) (cit. on p. 32).

[44] J. Strait et al. (DUNE), "Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 3: Long-Baseline Neutrino Facility for DUNE June 24, 2015", (2016), arXiv:1601.05823 [physics.ins-det] (cit. on pp. 33, 37, 53).

[45] P. Huber, M. Lindner, and W. Winter, "Simulation of long-baseline neutrino oscillation experiments with GLoBES (General Long Baseline Experiment Simulator)", Comput. Phys. Commun. **167**, 195 (2005), arXiv:hep-ph/0407333 (cit. on pp. 32, 38).

[46] L. Fields, *DUNE Fluxes*, https://home.fnal.gov/~ljf26/DUNEFluxes/ (visited on 2/7/2021) (cit. on pp. 32, 34, 38).

[47] K. Aamodt et al. (ALICE), "The ALICE experiment at the CERN LHC", JINST **3**, S08002 (2008) (cit. on pp. 34, 57).

[48] J.-M. Levy, "Kinematics of an off axis neutrino beam", (2010), arXiv:1005.0574 [hep-ex] (cit. on p. 36).

[49] T. Cai et al., *The DUNE-PRISM Near Detector Program*, tech. rep., DocDB 8106 (2018) (cit. on pp. 36, 37).

[50] A. Salles and C. Walter, "Construction begins on international mega-science experiment to understand neutrinos", Fermilab Newsroom (2017) (cit. on p. 39).

[51] E. Collett and J. Willhite, *Defining Constraints of the Ross Cage Dimensional & Weight Requirements*, tech. rep., DocDB 3582 (2020) (cit. on p. 40).

[52] Brookhaven National Laboratory, *Liquid Argon Properties*, https://lar.bnl.gov/properties/ (visited on 2/7/2021) (cit. on pp. 41, 66, 108, 126, 131).

[53] S. Agostinelli et al. (GEANT4), "GEANT4–a simulation toolkit", Nucl. Instrum. Meth. A **506**, 250–303 (2003) (cit. on p. 44).

[54] J. S. Marshall and M. A. Thomson, "The Pandora Software Development Kit for Pattern Recognition", Eur. Phys. J. C **75**, 439 (2015), arXiv:1506.05348 [physics.data-an] (cit. on pp. 44, 113).

[55] B. Abi et al. (DUNE), "First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN Neutrino Platform", JINST **15**, P12004 (2020), arXiv:2007.06722 [physics.ins-det] (cit. on pp. 45, 60, 70, 103, 105–110, 120, 129, 198, 211).

[56] G. P. Zeller et al. (NuTeV), "A Precise Determination of Electroweak Parameters in Neutrino Nucleon Scattering", Phys. Rev. Lett. **88**, 091802 (2002), arXiv:hep-ex/0110059 (cit. on p. 53).

[57] C. Adams et al. (LBNE), "The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe", (2013), arXiv:1307.7335 [hep-ex] (cit. on p. 53).

[58] R. M. Bionta et al., "Observation of a Neutrino Burst in Coincidence with Supernova SN 1987a in the Large Magellanic Cloud", Phys. Rev. Lett. **58**, 1494 (1987) (cit. on p. 54).

[59] S. Amerio et al. (ICARUS), "Design, construction and tests of the ICARUS T600 detector", Nucl. Instrum. Meth. A **527**, 329–410 (2004) (cit. on p. 57).

[60] D. L. Adams et al., "Design and performance of a 35-ton liquid argon time projection chamber as a prototype for future very large detectors", JINST **15**, P03035 (2020), arXiv:1912.08739 [physics.ins-det] (cit. on p. 57).

[61] L. Evans and P. Bryant, "LHC Machine", JINST **3**, S08001 (2008) (cit. on p. 57).

[62] G. Aad et al. (ATLAS), "The ATLAS Experiment at the CERN Large Hadron Collider", JINST **3**, S08003 (2008) (cit. on p. 57).

[63] S. Chatrchyan et al. (CMS), "The CMS Experiment at the CERN LHC", JINST **3**, S08004 (2008) (cit. on p. 57).

[64] A. A. Alves Jr. et al. (LHCb), "The LHCb Detector at the LHC", JINST **3**, S08005 (2008) (cit. on p. 57).

[65] G. Aad et al. (ATLAS), "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys. Lett. B **716**, 1–29 (2012), arXiv:1207.7214 [hep-ex] (cit. on p. 58).

[66] R. Herbst et al., "Design of the SLAC RCE Platform: A general purpose ATCA based data acquisition system", in 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (2014), 1–4 (cit. on pp. 61, 78).

[67] G. Unel (ATLAS Tdaq), "FELIX: the New Detector Readout System for the ATLAS Experiment", PoS **TWEPP2018**, 140 (2019) (cit. on pp. 61, 78, 80).

[68] D. Montanari et al., "Development of membrane cryostats for large liquid argon neutrino detectors", IOP Conf. Ser. Mater. Sci. Eng. **101**, 012049 (2015) (cit. on p. 62).

[69] B. Abi et al. (DUNE), "The Single-Phase ProtoDUNE Technical Design Report", (2017), arXiv:1706.07081 [physics.ins-det] (cit. on pp. 63, 66, 67, 71, 78, 83, 103).

[70] N. Charitonidis and I. Efthymiopoulos, "Low energy tertiary beam line design for the CERN neutrino platform project", Phys. Rev. Accel. Beams **20**, 111001 (2017) (cit. on pp. 69, 71).

[71] M. Potekhin, "The protoDUNE-SP experiment and its prompt processing system", PoS **EPS-HEP2017**, 513 (2017) (cit. on p. 69).

[72] The DUNE Collaboration, *Dunetpc*, https://cdcvs.fnal.gov/redmine/projects/dunetpc (visited on 2/7/2021) (cit. on p. 70).

[73] K. Wood, *NP04beamRunSum*, https://github.com/krwood/NP04beamRunSum (visited on 2/7/2021) (cit. on p. 73).

[74] G. Rameika, *ProtoDUNE-SP Status, LBNC Meeting*, (2018) https://indico.fnal.gov/event/19266/contributions/51196/attachments/31878/39074/LBNC-Rameika_December18.pdf (visited on 2/7/2021) (cit. on pp. 72, 74).

[75] M. Spanu, "The status and results from ProtoDUNE single phase", Journal of Physics: Conference Series **1312**, 012003 (2019) (cit. on p. 75).

[76] G. Aad et al. (ATLAS), "Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", JHEP **12**, 055 (2015), arXiv:1506.00962 [hep-ex] (cit. on p. 77).

[77] E. L. Snider and G. Petrillo, "LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors", J. Phys. Conf. Ser. **898**, 042057 (2017) (cit. on pp. 81, 113).

[78] E. Hazen and M. Worcester, *ProtoDUNE WIB Output Data Formats*, tech. rep., DocDB 1701 (2017) (cit. on pp. 82, 83).

[79] The DUNE Collaboration, *Dune-raw-data*, https://cdcvs.fnal.gov/redmine/projects/dune-raw-data (visited on 2/7/2021) (cit. on p. 84).

[80] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003) (cit. on pp. 89, 91).

[81] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression", IEEE Transactions on Information Theory **23**, 337–343 (1977) (cit. on p. 91).

[82] *Intel® QuickAssist Technology API Programmer's Guide*, Intel Corporation (2014) (cit. on p. 92).

[83] Intel Corporation, *Intel® QuickAssist Technology (QAT) QATzip Library* (cit. on pp. 92, 97).

[84] R. Acciarri et al. (MicroBooNE), "The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector", Eur. Phys. J. C **78**, 82 (2018), arXiv:1708.03135 [hep-ex] (cit. on pp. 101, 102, 110, 111).

[85] A. Higuera, "ProtoDUNE-SP Electron Analysis", DocDB 18355, 2020 (cit. on pp. 103, 128, 138).

[86] Y. Li et al., "Measurement of Longitudinal Electron Diffusion in Liquid Argon", Nucl. Instrum. Meth. A **816**, 160–170 (2016), arXiv:1508.07059 [physics.ins-det] (cit. on p. 103).

[87] C. Sarasty, *Signal processing in the ProtoDUNE-SP LArTPC*, (2019) https://indico.cern.ch/event/782953/contributions/3465962/attachments/1889003/3115676/Signal_processing_in_the_protoDUNE-SP_LArTPC.pdf (visited on 2/7/2021) (cit. on pp. 103, 104).

[88] C. Adams et al. (MicroBooNE), "Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE", JINST **13**, P07007 (2018), arXiv:1804.02583 [physics.ins-det] (cit. on p. 103).

[89] R. Veenhof, "Garfield, a drift chamber simulation program", Conf. Proc. C **9306149**, 66–71 (1993) (cit. on pp. 104, 105).

[90] C. Sarasty, *Coherent Noise in ProtoDUNE*, (2018) https://indico.fnal.gov/event/18925/contributions/49401/attachments/30811/37861/ProtoDUNE_Coherent_Noise.pdf (visited on 2/7/2021) (cit. on p. 104).

[91] H. Schellman (DUNE), "Computing for the DUNE Long-Baseline Neutrino Oscillation Experiment", EPJ Web Conf. **245**, 11002 (2020), arXiv:2004.09037 [physics.ins-det] (cit. on p. 109).

[92] J. Asaadi, *Gauss Hit Finder Class*, https://nusoft.fnal.gov/larsoft/doxsvn/html/GausHitFinder__module_8cc_source.html (visited on 2/7/2021) (cit. on p. 109).

[93] The Pandora Organisation, *Pandora Changes Since MCC10*, (2018) https://indico.fnal.gov/event/17650/contributions/44278/attachments/27430/33966/ProtoDUNE_SimReco_Green_2018_7_18.pdf (visited on 2/7/2021) (cit. on pp. 111, 112).

[94] The Pandora Organisation, *Pandora Pattern Recognition Performance*, (2017) https://www.hep.phy.cam.ac.uk/~marshall/PandoraDUNE_25.Jan.2017.pdf (visited on 2/7/2021) (cit. on p. 112).

[95] The Pandora Organisation, *Pandora Pattern Recognition Update*, https://www.hep.phy.cam.ac.uk/~marshall/DUNE_CMtalk_May2017.pdf (visited on 2/7/2021) (cit. on p. 112).

[96] C. Green et al., "The Art Framework", J. Phys. Conf. Ser. **396**, 022020 (2012) (cit. on p. 113).

[97] W. E. Brown et al., *The fhicl homepage*, https://cdcvs.fnal.gov/redmine/projects/fhicl (visited on 2/7/2021) (cit. on p. 114).

[98] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", Nucl. Instrum. Meth. A **389**, 81–86 (1997) (cit. on p. 114).

[99] J. Stock, *cheat::BackTrackerService Class Reference*, (June 26, 2020) `https://nusoft.fnal.gov/larsoft/doxsvn/html/classcheat_1_1BackTrackerService.html` (visited on 2/7/2021) (cit. on p. 114).

[100] D. R. Yocum et al., "FermiGrid", in TeraGrid '07: Broadening Participation in TeraGrid (May 2007) (cit. on p. 117).

[101] J. Pivarski, *The uproot homepage*, `https://pypi.org/project/uproot/` (visited on 2/7/2021) (cit. on p. 117).

[102] The pandas community, *The pandas homepage*, `https://pandas.pydata.org` (visited on 2/7/2021) (cit. on p. 117).

[103] A. Paudel, *DQdx and dEdx calibration instructions*, (2019) `https://wiki.dunescience.org/wiki/DQdx_and_dEdx_calibration_instructions` (visited on 2/7/2021) (cit. on p. 129).

[104] C. Anderson et al., "The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab", JINST **7**, P10019 (2012), arXiv:1205.6747 [physics.ins-det] (cit. on p. 130).

[105] R. Acciarri et al. (ArgoNeuT), "A Study of Electron Recombination Using Highly Ionizing Particles in the ArgoNeuT Liquid Argon TPC", JINST **8**, P08005 (2013), arXiv:1306.1712 [physics.ins-det] (cit. on pp. 130–132).

[106] A. Reynolds, *Hit Tagging Update*, (2019) `https://indico.fnal.gov/event/20654/contributions/58540/attachments/36646/44609/CNN_Update_May_2019.pdf` (visited on 2/7/2021) (cit. on p. 133).

[107] V. T. McGary (MiniBooNE), "NC π^0 Production in the MiniBooNE Antineutrino Data", AIP Conf. Proc. **981**, 250–252 (2008), arXiv:0801.0628 [hep-ex] (cit. on p. 146).

[108] J. Rak, *All you need to know about π^0 decay kinematics*, (2009) `https://trac.cc.jyu.fi/projects/alice/attachment/wiki/jan/attachments/piz_assym.pdf` (visited on 2/7/2021) (cit. on p. 147).

[109] E. James, *ProtoDUNE II*, (2020) `https://indico.cern.ch/event/862992/contributions/3682812/attachments/1980848/3298600/ProtoDUNE2_Feb042020.pdf` (visited on 2/17/2021) (cit. on p. 173).

[110] D. Christian, *WIB/FELIX Data Frame Format*, tech. rep., DocDB 14947 (2020) (cit. on p. 173).

Summary

Chapter 1: Neutrino oscillation

Over the ages, humans have employed ever-complicated models to describe the world around them. From water, earth, fire and air to the periodic table, they made steady progress in identifying all stable elements: atoms, considered fundamental and indivisible particles at the time. Nonetheless, in the middle of the last century, the atom was split to explosive effect. It turned out that atoms consist of a central nucleus and a cloud of electrons swarming around it. We know now that the tiny but dense nucleus is itself made up of protons and neutrons. These particles making up every nucleus, collectively called nucleons, are themselves made of quarks.

We currently consider quarks and electrons to be elementary particles, along with other fermions such as muons and the three neutrino flavours. Alongside these particles that make up all known matter in the universe, there are the force carriers that allow for interactions between particles, such as photons and gluons. The latest addition to the elementary particle zoo is the recently discovered Higgs boson, which gives particles their mass. Together, these particles and the interactions between them form the Standard Model, currently the best model to describe the basic constituents of matter as we know it. Figure S.1 shows a diagram of all known elementary particles as they fit into this Standard Model.

All matter that we interact with on a daily basis is made up of just three elementary particles: up-quarks, down-quarks and electrons, held together by photons and gluons. Nonetheless, there is a particle that is more abundant still than quarks or electrons: the neutrino. Neutrinos are created in interactions involving the weak nuclear force, which in nature most often occur in stellar nuclear fusion. These particles have the smallest nonzero mass of all particles in the Standard Model and interact only via the weak nuclear force and gravity. As the name suggests, the weak nuclear force is incredibly weak: whereas particles like electrons and photons interact readily within

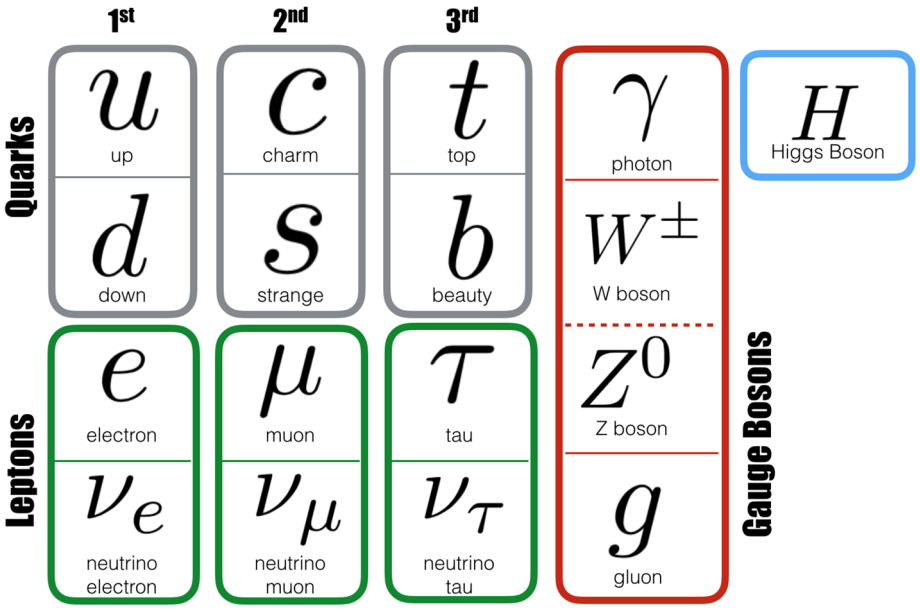

Figure S.1: All known elementary particles of the Standard Model. Taken from [1].

matter, neutrinos may travel through light-years of material before being stopped. The sun produces so many neutrinos that trillions of them pass through the human body every second without affecting it in the slightest.

Aside from their ghostly nature, neutrinos are notable for another feature: they change flavour mid-flight. There are three flavours of neutrinos, named for their associated charged leptons: the electron neutrino, muon neutrino and tau neutrino. In what is called neutrino oscillation, an initial electron neutrino may after some time be observed as a muon- or tau neutrino, and vice versa. This oscillation is periodic in nature and its rate depends on the energy of the neutrino as well as density of the material the neutrino is travelling through. Figure S.2 shows the oscillation of the three neutrino types as they travel through empty space. This is an interesting phenomenon to study in itself, but it also ties into some fundamental principles of physics.

Neutrino oscillation may explain why half of the universe is missing. According to Albert Einstein's famous equation of energy-mass equivalence, $E = mc^2$, particles of mass m may be created with a certain energy E at an 'exchange rate' of the speed of light squared c^2. Normally, the creation of a particle always goes paired with the creation of its antiparticle. However, this matter-antimatter symmetry is not visible in the universe around us, where we only observe regular matter in large amounts. One of the parameters that governs neutrino oscillation, δ_{CP}, determines the asymmetry

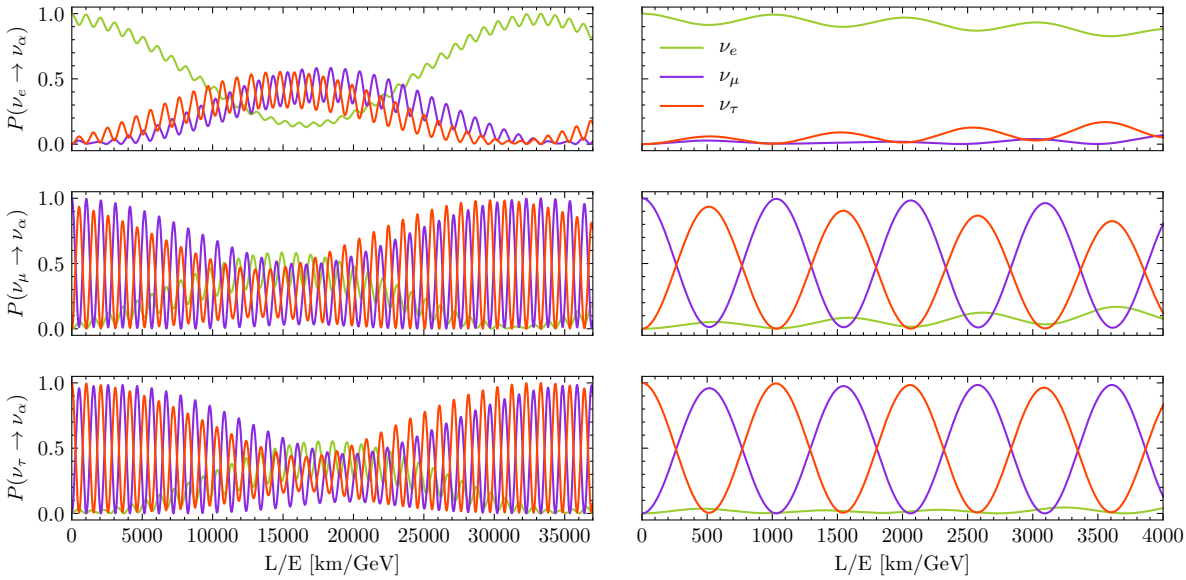

Figure S.2: Three-neutrino oscillation in vacuum for long (left) and short (right) travel length. Shown are an initial electron neutrino (top), muon neutrino (middle) and tau neutrino (bottom). The vertical axis of each graph denotes the probability of detecting a certain neutrino flavour.

between neutrinos and antineutrinos, and may give us an insight into processes that prefer matter over antimatter or vice versa. If δ_{CP} turns out to be nonzero, it could thus help to explain the large matter-antimatter asymmetry that we see in the universe around us.

Another fundamental neutrino property that influences neutrino oscillation is the neutrino mass ordering. We know that two of the three neutrinos masses are fairly close together, while the third deviates substantially more. We do not know whether this third neutrino mass is much higher or much lower than the other two, however, which is called the neutrino mass ordering problem. Since neutrino mass differences directly influence neutrino oscillation, we can use oscillation measurements to probe this neutrino mass ordering.

Chapters 2 and 3: DUNE and ProtoDUNE

Because neutrinos are able to travel through large lengths of matter unimpeded, it is a challenge in itself to catch any neutrino interaction within a particle detector. Neutrino detectors are therefore always as large and as quiet as possible. In order to create this quiet particle environment, they are often situated deep underground or underwater so

that cosmic rays cannot reach them, and care is taken to use low-radiation construction material where necessary.

The Deep Underground Neutrino Experiment, or DUNE, is an upcoming neutrino experiment that is currently under construction in the US. In essence, it consists of an intense neutrino beam, a 'near detector' that observes the initial conditions of the beam and a 'far detector' that observes the same beam 1300 km away. Both the neutrino beam and near detector are to be located at Fermilab near Chicago, and the far detector is being constructed in the Homestake mine, a former goldmine under the Black Hills of South Dakota that now houses the Sanford Underground Research Facility. Through this twin detector setup, DUNE measures both the neutrino beam's initial and final flavour composition to get a handle on its oscillation in between.

The DUNE far detector will be the largest detector of its type by far. Divided between four separate modules, 70 kt of liquid argon kept at -186 °C will be put under a strong electric field 1.5 km underground. Electrically charged high-energy particles rip the electrons from their argon atoms, leaving a trail of ionised particles in their wake as they travel through the detector. Driven by the electric field, the liberated electrons travel multiple metres through the detector material before they arrive at the detector's anode panels. These are made up of three planes of wires which carefully record the ionisation electrons' position and time of arrival. From these measurements, algorithms reconstruct the ionisation trails in three-dimensional space, yielding a high-resolution image of particle interactions within the detector. This type of detector projects one dimension of its measurement domain onto the anode plane, recording it instead in terms of time, as shown in figure S.3. For this reason, it is called a Time Projection Chamber, or TPC.

Although there have been liquid argon TPCs prior to DUNE, none of them come close to the scale of the DUNE far detector. The ProtoDUNE experiment aimed to test full-scale DUNE far detector components at CERN, subjecting them to a charged particle beam from the Super Proton Synchrotron. The modularity of the far detector makes it possible to scale down one of its modules from roughly 19 (W) $\times$ 18 (H) $\times$ 66 (L) m^3 to 11 $\times$ 11 $\times$ 11 m^3, holding 0.77 kt instead of 17.5 kt of liquid argon.

Together with the Neutrino Platform at CERN, the DUNE Collaboration built two ProtoDUNE prototypes named single-phase and dual-phase for the state of their respective argon contents. The single-phase prototype is the focus of this thesis and was constructed between 2016 and 2018 to take advantage of a scheduled charged

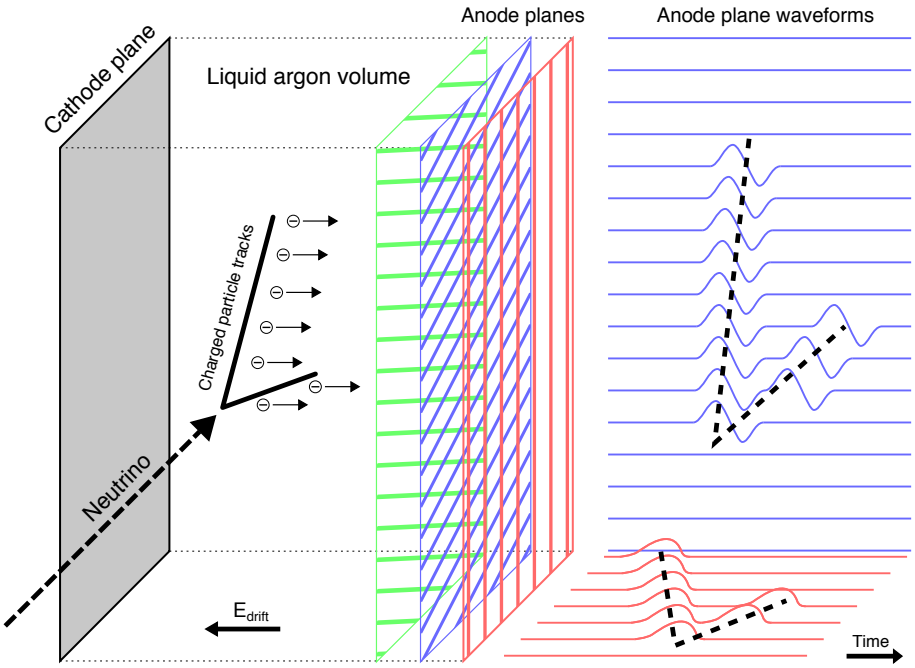

Figure S.3: The workings of the (Proto)DUNE TPC. Charged particles originating from a neutrino interaction create an ionisation track, the electrons of which are attracted into the anode planes. These record the electrons' position and time of arrival for a three-dimensional reconstruction of the event.

particle beam at the end of 2018. In what is called the 'beam run,' the prototype detector successfully recorded over 4 million beam particle interactions within the detector volume, achieving and in some aspects exceeding its data collection goals. Aside from this, ProtoDUNE proved to be a valuable test bed for the technologies that the DUNE far detector will use.

Chapter 4: Data access and compression in the ProtoDUNE DAQ

Experiments in modern particle physics rely on large numbers of particle interactions to draw conclusions about their underlying physical processes. To accumulate these volumes of interaction events, experiments record at as high a rate as possible. ProtoDUNE recorded 25 high-resolution images of beam particle interactions per second during its beam run. In order to take these images, its six anode planes each have 2560 channels that register ionisation electron signals at 2 MHz. To limit the total data volume that these 15,360 channels generate, their signals are only passed on within a

time window around beam particle interactions. Nonetheless, the raw data stream coming out of the detector during data taking is substantial: 91 Gbps per anode plane.

ProtoDUNE employed two data acquisition systems to manage its TPC data flow. One is based on the Reconfigurable Cluster Elements system and was used as a reliable baseline system, reading out five out of ProtoDUNE's six anode planes. The Front-End LInk eXchange system (FELIX) read out the final anode plane, serving as a prototype readout system. Originally developed for use in the ATLAS detector, FELIX is PCIe-based, which is a common standard that allows for it to be plugged into off-the-shelf computer hardware. The main advantage in doing so is that it does not require any custom hardware beyond the FELIX PCIe card and can thereby make use of rapid advances in commercial computing and telecommunication.

As it receives data from the detector, FELIX slices the relevant window out of the data stream and passes it to its host system. In order to manipulate and transfer this data, the values within need to be made accessible. In this chapter we develop a method to gain access to individual wire values from the data packets FELIX delivers to its host. We do this by constructing a so-called overlay out of C-native bit-fields that follow the pattern of structures within the data packets. Interpreting the data through this overlay yields an efficient way to extract individual or ranges of values. The primary use-case for these overlays is in the event builder, which reads the data from all detector systems and produces a unified package containing all data collected in the particle interaction. However, other systems make use of the overlay as well: the online monitor reads and visualises detector data while a run is ongoing, and data reordering and compression algorithms in the data acquisition chain likewise need direct access to the data through the overlay.

In this chapter we also develop compression algorithms for the ProtoDUNE data acquisition chain and study their performance on MicroBooNE data. Compression in software proves to be very effective, far surpassing the targeted compression ratio of 4. However, due in part to the complexity of the incoming data structures, compression in software was not able to keep up with time constraints. Instead, we reordered the data into a more structured form and fed it to a dedicated compression hardware card, powered by Intel® QuickAssist Technology. We show that this method succeeds in meeting the compression ratio and compression time goals. The overlays can read all data formats seamlessly, including compressed and reordered formats, thanks to the format information being stored alongside the data and automatically being applied by the overlay. The end result is a highly user-friendly and flexible system that will be

able to accommodate a large number of future data formats while retaining indefinite backward compatibility.

Chapter 5: Reconstruction

Reconstruction is the bridge between recorded data and physics. It is the task of the reconstruction chain to interpret the raw data and form them into particle tracks or showers. In ProtoDUNE, the reconstruction chain spreads these processes out over multiple stages. First, the signal processing step cleans up the data, removing the most predictable forms of noise. Then, it characterises peaks in wire current as discrete hit objects, each with its own height and width. Reconstruction algorithms find matches between hits from multiple wire planes to locate the charge deposit in three dimensions in the detector volume. Subsequent stages deal with the formation of higher-level structures. A clustering algorithm groups charge deposits into distinct objects that it believes belong to a single ionising particle. Finally, these clusters are assigned a type based on their characteristics.

The ProtoDUNE reconstruction chain identifies two types of single-particle clusters: tracks and showers. The liquid argon TPCs of ProtoDUNE and the DUNE far detector have a spatial resolution of roughly 0.5 cm, thanks to which they are able to resolve individual tracks of many types of charged particles. These track-like particles include hadrons such as protons, pions and kaons, but muons also form recognisable tracks in the detector. In ProtoDUNE, which sits on the surface, cosmic muons provide a constant backdrop to beam particle interactions. High energy electrons, positrons and photons undergo rapid electromagnetic interactions with liquid argon, creating successively more electrons, positrons and photons. These cascades are called showers and are too dense even for the high-resolution TPC of ProtoDUNE to resolve individual particles. The distinction between track and shower is important as it is directly linked to the type of the ionising particle. This distinction is often far from obvious, however, especially when considering small clusters of hits from relatively low-energy particles.

For the analyses shown in this thesis, we have made use of hundreds of thousands of simulated and real events. We are only interested in secondary particles created in interactions between the beam particle and the argon in the detector. For this reason, we select events that contain hadronic beam particles, as electromagnetic particles normally create one single particle shower without distinct secondary particles. In this

chapter, we show that key aspects of simulated beam interactions match reality well, although there are small discrepancies in the creation of secondary particles.

Chapter 6: Photon shower identification

In order to measure neutrino oscillation, DUNE will identify the type of the neutrinos that interact within its detectors. Neutrinos themselves are invisible, but they can be identified through their interaction products. The electron neutrino is especially important in DUNE and has one characteristic signature in which it produces an electron resulting in an electromagnetic shower. There is one interaction in particular that interferes with this measurement: a neutrino of any type may create a π^0-particle that immediately decays into two photons, both of which can produce electromagnetic showers of their own. It may seem as though the initial interaction was caused by an electron neutrino if one of these showers is not reconstructed for whatever reason. Distinguishing electron showers from photon showers is therefore vital to DUNE's physics goals.

Determining the energy deposition at the start of an electromagnetic shower is one method of determining the type of particle that caused it. Electrons and positrons are electrically charged and start ionising the argon in the detector immediately. Photons, on the other hand, have no electrical charge and are thus invisible until they interact with the argon and create an electron-positron pair. The start of a photon shower thus consists of two overlapping electron showers, depositing twice as much energy as a single electron shower.

In this chapter, we show that the initial energy deposition in photon showers is indeed twice that of electron showers. We begin by carefully selecting objects in beam interaction events to obtain a pure photon sample. Next, we place a cylinder around each shower start to define the region in which we expect to see a clear distinction between electron and photon showers. Figure S.4 shows this procedure graphically. Determining the energy deposition of all hits inside this shower start region and taking their median yields a clear value at double the electron shower energy deposition. Although there are other methods to distinguish electron showers from photon showers, this procedure provides a promising complementary handle on the shower's particle type.

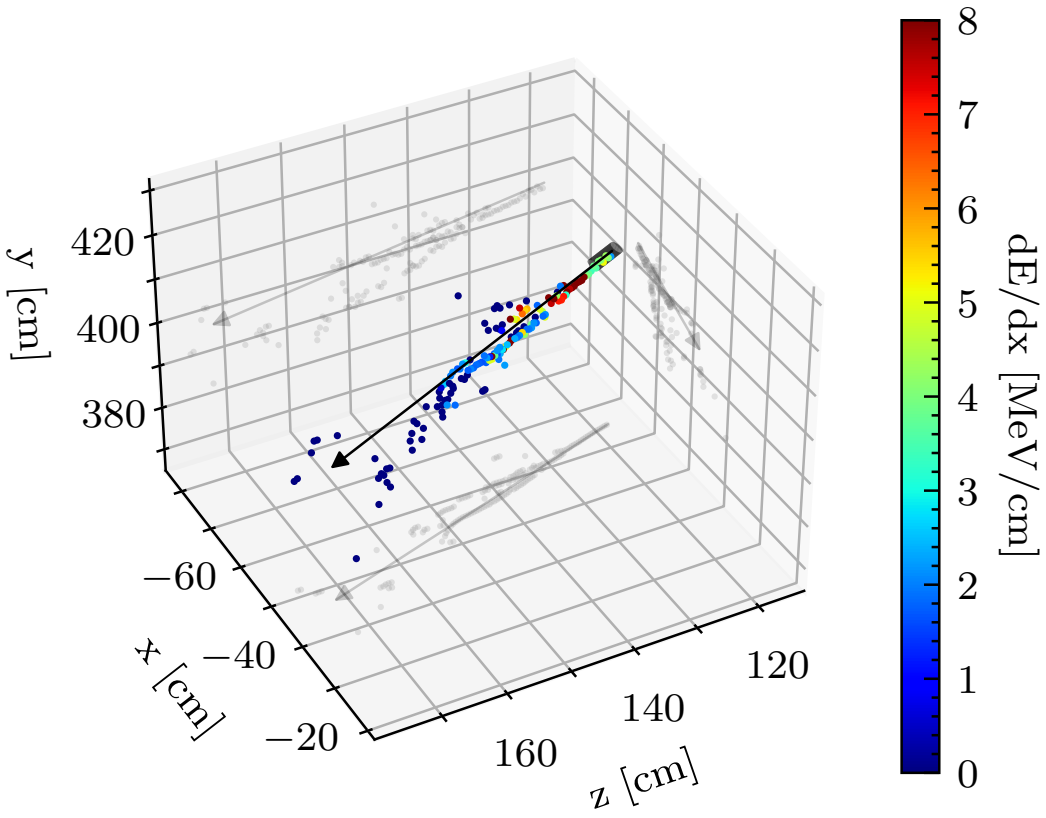

Figure S.4: A shower in isolation, shown in three dimensions as a cloud of hits (coloured markers) and as a reconstructed object with a start point, direction and length (solid arrow). Two-dimensional projections of the shower are shown as transparent grey markers and arrows. The colour of the markers corresponds to the size of their individual energy deposits. A transparent grey cylinder for use in characterisation is shown around the shower start.

Although the distinction between electron showers and photon showers is clear when using this method, it remains open to improvement. Due to the dense particle environment in ProtoDUNE, the method needs to make use of a small capture cylinder that risks missing good hits. A consideration of the energy deposition in the shower as a whole might be a more robust method and could be a good starting point for further research.

Chapter 7: Neutral pion reconstruction

As mentioned, π^0-particles are a major source of confusion in neutrino identification. However, they can also serve a useful purpose in energy calibration. They decay

almost instantly and almost always produce two photons, which it emits back-to-back if it is at rest. If the π^0-particle has some kinetic energy, it is inherited by its decay photons, which are then boosted in the same direction as their parent particle. The resulting angle between the photons (θ), in combination with their energy (E_1, E_2), determines a quantity that is known as the invariant mass of the π^0-particle:

$$m_\pi = \sqrt{2E_1E_2(1-\cos\theta)}.$$

This quantity is called invariant because it stays the same no matter the kinetic energy of the π^0-particle. This makes it identical to the rest mass of the particle, which is a well-known quantity. The invariant mass therefore provides a standard candle which shows how well the energy of photon showers is reconstructed. Figure S.5 shows an example of a particularly well-formed π^0-particle candidate.

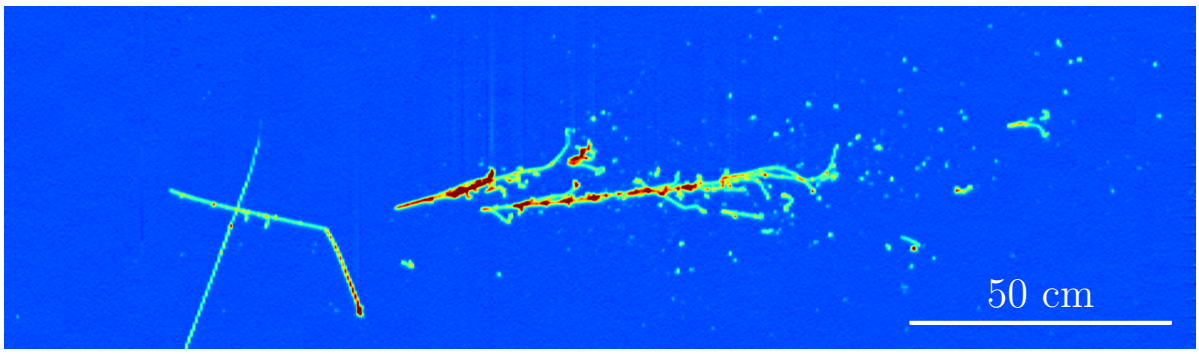

Figure S.5: An example of a π^0-particle event. The beam particle comes in from the left and creates a π^0-particle that instantly decays into two invisible photons which each create an electromagnetic shower. Taken from [55].

Since most photons in ProtoDUNE originate from π^0-particle decay, we are able to select a pure sample of π^0-decay products by simply collecting a pure photon sample. We can therefore recycle much of the previous chapter's sample selection methods as it focuses on photon showers as well. The showers that follow from π^0-decay almost always come in pairs of two and therefore our sample should consist of shower pairs as well. There are various matters that complicate the process of shower pair selection. π^0-decay showers often have too low of an energy to be reconstructed at all and tend to be created close together, which occasionally causes them to look like a single shower. Additionally, a particle interaction that creates multiple π^0-particles opens the door for mismatched showers to enter into pairs. We have taken care to mitigate these issues, but the fact remains that the π^0-particle does not let itself be easily reconstructed.

Initially, we perform the energy reconstruction of photon showers based on truth information in simulation. After we apply all corrections to obtain the collected charge within the shower, a single multiplicative factor R remains as the bridge between the charge and energy of the shower. In simulated events, we can set this factor so that the reconstructed shower energy lines up with the true energy of the photon that caused it. Using this value for R as we calculate the invariant π^0-mass in real beam interactions, we find a slight discrepancy with the simulation. In all, we determine the difference in R to be 1.05 ± 0.12. This discrepancy is small, but can be solved by setting the value of R in data to be a factor $1/1.05$ lower than that in simulation.

Conclusions and outlook

The analyses in this thesis deal with photon showers in ProtoDUNE, but all of the methods are applicable to the DUNE far detector. Having shown that photon and electron showers differ in their initial energy deposition in ProtoDUNE, we expect it to perform even better in DUNE, where shower objects are likely to be better defined against a lower background signal. Likewise, shower energy calibration using the invariant π^0-mass will be vital in DUNE, and various analyses have already started on this topic. Being both a blessing and a curse, the π^0-particle is sure to be in the analysis spotlight for the foreseeable future. In the meantime, the construction of DUNE presses on towards a bright future of precision neutrino physics.

Su

Samenvatting

Hoofdstuk 1: Neutrino-oscillatie

Door de eeuwen heen hebben mensen verscheidene modellen gebruikt om de wereld om hen heen te beschrijven. Van water, aarde, vuur en lucht tot het periodiek systeem maakte men gestaag voortgang in het identificeren van alle stabiele elementen: atomen, waarvan men toen dacht dat ze fundamenteel en ondeelbaar waren. Het atoom werd afgelopen eeuw met explosieve effecten gesplitst. Het bleek dat atomen bestaan uit een centrale kern en een wolk elektronen die om de kern heen zoemen. We weten vandaag de dag dat de kleine maar zware kern zelf uit protonen en neutronen bestaat. Deze deeltjes, die samen nucleonen worden genoemd, bestaan zelf uit quarks.

We beschouwen quarks en elektronen tegenwoordig als elementaire deeltjes, net zoals andere fermionen zoals muonen en de drie smaken neutrino's. Naast deze deeltjes, waar alle materie in het universum uit bestaat, zijn er de krachtvoerende deeltjes die ervoor zorgen dat deeltjes met elkaar kunnen interageren, zoals fotonen en gluonen. De laatste toevoeging aan de dierentuin van elementaire deeltjes is het recent ontdekte higgsboson, die deeltjes hun massa geeft. Samen vormen deze deeltjes en hun interacties het standaardmodel van de deeltjesfysica, wat op dit moment het model is dat de basiscomponenten van de natuur het best beschrijft. Figuur S.1 laat een diagram zien met alle deeltjes van dit standaardmodel.

Alle materie waar wij dagelijks mee in aanraking komen is gemaakt van slechts drie elementaire deeltjes: up-quarks, down-quarks en elektronen, bij elkaar gehouden door fotonen en gluonen. Er is echter een deeltje dat nog vaker voorkomt dan quarks of elektronen: het neutrino. Neutrino's worden geproduceerd in interacties van de zwakke wisselwerking, die in de natuur het vaakst voorkomt in de kernfusie van sterren. Deze deeltjes hebben de kleinste massa van alle massieve deeltjes in het standaardmodel en interageren alleen via de zwakke wisselwerking en de zwaartekracht. Zoals de naam al doet vermoeden is de zwakke wisselwerking ontzettend zwak: hoewel deeltjes

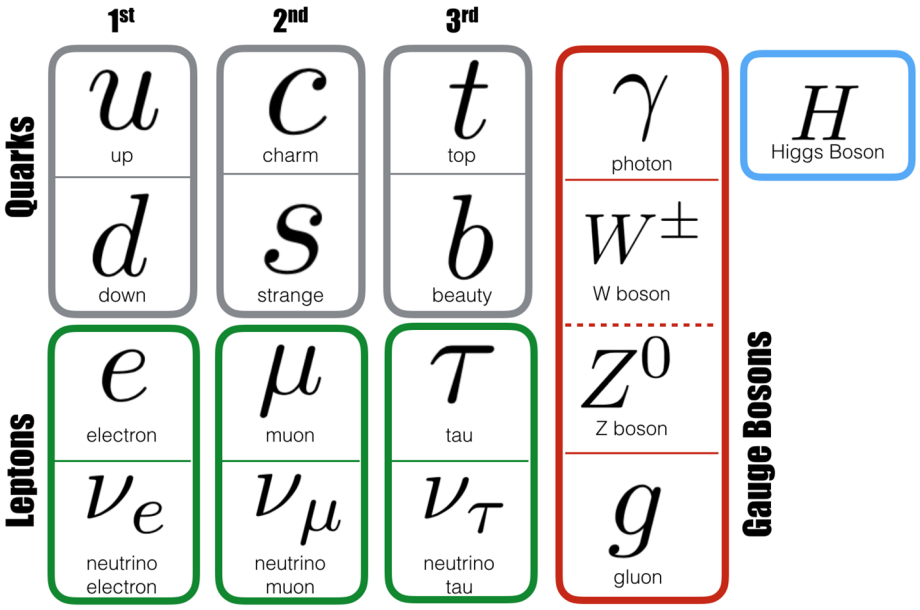

Figure S.1: Alle bekende elementaire deeltjes in het standaardmodel. Overgenomen van [1].

zoals elektronen en fotonen gemakkelijk interageren met materie, kunnen neutrino's door lichtjaren aan materiaal reizen zonder te worden gestopt. De zon produceert zoveel neutrino's dat miljarden elke seconde door het menselijk lichaam gaan zonder ook maar enig effect te hebben.

Naast hun spookachtige natuur zijn neutrino's speciaal om een andere eigenschap: ze veranderen van smaak terwijl ze door de ruimte vliegen. Er zijn drie smaken neutrino's, vernoemd naar het geladen lepton waar ze mee geassocieerd zijn: het elektron-neutrino, muon-neutrino en tau-neutrino. In zogenaamde neutrino-oscillatie kan een elektron-neutrino in een muon-neutrino of tau-neutrino veranderen, en vice versa. Deze oscillatie is periodiek en de frequentie van de oscillatie hangt af van de energie van het neutrino en de dichtheid van de materie waar het doorheen vliegt. Figuur S.2 laat zien hoe de drie neutrinosmaken oscilleren terwijl ze door lege ruimte reizen. Op zichzelf is dit al een interessant fenomeen om te bestuderen, maar het opent ook de deur naar een aantal fundamentele principes in de fysica.

Neutrino-oscillatie zou kunnen verklaren waarom de helft van het universum lijkt te ontbreken. Volgens de beroemde massa-energie equivalentie van Albert Einstein, $E = mc^2$, kan een deeltje van massa m met een energie E worden gemaakt volgens een wisselkoers van de snelheid van het licht in het kwadraat c^2. Normaliter gaat de creatie van een deeltje altijd gepaard met de creatie van zijn antideeltje. Die materie-

Sa

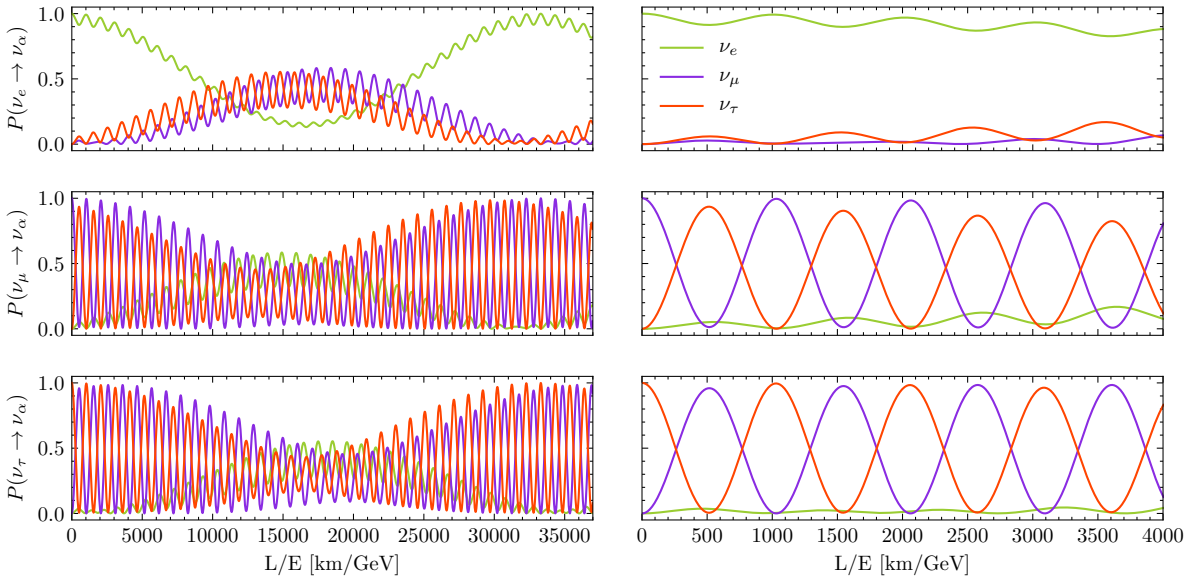

Figure S.2: Drie-neutrino oscillatie in vacuüm voor lange (links) en korte (rechts) afstand. Te zien zijn een aanvankelijk elektron-neutrino (boven), muon-neutrino (midden) en tau-neutrino (onder). De verticale as in elke grafiek geeft aan hoe waarschijnlijk het is om een bepaalde smaak waar te nemen.

antimaterie symmetrie is echter niet te zien in het universum, waar we alleen materie in grote hoeveelheden zien. Een van de parameters die neutrino-oscillatie beïnvloedt is δ_{CP}, die de asymmetrie tussen neutrino's en antineutrino's definieert en ons dus inzicht kan geven in processen die voorkeur geven aan materie boven antimaterie of vice versa. Als δ_{CP} niet nul blijkt te zijn, zou het dus de grote materie-antimaterie asymmetrie in het universum kunnen verklaren.

Een ander fundamenteel principe dat invloed uitoefent op neutrino-oscillatie is de neutrinomassahiërarchie. We weten dat twee van de drie neutrinomassa's dicht bij elkaar liggen en dat de derde significant afwijkt. We weten alleen niet of deze derde massa veel hoger of veel lager is dan de andere twee. Dit wordt het neutrinomassahiërarchieprobleem genoemd. Omdat de neutrinomassaverschillen direct invloed hebben op neutrino-oscillatie kunnen we oscillatie-experimenten gebruiken om de massahiërarchie te meten.

Hoofdstukken 2 en 3: DUNE en ProtoDUNE

Omdat neutrino's ongestoord door grote hoeveelheden materie kunnen reizen is het een uitdaging om neutrino-interacties te vangen in een deeltjesdetector. Neutrinodetec-

toren zijn daarom altijd zo groot en zo stil mogelijk. Om deze stille deeltjesomgeving te creëren worden ze vaak diep ondergronds of onderwater gebouwd zodat kosmische straling ze niet kan bereiken. Waar nodig wordt ook speciaal bouwmateriaal gebruikt met een laag stralingsgehalte.

Het Deep Underground Neutrino Experiment, of DUNE, is een toekomstig neutrino experiment dat op dit moment in de VS in opbouw is. In essentie bestaat het uit een intense neutrinobundel, een 'near detector' die de begintoestand van de bundel observeert en een 'far detector' die dezelfde bundel 1300 km verderop observeert. De neutrinobundel en near detector worden beide gebouwd op Fermilab nabij Chicago en de far detector is in opbouw in de Homestake mijn, een voormalige goudmijn onder de Black Hills van South Dakota waar vandaag de dag de Sanford Underground Research Facility gevestigd is. Door deze dubbeldetector opstelling meet DUNE de begin- en eindsmaak van bundelneutrino's om de oscillatie ertussenin te bepalen.

De DUNE far detector zal veruit de grootste detector van zijn soort zijn. Verdeeld over vier modules 1.5 km onder de grond zal 70 kt vloeibaar argon met een temperatuur van -186 °C onder een sterk elektrisch veld worden geplaatst. Elektrisch geladen deeltjes die door de detector vliegen trekken elektronen van hun argonatomen af en laten daardoor een spoor van geïoniseerde deeltjes achter. Aangedreven door het elektrische veld reizen de bevrijde elektronen meerdere meters door het vloeibare argon voordat ze de anodepanelen van de detector bereiken. Deze zijn gemaakt van drie vlakken van draden die precies de locatie en aankomsttijd van de ionisatie-elektronen bepalen. Uit deze metingen kunnen algoritmes de originele ionisatiesporen in drie dimensies reconstrueren zodat er een beeld kan worden gevormd van de deeltjesinteracties in de detector. Figuur S.3 laat zien hoe dit type detector een ruimtelijke dimensie op de anode projecteert en het in plaats daarvan registreert in termen van aankomsttijd. Om deze reden heet dit type detector een Time Projection Chamber, of TPC.

Hoewel er in het verleden meerdere vloeibaar argon TPC's hebben bestaan, kwamen er geen in buurt van de schaal van de DUNE far detector. Het ProtoDUNE experiment had als doel om de onderdelen van de DUNE far detector te testen op CERN door ze aan een geladen deeltjesbundel van de Super Proton Synchrotron bloot te stellen. De modulariteit van de far detector maakt het mogelijk om een van de modules te verkleinen terwijl de componenten op dezelfde schaal blijven: van ongeveer 19 (B) $\times$ 18 (H) $\times$ 66 (L) m^3 tot 11 $\times$ 11 $\times$ 11 m^3 in ProtoDUNE, welke 0.77 kt aan vloeibaar argon bevat in plaats van 17.5 kt.

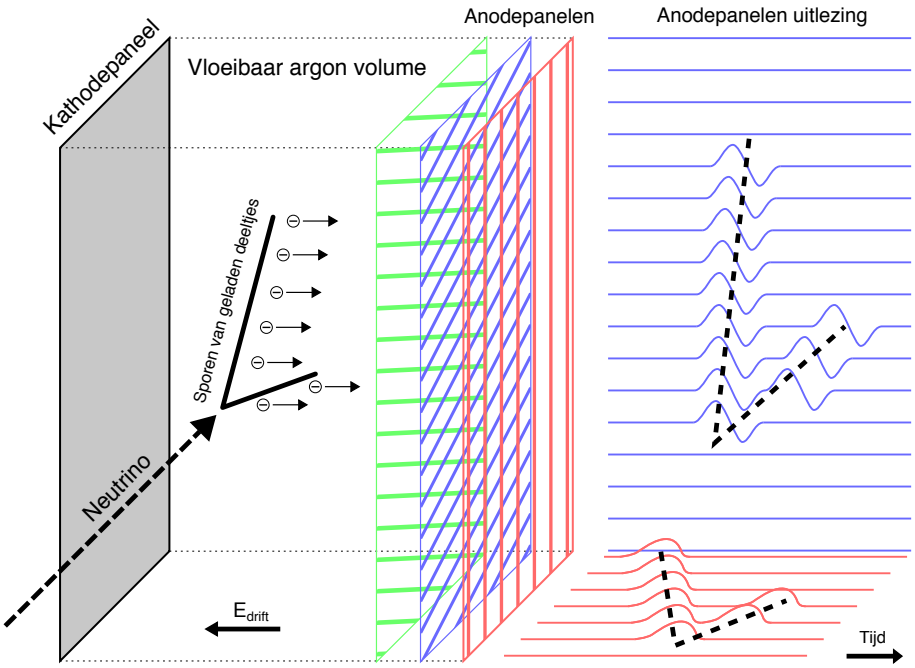

Figure S.3: De werking van de (Proto)DUNE TPC. Geladen deeltjes van neutrinointeracties creëren een ionisatiespoor waarvan de elektronen naar de anode worden geleid. Deze registreert de positie en aankomsttijd van de elektronen zodat de interactie in drie dimensies kan worden gereconstrueerd.

Samen met het Neutrino Platform van CERN heeft de DUNE collaboratie twee ProtoDUNE prototypedetectoren gebouwd, genaamd 'single-phase' en 'dual-phase' naar de toestand waarin hun argon inhoud zich bevindt. Het single-phase prototype is het onderwerp van dit proefschrift en was gebouwd tussen 2016 en 2018 om gebruik te maken van een geplande geladen deeltjesbundel aan het einde van 2018. In de zogenaamde 'beam run' legde ProtoDUNE meer dan 4 miljoen interacties van bundeldeeltjes vast, waarmee het zijn data-collectie doelen bereikte en in sommige gevallen zelfs overtrof. Daarnaast bleek ProtoDUNE een waardevol platform voor het testen van technologie die gebruikt zal worden in de DUNE far detector.

Hoofdstuk 4: Datatoegang en compressie in de ProtoDUNE DAQ

Moderne experimenten in de deeltjesfysica zijn afhankelijk van grote aantallen deeltjesinteracties om conclusies te trekken over hun onderliggende fysische processen. Om deze aantallen te bereiken moeten experimenten met een zo hoog mogelijke snelheid

interacties opnemen. ProtoDUNE nam met hoge resolutie 25 beelden op per seconde van interacties tussen bundeldeeltjes en argon. Elk van ProtoDUNE's zes anodepanelen heeft 2560 kanalen waarmee het met 2 MHz voltagemetingen uitvoert om 25 keer per seconde een beeld te vormen. Om de datastroom uit deze 15.360 kanalen te limiteren worden hun metingen alleen doorgelaten binnen een bepaald tijdsvenster rond interacties van bundeldeeltjes. De ruwe datastroom is niettemin substantieel: 91 Gbps per anodepaneel.

ProtoDUNE gebruikte twee data-acquisitiesystemen om zijn TPC datastroom te beheren. Een ervan is gebaseerd op het Reconfigurable Cluster Elements systeem en werd gebruikt als betrouwbaar basissysteem op vijf van de zes anodepanelen. Het Front-End LInk eXchange systeem (FELIX) werd gebruikt om de data van het laatste anodepaneel te verwerken en diende als een prototype systeem. Het was origineel ontworpen voor gebruik in de ATLAS detector en is gebaseerd op de PCIe interface, wat veel voorkomt in commerciële computerhardware. Het grootste voordeel hiervan is dat er geen speciaal ontworpen hardware nodig is om de detector uit te lezen naast de FELIX PCIe kaart en dat er dus gebruik kan worden gemaakt van de snelle ontwikkelingen binnen commerciële computers en telecommunicatie.

Terwijl het data ontvangt van de detector snijdt FELIX het relevante venster uit de datastroom en geeft het door aan zijn hostsysteem, de server waarin hij is geïnstalleerd. Om deze data te manipuleren en door te geven moeten de waardes binnen de data eerst toegankelijk worden gemaakt. In dit hoofdstuk ontwikkelen we een methode om toegang te krijgen tot waardes van individuele anodedraden uit de datapakketjes die FELIX doorgeeft aan zijn hostsysteem. We doen dit door een zogenaamde 'overlay' te ontwikkelen, gebruikmakend van bit-fields uit de programmeertaal C die het patroon volgen van structuren in de data. Het interpreteren van de data door dit overlay bleek een efficiënte manier om een individuele waarde of een reeks aan waardes uit te lezen. Het overlay wordt voornamelijk gebruikt door de event builder, die de data van een interactie uit alle detectorsystemen combineert in één makkelijk leesbaar pakketje. Andere algoritmes maken ook gebruik van dit overlay: de online monitor gebruikt het om de detectordata te lezen en te visualiseren terwijl metingen binnenkomen, en algoritmes om de data te herordenen en comprimeren hebben eveneens directe toegang tot de data nodig door het overlay.

In dit hoofdstuk ontwikkelen we daarnaast compressiealgoritmes voor de ProtoDUNE data-acquisitieketen en bestuderen hun prestaties toegepast op MicroBooNE data. Compressie geïmplementeerd in software blijkt zeer effectief en overtroeft de

beoogde compressiefactor van 4 ruim. Dit compressiealgoritme is echter niet snel genoeg om de rest van de data-acquisitie bij te houden, deels door de complexiteit van de datastructuur. In plaats van compressie in software hebben we daarom de data herschikt in een meer gestructureerde vorm alvorens het aan een hardwarekaart te voeren die specifiek op compressie gericht is, aangestuurd door Intel® QuickAssist Technology. We laten zien dat deze methode erin slaagt om de beoogde compressiefactor en compressietijd te behalen. De overlay kan alle dataformats zonder problemen lezen, inclusief herschikte en gecomprimeerde data, dankzij de format-informatie die met de data wordt bijgeleverd en wordt gelezen door de overlay. Het eindresultaat is een gebruiksvriendelijk en flexibel systeem dat een groot aantal toekomstige dataformats aan zal kunnen terwijl het alle formats uit het verleden blijft kunnen lezen.

Hoofdstuk 5: Reconstructie

Reconstructie is de brug tussen data en fysica. Het is de taak van de reconstructieketen om de ruwe data te interpreteren en om te vormen tot deeltjessporen. In ProtoDUNE spreidt de reconstructieketen deze processen uit over meerdere trappen. In de eerste stap vindt de signaalverwerking plaats, welke de data opschoont en de meest voorspelbare ruis wegneemt. De keten karateriseert vervolgens de pieken in de stroom op de anodedraden als discrete hit objecten, elk met zijn eigen hoogte en breedte. Reconstructiealgoritmes vinden vervolgens matches tussen hits van meerdere drade nvlakken om de lading in driedimensionale ruimte te plaatsen. Volgende stappen zorgen voor de ontwikkeling van objecten op een hoger niveau. Een clusteralgoritme verzamelt hits in afzonderlijke objecten waarvan het denkt dat het door één deeltje werd veroorzaakt. Als laatste worden deze clusters een type toegewezen afhankelijk van hun karakteristieken.

De reconstructieketen van ProtoDUNE identificeert twee types objecten die door een enkel deeltje kunnen worden gecreëerd: tracks en showers. De vloeibaar argon TPC's van ProtoDUNE en de DUNE far detector hebben een ruimtelijke resolutie van ongeveer 0.5 cm, waardoor ze in staat zijn om individuele sporen van geladen deeltjes te registreren. Onder de track-achtige deeltjes vallen hadronen zoals protonen, pionen en kaonen, maar ook muonen vormen herkenbare tracks in de detector. In ProtoDUNE, die op het oppervlak ligt, zorgen kosmische muonen voor een constante achtergrond bij bundeldeeltjesinteracties. Hoog-energetische elektronen, positronen en fotonen ondergaan snelle elektromagnetische interacties met vloeibaar argon waardoor ze steeds

meer elektronen, positronen en fotonen aanmaken. Deze cascades worden showers genoemd en zijn zelfs te dichtbevolkt voor de hoge resolutie TPC van ProtoDUNE om individuele deeltjessporen in te zien. Het onderscheid tussen tracks en showers is belangrijk omdat het direct is verbonden met het type van het ioniserende deeltje. Dit onderscheid is echter niet altijd duidelijk, vooral in het geval van kleine clusters gemaakt door relatief laag-energetische deeltjes.

Voor de analyses in dit proefschrift hebben we gebruik gemaakt van honderdduizenden gesimuleerde en echte deeltjesinteracties. We zijn alleen geïnteresseerd in secundaire deeltjes gemaakt in de interactie tussen het bundeldeeltje en het argon in de detector. Daarom selecteren we alleen interacties voortkomend uit hadronische bundeldeeltjes, aangezien elektromagnetische deeltjes normaliter een enkele shower creëren zonder afzonderlijke secundaire deeltjessporen. In dit hoofdstuk laten we zien dat belangrijke aspecten van de gesimuleerde bundelinteracties goed overeenkomen met de realiteit, al zijn er kleine verschillen in de creatie van secundaire deeltjes.

Hoofdstuk 6: Identificatie van fotonshowers

Om neutrino-oscillatie te meten zal DUNE de smaak moeten identificeren van de neutrino's die interacties aangaan in zijn detectoren. Neutrino's zijn zelf onzichtbaar, maar kunnen worden geïdentificeerd door hun interactieproducten. Het elektron-neutrino is vooral belangrijk in DUNE en heeft een karakteristieke manier van interageren waarbij het een elektromagnetische shower produceert. Er is een interactie die in het bijzonder deze meting belemmert: een neutrino van een willekeurige smaak kan een π^0-deeltje creëren dat meteen vervalt naar twee fotonen welke beide hun eigen elektromagnetische showers maken. Als een van deze showers niet als zodanig wordt herkend, kan het dus lijken alsof deze interactie specifiek door een elektron-neutrino was aangezet. Onderscheid kunnen maken tussen showers van elektronen en fotonen is dus essentieel voor de fysicadoelen van DUNE.

Door de energieafzetting rond het startpunt van showers te bepalen kan het type van het veroorzakende deeltje bepaald worden. Elektronen en positronen zijn elektrisch geladen en beginnen gelijk het argon in de detector te ioniseren. Fotonen, aan de andere kant, hebben geen elektrische lading en zijn dus onzichtbaar totdat ze interageren met het argon en een elektron-positron paar creëren. Het begin van een fotonshower bestaat dus uit twee overlappende elektronshowers en zet dus twee keer zoveel energie af als een enkele elektronshower.

In dit hoofdstuk laten we zien dat de energieafzetting rond het begin van fotonshowers inderdaad twee keer zo hoog is als dat in elektronshowers. We beginnen met het selecteren van bundeldeeltjesinteracties om een puur fotonsample op te bouwen. Daarna plaatsen we een cilinder rond het begin van elke shower om de regio te definiëren waarin we een duidelijk verschil verwachten tussen elektron- en fotonshowers. Figuur S.4 laat dit proces grafisch zien. We bepalen de energieafzetten voor elke hit in deze regio en nemen vervolgens de mediaan van de resulterende groep waardes, wat een duidelijk resultaat oplevert twee keer zo groot als dat voor elektronshowers. Hoewel er andere methodes bestaan om elektronshowers van fotonshowers te onderscheiden, verschaft deze procedure een veelbelovend aanvullend handvat op het deeltjestype van de shower.

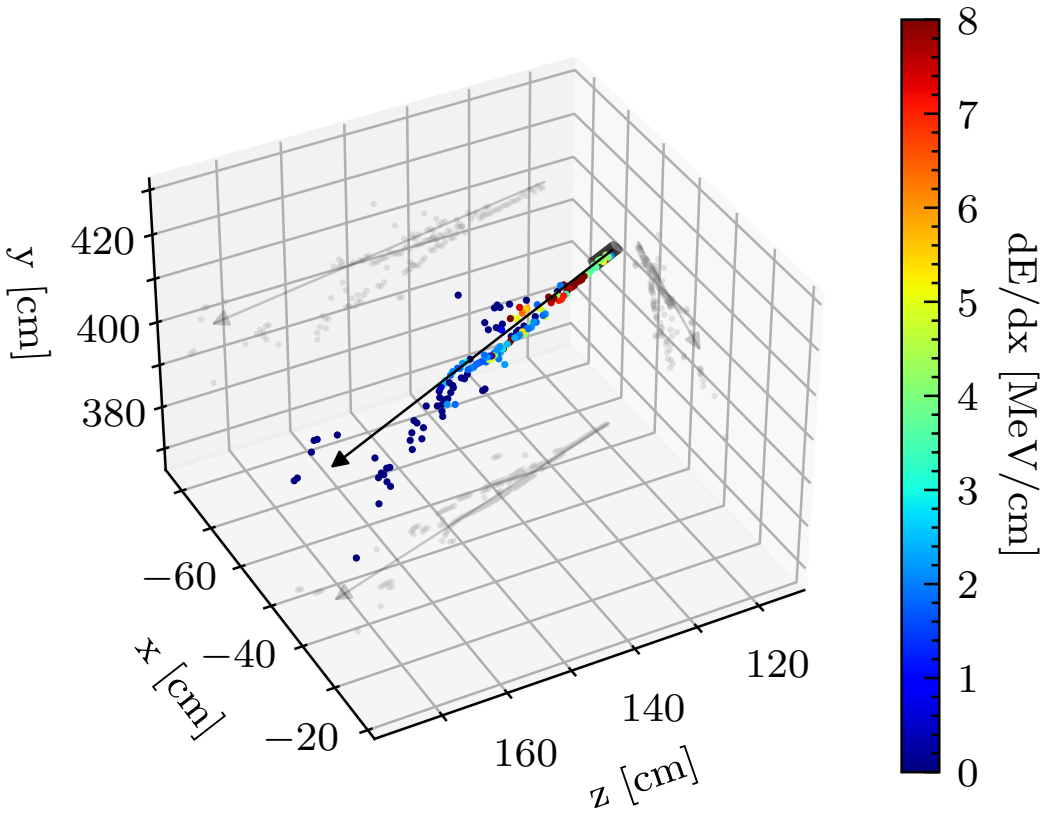

Figure S.4: Een enkele shower in drie dimensies als een wolk hits (gekleurde stippen) en als een gereconstrueerd object met een startpunt, richting en lengte (zwarte pijl). Tweedimensionale projecties van de shower zijn weergegeven als transparant grijze stippen en pijlen. De kleur van de stippen geeft de grootte van de energieafzetting in de hits aan. Een transparant grijze cilinder geeft de startregio van de shower aan.

Hoewel het onderscheid tussen elektron- en fotonshowers duidelijk is met deze methode, blijft het open voor verbetering. Door de dichtbevolkte deeltjesomgeving in ProtoDUNE moet deze methode gebruik maken van een kleine cilinder om de startregio te definiëren zonder ongerelateerde hits mee te nemen. Een robuustere methode zou kunnen voortkomen uit de energieafzetting van de gehele shower, wat een goed startpunt voor toekomstig onderzoek zou kunnen zijn.

Hoofdstuk 7: Reconstructie van neutrale pionen

Zoals gezegd zorgen π^0-deeltjes voor een groot deel van de ruis in neutrino-identificatie. Desalniettemin kunnen ze nuttig zijn in energiekalibratie. Dit deeltje vervalt bijna direct na hun creatie en produceert bijna altijd twee fotonen, welke het 180° van elkaar af uitzendt als het zelf niet in beweging is. Als het wel in beweging is, nemen zijn vervalsfotonen deze beweging over en krijgen een zogenaamde boost in dezelfde richting. De resulterende hoek tussen de fotonen (θ), in combinatie met hun energie (E_1, E_2), bepaalt een grootheid die de invariante massa van het π^0-deeltje wordt genoemd:

$$m_\pi = \sqrt{2E_1E_2(1-\cos\theta)}.$$

De invariante massa wordt zo genoemd omdat hij onveranderlijk is onder verschillen in de kinetische energie van het π^0-deeltje. Hierdoor is het identiek aan de rustmassa van dit deeltje, welke zeer nauwkeurig bekend is. De invariante massa is daardoor een waardevol handvat om te bepalen hoe goed de reconstructie van de shower-energie is. Figuur S.5 laat een bijzonder mooie interactie zien waarin een waarschijnlijk π^0-verval goed te zien is.

Omdat de meeste fotonen in ProtoDUNE van π^0-verval afkomen kunnen we een pure dataset van π^0-vervalproducten verkrijgen door simpelweg een pure dataset van fotonen te selecteren. We kunnen daardoor een groot deel van de sample-selectie overnemen van het vorige hoofdstuk omdat die zich ook op fotonshowers richt. De showers van π^0-verval komen bijna altijd voor in paren en ons sample moet daarom ook zijn opgebouwd uit paren van showers. Het proces om paren te vinden in een dataset van showers wordt bemoeilijkt op meerdere manieren. π^0-verval showers hebben vaak een te lage energie om überhaupt te worden gereconstrueerd en als ze wel worden gereconstrueerd is dat vaak dicht bij elkaar, waardoor ze er vaak als één shower uitzien. Daarnaast is er een kans dat er meerdere π^0-deeltjes in dezelfde interactie

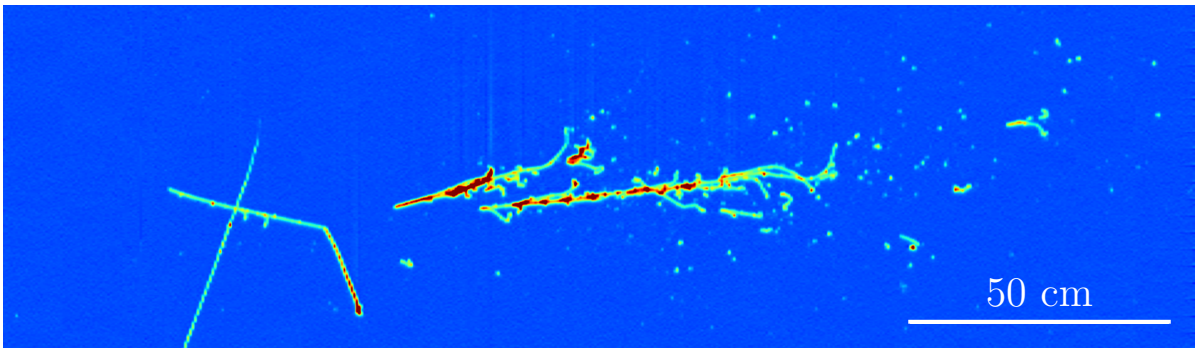

Figure S.5: Een voorbeeld van π^0-verval in ProtoDUNE. Het bundeldeeltje komt het beeld binnen van links en creëert een π^0-deeltje wat direct vervalt naar twee onzichtbare fotonen welke elk een elektromagnetische shower veroorzaken. Overgenomen van [55].

worden geproduceerd, wat ertoe kan leiden dat twee ongerelateerde fotonshowers in een paar worden geplaatst. We hebben deze kwesties zo goed mogelijk opgelost, maar het feit blijft dat het π^0-deeltje zich niet makkelijk laat reconstrueren.

We hebben allereerst de energiereconstructie van fotonshowers gebaseerd op informatie uit simulaties. Na alle correcties toegepast te hebben om de lading in de shower over te houden rest een enkele factor R als de brug tussen de lading en energie van de shower. In gesimuleerde events kunnen we deze waarde precies zo zetten dat de gereconstrueerde energie zo goed mogelijk overeen komt met de ware energie van het foton. Met deze waarde van R berekeken we de invariante π^0-massa in echte bundelinteracties en vinden slechts een klein verschil met de simulaties. We bepalen dat het totale verschil in R 1.05 ± 0.12 is. Dit verschil is klein, maar kan worden opgelost door de waarde van R in echte bundelinteracties te vermenigvuldigen met een factor $1/1.05$.

Conclusies en vooruitzicht

De analyses in dit proefschrift betreffen fotonshowers in ProtoDUNE, maar alle gebruikte en ontwikkelde methodes zijn toepasbaar op de DUNE far detector. Nu we hebben laten zien dat foton- en elektronshowers in ProtoDUNE al verschillen in hun initiële energieafzetting verwachten we dat deze methode nog beter zal werken in DUNE, waar shower objecten naar alle waarschijnlijkheid beter afgetekend zullen zijn tegen een achtergrond met minder ruis en andere deeltjes. De energiekalibratie van showers zal eveneens essentieel zijn in DUNE en meerdere projecten zijn al begonnen

dit onderwerp te bestuderen. Zowel een zegen als een vloek, het π^0-deeltje zal nog lang in het analyse-voetlicht staan. In de tussentijd vordert de bouw van DUNE voor een mooie toekomst van precisieneutrinofysica.

Acknowledgements

As the only person at Nikhef to work on DUNE full-time, I sometimes looked at people in a tight-knit working group with a tinge of envy. As the years of my PhD programme passed by, however, I found that my unique position in the institute allowed me to meet and work with many more people than I otherwise might have. At Nikhef I was placed in the ATLAS group initially, then had the privilege to spend a year at CERN, where I got to work with people in data acquisition, detector commissioning and ProtoDUNE physics. I joined the neutrino group after coming back to Nikhef and spent a several months in a small group of DUNE colleagues at the University of Houston. Most of my work contacts were already strictly online, but when the 2020 pandemic hit, all of them moved to the realms of the internet. I was fortunate to already be on a trajectory towards a finished thesis by that time, for which I have everyone along the way to thank.

My first thanks go to my promotor and co-promotor, Prof. Dr. Stan Bentvelsen and Dr. Frank Filthaut. You were brave enough to open a PhD position in an experiment where Nikhef had little presence and afforded me a lot of independence and freedom. At the same time, you never failed to impress me with your deep insight, knowledge and eye for detail on topics that you had only a fraction of your time to spend on.

My gratitude similarly goes out to my doctorate committee: Prof. Dr. Auke Pieter Colijn, Prof. Dr. Ir. Paul de Jong, Prof. Dr. Sijbrand de Jong, Dr. Giovanna Lehmann-Miotto, Dr. Marieke Postma, Prof. Dr. Wouter Verkerke and Prof. Dr. Marcel Vreeswijk. Thank you for taking the time to read my thesis and provide valuable feedback.

I have the Nikhef ATLAS group to thank for welcoming me into the institute with lots of table football and Friday evening borrels. Daan, Kees, Marc, Pepijn, Terry, Matteo, Lydia and many others, thanks for all the good times back when many of us were just starting out on our PhD journey. The LHCb group also made Nikhef feel like home. Thank you especially, Daniel, Michele, Maurício and Laís. Likewise at Nikhef, I would like to thank Andrea, Jos, Henk and Frans for patiently helping me understand

the FELIX DAQ. Eric and Tom, thank you as well for your input and indispensable help from the DUNE DAQ perspective.

I had the great privilege to work at CERN for a little over a year. Broos, thanks for forming a house of Vermeulens with me during this time. I enjoyed your company and I think we really made the apartment in the fields just outside of idyllic Ferney-Voltaire into our own. Nadia and Daniil, I want to thank you for warmly welcoming me to CERN. Giovanna, thank you for accepting me into the EP-DT group at CERN and for your patient yet urgent guidance. Roland and Enrico, in the last four years you were the people I have worked with most closely. When I started my work on DAQ at CERN I had little programming experience and I cannot overstate how much I learned from you. Aside from that, I fondly remember our late nights of drinks and video games in Saint-Genis. Marco, Danilo and all others in the group, thank you for the good times.

The construction of ProtoDUNE in so little time was a great challenge and I am glad and honoured to have made my contributions to it. The end result was made possible through the work of an amazing group, some individuals of which I would like to highlight in particular. Roberto and Geoff, thank you for the warm welcome into the group and continued support throughout the months of work, as well as the many, many barbecues. Jake and Chris, thanks for your friendship and company. Grazie Maura and Francesca for much the same. Aidan and Sarah, your work ethic and humility inspire me to this day. Serhan, Flor, Kevin, Stephen and many others, thank you for being the technical backbone of the experiment and an integral part of the group. Thank you for your company, Peter, Bryan, Andrew and all others from the American groups. Likewise to you, James, Marco, Alex, Fabio and all the others from the British groups. Ajib and Heng-Ye, you were some office-mates I could truly depend on in these times. On the DUNE side, thank you Erin, Alex and Jason for your friendship. Casandra, ProtoDUNE brought us together and I will always be grateful for that. Thank you for coming on this adventure with me.

Going into my analysis, I have many people to thank who helped me on my way. Robert and Dorota, thank you for so patiently introducing me to the larsoft analysis framework. Leigh, Tingjun and Stefania, thank you for all your advice and guidance on my analysis. Steve and Lorena, thanks for helping me to untangle the quirky world of Pandora particle reconstruction.

Ruud, Luc and Bjarni, thank you for the good times throughout the years. We have had many interesting voyages, but the one to Zermatt where we hiked over (and sometimes slid down) the mountains around the Matterhorn was particularly memorable. Here's to many more exciting trips in the future! Additionally, thank you Giovana and Erin for your longtime company and support from afar. I feel very grateful to have known you.

Leaving CERN after an amazing year was difficult, but the neutrino group at Nikhef made it easy by accepting me with open arms. As the only DUNE-collaborator in the group, I gave plenty of presentations that were mostly irrelevant to everyone else, so thank you to all for your interest and advice in spite of that. Thank you Paul, Maarten, Ronald, Dorothea and the other seniors of the group for your expert views on reconstruction issues. Likewise, thank you to Bruno, Alfonso, Valentin, Rodrigo, Karel, Rasa, Bouke, Thijs and Clara for the creative insights when talking about physics, but also for your company and friendship. In particular, thank you Jordan for involving me in your speedruns and fortnightly projects, Brían for teaching me to enjoy the finer Guinnesses in life and Lodewijk for the whisky tastings and Smash Bros. matches. My thanks also go Suzan and Fatih for being my wonderful office-mates during this time.

During the couple of months I spent at the University of Houston, I have Lisa to thank for graciously accepting me into her group and Aaron for helping in laying the groundwork for my analyses. Even if it was just for a couple of months, I am grateful for the close collaboration.

Finally, I would like to thank my family for their unwavering support. Thank you Bas, Edith and Luc for being the best uncle, aunt and cousin I could wish for. Most of all, I would like to thank my parents and (not so) little brother. Thank you Dorien, Jos and Tim for always being there when I needed you. You have been my mental bedrock and safe haven. I could not have done it without you.