Convolutional neural networks for the CHIPS neutrino detector R&D project

Josh Chalcraft Tingey
University College London

Submitted to University College London in fulfilment
of the requirements for the award of the
degree of **Doctor of Philosophy**

Declaration

I, Josh Chalcraft Tingey confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Josh Chalcraft Tingey

Abstract

The CHerenkov detectors In mine PitS (CHIPS) neutrino detector R&D project aims to develop novel strategies and technologies for very large yet 'cheap as chips' water Cherenkov neutrino detectors. Via deployment in a body of water, use of commercially available components, and instrumentation coverage optimisation for the study of exclusively accelerator beam neutrinos, CHIPS will enable megaton scale detectors to become a reality at the cost of $200k-$300k per kt of sensitive mass. During the summer of 2019 a prototype CHIPS detector, CHIPS-5, was deployed into the Wentworth 2W disused mine pit in northern Minnesota, 7 mrad off the NuMI beam axis. A novel data acquisition system was introduced using cheap single-board computers and open-source software.

This work presents a novel approach to water Cherenkov neutrino detector event reconstruction and classification. Three forms of a Convolutional Neural Network, a type of deep learning algorithm, have been trained to reject cosmic muon events, classify beam events, and estimate neutrino energies, all using only the raw detector event as input. When evaluated on the expected distribution of CHIPS-5 events, this new approach is shown to be robust and explainable as well as providing a significant performance increase over the standard likelihood-based reconstruction and simple neural network classification. Promisingly, the performance presented here is comparable to the more complex (and expensive) neutrino oscillation experiments within the field.

Impact Statement

The impact of this work inside the domain of academia is straightforward and explained in great detail throughout the thesis. A novel application of modern machine learning techniques is made to reconstruct and classify neutrino events within a water Cherenkov detector studying neutrino oscillations. Although other experiments within the field have conducted similar preliminary work, this is the first known comprehensive application of such methods. Hopefully, this work will promote the use of these techniques for water Cherenkov neutrino detectors more broadly.

Outside of academia, the impact of this work is less clear. Still, it is probably most pronounced in the field of machine learning, which promises to bring substantial societal advancements. Sometimes the application of machine learning to real-world problems is seen as somewhat secondary to achieving marginally incremental improvements on standardised challenges. Hopefully, this work contributes towards a broader effort to make the application of machine learning methods more central and broaden the scope of tasks considered. Currently, the field is dominated by the commercial needs of a small number of large technology companies, such as Google and Facebook, any work that applies machine learning to a novel task can only help to broaden the scope of the field and make it more applicable to a wider range of real-world tasks.

Acknowledgements

Firstly I would like to thank my supervisor Jenny for the independence she offered me throughout by PhD to pursue the research path that interested me the most, and furthermore, for leading the whole CHIPS collaboration through two intense summers of construction activity in Hoyt Lakes.

I would also like to thank everyone within the broader CHIPS collaboration. Foremost Stefano for his incredible support and advice throughout all the stages of my PhD, as well as the UCL group of Tom, Sim, Petr, and John for always being around and making the gulag bearable. I would also like to thank everyone else within the UCL HEP group, particularly my D14 office mates.

Above all, I must thank Becca and my family for all of their endless support.

Preface

Here I outline my contributions to the work described in this thesis on a chapter by chapter basis. Hopefully, this will make clear what is my own work and what is that of others. Due to the relatively small number of core people involved with the CHIPS project, I was somewhat unusually involved with a much broader scope of work than usual for a HEP PhD student, as were the majority of the CHIPS team.

In Chapter 3 the CHIPS project and the CHIPS-5 detector are described. I was personally heavily involved with developing the event generation, detector simulation, and event reconstruction frameworks for CHIPS, alongside extensive detector optimisation studies to inform the final configuration of the CHIPS-5 detector. I also contributed to the CHIPS-5 construction efforts, primarily instrumentation testing, calibration, and installation, amongst others.

Chapter 4 describes the data acquisition system developed and implemented for CHIPS-5. I played a major role in the development, construction, testing, and installation of this system. Specifically, I made a significant contribution to the development of the networking solution, the high-level hardware implementation, the control and monitoring software, and the development of the novel low-level Madison ţDAQ and Beaglebone system.

Chapter 5 and Chapter 6 describe the application of Convolutional Neural Networks to the characterisation of neutrino events within CHIPS-5. The work in both these chapters was solely conducted by me, representing the bulk of my efforts. I developed the complete pipeline, from event generation, simulation, model development and training, to comprehensive evaluation.

Contents

Chapter 1

Introduction

It is abundantly evident that machine learning and the broader field of artificial intelligence will play an ever-increasing role in the world around us over the coming years. Virtually every industry and human on the planet are promised to be touched by their proliferation. Within experimental particle physics, the application of modern machine learning methods should bring significant progress; however, as with any technique, the inherent weaknesses should always be kept in mind.

Already, a widespread revolution is underway across the field. From the way particle interactions are simulated to the way recorded events are analysed, machine learning techniques, usually deep learning algorithms, are yielding dramatic performance improvements. This step-change is particularly true for the study of the vastly abundant yet incredibly difficult to detect neutrino. Principally driven by the fact that the raw output from neutrino detectors is well suited to the algorithms at the forefront of computer vision research, many neutrino experiments now routinely use deep learning methods for event analysis.

However, to date, a thorough end-to-end implementation of such techniques is not yet in use for the reconstruction and classification of neutrino events within long-baseline water Cherenkov detectors studying accelerator beam neutrinos. Without redress, this lack of progress could have significant implications for the future of neutrino physics, especially when such experiments are deemed a highly promising (and potentially cheap) channel for answering some of the critical unsolved problems of the field.

This thesis presents a broad range of work conducted for one such experiment, the CHerenkov detectors In mine PitS (CHIPS) neutrino detector R&D project. CHIPS aims to develop very large yet 'cheap as chips' water Cherenkov detectors that can be deployed into deep bodies of water on the Earth's surface, allowing megaton scale detectors to become a reality. The thesis's principal work involves the novel application of Convolutional Neural Networks, a type of deep learning algorithm, to the reconstruction, classification, and energy estimation of neutrino events within the CHIPS-5 prototype detector module.

It is hoped that the work presented in this thesis will achieve three principal goals. Firstly, motivate the need for the CHIPS project and detail how it accomplishes its aims. Secondly, show that Convolutional Neural Networks can be used to fully reconstruct and classify neutrino events within water Cherenkov detectors in an explainable and robust manner whilst also providing significant performance improvements. Finally, inform the development of similar Convolutional Neural Network implementations for other water Cherenkov neutrino detectors.

To begin, Chapter 2 introduces the history of neutrino physics, as well as the theoretical background and current status of the field as motivation for the CHIPS project. A full description of the CHIPS detector concept follows in Chapter 3, with a principal focus on the CHIPS-5 prototype detector module deployed into the NuMI neutrino beam during the autumn of 2019. A particularly detailed overview of the CHIPS-5 data acquisition and monitoring systems is then given in Chapter 4.

Chapter 5 begins by introducing the standard reconstruction and classification methods to be replaced alongside a review of the relevant neural network theory and current deep learning applications within the field of neutrino physics. The chapter concludes by detailing the three Convolutional Neural Networks that have been developed for CHIPS-5, including that for cosmic muon rejection, beam event classification, and neutrino energy estimation.

The thesis closes in Chapter 6 with a comprehensive evaluation of the new Convolutional Neural Network approach. Firstly, the final combined performance is determined and compared with similar experiments. Secondly, the inner workings of the trained networks are explored. Thirdly, the robustness of the network outputs to distributional changes in the input is studied. Finally, alternative implementations are discussed to highlight the key factors driving performance.

Chapter 2

Neutrino physics

Vast numbers of neutrinos pass through everything that surrounds us each second, each one incredibly unlikely to interact even once. Nearly a century since they were first proposed [1], these mysterious particles have now conclusively been proven to undergo oscillations between their flavour states. This profound discovery has opened the door to physics beyond that initially conceived within the Standard Model, which may reveal new, fundamental insights into the universe. This chapter aims to outline the historical context, theoretical background, and open questions surrounding the neutrino to motivate the CHIPS R&D project.

2.1 A history

2.1.1 Discovery of the neutrinos

In the early 20th century, beta decays were assumed to follow the simple two-body process, $A \rightarrow B + e$, where A spontaneously emits a single electron. To conserve both energy and angular momentum, the ejected electron was believed to have discrete kinetic energy defined by the difference in binding energies between A and B. However, in 1914, J. Chadwick instead measured a continuous electron energy spectrum [2], placing this assumption in doubt.

W. Pauli proposed a 'desperate solution' to this paradox in 1930 [1]. If a light, neutrally charged, spin 1/2 particle was also produced in the interaction, the continuous energy distribution could be explained. Initially, this peculiar new particle was named the *neutron*. However, to avoid confusion with the heavy baryon of the same name discovered in 1932, E. Fermi renamed it the *neutrino* when he formalised his theory of beta decay in 1934 [3].

The same year, H. Bethe and R. Peierls [4] used Fermi's work to estimate the cross-section for the inverse beta decay process

$$\bar{\nu} + p^+ \rightarrow n + e^+. \tag{2.1}$$

They calculated an upper limit of $\sigma < 10^{-44}\mathrm{cm}^2$ for neutrinos of energy 2.3 MeV, an incredibly small value. When summarising, they declared that 'it seems highly improbable that, even for cosmic ray energies, σ becomes large enough to allow the process to be observed'. Although this estimation did not prove accurate, and extensive neutrino observations have since been made, it hinted at the great difficulties experimentalists would face measuring the neutrino in the years to come.

After an initial tentative identification in 1953, F. Reines and C. Cowan made the first confirmed observation of the neutrino in 1956 [5]. Electron antineutrinos produced within the Savannah River Plant nuclear reactor were detected via the inverse beta decay process outlined in equation (2.1). In an underground room of the reactor building, a 'club-sandwich' detector was constructed containing three 1500 litre liquid scintillator tanks interspersed with two 200 litre cadmium doped water target tanks. A total of 330 photomultiplier tubes then measured the characteristic signal for the interaction, a positron annihilation followed shortly afterwards by a gamma-ray burst from the neutron capture of the neutrino.

A second distinct neutrino, the muon neutrino, was discovered in 1962 at the Brookhaven based Alternating Gradient Synchrotron (AGS) [6]. Protons from the AGS beam incident upon a fixed target produced charged pions which then decayed into a beam of muons and neutrinos. After passing through steel and lead absorbers to remove the muons, neutrino interactions were detected in a series of downstream spark chambers.

If only a single neutrino existed, both interactions

$$\nu + p^+ \rightarrow \mu^+ + n, \tag{2.2}$$

$$\nu + p^+ \rightarrow e^+ + n, \tag{2.3}$$

were expected to occur at the same rate. However, predominantly single muons (alongside backgrounds), identified by a single long track in the spark chambers were detected,

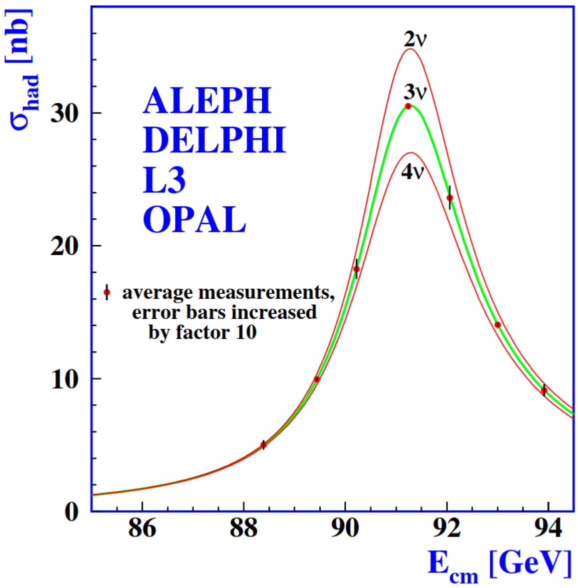


Figure 2.1: Combined hadron production cross-section measurements around the Z^0 resonance from the LEP experiments. The curves show the predicted cross-section for two, three, and four active neutrinos. Note how the data fits the three active neutrino hypothesis incredibly well. Figure taken from reference [10].

confirming the existence of the muon neutrino. Not only was this experiment the first to construct and use an artificial neutrino beam, but it also won the 1988 Nobel Prize in Physics.

The Z^0 and $W^\pm$ bosons were discovered at the CERN based Super Proton Synchrotron in 1983 [7, 8]. Crucially, as Z^0 bosons were expected to decay to neutrinos, measurements made to the decay width could strongly constrain the number of neutrino flavours. The ALEPH, DELPHI, L3, and OPAL experiments at the LEP e^+e^- collider (see Figure 2.1), as well as the Mark II experiment at SLAC [9], made such precise measurements in the 1990s, indicating that the number of light active neutrino flavours was 2.984 ± 0.008 [10].

This indication from LEP combined with the discovery of the charged tau lepton in 1975 [11] made the existence of a third tau neutrino extremely likely. The Fermilab based DONUT experiment finally discovered this particle in 2001 using 800 GeV protons from the Tevatron [12], completing the trio of neutrinos we know today. However, additional *sterile* neutrino flavours which do not couple to the weak force are still a possibility, with various hints to their existence having been observed [13].

2.1.2 Discovery of neutrino oscillations

At the Homestake mine in the 1960s, a 400 000 litre tank, placed 1.5 km underground, was filled with the dry cleaning fluid perchloroethylene (C_4Cl_8). Its goal was to measure the electron neutrino flux incident upon the Earth from the Sun via the interaction

$$^{37}Cl + \nu_e \rightarrow {}^{37}Ar + e^-. \tag{2.4}$$

This process (equation (2.4)) allowed the neutrino flux to convert the chlorine contained within the tank into the noble gas argon. Every few weeks the tank was purged with gaseous helium and the amount of argon generated, and indirectly the solar neutrino flux, measured.

After analysis, the number of electron neutrino interactions per ^{37}Cl per second, was found to be no greater than 3 [14]. When compared to the predictions made by the Standard Solar Model ranging between 4.4 and 22 [15], a deficit was observed. Dubbed the *solar neutrino problem*, this was initially believed to be due to an unexplained experimental flaw. However, other experiments, such as the water Cherenkov Kamiokande II [16] and the SAGE and GALLEX galium based capture tanks also observed this discrepancy [17, 18].

A second deficit, the *atmospheric neutrino anomaly*, was also indirectly observed for neutrinos generated in the Earth's upper atmosphere by cosmic rays. Such neutrinos formed a pivotal background to both the Kamiokande and IMB experiments, which were primarily designed to measure proton decay. When evaluating their backgrounds, a deficit in the number of muon neutrinos compared to electron neutrinos was observed [19, 20]. The successor to the Kamiokande experiment, Super-Kamiokande also measured a similar deficit [21].

The phenomenon of neutrino oscillations was put forward as a solution to this problem. If neutrinos could change flavour as they propagated, the measured deficits could be explained. The SNO experiment finally confirmed such oscillations in 2001 [22].

SNO consisted of a 1 kt tank of deuterium (heavy water), equipped with 9500 photomultiplier tubes. Light from three separate neutrino interaction channels:

$$\nu_i + e \rightarrow \nu_i + e, \tag{2.5}$$

$$\nu_i + d \rightarrow p + n + \nu_i, \tag{2.6}$$

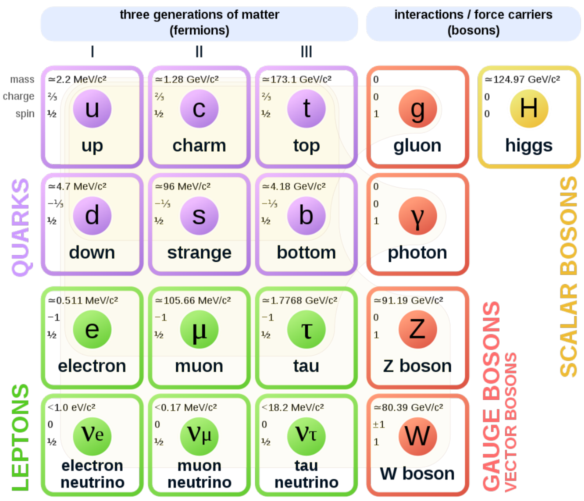


Figure 2.2: The particles of the Standard Model. Figure taken from reference [23].

$$\nu_e + d \rightarrow p + p + e, \tag{2.7}$$

was measured, where d is the deuterium nucleus. The first and second channels were sensitive to all three neutrino flavours, but crucially only electron neutrinos could interact via the third channel. By comparing the relative rates between the channels, SNO was able to prove to 5.3 σ that electron neutrinos had oscillated to other flavours.

2.2 The Standard Model and neutrinos

The Standard Model of particle physics describes the thirty-seven (ignoring colour) known fundamental subatomic particles and their interactions. Combining both quantum chromodynamics (describing the strong force) and electroweak theory (describing the electromagnetic and weak forces), the Standard Model is a gauge theory obeying the local gauge symmetries of $U(1) \times SU(2) \times SU(3)$. The particles, along with their various properties, are outlined in Figure 2.2, all have an associated anti-particle where some of the particles quantum numbers are inverted.

The twelve quarks (ignoring colour) and twelve leptons, all spin 1/2 particles, are named fermions. Further divided into three generations (or flavours), they make up all the Universe matter content. The quarks never exist in a free state and bind together into mesons or baryons, such as pions and protons, respectively. Three massive charged particles, the electron, the muon, and the tau and their three corresponding (initially assumed to be massless) neutrinos, make up the leptons alongside their antiparticles.

The spin 1 gauge bosons carry the electromagnetic, strong, and weak forces. The photon carries the electromagnetic force (affecting all charged particles), the eight gluons carry the strong force (binding the quarks together), and the massive Z^0 and $W^{\pm}$ carry the weak force. The final particle, the Higgs, is a massive scalar boson. Via the process of spontaneous symmetry breaking, the Higgs provides the mechanism to give the particles their mass.

Within the Standard Model, neutrinos only interact via the weak force, exclusively coupling to the Z^0 and $W^{\pm}$ bosons. As the weak interaction maximally violates parity, in that it only couples to left-handed chiral particles, the original Standard Model only contains left-handed neutrinos and their corresponding right-handed anti-neutrinos.

Although neutrinos were initially thought to be massless, and no direct detection of their mass has been made, it is now believed that neutrinos are massive (albeit very light) particles. This belief is due to the substantial amount of evidence in support of neutrino oscillations, such as the result from SNO amongst others detailed in Section 2.5. These oscillations require three distinct neutrino mass states, implying at least two are non-zero.

The Standard Model does not strictly rule out massive neutrinos, which are permitted as either *Dirac* or *Majorana* particles. If neutrinos are Dirac in nature, they acquire their mass via a Yukawa coupling, just like the other fermions. This coupling mixes both left-handed and right-handed fields, requiring the existence of right-handed neutrinos (left-handed antineutrinos) that do not interact via the weak force. Conversely, if neutrinos are Majorana in nature, a mass term can be introduced containing just the left-handed neutrino states. This term requires that the neutrino is its own anti-particle, which as neutrinos do not carry a charge, is possible [13].

Strongly model-dependent cosmological observations of the cosmic microwave background currently provide the best limit on the combined sum of the neutrino masses at $\sum m_\nu < 0.12$ eV [24]. However, more impressively, the KATRIN experiment has been able to make model-independent direct measurements of the upper mass limit by looking at the energy spectrum of electrons emitted from beta decay, giving a result of $m_\nu < 1.1$ eV [25].

2.3 Neutrino oscillations

B. Pontecorvo, Z. Maki, M. Nakagawa, and S. Sakata developed the theory of neutrino oscillations in the 1960s [26–28]. Fundamentally, it is a manifestation of the phenomenon of quantum interference. If neutrinos are massive, their mass eigenstates are not necessarily the same as their weakly interacting flavour eigenstates. Instead, the flavour states are a superposition of the three mass states, each propagating as distinct waves evolving differently with time. As a direct consequence, if a neutrino is created with a specific flavour α, its flavour composition will change with time, such that it may later be detected as having a flavour β. The probability of this change is found to be periodic (hence oscillations) and dependent on the neutrino energy, the propagation distance, and a rotational mixing matrix.

2.3.1 Neutrino mixing

The mixing between the flavour eigenstates $|\nu_\alpha\rangle$, where $\alpha = e, \mu, \tau$, and the mass eigenstates $|\nu_k\rangle$, where $k = 1, 2, 3$, is described by the *Pontecorvo-Maki-Nakagawa-Sakata* (PMNS) matrix U, such that

$$|\nu_\alpha\rangle = \sum_k^3 U^*_{\alpha k} |\nu_k\rangle \,, \tag{2.8}$$

and vice-versa

$$|\nu_k\rangle = \sum_\alpha^3 U_{\alpha k} |\nu_\alpha\rangle \,. \tag{2.9}$$

The PMNS matrix is a unitary, complex, 3×3 matrix, similar to the Cabibbo-Kobayashi-Maskawa (CKM) matrix for quark mixing, and has the form

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \\ |\nu_\tau\rangle \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \\ |\nu_3\rangle \end{pmatrix} . \tag{2.10}$$

Three mixing angles and six complex phases can generally describe a 3×3 matrix such as this. However, the majority of these phases can be removed without affecting any physical process. This leaves three mixing angles θ_{12}, θ_{23}, θ_{13} and a single phase δ which if non-zero produces CP violation and so is commonly denoted δ_{CP}.

With $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$, the standard parametrisation of U assuming neutrinos are Dirac particles is given by

$$\mathrm{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.11}$$

$$= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{pmatrix}. \tag{2.12}$$

If neutrinos are instead Majorana in nature two additional phases α_{21} and α_{31} are required, such that the mixing matrix needs to be multiplied by

$$\mathrm{diag}(1, e^{\frac{i\alpha_{21}}{2}}, e^{\frac{i\alpha_{31}}{2}}). \tag{2.13}$$

2.3.2 Oscillations in a vacuum

As the neutrino mass states are eigenstates of the Hamiltonian with energy eigenvalues E_k:

$$H\left|\nu_k\right\rangle = E_k\left|\nu_k\right\rangle, \tag{2.14}$$

their time evolution is described by the Schrödinger equation. Therefore, the neutrino flavour state evolves with time t, such that

$$\left|\nu_\alpha(t)\right\rangle = \sum_k^3 U^*_{\alpha k} e^{-iE_k t}\left|\nu_k\right\rangle. \tag{2.15}$$

After a time t, the probability of finding $|\nu_\alpha\rangle$ in the state $|\nu_\beta\rangle$ is then given by

$$P(\nu_\alpha \to \nu_\beta, t) = |\left\langle\nu_\beta|\nu_\alpha(t)\right\rangle|^2 = \left|\sum_k^3 U^*_{\alpha k} U_{\beta k} e^{-iE_k t}\right|^2 \tag{2.16}$$

$$= \sum_{k}^{3} \sum_{j}^{3} U_{\alpha k}^{*} U_{\beta k} U_{\alpha j} U_{\beta j}^{*} e^{i(E_k - E_j)t}. \tag{2.17}$$

As neutrinos are ultrarelativistic ($p \simeq E$) the approximation can be made that

$$E_k = \sqrt{\vec{p}_k^{\,2} + m_k^2} \simeq E + \frac{m_k^2}{2E}, \tag{2.18}$$

where $\vec{p}_k$ and m_k are the neutrino momentum and energy, and E is the neutrino energy without the mass. This allows the substitution

$$E_k - E_j \sim \frac{\Delta m_{kj}^2}{2E}, \tag{2.19}$$

where Δm_{kj}^2 is the squared mass difference (*mass splitting*) between the k and j mass states. Additionally, the relativistic limit allows the simplification $t = L$, where L is the distance from neutrino creation to detection. Combined, the oscillation probability can be written as

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta, t) = \delta_{\alpha\beta} &- 4 \sum_{k>j} \mathrm{Re}(U_{\alpha k}^{*} U_{\beta k} U_{\alpha j} U_{\beta j}^{*}) \sin^2\left(\frac{\Delta m_{kj}^2 L}{4E}\right) \\ &\pm 2 \sum_{k>j} \mathrm{Im}(U_{\alpha k}^{*} U_{\beta k} U_{\alpha j} U_{\beta j}^{*}) \sin\left(\frac{\Delta m_{kj}^2 L}{2E}\right), \end{aligned} \tag{2.20}$$

where the last term has a positive sign for neutrinos and a negative sign for anti-neutrinos.

From inspecting equation (2.20), the superposition of mass eigenstates is seen to drive the oscillation of the flavour state, with the amplitude arising from the elements of the PMNS matrix. Furthermore, as a fixed experimental location and neutrino source define a constant L/E, the period of oscillation is determined by the squared mass difference between the flavour states Δm_{kj}^2.

2.3.3 Oscillations in matter

As neutrinos propagate through matter, they undergo coherent forward scattering with nucleons. These interactions do not change the neutrino state or momentum, but they do impart an interaction potential onto the neutrinos. Two types of interaction can take place, either through the exchange of a Z^0 in a *neutral-current* (NC) interaction, or a

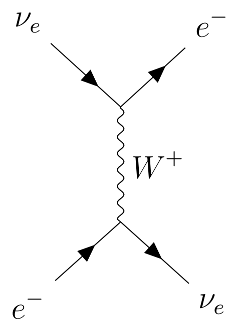


Figure 2.3: Feynman diagram of a charged-current coherent scattering interaction between an electron neutrino and an electron.

$W^{\pm}$ in a *charged-current* (CC) process. The two interaction potentials are given by:

$$V_{NC} = \pm \frac{1}{\sqrt{2}} G_F n_n, \tag{2.21}$$

$$V_{CC} = \pm \sqrt{2} G_F n_e, \tag{2.22}$$

were n_n and n_e are the number density of neutrons and electrons in the propagation medium respectively and G_F is Fermi's constant. The sign of the potential is positive for neutrinos and negative for anti-neutrinos.

Crucially, as matter is full of electrons but empty of muons or taus, only electron neutrinos will interact via the CC channel as shown in Figure 2.3. Consequently, only electron neutrinos are affected by the V_{CC} potential, while all flavours are affected by the V_{NC} potential. When quantitatively applied as an additional potential to the vacuum Hamiltonian, a resonance oscillation term is introduced which significantly modifies the vacuum oscillation probabilities. This phenomenon is known as the Mikheev, Smirnov, and Wolfenstein (MSW) effect [29, 30].

It is important to note two things. Firstly, the $\pm$ in the V_{CC} potential introduces a difference between neutrinos and anti-neutrinos. Secondly, the modified oscillation probability is found to be non-symmetric with respect to the sign of the Δm^2_{kj} parameters.

2.4 Neutrino interactions

As outlined in Section 2.3.3, there are two types of weak neutrino interactions with nucleons: charged-current (CC) and neutral-current (NC), occurring via the exchange of

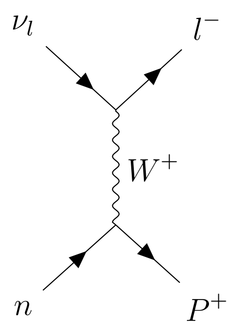


Figure 2.4: Feynman diagram of a charged-current quasi-elastic interaction between a neutrino and a neutron. By identifying the flavour of the final state charged lepton the flavour of the original neutrino can be determined.

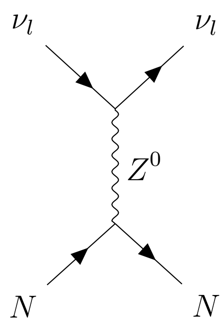


Figure 2.5: Feynman diagram of a neutral-current interaction between a neutrino and a nucleon. The flavour of the neutrino can not be identified as there is no charged lepton in the final state.

a Z^0 or $W^\pm$ respectively. During a CC interaction, the neutrino is transformed into a charged lepton of the same flavour, conserving charge, flavour, and lepton number as shown in Figure 2.4. Contrastingly, during a NC interaction, there is no charge or flavour exchange, and the neutrino continues into the final state as shown in Figure 2.5.

CC interactions provide the signal events for the majority of neutrino experiments as the original neutrino flavour can be determined via identification of the charged lepton. Conversely, NC interactions are commonly background events as there is no way to identify the original neutrino given the absence of a charged lepton in the final state.

For CHIPS the relevant neutrino interaction energy regime is commonly referred to as the *transition region*, covering a range of energies between 1 and 10 GeV. Both CC and NC interactions within this regime fall into one of five main categories:

- **Quasi-Elastic scattering (QEL)** is the dominant channel for energies below 1 GeV and involves the neutrino scattering off the entire nucleon. In the CC neutrino case, this converts the target neutron to a proton (reversed for antineutrinos). The same process can also occur in a NC interaction, where it is referred to as *elastic* scattering.

- **Meson Exchange Current (MEC)** provides an additional contribution to the energy range dominated by QEL interactions, involving two nucleons and producing two protons in the final state. Sometimes referred to as the *2p2h* interaction, it has become essential for explaining the discrepancy between theory and observations within many modern neutrino experiments, and is now included by default in the popular event generators [31].
- **Resonant pion production (Res)** is the dominant channel between 1 and 2 GeV and involves the neutrino exciting the nucleon into a resonant state. The resonance then decays the vast majority of the time to produce a single pion and nucleon. Resonant production is responsible for the majority of NC π^0 events within water Cherenkov detectors such as CHIPS. This channel, as discussed further in Section 3.1.2 forms an incredibly tricky component of the background as it can mimic the event topology of a single electron.
- **Coherent pion production (Coh)** is a much rarer interaction mechanism where the neutrino scatters coherently from the entire nucleus. This process transfers negligible energy to the target and produces a significantly forward scattered single pion with no nuclear recoil in the final state.
- **Deep Inelastic Scattering (DIS)** is dominant for neutrino energies above 3 GeV with the additional energy allowing for the neutrino to resolve the individual quark content of the nucleon. The subsequent scattering from an individual quark produces a hadronic shower in the final state containing multiple pions which are often a challenge to reconstruct.

The standard ν_e cross-sections (very similar to those for ν_μ) included with the GENIE event generator [32, 33] are shown in Figure 2.6 for both CC and NC interactions. For ν_τ interactions the CC cross-sections differ considerably due to the large tau lepton mass, as shown in Figure 2.7, only becoming non-negligible for energies above ~ 5 GeV.

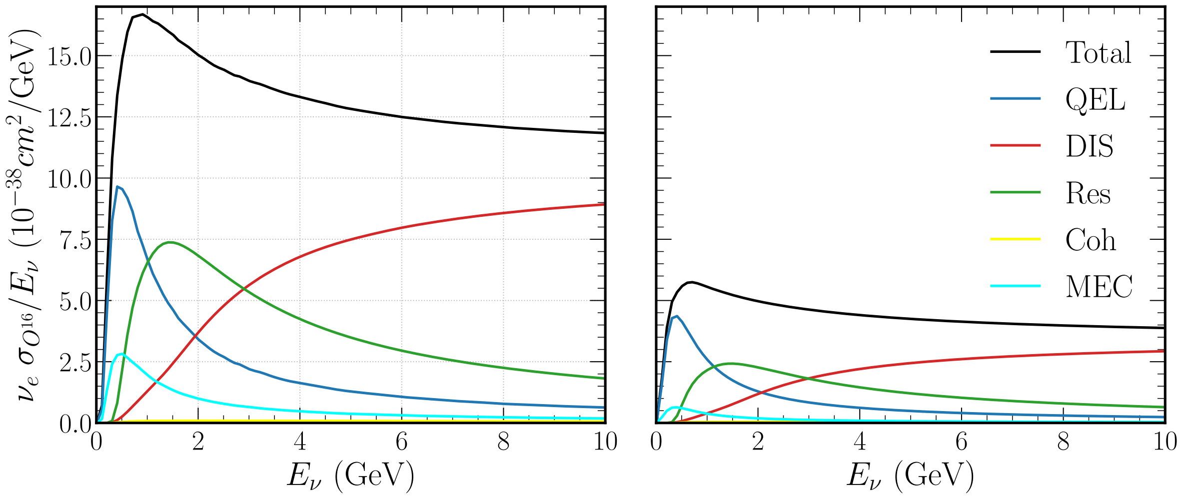


Figure 2.6: GENIE cross-sections on oxygen divided by neutrino energy for CC ν_e (left) and NC ν_e (right). Shown are the total, as well as the Quasi-Elastic (QEL), Deep Inelastic Scattering (DIS), Resonant (Res), Coherent (Coh), and Meson Exchange Current (MEC) cross-sections. Note how the total cross-section approaches a linear dependence on energy and the NC cross-sections are smaller than their CC counterparts. See reference [33] for generation details.

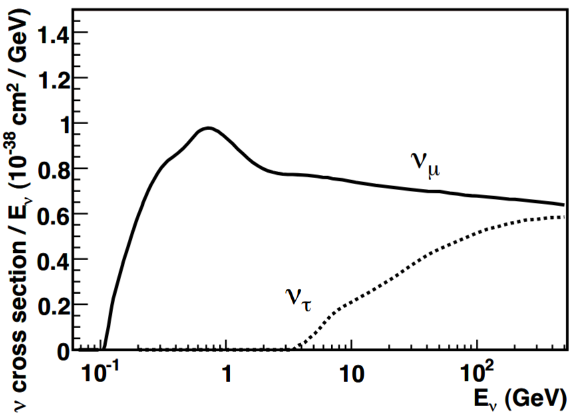


Figure 2.7: The total CC interaction cross-section per nucleon divided by the neutrino energy for ν_μ and ν_τ. Figure taken from reference [34].

2.5 Current status and the future

The three-component representation of the PMNS matrix presented in equation (2.11) splits our understanding of neutrino oscillations into three sectors. Historically, the sectors corresponded to experiments observing different sources of neutrinos, from either the Sun, atmosphere, or nuclear reactors. Hence the standard names given to each: the *solar sector*, *atmospheric sector*, and *reactor sector*. However, it is perhaps more rigorous to think of each sector as corresponding to the parameters they encompass: Δm^2_{21} and θ_{12} for solar, Δm^2_{32} and θ_{23} for atmospheric, and θ_{13} for reactor.

During the last couple of decades, the focus of neutrino experiments across all sectors has shifted from the discovery of neutrino fundamentals to the precise measurement of oscillation parameters. This change has led to a corresponding shift into using increasingly abundant neutrino sources (such as accelerator neutrino beams) and larger and larger detectors. Here we outline the current status, open questions, and future for neutrino physics.

2.5.1 Current status

Global fits to neutrino oscillation data best summarise the current state of neutrino physics. These aggregate the latest experimental results to constrain the neutrino oscillation parameters. The best-fit results from one such fit, NuFIT [35, 36], are shown in Figure 2.8. A detailed overview of other global fits is given in The Particle Data Group's 2020 Review of Particle Physics (see reference [13]). Below, the solar, atmospheric, and reactor best-fit parameters are presented alongside the latest experiments to contribute to their values.

Solar parameters: θ_{12} and Δm^2_{21}

Current best fit results for the solar sector parameters yield $\theta_{12} \sim 34^\circ$ with $\sim 1.5^\circ$ of uncertainty, and $\Delta m^2_{21} \sim 7.4 \times 10^{-5}\mathrm{eV}^2$ with $\sim 0.4 \times 10^{-5}\mathrm{eV}^2$ of uncertainty.

These results primarily come from a range of historical and currently running experiments studying neutrinos generated by the Sun: the radiochemical chlorine-based Homestake [37]; the gallium-based GALLEX [38] and SAGE [39]; the deuterium Cherenkov SNO [40]; the water Cherenkov Super-Kamiokande [41–44]; and the liquid scintillator

NuFIT 5.0 (2020)

		Normal Ordering (best fit)		Inverted Ordering ($\Delta\chi^2 = 2.7$)	
		bfp $\pm 1\sigma$	3σ range	bfp $\pm 1\sigma$	3σ range
without SK atmospheric data	$\sin^2\theta_{12}$	$0.304^{+0.013}_{-0.012}$	$0.269 \to 0.343$	$0.304^{+0.013}_{-0.012}$	$0.269 \to 0.343$
	$\theta_{12}/^\circ$	$33.44^{+0.78}_{-0.75}$	$31.27 \to 35.86$	$33.45^{+0.78}_{-0.75}$	$31.27 \to 35.87$
	$\sin^2\theta_{23}$	$0.570^{+0.018}_{-0.024}$	$0.407 \to 0.618$	$0.575^{+0.017}_{-0.021}$	$0.411 \to 0.621$
	$\theta_{23}/^\circ$	$49.0^{+1.1}_{-1.4}$	$39.6 \to 51.8$	$49.3^{+1.0}_{-1.2}$	$39.9 \to 52.0$
	$\sin^2\theta_{13}$	$0.02221^{+0.00068}_{-0.00062}$	$0.02034 \to 0.02430$	$0.02240^{+0.00062}_{-0.00062}$	$0.02053 \to 0.02436$
	$\theta_{13}/^\circ$	$8.57^{+0.13}_{-0.12}$	$8.20 \to 8.97$	$8.61^{+0.12}_{-0.12}$	$8.24 \to 8.98$
	$\delta_{\mathrm{CP}}/^\circ$	195^{+51}_{-25}	$107 \to 403$	286^{+27}_{-32}	$192 \to 360$
	$\frac{\Delta m^2_{21}}{10^{-5}\ \mathrm{eV}^2}$	$7.42^{+0.21}_{-0.20}$	$6.82 \to 8.04$	$7.42^{+0.21}_{-0.20}$	$6.82 \to 8.04$
	$\frac{\Delta m^2_{3\ell}}{10^{-3}\ \mathrm{eV}^2}$	$+2.514^{+0.028}_{-0.027}$	$+2.431 \to +2.598$	$-2.497^{+0.028}_{-0.028}$	$-2.583 \to -2.412$
		Normal Ordering (best fit)		Inverted Ordering ($\Delta\chi^2 = 7.1$)	
		bfp $\pm 1\sigma$	3σ range	bfp $\pm 1\sigma$	3σ range
with SK atmospheric data	$\sin^2\theta_{12}$	$0.304^{+0.012}_{-0.012}$	$0.269 \to 0.343$	$0.304^{+0.013}_{-0.012}$	$0.269 \to 0.343$
	$\theta_{12}/^\circ$	$33.44^{+0.77}_{-0.74}$	$31.27 \to 35.86$	$33.45^{+0.78}_{-0.75}$	$31.27 \to 35.87$
	$\sin^2\theta_{23}$	$0.573^{+0.016}_{-0.020}$	$0.415 \to 0.616$	$0.575^{+0.016}_{-0.019}$	$0.419 \to 0.617$
	$\theta_{23}/^\circ$	$49.2^{+0.9}_{-1.2}$	$40.1 \to 51.7$	$49.3^{+0.9}_{-1.1}$	$40.3 \to 51.8$
	$\sin^2\theta_{13}$	$0.02219^{+0.00062}_{-0.00063}$	$0.02032 \to 0.02410$	$0.02238^{+0.00063}_{-0.00062}$	$0.02052 \to 0.02428$
	$\theta_{13}/^\circ$	$8.57^{+0.12}_{-0.12}$	$8.20 \to 8.93$	$8.60^{+0.12}_{-0.12}$	$8.24 \to 8.96$
	$\delta_{\mathrm{CP}}/^\circ$	197^{+27}_{-24}	$120 \to 369$	282^{+26}_{-30}	$193 \to 352$
	$\frac{\Delta m^2_{21}}{10^{-5}\ \mathrm{eV}^2}$	$7.42^{+0.21}_{-0.20}$	$6.82 \to 8.04$	$7.42^{+0.21}_{-0.20}$	$6.82 \to 8.04$
	$\frac{\Delta m^2_{3\ell}}{10^{-3}\ \mathrm{eV}^2}$	$+2.517^{+0.026}_{-0.028}$	$+2.435 \to +2.598$	$-2.498^{+0.028}_{-0.028}$	$-2.581 \to -2.414$

Figure 2.8: Three-flavour results from the NuFIT v5.0 global fit as of July 2020, showing both 1σ uncertainties and $\pm 3\sigma$ ranges. The first column shows results for an assumed normal hierarchy while the second for an inverted hierarchy, both of which are introduced in Section 2.5.2. The lower section includes atmospheric neutrino data from the Super-Kamiokande experiment, while the upper section does not. Figure taken from reference [36].

detector Borexino [45–47]. Most of which were mentioned previously in Section 2.1.2, originally involved in studying the solar neutrino problem.

Additionally, by locating a 1 kt liquid scintillator detector between several nuclear reactors, the KamLAND experiment has probed the solar sector in a terrestrial setting using reactor antineutrinos [48]. Unlike solar neutrinos, which are heavily influenced by the high electron density of the Sun and the subsequent MSW effect, the antineutrinos detected by KamLAND were not influenced, removing any measurement ambiguity.

Atmospheric parameters: θ_{23} and Δm^2_{32}

Current best fit results for the atmospheric sector parameters yield $\theta_{23} \sim 48^\circ$ with $\sim 2^\circ$ of uncertainty, and $\Delta m^2_{32} \sim 2.45 \times 10^{-3}\text{eV}^2$ (assuming normal hierarchy as discussed in Section 2.5.2) with $\sim 0.03 \times 10^{-3}\text{eV}^2$ of uncertainty. These results come from one of two primary sources.

Firstly, experiments studying atmospheric neutrinos generated in the upper atmosphere by cosmic rays, such as IceCube [49, 50] and the previously mentioned Super-Kamiokande [51]. IceCube is a neutrino observatory based at the Amundsen-Scott South Pole Station in Antarctica, consisting of strings of PMTs embedded deep within the ice. Whilst, Super-Kamiokande is a large 50 kt water Cherenkov detector, equipped with 11000 PMTs, and located 1 km underground in Gifu Prefecture, Japan.

Furthermore, experiments measuring $P(\nu_\mu \to \nu_\mu)$ using accelerator beam generated muon neutrinos over a long-baseline of many hundreds of kilometres are sensitive to the atmospheric parameters. Such experiments include: MINOS [52, 53] (now complete) and NOvA [54, 55] using a beam generated at Fermilab with detectors in northern Minnesota, and T2K [56] using a beam generated at J-PARC (on the east coast of Japan) and using the Super-Kamiokande detector described above.

Reactor parameter: θ_{13}

The most recent parameter to be measured and found to be non-zero is θ_{13}, with current best fit results yielding a value of $\theta_{13} \sim 8.5^\circ$ with $\sim 0.25^\circ$ of uncertainty (note that this is now the most tightly constrained of all the PMNS mixing angles).

Primarily, θ_{13} measurements come from reactor experiments measuring electron antineutrino disappearance, $P(\bar{\nu}_e \to \bar{\nu}_e)$. The Daya Bay [57, 58], RENO [59, 60], and

Double Chooz [61] experiments located in China, South Korea, and France respectively all published the strongest results for this measurement at a similar time.

Long-baseline accelerator experiments are also sensitive to θ_{13} through the electron neutrino appearance channel $P(\nu_\mu \to \nu_e)$. Measurements have been made by MINOS [53], T2K [62], and NOvA [63] but with much greater uncertainties than reactor experiments.

2.5.2 Open questions

There are still many open questions (mainly ambiguities) regarding the parameters of the PMNS matrix. CHIPS and other long-baseline accelerator experiments can help resolve those discussed below by studying the appearance of electron neutrinos in a muon neutrino beam. The approximate oscillation probability for this transition, including the MSW effect, takes the form [13]:

$$\begin{aligned} P(\nu_\mu \to \nu_e, \bar{\nu}_\mu \to \bar{\nu}_e) \sim\ & 4\sin^2\theta_{13}\sin^2\theta_{23}\frac{\sin^2\Delta}{(1-A)^2} \\ & + \alpha^2\sin^2 2\theta_{12}\cos^2\theta_{23}\frac{sin^2 A\Delta}{A^2} \\ & + 8\alpha J\cos(\Delta \pm \delta_{CP})\frac{\sin\Delta A}{A}\frac{\sin\Delta(1-A)}{(1-A)}, \end{aligned} \tag{2.23}$$

with

$$J = \cos\theta_{12}\sin\theta_{12}\cos\theta_{23}\sin\theta_{23}\cos^2\theta_{13}\sin\theta_{13}, \tag{2.24}$$

$$\Delta = \frac{\Delta m_{31}^2 L}{4E_\nu}, \tag{2.25}$$

$$A = \pm\frac{2E_\nu V}{\Delta m_{31}^2}, \tag{2.26}$$

$$\alpha = \frac{\Delta m_{21}^2}{\Delta m_{31}^2}, \tag{2.27}$$

where V is the effective matter potential of the Earth's crust, and $\pm$ is positive for neutrinos and negative for antineutrinos.

Octant of θ_{23}

The mixing angle θ_{23} is known to be non-zero and large. Initially thought to be maximal such that $\theta_{23} = \pi/4 = 45^\circ$, there are now indications that this may not be the case [54]. Naturally, this possibility raises the question of whether $\theta_{23} < 45^\circ$ or $\theta_{23} > 45^\circ$, commonly referred to as the octant. Although θ_{23} is easily measured be studying ν_μ disappearance, a degeneracy in the oscillation probability removes the ability to determine the octant. However, ν_e appearance via the first term in equation (2.23) allows for the octant to be resolved. Current results suggest the $\theta_{23} > 45^\circ$ octant is preferred; however, a significant measurement is yet to be made.

Mass hierarchy

The value of Δm^2_{21} is known to be small and positive, meaning that $m_2 > m_1$ where m_1 is defined to be the dominant mass state of the electron neutrino [40]. However, the sign of Δm^2_{32} is still currently unknown. This allows for two possible scenarios, either $m_1 < m_2 < m_3$, known as the *normal hierarchy*, or $m_3 < m_1 < m_2$, known as the *inverted hierarchy*, as illustrated in Figure 2.9. Resolving this ambiguity will significantly improve the ability of experiments to both determine the value of δ_{CP} (discussed next) and discover if neutrinos are Dirac or Majorana in nature.

Oscillations in vacuum are not sensitive to the sign of the Δm^2_{kj} parameters. However, matter effects allow for the sign to be determined. Long-baseline ν_e appearance is sensitive to the sign of Δm^2_{32} via the Δ, A, and α parameters in equation (2.23). Alternatively, determining the hierarchy is also possible using future reactor experiments such as JUNO [64].

CP violation

The current level of CP violation observed in the quark sector does not fully explain the matter-antimatter asymmetry of the universe. However, if CP violation is shown to exist in the leptonic sector, leptogenesis models of the early universe allow this to go some way to resolving the issue. Neutrino CP violation, when $P(\nu_\alpha \to \nu_\beta) \neq P(\bar{\nu}_\alpha \to \bar{\nu}_\beta)$, is possible in the PMNS matrix when δ_{CP} is not 0 or π. ν_e appearance, sensitive to δ_{CP} via the third term in equation (2.23) is the most promising channel for measurement. The oscillation probabilities are seen to change significantly with δ_{CP}, as shown in Figure 2.10.

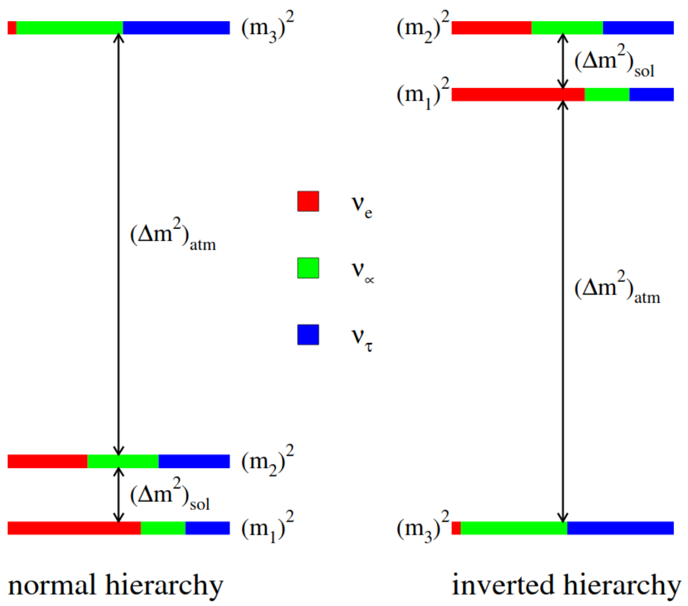


Figure 2.9: Diagram of the two possible neutrino mass hierarchies. The atmospheric (atm) and solar (sol) naming convention is used for the mass differences. Δm^2_{21} (sol) is small and positive, while Δm^2_{32} (atm) is large and its sign is unknown. Figure taken from reference [65].

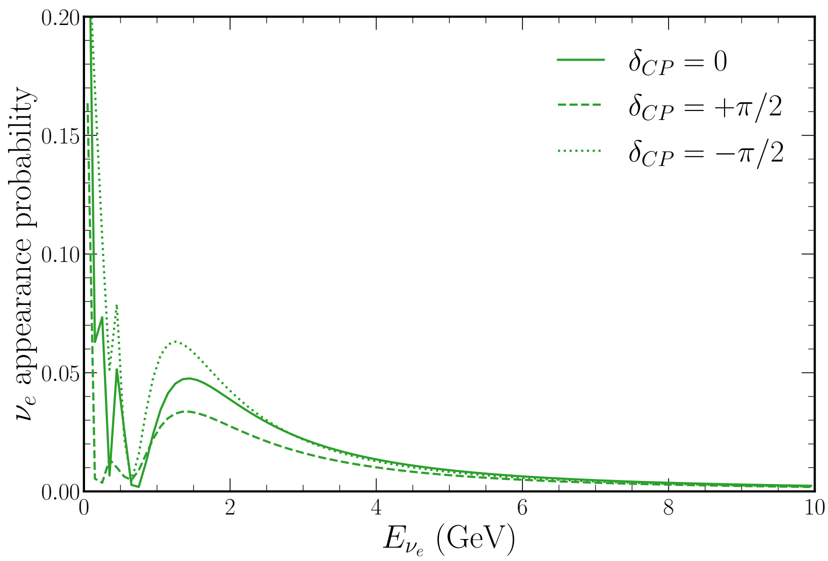


Figure 2.10: Appearance probability for ν_e oscillating from ν_μ at a baseline of 712 km. The oscillations are show for three different values of δ_{CP}.

Current results from both NOvA [54] and T2K [66] hint at a possible non-zero value; however, a significant measurement is yet to be made.

2.5.3 The future

The current generation of long-baseline accelerator experiments (NOvA and T2K) are not expected to unambiguously determine the octant of θ_{23}, the mass hierarchy, or the value of δ_{CP}. To finally answer these questions, primarily two next-generation experiments are planned. The USA based Deep Underground Neutrino Experiment (DUNE) and Hyper-Kamiokande in Japan. Both will have rich neutrino physics programs alongside their primary goals, studying atmospheric and supernova neutrinos, as well as nucleon decay.

DUNE [67, 68], will consist of four 10 kt liquid argon time projection chambers (LArTPCs) located 1.5 km underground at the Sanford Underground Research Facility, South Dakota. This location is the same as the Homestake mine chlorine experiment previously mentioned in Section 2.1.2. LArTPCs have been chosen as they provide excellent track reconstruction and particle identification performance.

The DUNE experiment will also use a new high intensity beam constructed at Fermilab (1285 km away) named the Long-Baseline Neutrino Facility (LBNF). Replacing the current 700 kW NuMI (Neutrinos at the Main Injector) beam, LBNF is initially planned to run at 1.2 MW, but will eventually reach 2.4 MW.

The initial cost estimate for DUNE combined with LBNF was forecast at between $1.26 billion and $1.86 billion [69]. However, due to rising underground excavation costs, this has recently risen to approximately $2.6 billion [70], an extremely high cost for a neutrino experiment.

Hyper-Kamiokande [71, 72], will be the next-generation replacement for the currently running Super-Kamiokande detector. When completed, it will consist of a cylindrical water Cherenkov detector, 60 m in height and 74 m in diameter, built under 650 m of rock at the Tochibora mine in Japan, close to the current Super-Kamiokande location. New high-efficiency photomultiplier tubes will be used, covering 40% of the detector walls to record the Cherenkov radiation emitted from neutrino events.

Moreover, there will be upgrades to the existing J-PARC neutrino beam currently used by T2K. Primarily this will increase the beam power from the current record of 515 kW to 1.3 MW. In total, the Hyper-Kamiokande project is expected to cost nearly $700 million [73].

Chapter 3

The CHIPS R&D project

In pursuit of answers to the open questions presented in the previous chapter, neutrino experiments are becoming increasingly, and possibly prohibitively, expensive and impractical. This trend is particularly true of the next generation of long-baseline experiments, DUNE and Hyper-Kamiokande, with cost estimates reaching billions of dollars and construction times of greater than half a decade. It is also telling that the vast majority of global research effort goes into just these two future projects, such is their complexity, cost, and lead time.

It is evident that for detectors to remain practical and affordable into the future, a novel design strategy is highly desirable. This approach is especially the case if megaton scale detectors are ever to become a reality. While instrumentation will continue to improve with time steadily, the statistics of low event counts will always limit neutrino experiments until vastly larger detectors can be built. Therefore, R&D efforts must focus on such detectors now, whilst also attempting to complement the current and upcoming generation of experiments.

The CHIPS R&D project aims to develop novel strategies and technologies for very large yet practical 'cheap as chips' water Cherenkov detectors [74]. Primarily aimed for deployment in long-baseline accelerator beam scenarios, CHIPS aims to lower the cost per kt of sensitive mass to between \$200k-\$300k. For comparison, the Super-Kamiokande detector cost approximately \$4 million per kt to build. As physics sensitivity depends on more than just sensitive mass, this comparison is not entirely rigorous; however, it highlights the scale of possible cost savings.

This chapter aims to describe the fundamental aspects of the CHIPS R&D project in detail and explain how it achieves its goals. Firstly, the CHIPS concept will be outlined along with both neutrino beam and Cherenkov detector physics for context. The design, construction, deployment, and status of the Chips-5 prototype detector will then follow. Finally, a description of the Monte Carlo methods used to simulate Chips detectors is given, alongside how the input neutrino events are generated.

3.1 The CHIPS concept

The Chips concept is to deploy cylindrical water Cherenkov detector modules into deep bodies of water on the Earth's surface such as lakes, reservoirs, and flooded mine pits. Initially constructed on land, Chips detectors can be floated into position before being sunk. The water above the sunken detector provides a modest overburden from cosmic rays, whilst the surrounding water provides support for a lightweight structure. By removing the need for underground excavation and expensive structural support, the cost of construction can be dramatically reduced.

Additionally, the common practice of building majority bespoke components is replaced by using modern commercially available components wherever possible. The number of expensive elements, such as photomultiplier tubes are also reduced by only considering multi-GeVaccelerator beam neutrino events, such that full high-density detector instrumentation is not required.

Furthermore, Chips detectors are not only designed to be cheap, but practical. Easy to build, quick to deploy, and upgradable once operational, multiple detector modules can be flexibly combined depending on available resources and funding. When compared to DUNE and Hyper-Kamiokande both which require a large upfront budget and many years to construct, cheap Chips detector modules can be deployed as needed in under a year by a relatively small team.

To date, Chips R&D efforts have been based in the USA to exploit the NuMI beam before the end of its lifetime. Plans are focused on the scaling of Chips detectors for the deployment of multiple modules within the LBNF beam once operational. Collaborators from primarily University College London, The University of Wisconsin Madison, and Nikhef are focused on multiple R&D efforts, each aiming to prove the viability of a crucial component of the Chips concept:

- **Detector construction:** Aiming to prove that the construction and deployment of Chips concept detector modules are possible. Two prototype detectors have so far been deployed. Firstly, the small CHIPS-M module shown in Figure 3.1, deployed into a flooded mine pit in northern Minnesota during the summer of 2014 [75–77]. Secondly, the much larger 5 kt CHIPS-5 module, deployed into the same pit during the summer of 2019 and detailed in Section 3.2.

- **Water filtration:** Aiming to prove that adequate water purity can be achieved using cheap, commercially available filtration. Extensive studies have proven that by filtering water directly from bodies of water on the Earth's surface (including flooded mine pits), adequate photon attenuation lengths of greater than 100 m are achievable [78, 79].
- **Physics sensitivity:** Aiming to prove that CHIPS concept detector modules (even the prototypes) can provide significant physics contributions alone or alongside the current and next generation of experiments. Single modules in the current NuMI beam (discussed in Section 3.1.1) and multiple modules in the future LBNF beam have been studied [77, 80, 81].
- **Data acquisition:** Aiming to prove that a cheap data acquisition (DAQ) system using commercially available components and software is viable [82]. Outlined in Chapter 4, CHIPS implements a novel use of cheap single-board computers to collect photomultiplier tube data.
- **Event reconstruction and classification:** Aiming to prove that deep learning techniques can be successfully applied to the reconstruction and classification of neutrino events from large water Cherenkov concepts such as CHIPS. The primary contribution of this thesis (detailed in Chapter 5 and Chapter 6), this work feeds directly into both the physics sensitivity studies mentioned above and detector design optimisation.

3.1.1 The neutrino beam

CHIPS detectors will primarily study the appearance of ν_e oscillating from ν_μ over a long-baseline. To generate a sufficient number of GeV scale ν_μ, a high-intensity accelerator beam is required. Currently, only two such beams exist, the J-PARC based beam in Japan used by the T2K experiment and the NuMI beam in the USA used by NOvA. Here we describe the NuMI beam [83] as it is directly relevant to current CHIPS efforts. However, it is essential to note that CHIPS detectors are designed to be deployed into any high-intensity neutrino beam, including the future NuMI replacement, LBNF.

Figure 3.1: Picture of the 3.3 m high CHIPS-M detector just before deployment. Temporary floatation bags are attached to the top rim of the detector, while the umbilical cord carrying data, power, and filtered water is attached to the base.

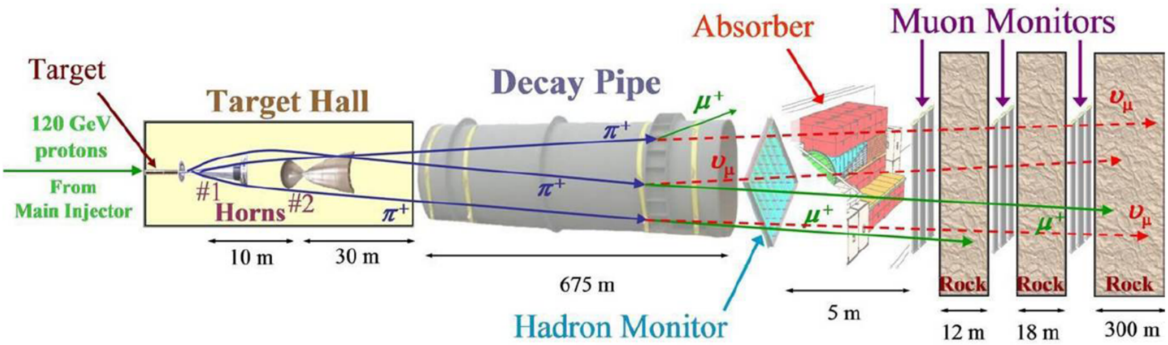


Figure 3.2: Schematic of the main components of the NuMI beamline. The MINOS and NOvA near detectors and the MINERvA experiment are located just to the right of what is shown. Figure taken from reference [83].

The NuMI beam is an accelerator muon neutrino beam produced at Fermilab near Chicago in the USA. Beginning operation in 2005 for the MINOS experiment, NuMI was upgraded in 2013 to provide a higher intensity and energy beam, principally to achieve a peak in neutrino energy near the ~1.5 GeV $\nu_\mu \rightarrow \nu_e$ oscillation maximum for NOvA. Currently, the NuMI beam achieves an intensity of 700 kW (740 kW at peak) making it the most powerful such beam in the world. A schematic of the NuMI beamline configuration is shown in Figure 3.2.

Every 1.33 seconds a 10 ţs long spill of protons accelerated to 120 GeV by the Main Injector ring is directed towards a stationary graphite target. The resulting interactions create a shower of hadrons containing predominantly pions and kaons. The hadrons are passed through a focusing system of two magnetic horns tuned principally to focus positively charged pions along the beamline while rejecting other particles. After focusing, any surviving hadrons are allowed to decay in flight to a beam of muon neutrinos in a 675 m long decay pipe via the processes:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \tag{3.1}$$

$$K^+ \rightarrow \mu^+ + \nu_\mu. \tag{3.2}$$

The resulting muons also decay such that $\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$, producing an intrinsic ν_e component as well as wrong sign ν_μ contamination.

Alternatively, the polarity of the horns can be used to switch the dominant sign of the hadrons focused, allowing NuMI to operate as either a neutrino or antineutrino beam. These two modes of operation are called *forward horn current* and *reverse horn current*, for primarily a neutrino or antineutrino beam composition respectively. Any remaining hadrons, alongside electrons, muons, and surviving primary protons are absorbed by rock downstream of the decay pipe, leaving just the neutrino components of the beam.

Long-baseline neutrino experiments typically consist of a *near* detector to measure the neutrino composition at source and a much larger *far* detector to measure the oscillated composition after many hundreds of kilometres. The NuMI beamline contains three detectors after 300 m of rock: The MINERvA spectrometer [84], the near detector for MINOS (now used by MINERvA), and the near detector for NOvA. CHIPS prototypes within the NuMI beam (such as CHIPS-5) will not have a dedicated near detector; therefore, data from the above detectors will be crucial for physics analysis in order to constrain the beam composition and flux.

The NuMI neutrino beam passes through the Earth's crust until it finally emerges in northern Minnesota. This is where the MINOS, NOvA, and prototype CHIPS far detectors are located (used to be located in the MINOS case), as shown in Figure 3.3. The Minnesota state nickname 'land of 10,000 lakes' is not an overstatement, with a vast number of potential lakes for CHIPS detector deployment. Additionally, intense iron ore mining on the 'Iron Range' provides many suitable, disused (and now flooded) mine pits. The exact CHIPS-5 prototype detector location is discussed in greater detail within Section 3.2.1.

Due to the kinematics of pion decay, whether the far detector is placed directly on the beam axis or not can have a significant impact on the observed energy spectrum of

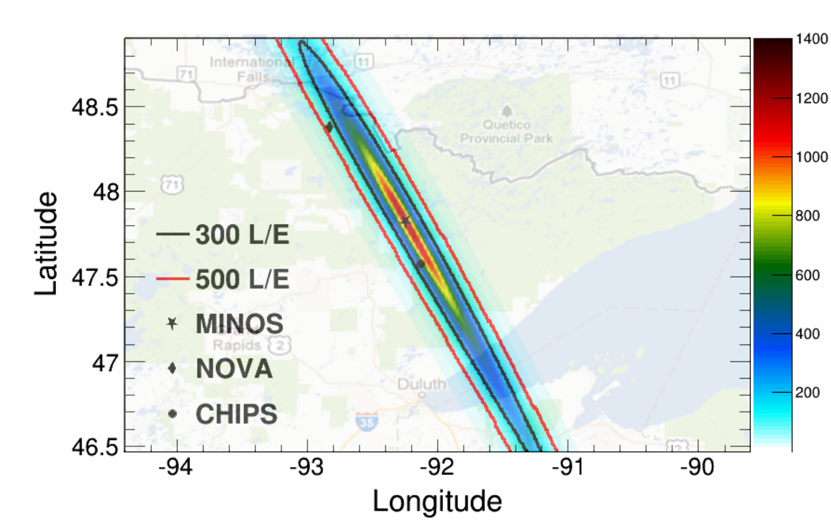


Figure 3.3: Map of the MINOS, NOvA and CHIPS locations in the NuMI beam as it surfaces in northern Minnesota. Shown (the z-axis) is the expected neutrino event rate per kt per year assuming no oscillations, with lines of constant L/E indicated with contours. The western extent of Lake Superior can be seen in the lower right of the map for reference. Image taken from reference [74].

beam neutrinos, as shown in Figure 3.4. For neutrinos directly on the beam axis, there is a strong dependence on the energy of the parent pion within equation (3.1). However, as the off-axis angle increases, the neutrino energy becomes less and less dependent on the parent pion energy and is restricted to a narrowing range of decreasing energies.

Known as the *off-axis effect* this phenomenon is used by both NOvA and T2K to create a narrow energy peak focused on the important $\nu_\mu \to \nu_e$ oscillation maximum. By reducing the tail of high energy neutrinos, the number of background NC events can also be greatly reduced, as is the case for the 7 mrad off-axis CHIPS-5 detector module.

3.1.2 Water Cherenkov detectors

The CHIPS detector concept is based upon the water Cherenkov technique for neutrino detection. A large body of target water is instrumented with photomultiplier tubes (PMTs) to record the Cherenkov radiation produced by sufficiently relativistic charged particles produced in neutrino interactions. By using readily available water as the target material and only instrumenting the surface of the volume, water Cherenkov detectors provide the best detection methodology for maximising volume and reducing cost.

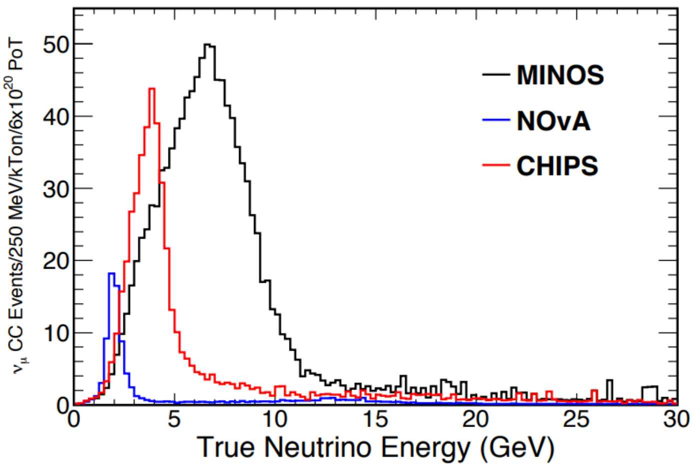


Figure 3.4: Muon neutrino flux for different NuMI detectors at different off-axis angles. Shown are the neutrino energy spectrums for MINOS (on-axis), CHIPS-5 (7 mrad off-axis), and NOvA (14 mrad off-axis). Figure taken from reference [74].

Cherenkov radiation is emitted by all electrically charged particles travelling faster than the local phase velocity of light in a dielectric medium. Similar to the sonic boom created by a supersonic aircraft, Cherenkov radiation forms a shock wave of coherent light, as shown in Figure 3.5. Typically, the emitted light has wavelengths in the optical to ultraviolet range (~ 400 nm). When projected onto the detector wall, the resulting cone of radiation generates a distinctive ring shape. The cone opening angle (the angle at which light is emitted) θ_c, is given by

$$\cos\theta_c = \frac{1}{\beta n(\lambda)}, \tag{3.3}$$

where $\beta = v/c$ and n is the refractive index of the medium [13]. Note that n is a function of the wavelength of emission λ, and so is the opening angle. As the refractive index of water is ~ 1.33 for typical wavelengths of emission, and using the ultrarelativistic limit $\beta \sim 1$, the opening angle is found to be $\sim 41°$.

In equation (3.3) there is no Cherenkov emission when $\cos\theta_c > 1$, which is the case when $\beta n(\lambda) < 1$. Therefore, a Cherenkov energy threshold E_t, exists for charged particles. When expressed in terms of the particle mass m, the threshold is given by

$$E_t = \gamma m = \frac{m}{\sqrt{1-(1/n)^2}}. \tag{3.4}$$

Again using $n \sim 1.33$, a Cherenkov emission threshold energy of $\sim 1.5\ m$ is typical.

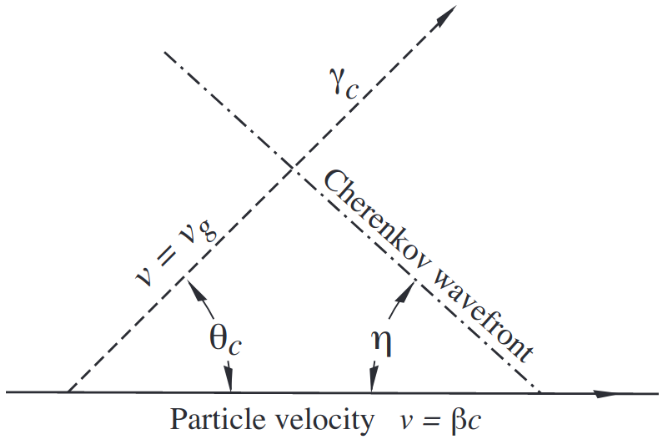


Figure 3.5: Diagram of Cherenkov radiation emission and wavefront angles. In a dispersive medium photons propagate at the group velocity v_g and the Cherenkov emission opening angle η is not equal to θ_c. Figure taken from reference [13]

The number of Cherenkov photons emitted by a particle of charge ze, per unit wavelength per unit path length is given by

$$\frac{d^2N}{d\lambda dx} = \frac{2\pi\alpha z^2}{\lambda^2}\left(1 - \frac{1}{1-\beta^2 n^2(\lambda)}\right), \tag{3.5}$$

where α is the fine structure constant ($\sim 1/137$) [13]. Integrating over the range of wavelengths for which PMTs are typically sensitive, 350 nm to 650 nm, and using $\beta \sim 1$ and $n \sim 1.33$ gives approximately 240 photons emitted per cm travelled by the charged particle [85].

By analysing the Cherenkov rings of light recorded by the PMTs on the walls of the detector, information about the charged particles within an event can be determined. The underlying neutrino interaction can then be understood indirectly. Primarily, the challenge for accelerator beam water Cherenkov detectors is the identification and reconstruction of an electron or muon ring likely to have been produced from a beam neutrino and not a cosmic ray. This event topology indicates a beam CC ν_e or CC ν_μ event respectively, rejecting NC and cosmic events.

The basic shape of a Cherenkov ring can be used to tell which charged particle created it. Muons are long-lived and typically travel many metres within the detector, producing a clean ring with sharp edges as shown in Figure 3.6. Conversely, electrons almost immediately initiate an electromagnetic shower; therefore, the observed ring is the sum of multiple rings produced from the individual electrons and positrons within the shower.

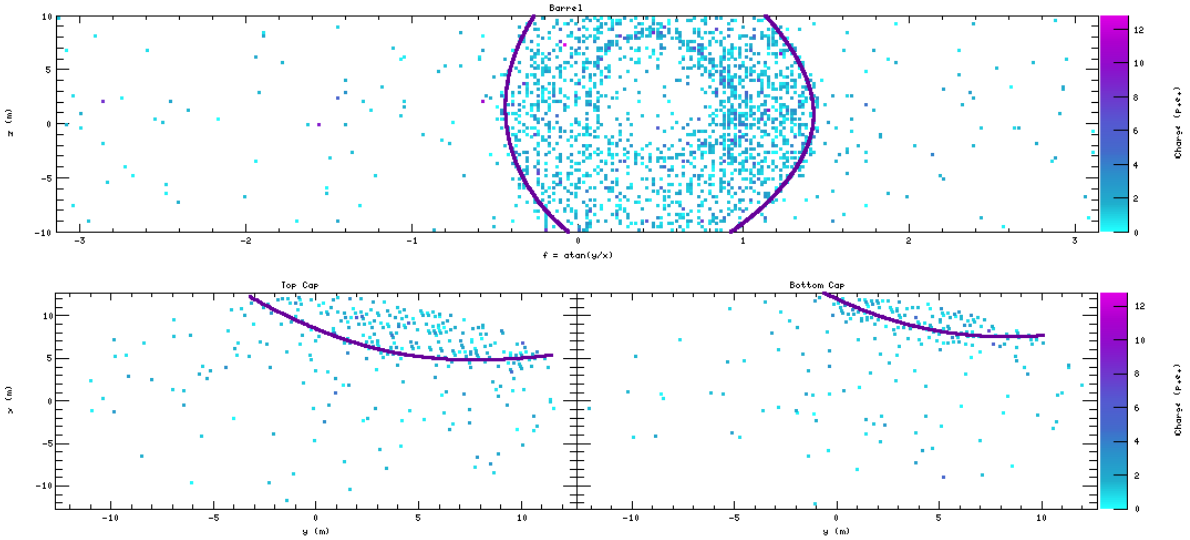


Figure 3.6: Event display of a simulated beam CC ν_μ quasi-elastic event with a single muon in the final state of energy 2.36 GeV. Both the unrolled barrel and endcaps are shown with every entry representing a hit PMT with the colour indicating the collected charge. The purple ring indicates the true muon Cherenkov cone boundary projected onto the detector walls.

As a consequence, when compared to a muon ring, electron rings are characteristically fuzzy, as shown in Figure 3.7. Additionally, a factor of this difference can be seen in Figure 3.8, which shows how the total amount of Cherenkov radiation is emitted for both electrons and muons as a function of distance from the interaction vertex.

The situation quickly becomes complicated when multiple charged particles above the Cherenkov threshold are involved, common at multi-GeV energies. In this case, multiple overlapping rings are observed, making reconstruction difficult. The worst-case scenario is when two rings overlap entirely, removing any ability to tell them apart. This topology is often the case for NC interactions producing a π^0 in the final state, forming the primary background for the signal CC ν_e appearance channel.

π^0 particles decay to a pair of photons with a 98.82% branching ratio, both of which almost immediately initiate an electromagnetic shower, just like an electron [13]. This process leads to the formation of two electron like rings separated by an angle θ_{ij}, given by

$$(1 - \cos\theta_{ij}) = \frac{m_\pi^2}{2E_i E_k}, \tag{3.6}$$

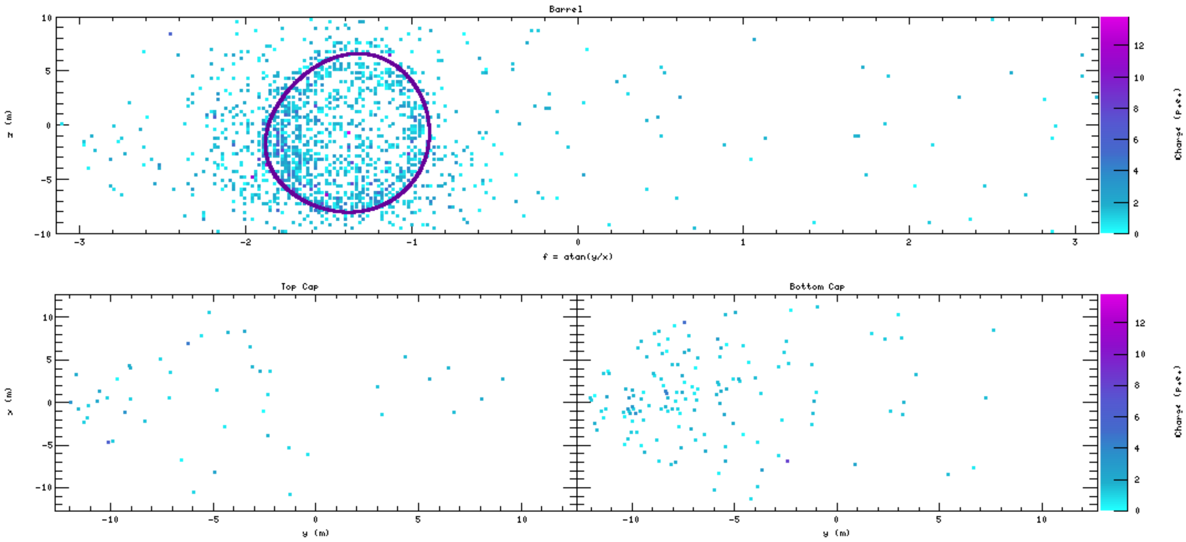


Figure 3.7: Event display of a simulated beam CC ν_e quasi-elastic event with a single electron in the final state of energy 1.05 GeV. Both the unrolled barrel and endcaps are shown with every entry representing a hit PMT with the colour indicating the collected charge. The purple ring indicates the true electron Cherenkov cone boundary projected onto the detector walls.

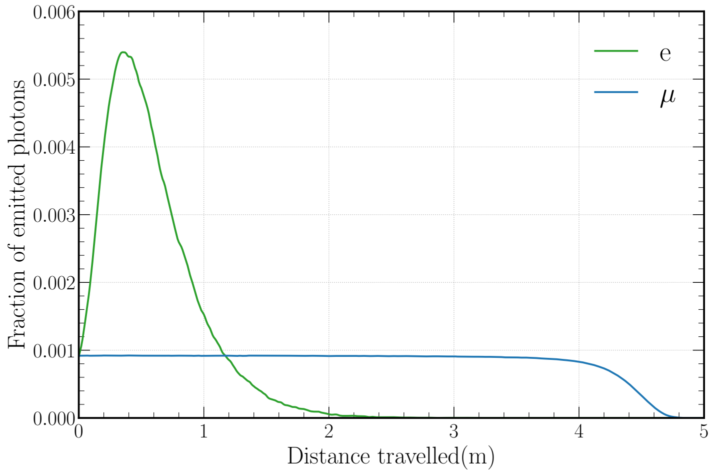


Figure 3.8: Fraction of the total number of Cherenkov photons emitted as a function of the distance from the interaction vertex for both electrons and muons with an initial energy of 2.5 GeV. Multiple particles within the electron-induced electromagnetic shower emit their Cherenkov radiation over a short distance and in slightly different directions (fuzzy). Conversely, a muon travels relatively much further emitting an approximately constant level of Cherenkov radiation as it does so (clean).

where m_π is the invariant mass of the π^0 and E_i and E_j are the energies of the two photons in the lab frame, respectively. Therefore, for a π^0 decaying to two 1 GeV photons, there is just 8° of separation between the rings, making them difficult to tell apart. This distinction is especially hard when electron like rings are also fuzzy. Alternatively, if the two photons have an unequal energy distribution, such that one is much more energetic than the other, the higher energy photon ring can dominate, and the other can not be identified, leading to what looks like a single electron ring, producing a misidentification.

3.2 CHIPS-5

CHIPS-5 is the first large scale prototype detector module for the CHIPS project. Cylindrical in shape with dimensions of 25 m in diameter and 12 m in height, CHIPS-5 has an inner surface area of 1924 m^2 and a total target mass of 5.9 kt. Via the process of design, construction, deployment, and data taking, CHIPS-5 primarily aims to refine the CHIPS concept for future full-scale (∼15 kt) modules. Consequently, CHIPS-5 is designed such that the details outlined in this section are entirely characteristic of what a full-sized CHIPS module is envisioned to be.

First, the location, structure, instrumentation, and water filtration are detailed for the complete detector. A discussion of the construction and deployment procedure follows before the current status is presented. The CHIPS-5 DAQ implementation is detailed separately in Chapter 4.

For clarity, neutrinos from the NuMI beam are assumed to enter the CHIPS-5 detector through the *upstream* wall of the cylindrical barrel, parallel to the floor and travelling towards the opposite *downstream* wall. Their direction of travel (the direction of the beam) is referred to as being along the x-axis for the rest of this work. The corresponding y-axis is the perpendicular horizontal axis, and the z-axis is vertical through the barrel, with the coordinate origin located at the centre of the detector.

3.2.1 Location

CHIPS-5 is located at the Wentworth 2W pit in northern Minnesota, USA, near the small town of Hoyt Lakes. A disused and flooded surface Taconite ore (a type of iron ore) mine pit, Wentworth 2W is located 7 mrad off the NuMI axis at a distance of 712 km from the

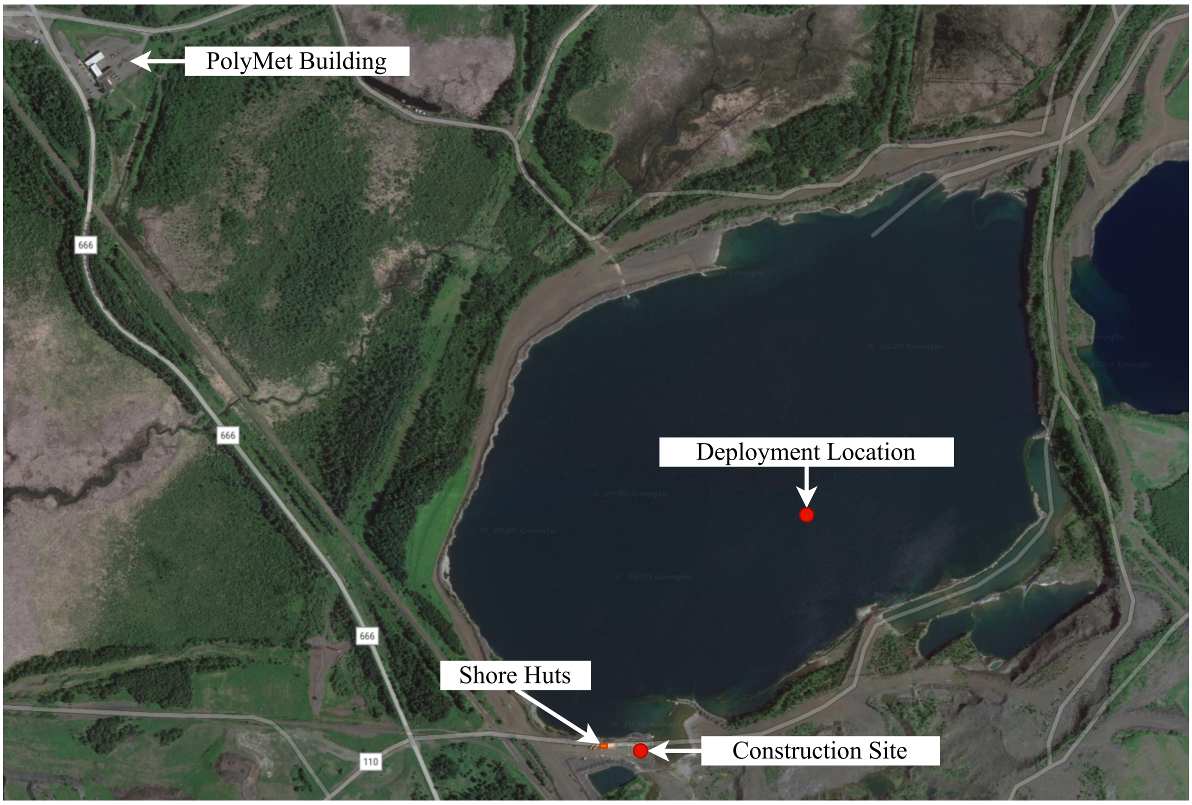


Figure 3.9: Satellite view of the Wentworth 2W flooded mine pit in northern Minnesota, showing key CHIPS-5 locations. The PolyMet building, shore huts, construction site, and deployment location are shown. For both the construction site and deployment location, the red circle shows the true CHIPS-5 detector size to scale.

beam target. Roughly 0.8 km by 1.2 km in size with a maximum depth of 60 m ($\pm$3 m throughout the year), the pit allows for a water overburden of approximately 50 m above CHIPS-5 when resting on the bottom. With an average daily low temperate of -24°C in January, the pit freezes over during winter, therefore, work is only possible during the summer months of May to October.

A sizeable earthen ramp on the south side of Wentworth 2W is used for detector construction. The construction site is easily accessible by road and well connected to power due to the heavy infrastructure in place for mining. Additionally, the nearby PolyMet mining administration building is used as a laboratory environment for the construction and testing of individual components before their installation within the detector. A labelled satellite view of Wentworth 2W is given in Figure 3.9 for context, with an aerial picture of the construction site shown in Figure 3.10.

Figure 3.10: Aerial picture of the CHIPS-5 construction site facing south. The Wentworth 2W pit is in the lower half of the image, with the part built CHIPS-5 detector visible at the bottom of the earthen construction ramp. The two white shore huts can be seen halfway up the ramp.

Figure 3.11: Picture of the CHIPS-5 structural frame with humans for scale. The endcaps can be seen separated by steel struts. Rows of stainless steel *stringers* are attached to the inside of each endcap for instrumentation mounting.

3.2.2 Structure

The structure of CHIPS-5 consists primarily of two 26 m diameter and 1.3 m high lightweight stainless steel circular *endcaps* that form the top and bottom of its cylindrical shape. During construction the conveniently named *top-cap* is held above the *bottom-cap* by 1.5 m long steel struts as shown in Figure 3.11. This configuration allows for the endcap instrumentation, detailed in Section 3.2.3, to be easily installed.

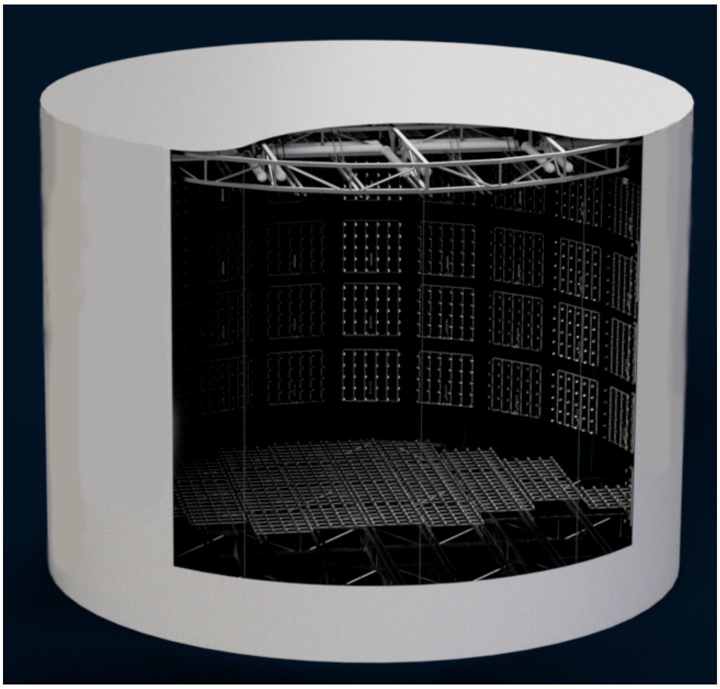


Figure 3.12: Graphical rendering of the fully deployed and expanded CHIPS-5 detector module with a section of the liner cutaway. The bottom endcap and wall instrumentation is visible, as well as the top endcap structure and floatation. The green lines indicate the Dyneema cables holding the endcaps together.

The two endcaps are connected by 28 Dyneema cables attached around their perimeter, each 12 m in length. Additionally, a total of 48 air-filled Polyvinyl Chloride (PVC) pipes, each 40 cm in diameter, are attached to the frame of the top-cap to make it buoyant. Once deployed into the pit, the bottom-cap sinks while the top-cap floats, pulling the Dyneema cable taut and forming the final expanded detector shape shown in Figure 3.12.

A lightproof and watertight liner is also installed to surround the fully expanded structure. Designed to isolate the clean internal water from the external pit water and to prevent non-Cherenkov light from reaching the PMTs, the liner is made from *geomembrane*, a flexible reinforced polymer membrane. Commercially available in large rolls, the liner is welded together during construction to form the top, bottom, and sides. Note that when fully deployed, the liner does not take any of the structural strain.

3.2.3 Instrumentation

CHIPS-5 is instrumented with PMTs arranged within distinct plane like structures called Planar Optical Modules (POMs), which take inspiration from the Digital Optical Modules used by IceCube and KM3NeT [86, 87]. Each POM is a roughly 2 m by 3 m array of watertight PVC tubing equipped with between 15 to 30 PMTs, as well as the lowest level of DAQ electronics. Commercially available PVC piping and connectors are used to form the structure of each plane, glued together with standard PVC primer and cement.

There are two types of POM used within CHIPS-5, differentiated by the PMTs and DAQ electronics they use and named after the institution at which they were primarily developed. Firstly, *Nikhef* POMs use 88 mm in diameter HZC PMTs [88], with electronics developed for the KM3NeT experiment [89, 90]. Secondly, *Madison* POMs use 76 mm in diameter Hamamatsu PMTs [91], donated from the NEMO-3 experiment [92], with electronics developed by CHIPS in collaboration with the Wisconsin IceCube Particle Astrophysics Centre (WIPAC) in Madison, Wisconsin.

The HZC PMTs have a high ratio of output electrons to incident photons (quantum efficiency) of 24.4% at a wavelength of 400 nm, compared to the low 12.0% ratio achieved by the Hamamatsu PMTs. Furthermore, the photon hit time resolution is $\sim$2 ns and $\sim$5 ns for the HZC and Hamamatsu PMTs respectively.

In total 6114 HZC, and 450 Hamamatsu PMTs are arranged into 226 Nikhef and 30 Madison POMs. Every PMT is housed within an assembly as shown in Figure 3.13 for the Nikhef case and Figure 3.14 for the Madison case. Importantly, to increase the level of light collection, each Nikhef PMT is equipped with a *light-cone* consisting of a circular reflective surface at 45ř to the PMT normal. The Madison PMT assembly is similar but has no light-cone. For POMs attached to either endcap, their PMTs are angled at 45ř facing the direction of the beam to optimise light collection.

All PMTs within a POM are connected to the lowest level of DAQ electronics contained within a dedicated onboard electronics box, either made from aluminium or PVC in the Nikhef or Madison case, respectively. A flexible PVC *pigtail* is attached to each POM electronics box containing connections to the higher level DAQ and power supply. A *water-block* within each pigtail ensures that if either the external link or POM becomes flooded, the other is still capable of withstanding the 6 atm of water pressure at the bottom of the pit. A fully assembled and installed Nikhef POM is shown in Figure 3.15 for reference.

The POMs are tiled next to each other on the detector walls, attached to either the stainless steel stringers on the top-cap and bottom-cap, or clipped to the Dyneema cables on the vertical walls of the *barrel*. As mentioned previously, full high-density detector instrumentation is not required for CHIPS detector modules, for two main reasons.

(a) Disassembled **(b)** Assembled

Figure 3.13: Disassembled (a) and assembled (b) Nikhef PMT assembly components. The assembly comprises of a black PVC insert, a HZC PMT, a transparent acrylic cover, and a reflective light-cone. The PMT is glued to the inside surface of the cover using a silicone-based optical gel, and a watertight seal is made between the insert and cover using an O-ring. The reflective light-cone is clipped to the front of the cover, and the whole assembly is glued into the POM PVC structure.

(a) Outside **(b)** Inside

Figure 3.14: A Hamamatsu PMT outside (a) and inside its insert (b). The PMT is potted inside its black PVC insert to create a watertight seal that can withstand the 6 atm of water pressure at the bottom of the pit. A ţDAQ detailed in Chapter 4, is seen attached to the base of the PMT in (a).

Figure 3.15: Picture of a single Nikhef full-density POM installed on the top-cap of the CHIPS-5 detector. Both the inward-facing and veto PMTs are visible as well as the aluminium electronics container and pigtail, whose end is covered in green tape.

Firstly, only highly directional accelerator beam events are to be studied. Therefore, the vast majority of neutrino interaction Cherenkov radiation is deposited on a relatively small downstream region of the detector walls. Secondly, beam neutrinos predominantly have multi-GeV energies, yielding a relatively large amount of Cherenkov radiation. Therefore, a smaller number of PMTs are required to capture adequate Cherenkov radiation from each interaction.

Consequently, the *photocathode coverage*, defined as the fraction of the complete detector wall area covered by the PMT photocathodes (the elements which the photons initially strike), is optimised within CHIPS-5 to reduce the total number of PMTs used. The detector is split into three distinct regions of PMT photocathode coverage, whose boundaries are defined by their azimuth angle ϕ from the centre of the downstream wall (where $\phi = 0^\circ$) using the centre of the detector as the origin. Each boundary is chosen by inspecting the fraction of beam Cherenkov photons that hit the detector walls as a function of ϕ. This distribution, generated by the detector simulation (outlined in Section 3.3.1), is shown in Figure 3.16 alongside the chosen boundaries.

Firstly, there is a *full-density* Nikhef POM region in the most downstream $\phi = \pm 75^\circ$ region of the detector with a $\sim 3\%$ photocathode coverage. Secondly, a *half-density* Nihkef POM region covering the $\phi = \pm 75^\circ$ to $\phi = \pm 180^\circ$ region of the endcaps and

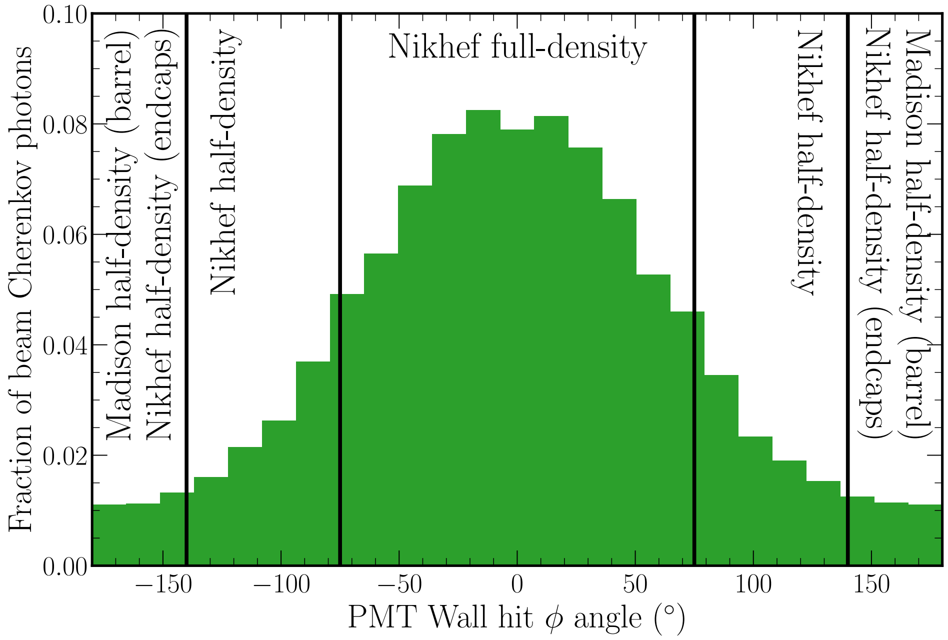


Figure 3.16: Fraction of beam event Cherenkov photons that hit the detector walls as a function of the hit PMT position angle ϕ. The different photocathode coverage regions are indicated.

the $\phi = \pm 75°$ to $\phi = \pm 140°$ region of the barrel with a $\sim 1.5\%$ photocathode coverage. Finally, a *half-density* Madison POM region covering the $\phi = \pm 140°$ to $\phi = \pm 180°$ upstream region of the barrel with a $\sim 0.8\%$ photocathode coverage.

Simulation studies have shown that this configuration, requiring just 60% of the PMTs compared to full high-density coverage, leads to only an 8% reduction in the average amount of total collected photoelectrons per event. This reduction corresponds to a decrease in the signal CC ν_e selection efficiency of just $\sim 2\%$ and a degradation in the CC ν_e energy reconstruction resolution of $\sim 5\%$, acceptable given the vast reduction in cost. It is important to note that these studies were performed using the standard event reconstruction and classification presented in Section 5.2 and should be revisited in the future with the new CNN implementation.

The CHIPS-5 detector module is equipped with a veto region within the top-cap frame structure to aid cosmic muon rejection. Separated from the main detector volume by a geomembrane liner, the 1.3 m high region allows for the rejection of predominantly downward going cosmic muons by detecting the Cherenkov radiation they produce. A total of 324 upward-facing HZC veto PMTs within the top-cap Nikhef POMs give a veto photocathode coverage of $\sim 0.6\%$. A graphical rendering of the full top-cap

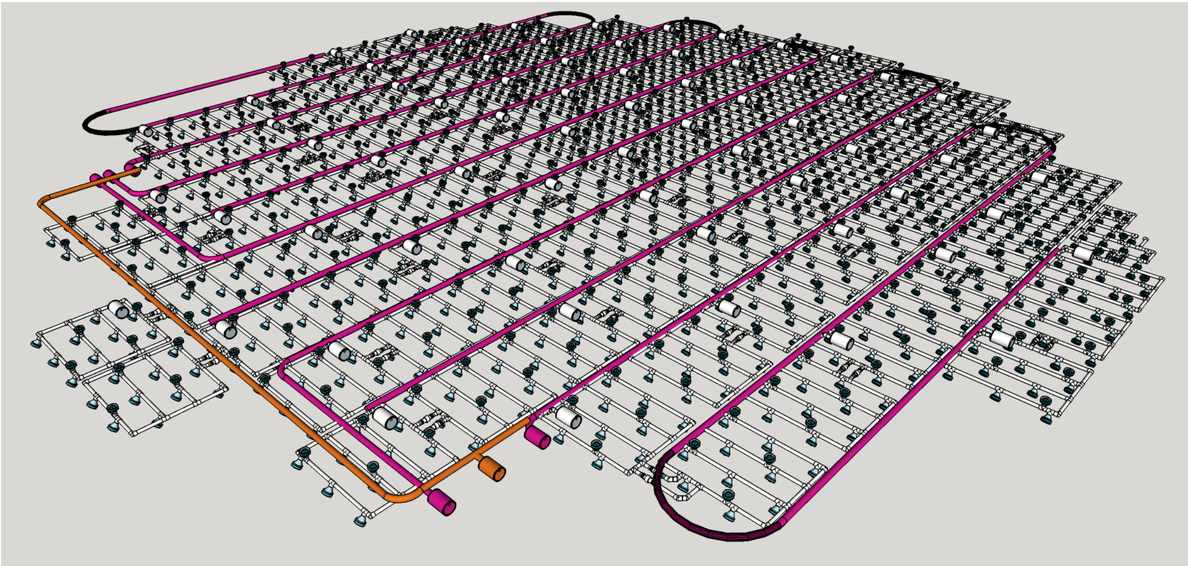


Figure 3.17: Graphical rendering of the top-cap POMs. Both the different photocathode coverage regions and the veto PMTs (contained within the top-cap POMs) are visible. The *manifolds* connecting the POMs with the higher level DAQ components are also shown in pink and orange.

instrumentation, showing the different density regions and veto PMTs, is shown in Figure 3.17.

All CHIPS-5 instrumentation receives power and is connected to the highest level DAQ systems through a 500 m long flexible PVC umbilical resting on the pit floor. The umbilical contains a single optical fibre for data and two shielded gauge 10 cables for power. One end of the umbilical is attached to the bottom-cap of the detector while the other enters a hut onshore containing the master power supply and onshore DAQ equipment.

3.2.4 Filtration

Though surprisingly clear, the Wentworth 2W pit water requires filtration to reach the necessary 30 m photon attenuation length for adequate light collection. Therefore, a high volume pump is used to continually pull the internal detector water through a 500 m long pipe to a filtration hut on the shore. After filtration, the water is returned to the detector through a second such pipe.

Within the hut, ten parallel sets of filters are installed in order to achieve a reasonable flow rate, allowing for the full detector volume to be filtered every ten days. Each filter

Figure 3.18: Picture of the CHIPS-5 filtration system within one of the shore huts. The pipes to and from the detector can be seen entering and exiting the hut respectively via a gap in the back wall.

set consists of a 51 cm long 10 μm carbon block filter followed by a 51 cm long 0.5 μm polypropylene filter, as shown in Figure 3.18. This configuration is found to achieve a photon attenuation length of 133(2) m after approximately two months of constant filtering [79].

3.2.5 Construction and deployment

The construction and deployment of any CHIPS detector module is a bespoke process that is relatively complex when compared to other experiments. This complexity is primarily due to a body of water being used for activities rather than solid ground. Below a simplified version of the CHIPS-5 construction and deployment procedure is outlined.

1. An earthen barrier is built separating the main body of Wentworth 2W from the construction area. As the pit water level rises during the summer, this acts as a dam preventing the construction area from flooding (at least in theory).

2. The bottom-cap liner is welded together before the endcap structural frames are constructed above, resulting in that shown in Figure 3.11.

3. The barrel liner is partly constructed starting from the detector base. Using a vertical liner roll for each of the 28 sides, the barrel liner is welded up to the height

of the top-cap, as is shown in Figure 3.19. The top-cap liner is also installed but not welded to the barrel liner.

4. A strongly buoyant *floating-dock*, made from steel and large plastic floats, is constructed surrounding the detector. The bottom-cap is attached with metal chains to winches on the floating-dock and with the Dyneema cables to winches on the top-cap.

5. The pre-assembled and tested endcap POMs are installed along with their associated power and data connections as well as other high-level DAQ components.

6. The earthen barrier is removed and the construction area flooded, causing the detector and floating-dock structure to float. The detector is towed by a boat to its deployment location.

7. The detector is slowly filled with water at the same time as being lowered using the winches on the floating-dock. This process continues until the top-cap reaches the surface of the water and begins the float. At this point, the steel struts separating the endcaps are removed.

8. The barrel POMs are installed in layers, brought to the detector location by boat. The gap between the barrel and top-cap liner allows this to happen. After a complete layer has been installed the bottom-cap is lowered using the winches on the floating-dock and top-cap and the barrel liner is correspondingly welded to a greater height before the procedure repeats.

9. After all the POMs layers have been installed, the barrel and top-cap liner are welded together, the detector umbilical attached, and the whole detector lowered until it rests at the bottom of the pit.

10. The water within the detector volume is filtered for approximately two months until a stable photon attenuation length is reached. Once complete, filtering continues to maintain the water clarity and data taking runs begin.

3.2.6 Current status

During the summer of 2018 and 2019 extensive CHIPS-5 construction work was carried out by a team of 10 to 15 collaborators at any one time. Given the novel nature of the

Figure 3.19: Picture of the Chips-5 detector module with liner welded up to the height of the top-cap. A section of the floating-dock can also be seen in the foreground.

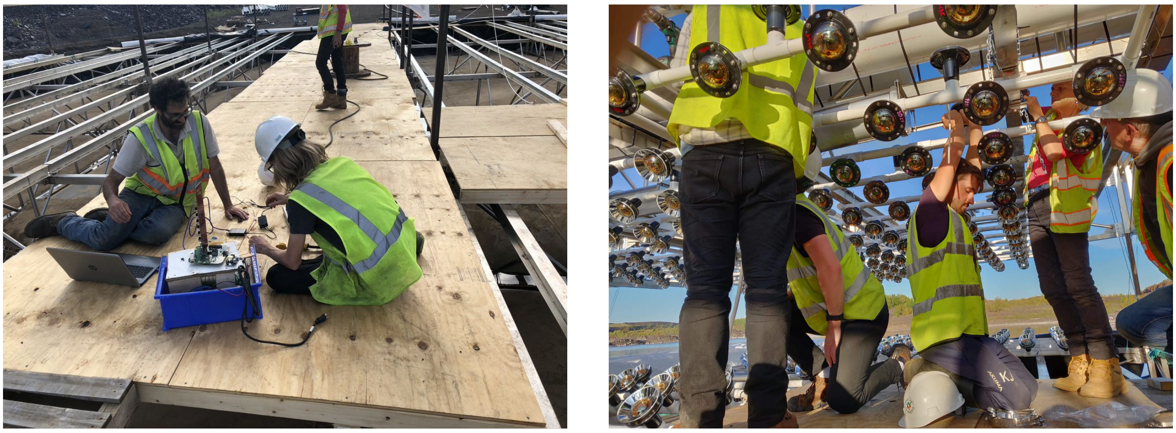
(a) Cornerstone placement.

(b) Floatation production.

Figure 3.20: Some Chips-5 construction work.

project, most tasks, some of which are pictured in Figure 3.20 and Figure 3.21, proved both challenging and time consuming.

By late summer 2019, it became apparent that deployment of a fully instrumented Chips-5 would be impossible before the pit started to freeze over in October. Therefore, the decision was taken to only partially instrument the endcaps, leave the barrel walls bare, not include the veto separation liner, and reduce the module height to 8 m. In October 2019 this version of Chips-5, shown in Figure 3.22, was deployed into the Wentworth 2W pit.



(a) Connection testing. (b) POM installation.

Figure 3.21: More CHIPS-5 construction work.

Analysis of preliminary data was successful in isolating individual cosmic muon events, one of which is shown in Figure 3.23. However, extensive data taking was paused due to a deployment induced tear in the liner preventing filtration of the internal water. Given the outbreak of the worldwide SARS-CoV-2 pandemic, all plans for work over the summer of 2020 were suspended, including essential repairs and instrumentation of the barrel walls. Plans for summer 2021 are currently under review.

3.3 Detector simulation and event generation

To describe CHIPS-5 and other CHIPS concept detectors in a simulated environment, a Monte Carlo simulation framework is used. Monte Carlo detector simulation and event generation methods are an indispensable tool within high energy particle physics. This is particularly true during the design and prototyping phase of an experimental project when no real-world data is available, as is predominantly the case with CHIPS. By matching the observables of a real-world detector as close as possible, these methods allow for the optimisation of detector design, the testing of reconstruction techniques, and the study of physics sensitivities.

Here, a description of the detector simulation and the beam and cosmic event generation procedures used by CHIPS are described. All are employed extensively for the principal work of this thesis presented in Chapter 5 and Chapter 6.

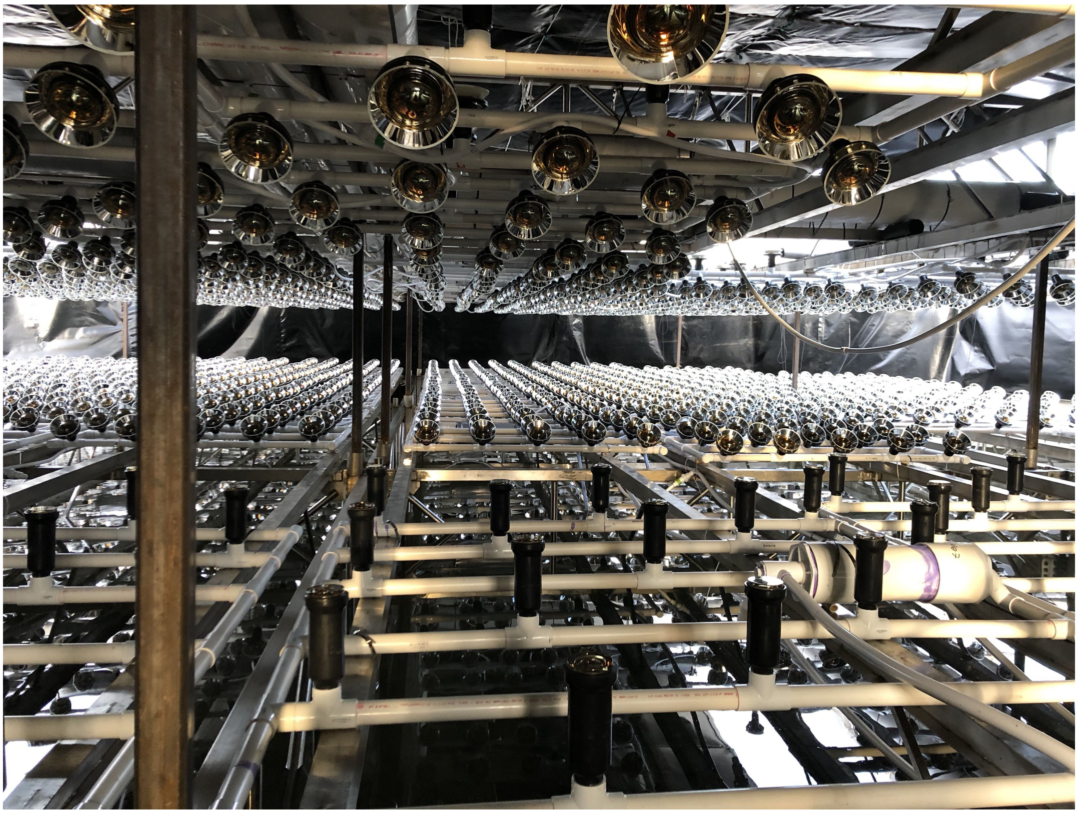


Figure 3.22: Inside picture of CHIPS-5 just before deployment, with a section of the total instrumentation shown. In total 56 Nikhef and 6 Madison POMs were installed. Instead of being mounted on the upstream detector walls the Madison POMs were mounted on the bottom-cap and can be seen in the foreground of the image. The flexible PVC tube pigtails and manifolds can also be seen connecting each POM to the higher level DAQ electronics and power supply.

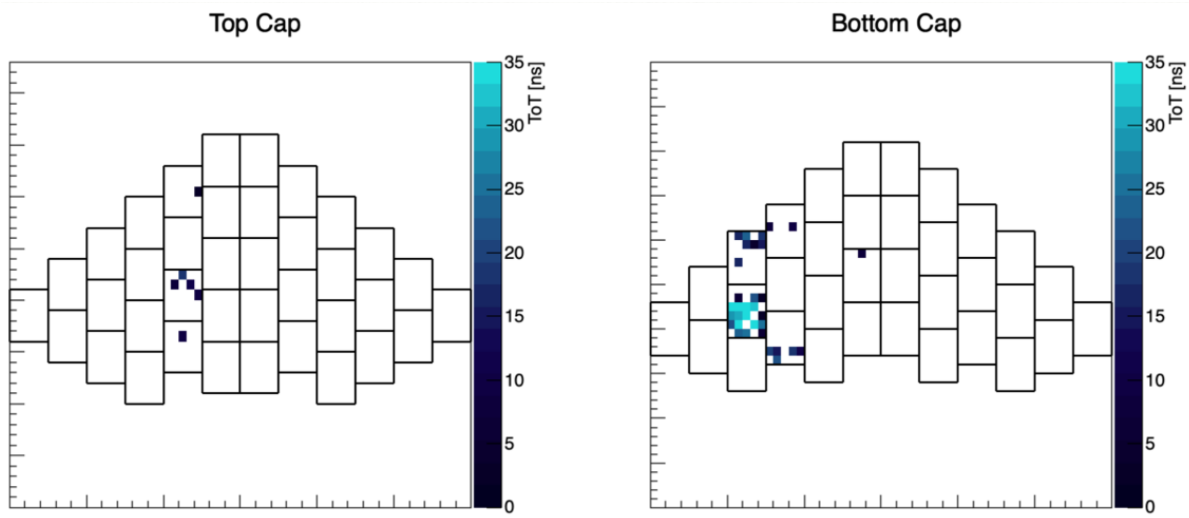


Figure 3.23: A cosmic muon event recorded within CHIPS-5, found by clustering hits in time. PMT hits from both the top-cap (left) and bottom-cap (right) are shown with the colour indicating the measured ToT value (described in Section 4.2). Due to the small attenuation length of unfiltered water, only a single bottom-cap POM has significant activity.

3.3.1 Detector simulation

The detector simulation uses the WCSim water Cherenkov simulation package [93] built on top of the Geant4 simulation framework [94–96]. Developed initially to simulate possible water Cherenkov detectors in the LBNF beam, WCSim is now used more widely in the field. The CHIPS version of WCSim allows for generic water Cherenkov detector geometries to be easily loaded at runtime via a series of simple XML configuration files. These changes allow for a broad range of detector geometries to be quickly considered without recompilation of the code.

The simulation builds an n-sided, regular polygonal prism consisting of two endcaps and a barrel, filled with water and lined with a low reflectivity *blacksheet*. The geometry is separated into *regions* within both the barrel and endcaps, defined either by a list of barrel sides or an opening angle, respectively. Each region is filled with a unique base unit of geometry known as the *unit cell*, as shown in Figure 3.24.

The unit cell defines a pattern of any number of PMTs, including their relative positions and in which direction they face. The final geometry is built by tiling the defined regions with their respective unit cell scaled to match the required regional photocathode coverage. Note that although exact detector PMT positions are not explicitly defined using this procedure, a given configuration will always generate the

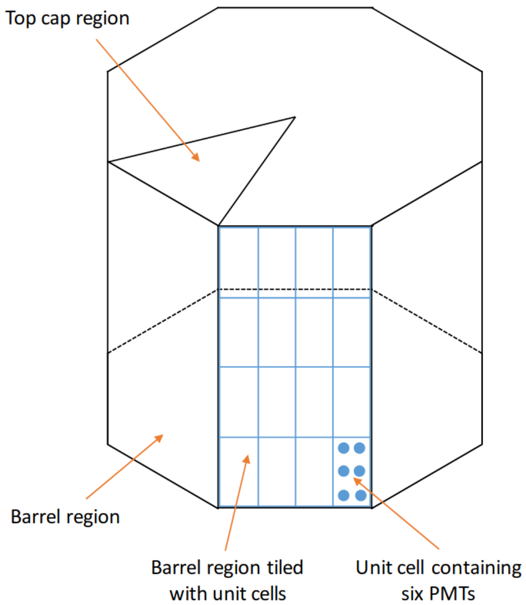


Figure 3.24: Illustrative diagram of a WCSim detector geometry. The endcap and barrel regions, tiled unit cells, and PMTs within a unit cell are all shown. Figure taken from reference [97].

same geometry (it is deterministic). In this work, the CHIPS-5 geometry is generated with 28 sides and regions matching the boundary angles and photocathode coverage detailed in Section 3.2.3.

The geometry shape, regions, and unit cells are defined in a configuration file. Additionally, a file for PMT definitions containing their shape, time resolution, and quantum efficiency is also used. Light-cones are described by a list of radial profile points in a further file. Although the underlying Geant4 material properties are mostly hardcoded (taken from the original WCSim implementation using proven Super-Kamiokande values), another configuration file is used to scale them. This scaling controls the water absorption, scattering (Rayleigh and Mie) lengths, and both the blacksheet and PMT glass reflectivity. Throughout this work (including that presented in Chapters 5 and 6) a conservative photon attenuation length of 50 m at 405 nm is used alongside negligible scattering [79], with the PMT glass reflectivity set to 24% [98], and the blacksheet reflectivity kept at the WCSim default of 4%.

A veto volume can also be defined. The veto is built as either a concentric shell around the whole inner volume with a given thickness or solely above the top-cap with a given height. Any PMTs defined as facing outwards within a unit cell look into the veto volume instead of the inner volume. In this work, the CHIPS-5 geometry is given a top-cap veto of height 1.3 m with a photocathode coverage matching that described in Section 3.2.3.

Once the full generation of the Geant4 geometry is complete, the final state tracks for each successive event to be simulated are loaded from either the beam or cosmic event generator files described in Section 3.3.2 and Section 3.3.3. Beam event vertices are randomly placed within the inner detector volume, while cosmic vertices are set at 1 m above the detector volume. WCSim then simulates the passage of all particles through the detector materials, with interactions, decays, and Cherenkov emission all considered.

Whenever a photon is calculated to have hit the photocathode of a PMT, an angular dependent acceptance efficiency is applied to see if it is recorded [98]. If accepted, all hits within 200 ns windows are grouped to form a single recorded hit, with a smeared first hit time used as the recorded time. The standard WCSim methodology is used to determine the total output charge of the hit, given the number of incident photons. This procedure involves a single photoelectron charge distribution being repeatedly probed for each photon, before the combined sum is returned [99].

By sampling this procedure multiple times, the output charge probability distribution given the number of incident photons can be generated, as shown in Figure 3.25. The reverse likelihood function of an actual number of incident photons given a measured digitised charge is also shown (with a lower value being more likely). The reverse likelihood function is used for the standard reconstruction methods outlined in Section 5.2.1.

The simulated PMT hits for each event are stored in an output file along with truth information and track descriptions. A full event takes approximately three seconds to simulate on a standard batch farm computing node.

3.3.2 Beam event generation

The expected flux of beam neutrinos at the CHIPS-5 detector location, shown in Figure 3.26, is generated using the existing beam simulation written for the NuMI experiments. Using the generated fluxes as input, the GENIE neutrino event generator (version 3.0.6) [32, 33] is used to generate beam neutrino events. Default neutrino cross-sections on water provided by GENIE are used during generation. All initial, intermediate, and final state particle tracks for each event are stored as output in a NUANCE formatted file for use in the detector simulation. Note that unoscillated input fluxes are used such

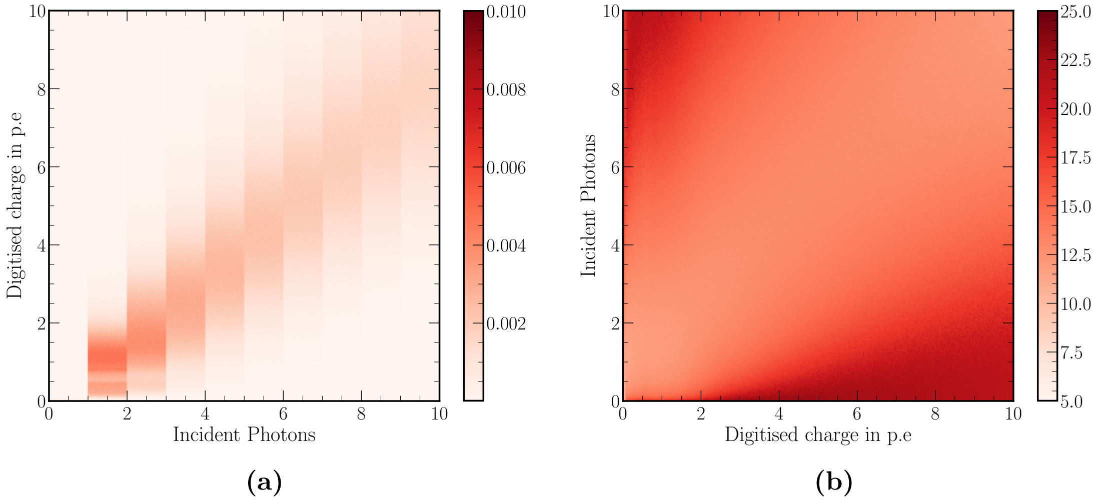


Figure 3.25: The probability of a digitised output charge given an incident number of photons used within the detector simulation is shown in (a). The reverse function, showing the likelihood of a number of incident photons given the digitised charge, is shown in (b). The digitised charge is given in units of p.e (photoelectrons), representing the 'measured' PMT charge pulse scaled to a corresponding number of initial electrons given the known parameters of the PMT.

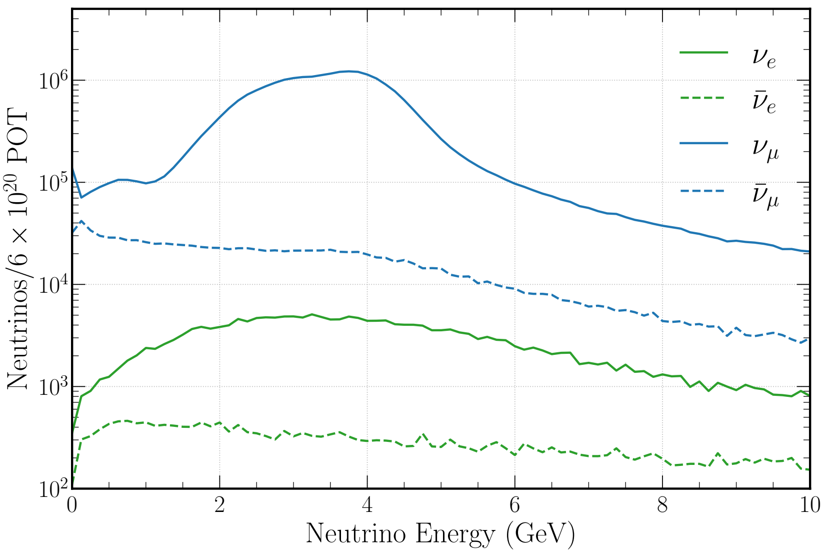


Figure 3.26: The neutrino mode (forward horn current) NuMI beam neutrino energy spectrum at the CHIPS-5 detector module location. Shown are the individual contributions from the different neutrino types and signs. No cross-sections or oscillations have been applied.

that analyses samples must be later weighted to match the desired oscillated neutrino composition.

3.3.3 Cosmic event generation

The Cosmic-Ray Shower Library (CRY) [100, 101] is used for cosmic ray event generation. Both the solar cycle and Earth's geomagnetic field are taken into account, with the CHIPS-5 latitude (47.56° N) and deployment date (30th October 2019) used as input. Single muons are generated at sea level by CRY within a 1 km by 1 km area, with the detector at it's centre. Note that only single muon events (the dominant cosmic component) are considered for simplicity.

Assuming a CHIPS-5 overburden of 50 m and a 2.2 MeV/cm^2 muon energy loss in water as suggested by reference [102], the muon parameters are updated to estimate their values at 1 m above the detector. All muons whose path do not cross the detector volume or do not have sufficient energy to reach the detector are discarded [103]. All accepted muon tracks are stored as output in a NUANCE formatted file for use in the detector simulation.

To use generated cosmic events in analysis, studies have looked at the likely cosmic rate for CHIPS detector modules at different water overburden depths [104]. In this work, the fits shown in Figure 3.27 for a cylindrical detector of both height and diameter equal to 24 m are used to estimate a CHIPS-5 cosmic muon rate of 11.8 KHz at 50 m of overburden.

Given the 10 ţs long NuMI beam spill occurring every 1.33 seconds, an in spill cosmic rate of 2.1 ± 0.2 million events per year is calculated with an in spill occupancy of 9%. Considering a typical event takes ∼100 ns to unfold, there is approximately a 0.3% chance that any beam event overlaps with a cosmic muon. This low coincidence shows just how powerful a short beam spill can be at reducing the significant cosmic background.

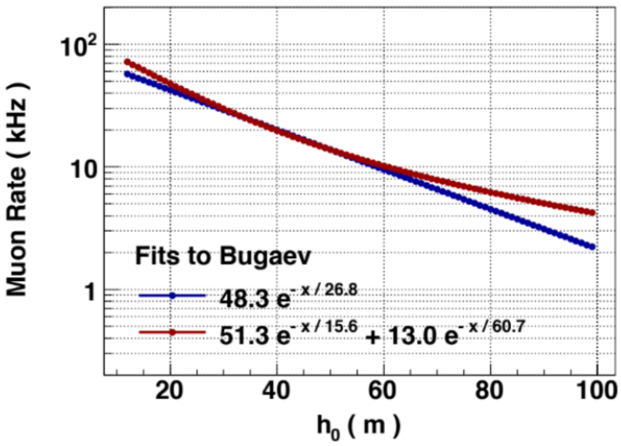


Figure 3.27: The expected cosmic muon rate as a function of water overburden depth (h_0) for a 24 m in height and 24 m in diameter CHIPS detector module. Shown are single (blue) and double (red) exponential function fits to work originally conducted in reference [105]. Importantly, in the region of interest at $h_0 \sim 50$ m both models are consistent. Figure taken from reference [104].

Chapter 4

Data acquisition for CHIPS

The primary task of any data acquisition system is the processing of low-level signals measuring real-world physics and their transfer to permanent storage for further analysis. Commonly, this procedure also includes decision making as to whether the signal is deemed interesting enough to record, known as a *trigger*. Both of these tasks can make DAQ systems incredibly complex, especially when they must operate efficiently and resiliently for vast amounts of data in real-time, whilst also providing detector control and monitoring.

In the context of the CHIPS project, the DAQ system records all PMT hits, timestamps them using a common clock, and transfers them out of the detector for processing. A trigger is applied to select those hits that fall within the interesting NuMI beam spill time window before the selected hits are moved to permanent storage for further analysis. Alongside these processes, the DAQ system also configures the detector and provides hardware and data quality monitoring.

Although relatively simple when compared to the incredibly complex and time-pressured DAQ systems of the LHC experiments, the DAQ system developed for the CHIPS project contains some unique approaches to solve the goals of the CHIPS concept. Namely, deployment within a body of water and a limited resource budget. In this chapter, the DAQ system for CHIPS as applied to the CHIPS-5 prototype detector module is described. The description is presented in two broad categories, hardware and software, with a short description of the timing system to begin.

4.1 White Rabbit timing

To ensure PMT hit times are synchronised throughout CHIPS detectors, a common clock must be shared across all timestamping electronics. For this purpose, CHIPS uses a *White Rabbit* (WR) network [106]. Initially developed at CERN, the open-source WR project provides an ethernet-based time distribution network with sub-nanosecond synchronisation accuracy between nodes. By using two-way exchanges of WR messages, precise adjustment of individual node clock phases and offsets are possible, across thousands of devices separated by tens of kilometres. All of this is achieved alongside a standard data transfer network capable of 1 Gb speeds.

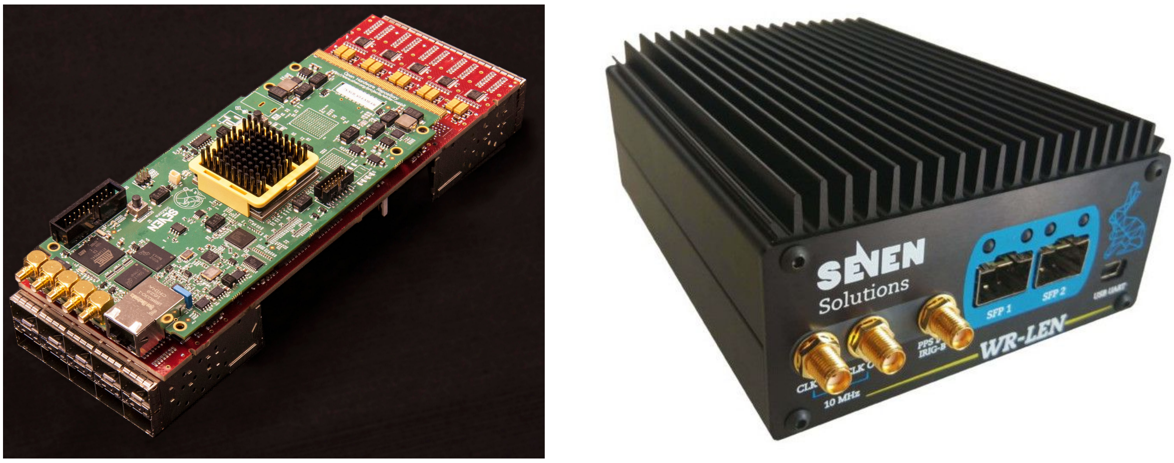
All nodes are synchronised to the clock of a *GrandMaster* node, typically a WR *switch*, the most common WR hardware component. As input, the GrandMaster switch receives an IRIG-B (Inter-Range Instrumentation Group timecode B) and a 10 MHz signal from a GPS disciplined oscillator. These inputs allow for synchronisation of the GrandMaster clock to International Atomic Time. As CHIPS detector modules require synchronisation to accelerator clocks many hundreds of kilometres away to determine the arrival time of beam spills, this GPS disciplined timing is essential.

WR hardware is commercially available from many vendors. Within CHIPS-5, two WR devices are used for time synchronisation and data transfer, both shown in Figure 4.1. Firstly, a compact version of the standard WR switch [107], specially developed for the CHIPS project at Nikhef [108]. Secondly, a WR-LEN (Lite Embedded Node) from Seven Solutions [109].

All WR components within CHIPS-5 are connected using 1 Gb bi-directional optical fibre connections, using the 1310 nm and 1550 nm wavelengths via Small Form-Factor Pluggable Transceivers (SFPs). Figure 4.2 shows the WR synchronised pulse per second rising edges for two CHIPS-5 WR switches separated by 500 m of optical fibre. With the vertical ticks representing single nanoseconds, sub-nanosecond time synchronisation accuracy between the switches is observed.

4.2 Hardware

The hardware of the CHIPS-5 DAQ system is split into two distinct implementations at its lower levels (closest to the PMTs), corresponding to the Nikhef and Madison POM



(a) White Rabbit Switch **(b)** White Rabbit LEN

Figure 4.1: Pictures of the White Rabbit timing hardware used within CHIPS-5. The compact WR switch specially designed for CHIPS is shown in (a), while the White Rabbit Lite Embedded Node (WR-LEN) from Seven Solutions is shown in (b).

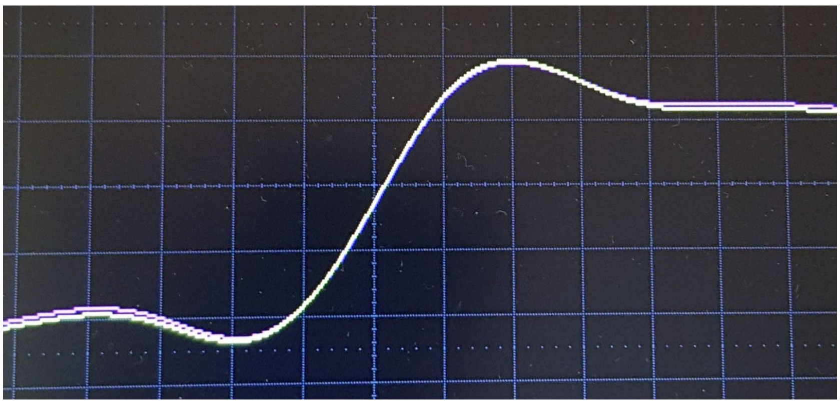


Figure 4.2: Oscilloscope display measuring the pulse per second output signal from two WR switches shown in pink and yellow at either end of a 500 m long optical fibre. The vertical ticks are in nanoseconds showing the sub-nanosecond synchronisation possible with the WR timing network.

types. CHIPS R&D efforts have principally developed the novel Madison implementation with the view of using this hardware exclusively. However, as a safe stepping stone, while development and testing are still ongoing, CHIPS-5 mainly contains proven Nikhef hardware developed for the KM3NeT experiment [90].

The complete DAQ and power distribution system for CHIPS-5 is diagrammatically shown in Figure 4.3. The following subsections describe each component, starting from the lowest level and working upwards. The Nikhef and Madison descriptions are separated for clarity in addition to the high-level combined hardware systems, part of which is not physically located within the detector but in an electronics hut onshore.

As a general overview, the DAQ system within the detector follows a tree-like structure. Starting from the lowest level (the leaves), individual PMTs are arranged into POMS. Groups of POMS, spatially close within the detector, are then connected to a small number of group-specific interface containers (either a *Nikhef-container* or *Madison-container*), holding networking and power supply devices for each group. All interface containers are attached to a central aggregation box (*junction-box*) which feeds connections into a single umbilical (the tree trunk) to shore.

Common to both low-level hardware implementations is the Time over Threshold (ToT) method for PMT signal digitisation. Each analogue PMT pulse is fed to a ToT discriminator coupled with a Time to Digital Converter (TDC) to generate each digitised recorded hit, as shown in Figure 4.4. Compared to the more common Analogue to Digital Converter (ADC) readout, ToT values do not scale linearly with charge. Instead, they follow a logarithmic-like function. Due to measurement uncertainties, this prevents the ToT method from distinguishing between high values of deposited charge, corresponding to approximately ten incident photons or greater per PMT per event for the PMTs used by CHIPS-5.

However, as the maximum expected incident photons within CHIPS-5 are relatively low (only a few), the impact of the ToT non-linearity is small, and the values are very consistent with those recorded using an ADC readout. Consequently, the use of ToT electronics is expected to have only a limited impact on the detector's performance whilst being much simpler to implement and notably cheaper.

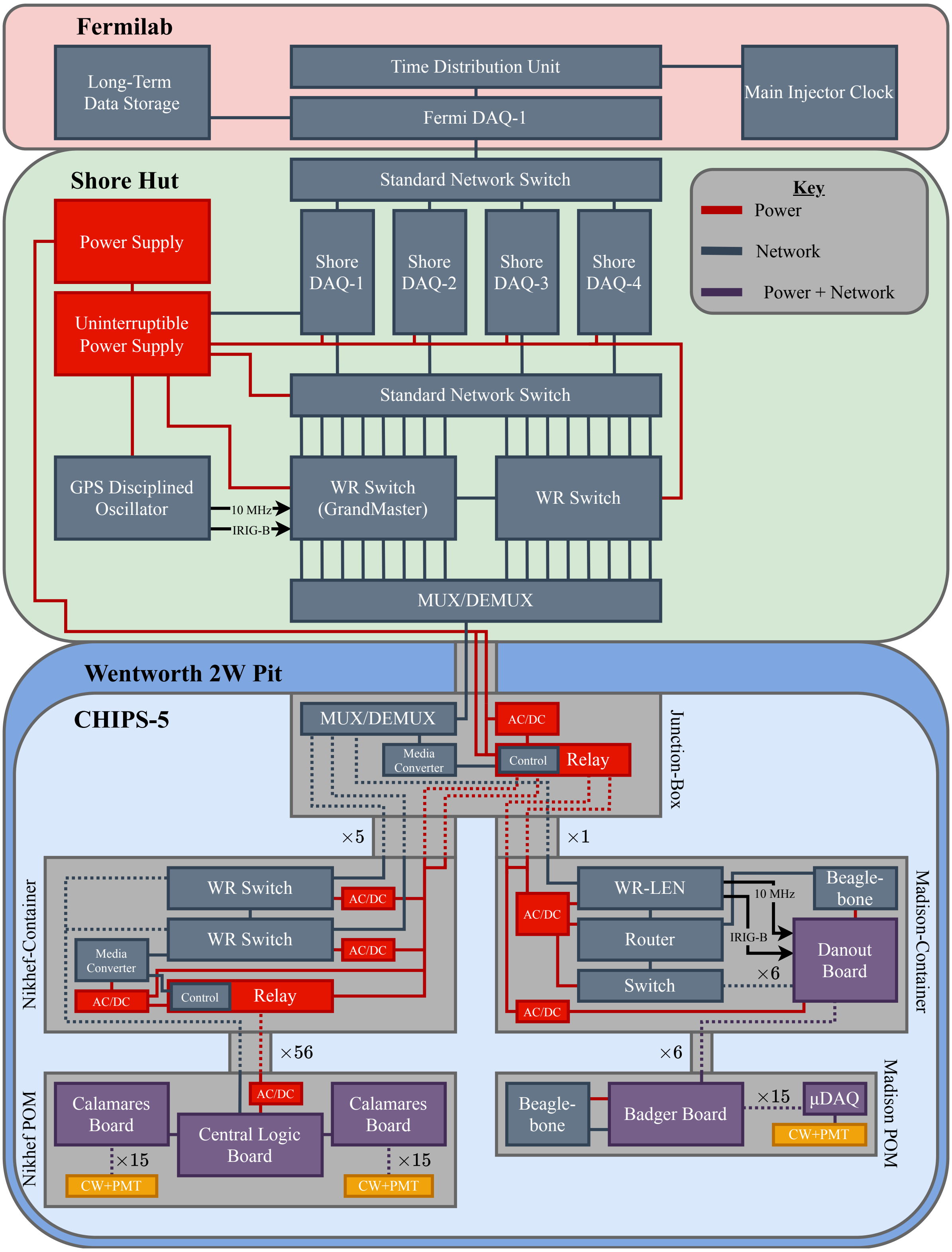


Figure 4.3: Diagram of the complete CHIPS-5 DAQ and power distribution system.

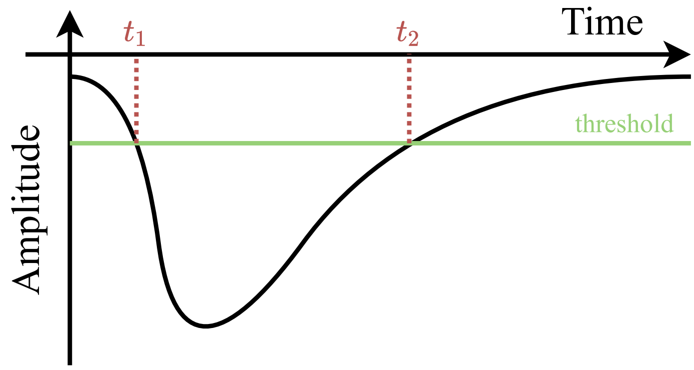


Figure 4.4: Illustrative diagram showing how a ToT value is measured. As soon as the rising edge of a PMT charge pulse rises above a given threshold (goes below in the negative charge case) a time is recorded t_1, when the falling edge later falls below the threshold a second time t_2 is recorded. The difference in time between t_1 and t_2 is output by the electronics as a digitised ToT value.

4.2.1 Nikhef hardware

All Nikhef HZC PMTs are attached directly to a simple readout board containing a high-voltage generating Cockcroft-Walton circuit [110]. Up to 30 such PMTs are connected to two fanout *Calamares* boards within the electronics box of each Nikhef POM via standard cables with RJ45 connectors, as shown in Figure 4.5. Both Calamares boards are directly attached to a *Central Logic Board* (CLB) [87, 111]. The CLB contains the ToT discriminators and TDCs used to digitise the recorded signals, as well as an FPGA (Field-Programmable Gate Array) programmed to provide additional logic and WR clock synchronisation for timestamping. Each Nikhef POM electronics box also contains an AC to DC power converter (AC/DC) whose output is fed into the CLB for distribution.

Every Nikhef POM is connected via a single optical fibre and a single power connection to an interface Nikhef-container, the contents of which are labelled in the bottom half of Figure 4.6. An aluminium structure holds two WR switches within each container specially designed to remove heat from the switch components and transfer it to the colder outer shell. Both switches are powered by independent AC to DC converters and connected via a single optical fibre each to the higher level DAQ systems. An additional connection is made between each switch to ensure that if one higher-level connection fails, the other can still be used for networking.

Each Nikhef-container also holds a relay board to control the power supply to individual POMs. The relay board control electronics are powered via an AC to DC

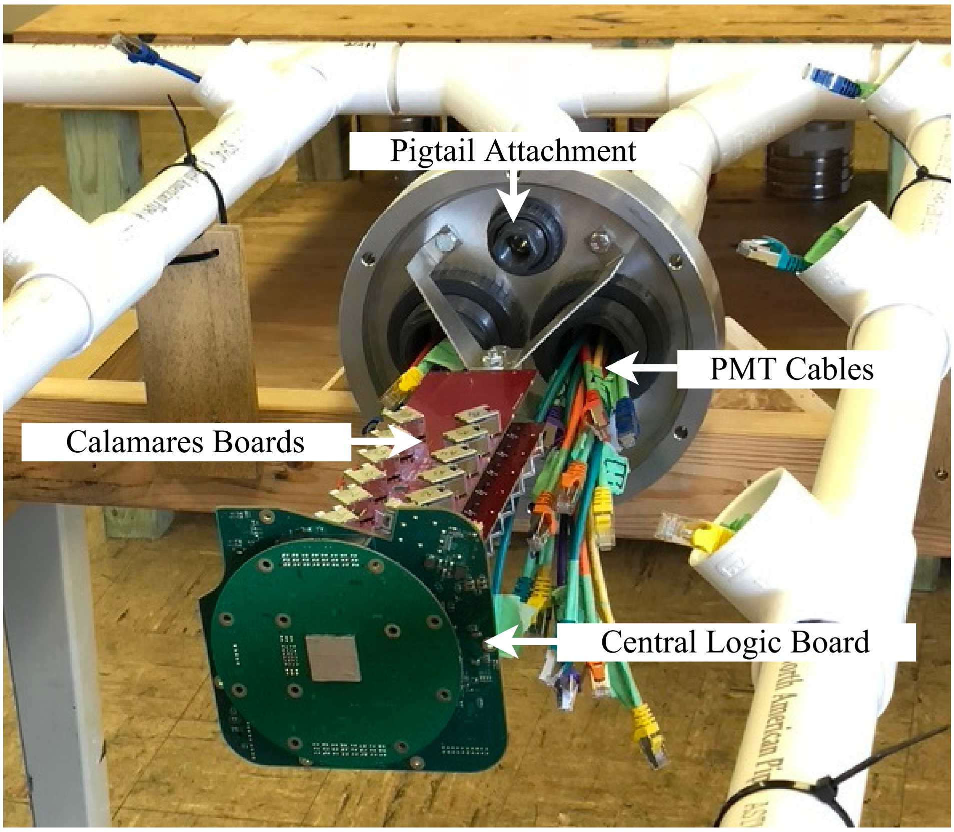


Figure 4.5: Labelled picture of the Nikhef POM electronics box without its aluminium casing. Both ends of the PMT cables can be seen, either at the PMT mounting points or entering the electronics box and not yet plugged into Calamares boards.

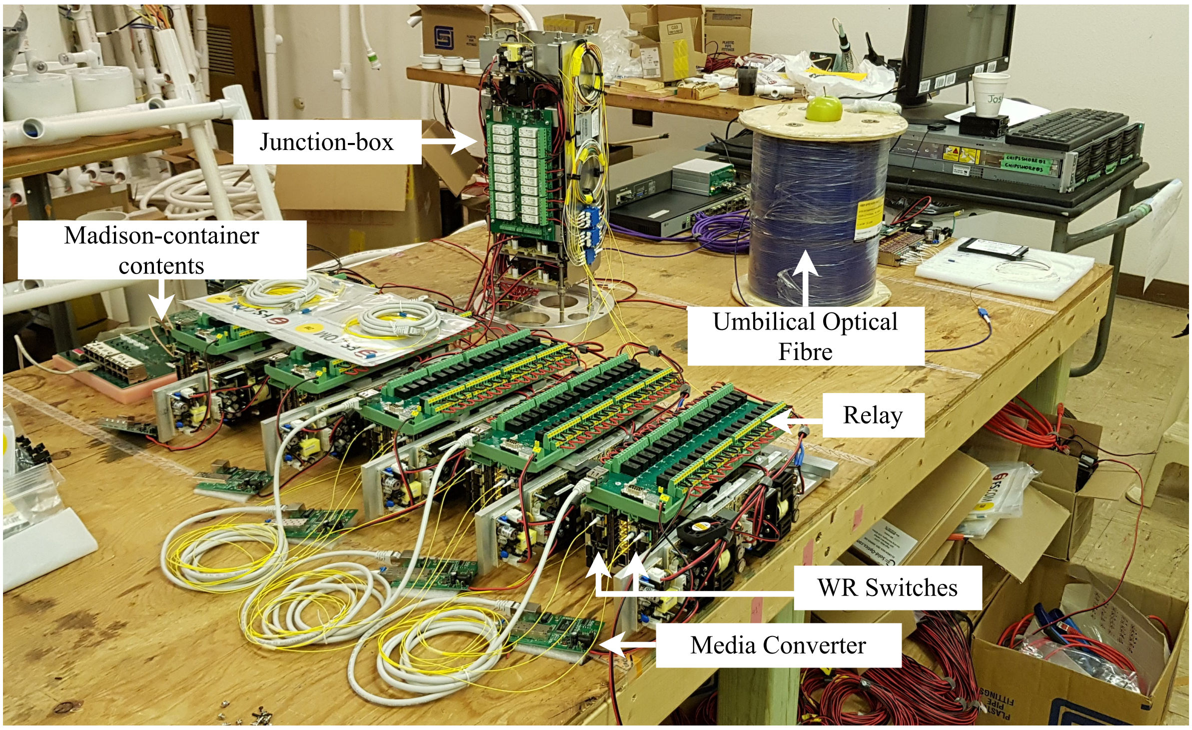


Figure 4.6: Labelled picture of the high-level non POM components of the CHIPS-5 DAQ system arranged on a table at the PolyMet mining administration building.

converter and connected to one of the switches via a media converter for networking. The media converter is required to convert the optical fibre WR switch connection to a standard RJ45 copper cable connection. A total of five Nikhef-containers are present within CHIPS-5.

4.2.2 Madison hardware

Every Madison Hamamatsu PMT is directly attached to a high-voltage generating Cockcroft-Walton board followed by a signal processing ţDAQ, as shown in Figure 3.14. The ţDAQ is a small microcontroller developed for both IceCube and CHIPS at WIPAC in Madison. Capable of timestamping and digitising signals directly at the PMT level, the ţDAQ also sets the PMT operating voltage by controlling the driving frequency of the Cockcroft-Walton board [82].

Up to 16 ţDAQs receive power, networking, and WR synchronised IRIG-B and 10 MHz timing signals from a *badger-board*, as shown in Figure 4.7. Standard cables with RJ45 connectors are used for these connections. The badger-board is located within the

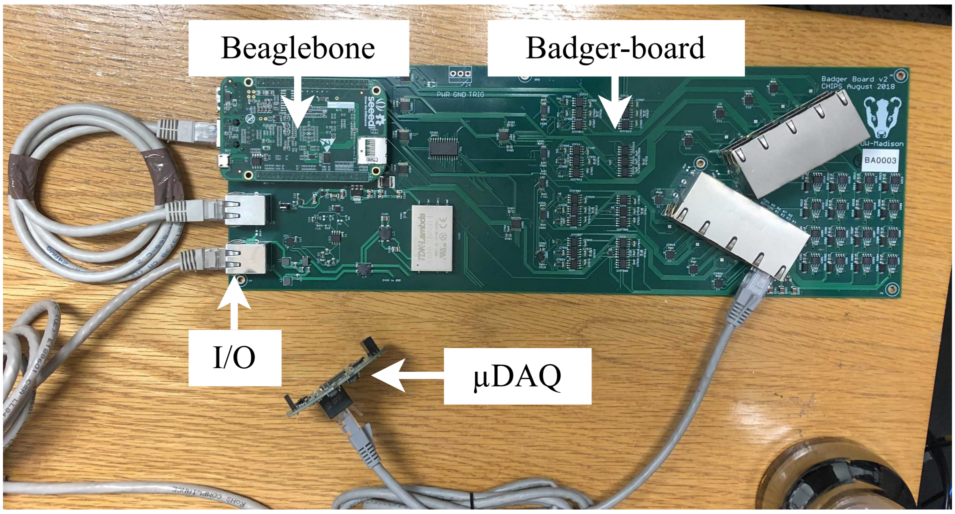


Figure 4.7: Labelled picture of the components of the Madison PoM electronics box.

electronics box of each Madison PoM and acts as a simple fanout and power control board. For logic, each badger-board has an attached mezzanine Beaglebone [112]. This single-board Linux machine (very similar to a Raspberry Pi) controls the power supply to, and receives hits from, the attached ţDAQs.

Similarly, up to 16 Madison PoM badger-boards receive power, networking, and WR synchronised IRIG-B and 10 MHz timing signals from a *danout-board* located within a single interface Madison-container. Again, standard cables with RJ45 connectors are used for these connections. The full contents of the Madison-container are shown in Figure 4.8. Similar to the badger-board, the danout-board acts as a simple fanout and power control board with an attached mezzanine Beaglebone. However, in this case, the attached Beaglebone acts only to control the power provided by the danout-board.

PMT hits and other data are instead routed through the danout-board into a networking stack. Consisting of a WR-LEN, a router (required due to the limited WR-LEN routing table size), and a switch (non-WR), the stack provides networking to the higher-level DAQ via a single optical fibre. The WR clock synchronised IRIG-B and 10 MHz timing signals are output by the WR-LEN to the danout-board for forwarding to the lower-level components. Additionally, two AC to DC converters provide power for both the devices within the container and all lower-level components via the danout-board.

Compared to the Nikhef hardware the CHIPS developed Madison implementation has a few key advantages, primarily driven by the core CHIPS concept of reducing cost:

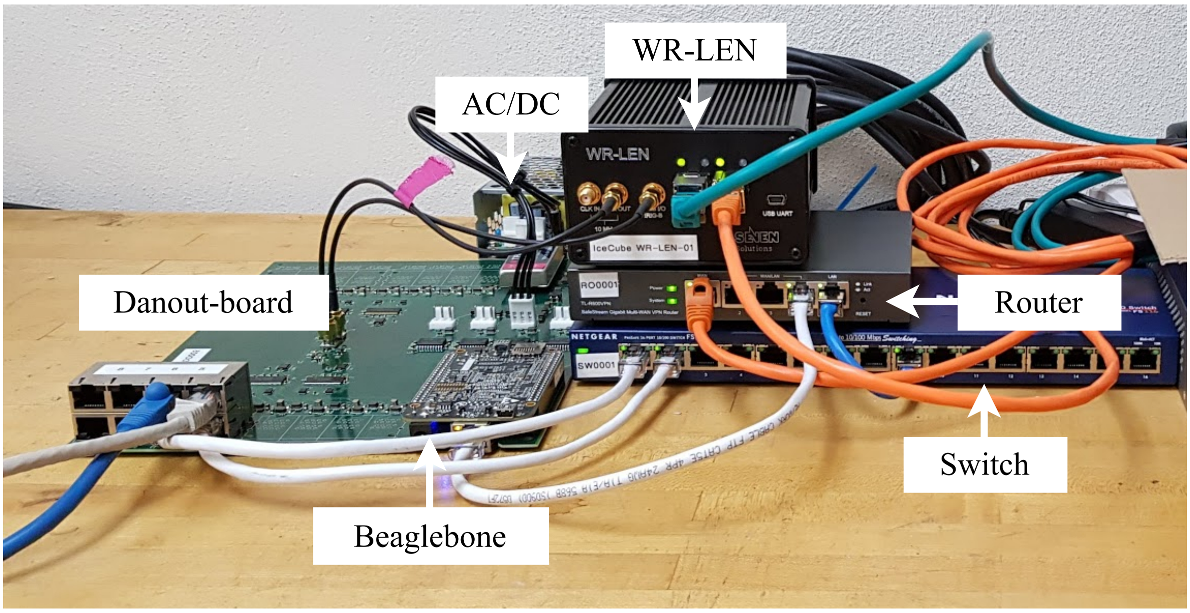


Figure 4.8: Labelled picture of the Madison-container components. The blue and grey cables exiting the left hand side of the image go to individual Madison PoMs connecting to the I/O port shown in Figure 4.7. An optical fibre connection into the left SFP port of the WR-LEN is used in reality rather than the copper connection shown here.

- Commercially available and cheap ($\sim$ £50 each) Beaglebones are leveraged for PoM level processing and logic instead of expensive, FPGA based CLBs. Not only are they cheap but Beaglebones provide a fully configurable general-purpose Linux machine that can easily be enhanced with new features once deployed. This is in comparison to CLBs which can be only be reconfigured via remote reprogramming of the FPGA.

- PMT hit processing and logic is pushed right to the lowest-level PMTs themselves via the use of ţDAQs. Although not currently fully realised this can have significant implications for the DAQ system as a whole. For example, future updates will allow for beam spill hit triggering to be performed on the ţDAQ, drastically reducing the capacity requirements of the rest of the DAQ system, further reducing costs.

- The number of expensive WR switches is reduced, with much cheaper WR-LENs taking their place. Although this change removes WR clock time synchronisation from the PoM level, tests have shown that achievable cable length time corrections still allow for sub-nanosecond synchronisation accuracy to be reached.

Figure 4.9: Picture of a Nikhef POM to Nikhef-container manifold (in white) attached to the top-cap of the CHIPS-5 detector. In the left of the image, two unattached Nikhef POM pigtail connections are seen, both covered in green tape and a plastic bag. The manifolds can also be seen in the graphical rendering of the top-cap shown in Figure 3.17.

4.2.3 Combined systems

Each interface container is connected to a single junction-box, labelled in Figure 4.6. This central container acts as the interface between the detector electronics and the umbilical, carrying data and power between the shore and the detector.

Due to the unique CHIPS constraint of complete submersion within water, all connections to the junction-box (as well as those between POMs and the interface containers) are made within watertight, flexible PVC tubing called *manifolds*. These tubes span all corners of the CHIPS-5 detector, as shown in Figure 4.9, usually attached to the endcap frames. Additionally, as is done for the individual POMs, water-blocks are placed between all interface containers and the junction-box to compartmentalise the higher-level DAQ components and prevent a single leak from taking down the whole system.

For networking the junction-box contains a Coarse Wavelength Division Multiplexing (CWDM) multiplexer/demultiplexer (MUX/DEMUX). This device supports 32 wavelengths for a total of 16 bi-directional 1 Gb connections over the single 500 m long umbilical optical fibre. Each WR-LEN or WR switch within the detector uses one of these channels exclusively with the corresponding wavelength SFP.

The two umbilical power connections are distributed via two thick copper plates to all the relay channels within the junction-box. Two relay boards are used to provide a

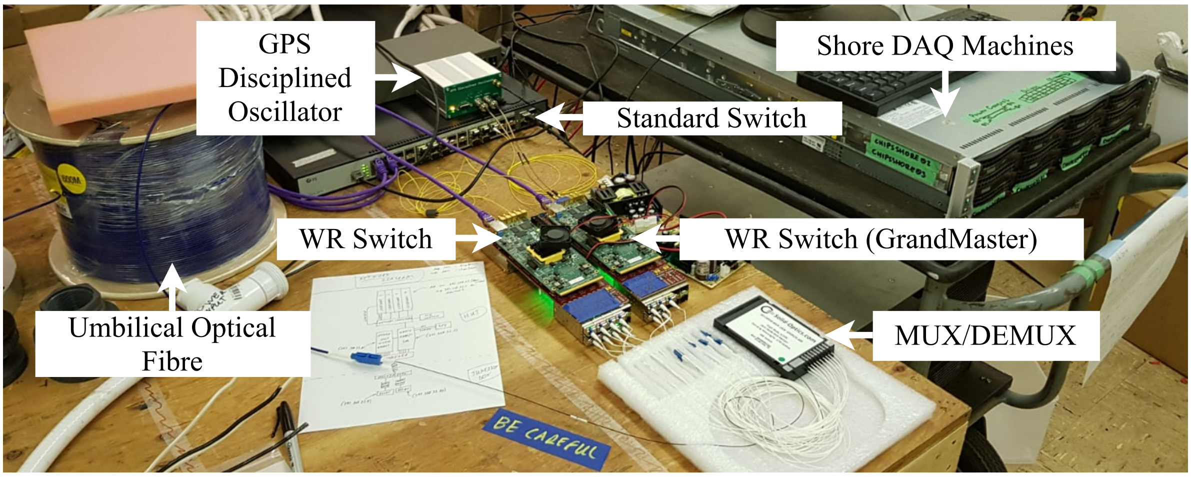


Figure 4.10: Labelled picture of the high-level onshore components of the CHIPS-5 DAQ system arranged on a table at the PolyMet mining administration building.

sufficient number of output channels, with their control electronics powered by separate AC to DC converters and each connected to one of the MUX/DEMUX networking channels via a media converter (optical fibre to RJ45). Each relay channel also has a built-in *trip gate* to immediately power-off the channel if a current surge is detected. This protection is particularly crucial for CHIPS-5 as water leaks are possible.

The contents of the DAQ electronics *Shore Hut* are shown in Figure 4.10. The single umbilical optical fibre connection passes through a MUX/DEMUX before each of the wavelength-specific channels are passed into one of two WR switches. Multiple Virtual Local Area Networks (VLANs) are configured on each switch such that for each wavelength channel only a single paired port on the other physical side of the switch carries that channels data to and from the standard networking switch (these connections are not present within Figure 4.10). Of the two WR switches, one is configured to be the GrandMaster with connections to a GPS disciplined oscillator. A single connection is also made between the WR switches for clock synchronisation.

The standard network switch provides 10 Gb connections to each of the Shore DAQ computing machines whose specific roles are detailed in Section 4.3. Each machine is also connected to a second switch providing connections to the external internet and DAQ components located at Fermilab. At Fermilab, a DAQ machine (Fermi DAQ-1) performs two main roles. Firstly, it forwards NuMI beam spill timing information from a Time Distribution Unit attached to the Main Injector clock to CHIPS-5. Secondly, it receives recorded detector data and places it into long-term storage.

An uninterruptible power supply provides power to all devices within the Shore Hut, supplying power for up to 15 minutes after a power cut, sadly quite a common occurrence. This protection gives all Shore Hut devices plenty of time to appropriately close any open data files and power down safely. The two detector umbilical power connections do not use the uninterruptible supply and instead draw power directly from the master supply.

4.3 Software and the flow of data

The software of the Chips-5 DAQ system provides three main functionalities: control of the detector instrumentation, the handling of recorded PMT hits, and the monitoring of hardware and data quality. Each of these functions are discussed within a specific subsection below alongside the corresponding software processes (applications) that perform them. Only the high-level software components (physically processed onshore) are detailed, with the low-level CLB, Beaglebone, and ţDAQ software implementations omitted for brevity. An in-depth discussion of the CLB software can be found in reference [113], while the Beaglebone and ţDAQ software implementation can be found at reference [114].

The DAQ software itself is mainly written in C++ and can be found at reference [115]. The system is comprised of multiple processes containing multiple components each (usually processed on separate threads), as is shown within Figure 4.11, expressed in terms of the flow of data between components. All DAQ processes make extensive use of multithreading and asynchronous communication, principally implemented using the low-level Boost Asio (asynchronous input/output) library [116].

All high-level processing takes place on one of three machines: Shore DAQ-1 for hit handling, Shore DAQ-2 for control and monitoring, and Fermi DAQ-1 for NuMI spill forwarding and storage. As the processing of PMT hits is the principal software task, hit handling takes place exclusively on Shore DAQ-1 to ensure maximum available processing power. Both Shore DAQ machines have a corresponding backup machine (Shore DAQ-3 and Shore DAQ-4) to take over their functions immediately in case of a fault.

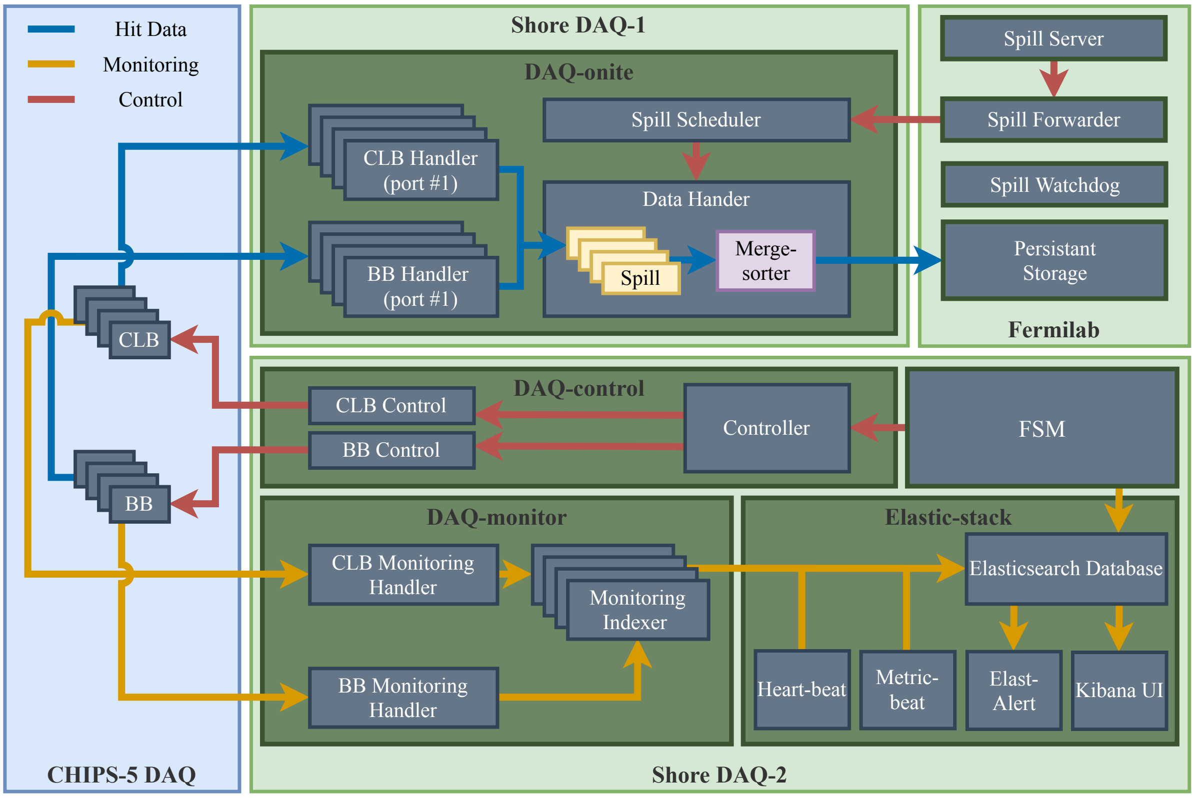


Figure 4.11: Diagram of the CHIPS-5 software system in terms of the flow of data between components. Beaglebone is abbreviated to BB.

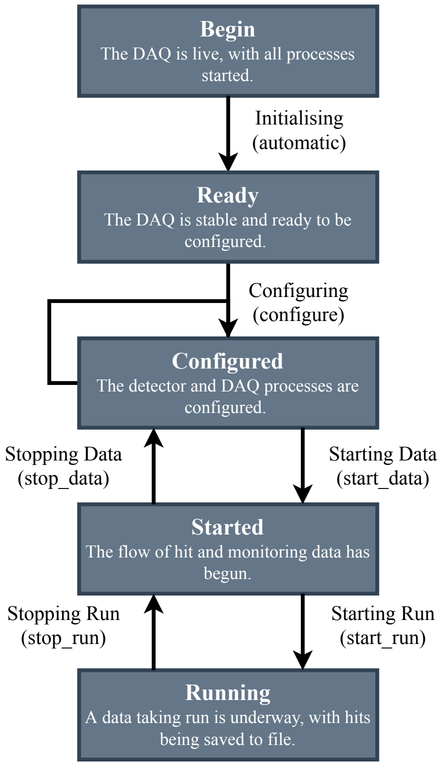


Figure 4.12: Diagram of the allowed states and transitions of the CHIPS-5 FSM. The command signal names for each transition are shown in brackets.

4.3.1 Control

All DAQ processes and low-level DAQ devices (CLBs and Beaglebones) conform to a global Finite State Machine (FSM) schema. The schema defines a finite set of states, shown in Figure 4.12, for which a DAQ process can only be in one at any given time. Specifically allowed transitions, triggered by a command signal, allow each process to move between states. During transition, the process performs the specific actions required to achieve the desired state. Every DAQ process makes its current internal state publicly known via constant publishing over a standard inter-process communication socket (Unix domain socket).

A singleton *FSM* process both issues the transition signals to and monitors the state of all DAQ processes. The FSM process defines the global state of the DAQ system, only transitioning to a new state once all children DAQ processes have reached the desired state. Currently, no control scheduling system is implemented to automatically configure the detector and conduct a given set of runs. Therefore, a simple command-line interface is used to issue state transition commands to the FSM process manually. Viewing the DAQ system as a whole, the actions performed at each state transition are as follows:

- **Initialising** occurs automatically when the DAQ system is started. During this transition all processes conduct the basic actions required for them to be able to function, such as opening inter-process communication channels, validating CLB and Beaglebone connections, and starting worker threads.

- **Configuring** occurs when the *configure* command is issued. During this transition the desired detector configuration is read from a human-readable file and all DAQ devices configured accordingly. This procedure primarily consists of setting the desired high-voltage level and ToT thresholds for each PMT via communication with all CLBs and Beaglebones.

- **Starting Data** occurs when the *start_data* command is issued. During this transition commands are sent to all CLBs and Beaglebones to start the flow of PMT hit and monitoring data.

- **Stopping Data** occurs when the *stop_data* command is issued. During this transition commands are sent to all CLBs and Beaglebones to stop the flow of PMT hit and monitoring data.

- **Starting Run** occurs when the *start_run* command is issued. During this transition the handling of PMT hit data is started, with the resulting data saved to file. This procedure includes the selection of hits according to the NuMI beam spill trigger, as detailed in Section 4.3.2.

- **Stopping Run** occurs when the *stop_run* command is issued. During this transition the handling of PMT hit data is stopped, and the run data file closed.

For control of the detector devices, a *DAQ-control* process issues commands to and monitors the state of all CLBs and Beaglebones. For communication, all messages are exchanged using standard TCP (Transmission Control Protocol) packets over ethernet. A DHCP server on Shore DAQ-2 assigns predefined static IP addresses to all devices to allow this to happen easily.

A single DAQ-control controller thread receives commands from the FSM process and instructs separate CLB and Beaglebone controller threads to issue the appropriate implementation-specific ‘slow-control’ commands for the given transition. A simple retry mechanism ensures that multiple attempts are made to achieve the desired state on each device. However, a CLB or Beaglebone is dropped after a maximum number of attempts have been made in order not to block the global DAQ transition due to a single faulty device.

4.3.2 Hit handling

As is done within the KM3NeT experiment, an ‘all-data-to-shore’ approach to PMT hit acquisition is taken, where all recorded hit data is sent via the WR network to the machines onshore. No triggering takes place within the detector. This is conducted by collecting all PMT hits within 10 ms long time windows. At the end of each window, collected hits are packaged along with a header into UDP (User Datagram Protocol) packets and sent to shore. This process occurs on both the CLBs and Beaglebones when in either the *Started* or *Running* FSM state.

Jumbo UDP packets of maximum size 9000 bytes are used instead of typical 1500 byte packets. This is done to both reduce the proportion of bandwidth taken up by headers, and reduce the total network traffic (number of packets), leading to a decrease in the overall proportion of dropped packets. Although UDP does not contain the error-checking and correction features built into TCP, the proportion of lost packets is negligible and the increased bandwidth desirable.

PMT hit packets are handled by the *DAQ-onite* process, named after the Taconite ore that used to be extracted from the Wentworth 2W pit. The primary function of DAQ-onite is to store incoming PMT hits whose timestamp matches a NuMI beam spill trigger window and discard those that do not. This is complicated by the fact that the exact NuMI spill time is not known in advance.

Whenever the Main Injector at Fermilab releases a spill of protons for the NuMI beam, the accurate timestamp of this event is only published as it happens. Therefore, by the time the Fermilab based *Spill Server* and *Spill Forwarder* processes have received and forwarded the signal to the Shore DAQ-1 machine running DAQ-onite, approximately 0.5 seconds have passed since the neutrino spill passed CHIPS-5.

To counter this problem DAQ-onite implements a *Spill Scheduler* prediction mechanism which uses the periodicity of the spill trigger signal to continually predict the time of multiple spill windows (each 100 ms long) in advance. As PMT hits are processed, they are either attached to a matching predicted *Spill* or immediately discarded. This process is conducted by a series of CLB and Beaglebone specific handling threads which decode the raw PMT hit packets across a range of input ports to increase throughput.

Once the accurate timestamp for a beam spill has been received and a configurable period has passed to catch any late-arriving hits, the spill window is closed. Once closed, the attached PMT hits are combined and sorted using a *Merge-sorter* algorithm before being saved to file. In addition to the PMT hit data, each data file includes context metadata and flags indicating data faults. All new data files are periodically forwarded to long-term storage at Fermilab for further analysis.

During the initial CHIPS-5 DAQ commissioning in late 2019, a very high average PMT hit rate of 100 KHz was observed across all PMTs. This was expected due to both the extended period the PMTs had been exposed to direct sunlight during construction, and the presence of a light leak through a tear in the detector liner. Somewhat beneficially, this allowed for load testing of the DAQ hit handling system with to a peak WR network throughput of 5.66 Gb per second of hit data. DAQ-onite was found to behave as expected, with only negligible data loss.

4.3.3 Monitoring

CHIPS-5 DAQ monitoring is comprised of two principal functions: the monitoring of hardware components and the monitoring of PMT hit data quality. The first of these aims to check that all hardware is alive and performing efficiently without error, whilst the second ensures that the recorded hit data is consistent with what is expected. Both of these functions are conducted within CHIPS-5 using an implementation built around a central Elasticsearch database [117].

Elasticsearch is an open-source JSON (JavaScript Object Notation) document based search engine and database. One of its primary applications is for logging, infrastructure observation, and application performance monitoring, making it a perfect fit for use here. Elasticsearch sits within the broader range of open source tools called the Elastic Stack [118], including those to both ingest data and produce visualisations. All Elasticsearch data is stored within collections of related JSON documents known as *indices*, which share a common set of key to value pairs.

Monitoring data is continually ingested into Elasticsearch from a wide range of sources across the CHIPS-5 DAQ system:

- *DAQ-monitor* receives and handles the monitoring data produced by the CLBs and Beaglebones. At the end of every 10 ms long PMT hit data taking window both CLBs and Beaglebones produce UDP monitoring data packets. Included within the packets are general POM wide status indicators such as temperate and humidity, as well as PMT specific hit rate values. DAQ-monitor decodes this data and continuously pushes it to Elasticsearch using a set of asynchronous *indexing* threads.
- The current FSM state of all the DAQ software processes in addition to the global state is frequently pushed to Elasticsearch by the FSM process.
- The status of the Fermilab based Spill Server and Spill Forwarder alongside general NuMI beam metrics are forwarded to Elasticsearch using the *Spill Watchdog* process operating on the Fermi DAQ-1 machine.
- All DAQ network devices are constantly 'pinged' to check they are alive by the *Heartbeat* process which forwards this data to Elasticsearch.
- Shore DAQ machine system metrics (CPU usage for example) as well as system logs and data input and output rates are frequently forwarded to Elasticsearch from instances of the *Metricbeat* process running on each DAQ machine.
- Any logs produced by the DAQ software are asynchronously logged to Elasticsearch.

The Kibana browser-based user interface application is used to visualise the monitoring data stored within Elasticsearch [119]. A series of preconfigured dashboards, one of which is shown in Figure 4.13, summarise the current and historical monitoring data in an easy to digest fashion. The primary advantage of using Kibana over the more typical bespoke implementations used by HEP experiments is that it is both easily configurable by non-experts and only requires a browser and an internet connection to use from anywhere (in addition to authentication).

Furthermore, an alerting system using the ElastAlert framework [120] provides automatic notification of events for human attention. The ElastAlert process continuously checks the Elasticsearch monitoring data for any breaches of predefined rules (x events in y time, for example). Upon the occurrence of a rule trigger, multiple forms of notification (email and Slack messages) are sent to appropriate recipients for further analysis.

Not only is the CHIPS-5 monitoring system extensible and easy to use it also has the advantage of leveraging the efforts of an active open source community. With no additional

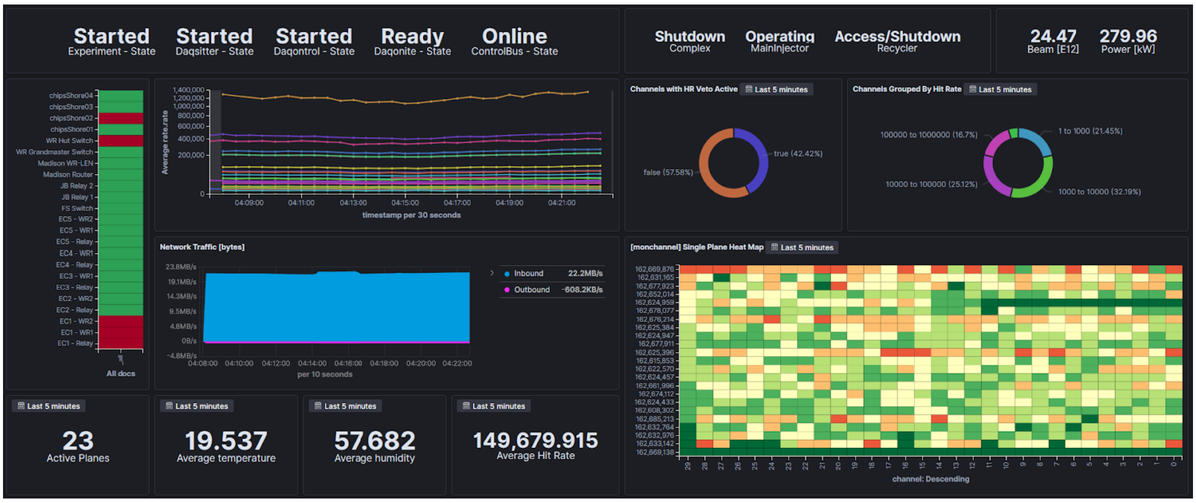


Figure 4.13: Screenshot of a Kibana monitoring dashboard for CHIPS-5. The dashboard pulls live monitoring data from the Elasticsearch database for its visualisations. The status of the various DAQ software components are shown in the top left above an average hit rate plot for each POM and the colour coded status indicators for various DAQ devices. A plot of the network data rates is also shown alongside global average indicators. The right hand side of the dashboard contains visualisations to monitor the overall status and specific hit rates of PMTs.

effort by the members of the CHIPS collaboration, the monitoring implementation will continue to improve with time as updates are made to the underlying Elastic Stack software. This is likely to include both performance improvements alongside new ways of quickly ingesting data and creating visualisations. The monitoring approach firmly achieves one of the driving principles of the CHIPS project, using commercially available, cheap (free in this case) components, wherever possible.

Chapter 5

Convolutional neural networks for CHIPS

For the majority of HEP experiments, event analysis entails the separation of signal from background, the identification of particle types, the discovery of spatial properties, and the estimation of energies. The same is true for CHIPS detectors, with the primary aims being the selection of appeared CC ν_e signal events from a sizeable background, and the estimation of associated neutrino energies.

For this purpose, the CHIPS project has so far relied on a likelihood-based reconstruction algorithm and a simple classification neural network driven by hand-engineered features. Both suffer from only considering what has been implemented in software and consequently what features are explicitly extracted from the data. This restriction makes them prone to ignoring the wide range of edge cases not contained within the bulk of neutrino events and unable to use all the underlying informative features of the data.

The work outlined in this chapter presents a replacement event analysis methodology for CHIPS. As with any implementation, this new methodology comes with its own limitations and difficulties; however, it is found to provide a significant performance improvement. Three *Convolutional Neural Networks* (CNNs) [121], a type of *deep learning* [122] neural network have been developed to achieve the primary aims outlined above, amongst others. One for cosmic muon rejection, one for beam event classification, and one for neutrino energy estimation. For evaluation purposes, only the implementation as applied to the CHIPS-5 detector module is considered in this work.

After mentioning previous implementations of deep learning for neutrino experiments, a description of the current (standard) techniques are given, before the theoretical

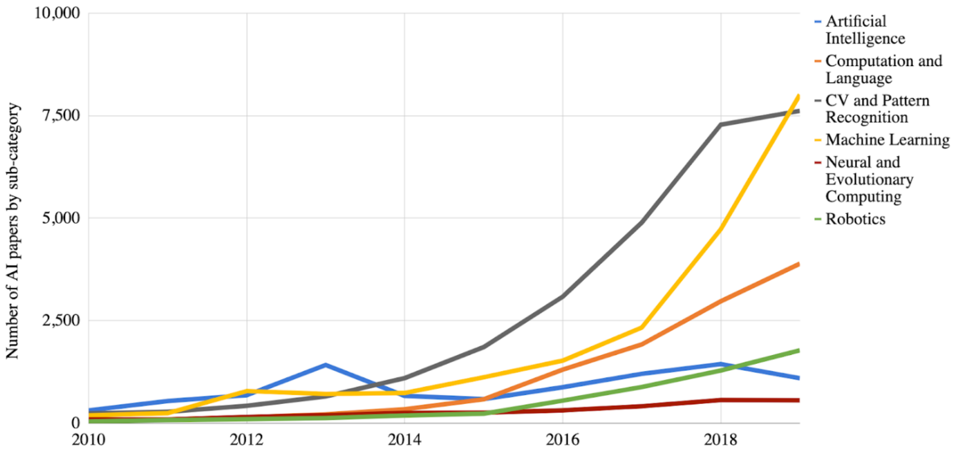


Figure 5.1: The number of artificial intelligence papers submitted to arXiv, broken down by sub-category. Note the particularly large increase in Computer Vision (CV) and pattern recognition papers. Figure taken from reference [124].

background of CNNs is outlined. The baseline implementation for CHIPS is then described, followed by the specific implementations for each of the three networks. A comprehensive evaluation of the trained networks is presented separately in the following chapter (Chapter 6).

5.1 Previous applications of deep learning for neutrino experiments

Over the last few years, neutrino experiments have started to adopt deep learning techniques for a range of event analysis tasks [123]. This trend has closely followed the general explosion of interest in the field amongst the global research community, especially within the sub-field of computer vision, as can be seen in Figure 5.1.

In 2016 the NOvA experiment applied a CNN to the task of classifying the interaction type of events within their sampling calorimeter detector [125]. Two views of raw detector data were used as input to train a network based on the popular GoogLeNet architecture [126] (discussed in Section 5.3.3). Further NOvA iterations have since been applied to both the classification of individual energy deposit clusters [127] and ν_e and e^- energy reconstruction [128].

CNNs have also been applied to liquid argon time-projection chambers. The MicroBooNE experiment [129] has shown that in addition to classification tasks, the spatial localisation of single particles within events is possible [130]. Furthermore, the DUNE collaboration has designed a network to output both the interaction class and counts of different particle types within an event [68, 131]. This approach is called *multi-task* learning and is discussed in detail within Section 5.4.3.

Applications to water Cherenkov detectors have also been made by both the Daya Bay experiment [132] and the KM3NeT/ORCA collaboration [133]. Furthermore, a type of CNN known as a *variational autoencoder* has been shown to approximate the distribution of simulated water Cherenkov data well [134]. If further such studies prove successful, this could allow for training on 'real' data to mitigate experimental uncertainties and vastly increase the speed of simulated data generation.

5.2 Standard event reconstruction and classification

It is essential to outline the standard event reconstruction and classification methods used by the CHIPS project until now. This is for two reasons. Firstly, to highlight their main weaknesses as motivation for the new CNN approach. Secondly, to provide context for the performance comparisons made in Chapter 6.

A likelihood method based on that implemented by MiniBooNE [135] is used for event reconstruction, while a simple neural network built using the TMVA package [136] is used for event classification. Both methods are representative of the mainstream approach used by the majority of water Cherenkov neutrino experiments for event analysis. A prime example is the fiTQun algorithm developed for the Super-Kamiokande detector, used for both atmospheric [137] and T2K [138] analyses.

5.2.1 Likelihood-based reconstruction

The event reconstruction methodology is simple in theory: for a given set of hypothesised charged particle tracks, the number of Cherenkov photoelectrons and the time at which the first of these is recorded for each PMT in the detector is predicted. By comparing this prediction with the measured hit charges and times the likelihood that the given track hypothesis produced the measured signals can be calculated. The parameters that describe the hypothesised tracks are then varied until the negative logarithm of the likelihood is minimised, identifying the best-fit parameters. A brief description of the full procedure is given below, however, for a detailed description see reference [97] and reference [85]. The full C++ software implementation can also be found in the repository at reference [139].

Seeding

The first stage of event reconstruction is the effective *seeding* of tracks that are then used in the full likelihood fit. The seeding methods aim to provide a good starting point for the minimisation, both to increase the efficiency of finding the optimal track parameters and also to avoid a false local minimum from being returned.

Firstly, the PMT hits are sliced in both space and time. Gaps in the time ordering of hits are used to separate the event into time slices. Each of these slices then undergoes basic clustering to remove outlying hits and ensure only the dominant collections of hits are considered. This procedure involves removing all hits that fall outside of clusters containing at least 50 hits. Clusters are generated by finding all hits that fall within 2 m of each other. Each cleaned slice is then run through simple geometric vertex finding algorithms to estimate the interaction position and time, in addition to the initial direction of the track.

A circular *Hough transform* algorithm, traditionally used for water Cherenkov ring finding is then applied [140]. As output, the voting-based transformation produces a space within which rings of PMT hits exist as peaks. The track direction values are further refined using this space before a search for secondary peaks is carried out to indicate if multiple particles are likely to be involved. This process results in a list of seeds, each with a score related directly to the height of the associated peak in Hough transform space.

Likelihood fit

Each particle track in a fit hypothesis comprises a vector of parameters $\vec{x}$, containing the following:

- the track interaction vertex position (x_0, y_0, z_0) and time (t_0);
- the initial track direction (d_θ, d_ϕ);
- the initial kinetic energy of the particle (E); and

- the particle type (muon, electron or photon).

For a photon hypothesis the distance between the interaction vertex and the beginning of the electromagnetic shower is also included as a parameter.

The hypothesised tracks are then initialised using the list of seeds found in the seeding procedure in descending order of Hough peak height score. As the seeding algorithms do not estimate the particle energy, a default value equal to the average particle energy observed within a sample run of the simulation is assigned. Additionally, constraints can be placed on a multi-track hypothesis to reduce the number of free parameters.

As an example, consider the NC π^0 case, where a multi-track two-photon hypothesis is used. Firstly, the initial parameters for the two photons are assigned from the two highest-scoring seeds from the seeding procedure. Secondly, the vertex position for both tracks is constrained to remain the same, and the directions and energies are set to be constrained by the invariant mass of the π^0.

In its simplest form the likelihood $\mathcal{L}(\vec{x})$, is a simple product of two terms:

$$\mathcal{L}(\vec{x}) = \mathcal{L}_{unhit}(\vec{x})\mathcal{L}_{hit}(\vec{x}) = \prod_{unhit} P_{unhit}(\vec{x}) \prod_{hit} P_{charge}(\vec{x}) P_{time}(\vec{x}), \tag{5.1}$$

where the first (*unhit*) term gives the likelihood that the hypothesis $\vec{x}$ will not predict a hit on the PMTs that do not have a measured hit, and the second (*hit*) term gives the likelihood that $\vec{x}$ produces the observed photoelectrons and hit times on the measured hit PMTs.

By considering the negative logarithm of the likelihood the computation can be simplified into a sum of logarithms over the PMTs, such that

$$-\log \mathcal{L}(\vec{x}) = -\sum_{unhit} \log(P_{unhit}(\vec{x})) - \sum_{hit} \log(P_{charge}(\vec{x})) - \sum_{hit} \log(P_{time}(\vec{x})). \tag{5.2}$$

This form has the effect of separating the charge (number of photoelectrons) and time prediction components, which can then be dealt with separately computationally. In the actual likelihood calculation the $P_{unhit}(\vec{x})$ and $P_{charge}(\vec{x})$ components are combined, where the probability of an unhit PMT is treated as a PMT with an observed charge equal to zero.

The Minuit2 algorithm contained within the ROOT software package [141] is used for the minimisation process. At each iteration, the charge and hit time predictions are made, and the negative logarithm of the likelihood is calculated. The track parameters are then varied to minimise the likelihood before the next iteration begins. Through a series of stages, each fixing and freeing specific parameters, the minimisation process converges to the best-fit parameters for the given hypothesis. This procedure typically takes two minutes on a standard batch farm computing node.

Downsides

The charge and hit time predictions and their associated likelihood contributions depend on a series of low-level inputs. Generally, these inputs describe how Cherenkov light is emitted from specific particles and how it propagates through the detector to be detected by the PMTs. Examples of these inputs include:

- the number of Cherenkov photons emitted by a particle of a specific type and energy;
- the fraction of Cherenkov light emitted at each step along a specific particle's track length;
- the angular distribution of Cherenkov photon emission for each type of particle;
- the survival probability of photons within the detector medium as a function of distance and energy;
- a detailed description of the PMT positions and directions;
- the angular efficiency of each PMT relative to the incident photon angle; and
- the probability of a measured charge given the predicted number of photoelectrons (see Figure 3.25b).

The above list demonstrates a fundamental problem with the likelihood-based approach. Namely, it is dependent on a finite list of inputs that must be implemented in software. If a physical process is overlooked, then the algorithm has access to a reduced amount of information with which to predict the PMT hit charges and times, impacting the performance of the best-fit track parameters. Additionally, the large amount of effort required to implement an input correctly and determine how it is used can make a likelihood-based approach time-consuming.

Moreover, the likelihood-based approach requires a predefined track hypothesis per fit. This restriction is at odds with the broad array of neutrino events expected within CHIPS detector modules. Many possible combinations of final state particles are possible, making the implementation of an approach where all possible events are considered, challenging. For example, the very similar Super-Kamiokande fiTQun algorithm attempts multiple fits for each event, sequentially adding charged particles to the hypothesis until the best-fit is found. This technique not only vastly increases the time required to analyse a single event, but can never be fully rigorous as it still ignores some scenarios.

5.2.2 Event classification

As the standard event reconstruction is based on the calculation of a likelihood (analogous to a goodness-of-fit), the likelihood ratio between different hypotheses can be used for event classification tasks. Additional hand-engineered features derived from the reconstruction outputs are also found to have power in classifying the event type.

Two simple neural networks are used, the first for CC ν_e - CC ν_μ separation and the second for CC ν_e - NC separation. Both contain a single hidden layer (as described in the following Section 5.3) with the number of neurons equal to the number of input parameters plus five. Output variables from both a single electron track and single muon track hypothesis fit are used for both networks, including:

- the $\Delta \log \mathcal{L}$ between e and μ hypothesis for both time and charge components;
- the total number of hit PMTs (N_{hits}) and total collected charge;
- $\frac{\Delta \log \mathcal{L}_{charge}}{N_{hits}}$;
- the fraction of hits inside, within, and outside the ring for both the e and μ hypotheses;
- the fraction of predicted charge outside the ring for both the e and μ hypotheses;
- the ratio of the total predicted charged to the total measured charge for both the e and μ hypothesis;
- the ratio of the reconstructed energy to the total measured charge for both the e and μ hypotheses;
- the reconstructed track direction under the e hypothesis;

- the fraction of hits in the downstream half of the detector;
- the number of seeds generated by the Hough transform seeding algorithm; and
- the peak height score of the first and last seeds found by the Hough transform seeding algorithm.

A sample of CC ν_e and CC ν_μ beam events characteristic of those expected to be seen within CHIPS-5 are used to train the first classifier, and a corresponding sample of CC ν_e and NC events for the second. The output values from both networks can then be used to select CC ν_e events from the background. Note that only the selection of CC ν_e events has been implemented, no CC ν_μ selection has been developed.

The principal limitation of this approach is that the input features are restricted to those that have been imagined (requiring extensive domain knowledge) and then implemented in software. The current list is undoubtedly non-exhaustive of all the possible variables and combinations of variables that can, in theory, be used for discrimination between events. Additionally, any systematic uncertainties in the likelihood-based reconstruction and, therefore, input variables to the neural networks, can lead to incorrect classification of events.

5.3 The theory of neural networks

There are many machine learning techniques: linear regression, logistic regression, k-nearest neighbours, decision trees, random forests, support vector machines, amongst others, all of which learn to make predictions about data. However, none has been as successful, especially in recent years, as the deep neural network. As both the size of datasets and the amount of available computing power has increased, deep neural networks have proved incredibly powerful for many tasks, as they are well suited to this paradigm.

Here we discuss the application of neural networks for *supervised learning*, one of two broad machine learning categories concerned with using labelled example data to train algorithms. The other broad category of *unsupervised learning*, where the properties of the dataset are inferred without labelled data is not discussed, however, will be used for network *explainability* in Section 6.2.

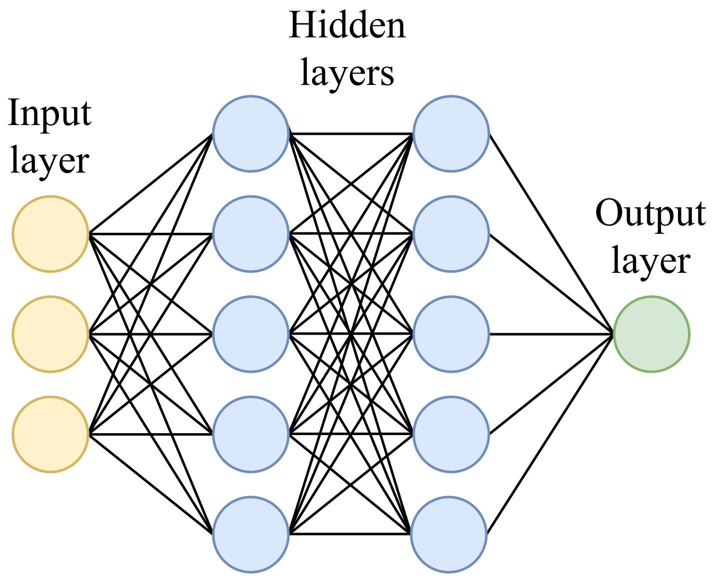


Figure 5.2: Illustration of a simple neural network. There is a single input layer (yellow), two hidden layers (blue), and an output layer (green). Each node corresponds to a neuron except for the input layer.

5.3.1 Neural network basics

A neural network is a type of algorithm inspired by the repeating cell structure of neurons within our brains. The basic building block of a neural network is a *neuron*, which takes a vector of k inputs $\vec{x} = (x_1, x_2, \ldots, x_k)$ and outputs a scalar $a(\vec{x})$. Individual neurons are arranged into layers, with the input of one layer being the output from the previous layer. The first layer is commonly referred to as the *input layer*, the middle layers as *hidden layers*, and the final layer as the *output layer*, as illustrated in Figure 5.2. In general, this simple neural network structure is referred to as *fully-connected*, as all the neurons in each layer have connections to all the neurons in the previous and following layers.

Input variables (traditionally hand-engineered features extracted from the raw data) are passed into the network via the input layer. Any number of hidden layers containing any number of neurons can then follow. The neurons contained within these layers are trained so that collectively their $a(\vec{x})$ functions solve the task at hand. For a regression task, the output layer returns a continuous decimal value. Conversely, for a classification task, a probability value between zero and one is output for each class. The forward passing of information from one layer to the next is why neural networks can also be referred to as *feed-forward graphs*.

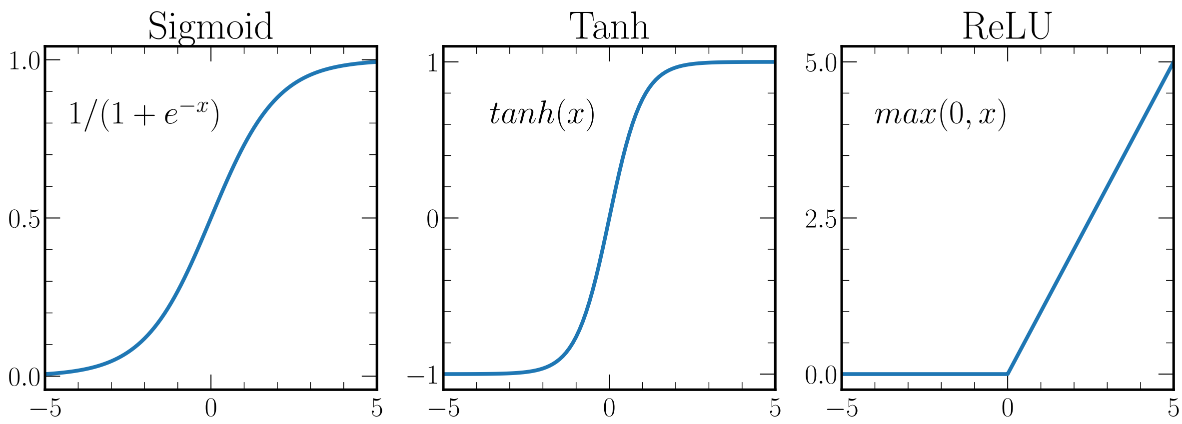


Figure 5.3: Common non-linear activation functions used for the neurons within neural networks.

For a neuron i, $a_i(\vec{x})$ can be decomposed into a neuron specific linear operation, followed by a non-linear operation which is the same across all neurons. The linear operation consists of the dot product of the input vector $\vec{x}$ with a vector of weights $\vec{w}^{(i)} = (w_1^{(i)}, w_2^{(i)}, \ldots, w_k^{(i)})$, plus a bias term $b^{(i)}$:

$$z^{(i)} = \vec{w}^{(i)} \cdot \vec{x} + b^{(i)}. \tag{5.3}$$

After applying the non-linear operation σ_i, commonly referred to as the *activation function*, the final neuron output can be written as

$$a_i(\vec{x}) = \sigma_i(z^{(i)}). \tag{5.4}$$

Traditionally, a step-function (for networks called *perceptrons*) was used for the activation function. However, as is shown in Section 5.3.2, a non-zero gradient (only valid at $x = 0$ for a step-function) is required for the practical training of neural networks. In fact, the choice of activation function can greatly affect how the network trains and performs.

Therefore, common activation function choices have been the *hyperbolic tangent* and the *sigmoid* functions, primarily because they are bounded and differentiable at all points. Recently, the *ReLU* and other similar functions have become popular, mainly due to their avoidance of *vanishing* gradients which occur when the tanh and sigmoid functions become saturated at large values of x. All of these functions are shown in Figure 5.3 for reference.

5.3.2 Training neural networks

Supervised neural network training uses labelled example data to iteratively find the optimal weights and biases (the network parameters) to maximise output performance. A *loss function* $E(\vec{w})$, is defined to quantify the difference between the network output and truth label for each example, where $\vec{w}$ is the vector of network parameters. For a given input example $(\vec{x}_i, y_i)$, with $\vec{x}_i$ being the input parameters and y_i the known truth label, the network generates an output $\hat{y}_i(\vec{w})$. Using this notation, we can construct loss functions suitable for different tasks.

In the case of simple binary classification the most commonly used function is the *binary cross-entropy*, where

$$E(\vec{w}) = -\sum_{i=1}^{n} y_i \log \hat{y}_i(\vec{w}) + (1 - y_i) \log[1 - \hat{y}_i(\vec{w})], \tag{5.5}$$

with the number of examples given by n. For a classification task where the number of classes is greater than two y can instead take on M values. In this case we redefine each example so that y is instead a vector y_{im}, such that

$$y_{im} = \begin{cases} 1 & \text{if } y_i = m, \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

This is commonly named a *one-hot* vector. The cross-entropy then becomes the *categorical cross-entropy*, given by

$$E(\vec{w}) = -\sum_{i=1}^{n} \sum_{m=0}^{M-1} y_{im} \log \hat{y}_{im}(\vec{w}) + (1 - y_{im}) \log[1 - \hat{y}_{im}(\vec{w})]. \tag{5.7}$$

For a regression task predicting a continuous output variable, the *mean-squared error* is most often used as the loss function, with

$$E(\vec{w}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i(\vec{w}))^2. \tag{5.8}$$

To find the optimal network parameters for the given task the loss function output (the *loss*) is iteratively minimised until it converges to the minimum (or in reality a local minimum that performs well). This is done by updating the network parameters at each

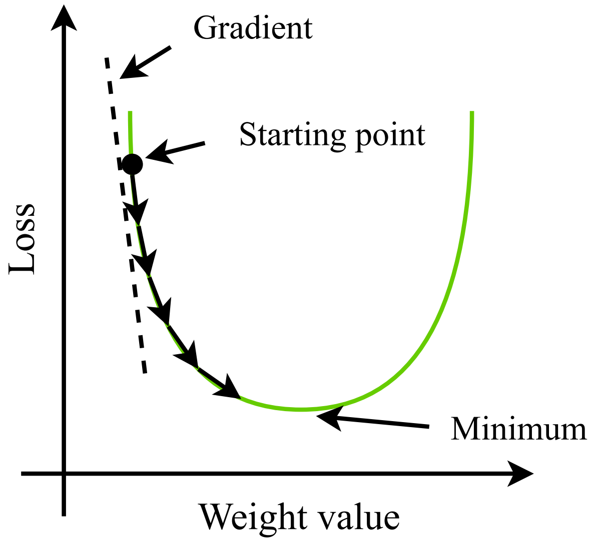


Figure 5.4: Illustrative diagram of the gradient descent procedure. Shown is the case for a loss function dependent on a single weight.

iteration t, to move in the direction of the loss gradient, using the update rule

$$\vec{w}_{t+1} = \vec{w}_t - \eta_t \nabla_{\vec{w}} E(\vec{w}), \tag{5.9}$$

where η_t is the *learning rate* which determines the size of the step taken at each iteration. This methodology is known as *gradient descent* and is illustrated in Figure 5.4.

Therefore, to use gradient descent, we require that the gradient of the loss function with respect to the parameters of the network can be calculated. Doing this for each parameter at every iteration would render neural networks impossible to train due to the vast computational requirements. Instead, an innovative application of the chain rule, in an algorithm called *backpropagation* is used [142]. Here we follow the derivation of the four main backpropagation equations given in reference [143].

For a network containing L layers, we can index the individual layers using $l = 1, \ldots, L$. The weight associated with the connection between the k-th neuron in layer $l-1$ and the j-th neuron in layer l can be denoted as w^l_{jk}, with the bias of the layer l neuron correspondingly written as b^l_j. The activation of the j-th neuron in layer l is then related to the outputs from the previous layer by

$$a^l_j = \sigma(z^l_j) = \sigma\left(\sum_k w^l_{jk} a^{l-1}_k + b^l_j\right), \tag{5.10}$$

where σ is the non-linear activation function.

The change in the loss function with respect to the linear weighted sum z_j^L of the j-th neuron in the last layer L, defines the error Δ_j^L, such that

$$\Delta_j^L = \frac{\partial E}{\partial z_j^L}. \tag{5.11}$$

Using the chain rule and the fact that $a_j^L = \sigma(z_j^L)$, the error is found to be equivalent to

$$\Delta_j^L = \frac{\partial E}{\partial a_j^L} \sigma'(z_j^L), \tag{5.12}$$

where $\sigma'(z_j^L)$ denotes the derivative of the non-linear activation function at z_j^L. This (equation (5.12)) is the first of the backpropagation equations.

The error Δ_j^l, of the j-th neuron in layer l can be expressed in terms of the error in the following layer $l+1$, by using

$$\begin{aligned}
\Delta_j^l &= \frac{\partial E}{\partial z_j^l}, \\
&= \sum_k \frac{\partial E}{\partial z_k^{l+1}} \frac{\partial z_k^{l+1}}{\partial z_j^l}, \\
&= \sum_k \Delta_k^{l+1} \frac{\partial z_k^{l+1}}{\partial z_j^l}, \\
&= \left(\sum_k \Delta_k^{l+1} w_{kj}^{l+1} \right) \sigma'(z_j^l).
\end{aligned} \tag{5.13}$$

The chain rule has again been used in addition to the fact that

$$z_k^{l+1} = \sum_j w_{kj}^{l+1} \sigma(z_j^l) + b_k^{l+1}, \tag{5.14}$$

to give the second of the backpropagation equations (equation (5.13)).

As $\partial b_j^l / \partial z_j^l = 1$, the error can also be viewed as the partial derivative of the loss function with respect to the bias, such that

$$\Delta_j^l = \frac{\partial E}{\partial z_j^l} = \frac{\partial E}{\partial b_j^l} \frac{\partial b_j^l}{\partial z_j^l} = \frac{\partial E}{\partial b_j^l}, \tag{5.15}$$

giving the third of the backpropagation equations (equation (5.15)).

The final backpropagation equation is given by the differential of the cost function with respect to the weight w_{jk}^l, which can be written as

$$\frac{\partial E}{\partial w_{jk}^l} = \frac{\partial E}{\partial z_j^l}\frac{\partial z_j^l}{\partial w_{jk}^l} = \Delta_j^l a_k^{l-1}. \tag{5.16}$$

The full backpropagation algorithm then proceeds as follows:

1. After calculating the activations a_j^1 for all neurons in the input layer, use the feed-forward architecture of the network to calculate all the activations at every layer using equation (5.10).
2. Use equation (5.12) to calculate the errors of the last layer neurons, requiring both the derivatives of the loss and activation functions.
3. Use equation (5.13) to 'backpropagate' the error through the network from the last layer to the input layer, calculating all Δ_j^l values.
4. Calculate the gradient of the loss function for all the weights and biases using equation (5.15) and equation (5.16).

A single activation finding *forward pass* followed by a single error propagating *backward pass* is all that's required to calculate the gradients for all weights and biases within the network. This incredibly efficient procedure allows for the use of gradient descent when training neural networks.

5.3.3 Convolutional neural networks

The broad category of deep learning covers multiple neural network techniques spanning a range of application fields such as computer vision, speech recognition, and natural language processing. By stacking many layers on top of each other to form a 'deep' network, these methods offer increased problem solving capacity by allowing higher-order non-linear functions to form throughout the network. As a direct consequence, instead of requiring hand-engineered features as input, deep networks can learn to extract the most powerful features for a given task from raw data. Here we outline just the specific deep learning technique used in this work; the CNN.

CNN operations

At their core, CNNs make use of a mathematical operation called *convolution*, which either entirely or in part replaces the simple vector multiplication seen in the fully-connected networks of Section 5.3.1. This difference makes CNNs incredibly powerful for applications with grid-like input data such as computer vision tasks.

Using standard CNN terminology, the discrete convolution between the *input* x, and the *kernel* w, is given by

$$f_i = (x * w)_i = \sum_{j=-\infty}^{\infty} x_j w_{i-j}, \tag{5.17}$$

where i and j denote the set of discrete values and f is commonly referred to as the *feature map*. In typical applications, the input is a two-dimensional array X. Therefore, both the kernel W, and the resulting feature map F, also become two-dimensional. In this case the convolution operation becomes

$$F_i = (X * W)_{i,j} = \sum_m \sum_n X_{i+m,j+n} W_{m,n}, \tag{5.18}$$

where the infinite sum in equation (5.17) has been replaced with a discrete sum over two-dimensional elements. Analogous to the simple neural network weights $\vec{w}$, first described in equation (5.3), the elements of $W_{m,n}$ are trained to minimise the loss, with the output feature maps passed through a non-linear activation function at each layer.

To illustrate the convolution operation, Figure 5.5 displays examples of a 4×4 input grid and a 3×3 kernel. The output feature map is generated by sliding the kernel across both dimensions of the input grid, summing the products of all associated elements at each step according to equation (5.18), as shown in Figure 5.6.

Two additional parameters impacting the feature map output size are introduced in Figure 5.6. The *stride* and *padding*. The stride S, governs how far the kernel moves at each step while the padding P, decides how the input grid is padded with zeros around its border. If L is the size of the input (both height and width) and K is the kernel size, the output feature map size O, is given by

$$O = \frac{(L - K + 2P)}{S} + 1. \tag{5.19}$$

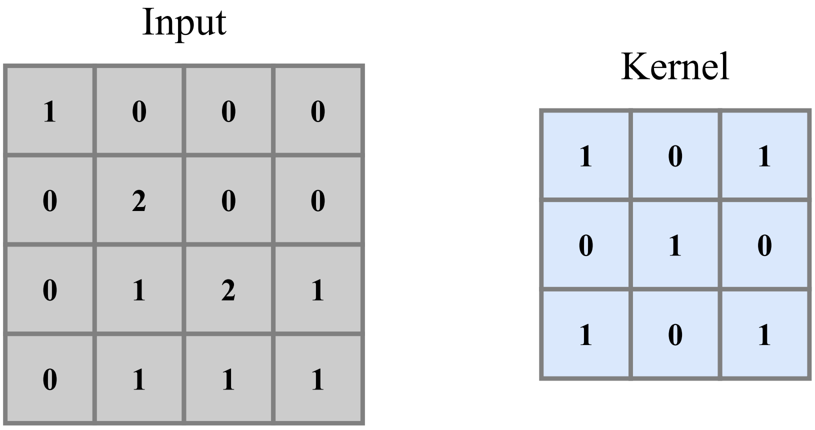


Figure 5.5: Example of an input grid (left) and kernel (right). The specific kernel shown is sensitive to x-shaped features

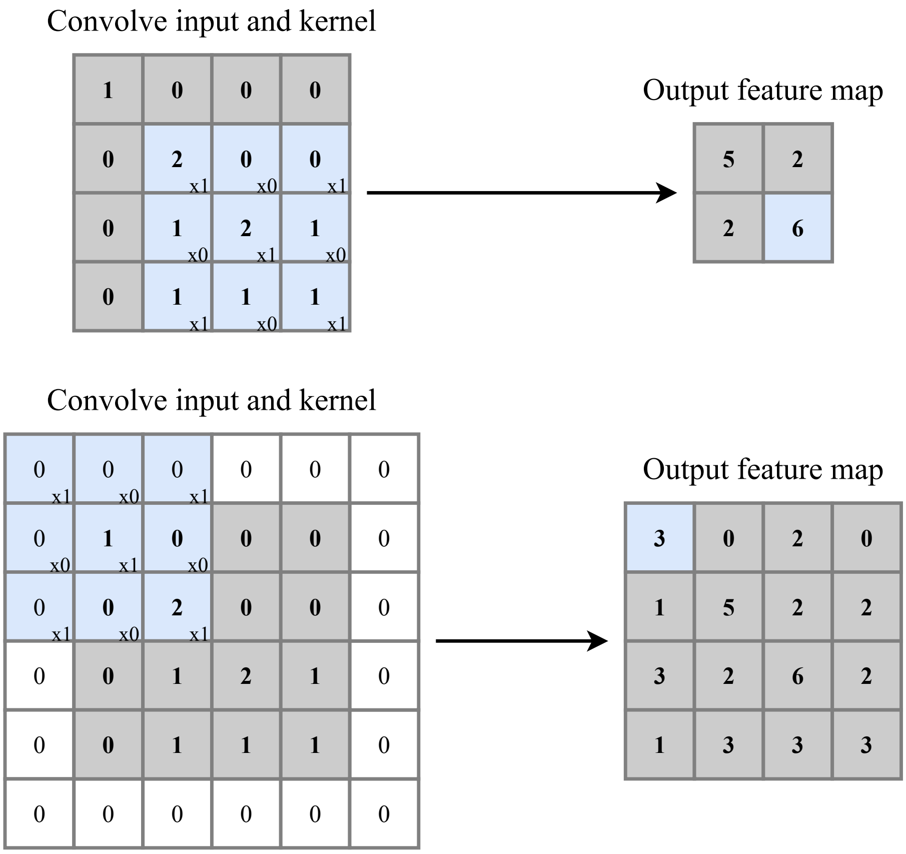


Figure 5.6: Example of a stride = 1 convolution operation involving the input grid and kernel from Figure 5.5. Both the operation for the case of *valid* (top) and *same* (bottom) padding are shown. The blue square in the top-left of the output feature maps indicates the output generated from the specific operation shown on the left, also in blue.

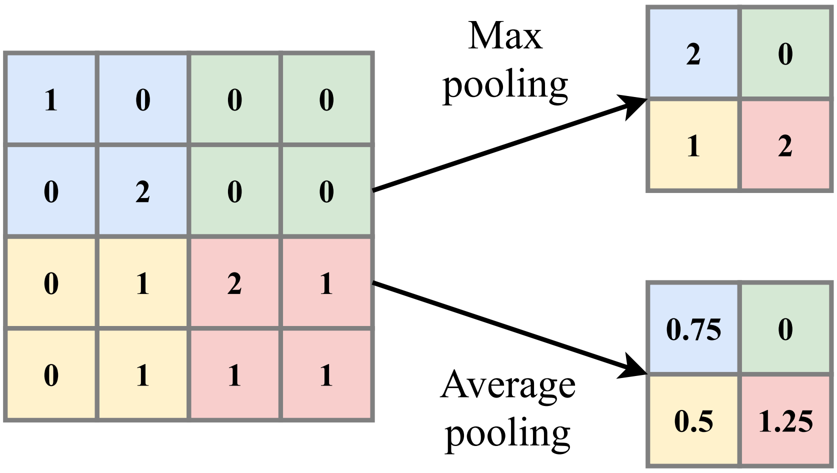


Figure 5.7: Example of both a max and average 2×2 pooling operation with stride = 2.

The other essential operation used within CNNs is pooling. Pooling layers coarse-grain the spatial information of the input to reduce the number of network parameters. *Max pooling* or *average pooling* are the two common ways this is achieved. In both cases, the input is first divided into rectangular regions, and then either the maximum or average value of the region is returned as output, for max or average pooling, respectively. Both pooling procedures are illustrated in Figure 5.7.

Taking inspiration from how neurons behave in the visual cortex of animals [144], small kernels are typically used to only scan over a small patch of the input at a time. Combined with the loss of absolute position information from pooling, a key feature of CNNs is highlighted. They exhibit translational invariance and respect the local structure contained within the input. In simpler terms, they do not care wherein the input image a particular feature exists, just that it exists.

CNN architectures

In 2012 the AlexNet CNN lowered the error rate of the ubiquitous ImageNet classification task [145] from 28% to 16% [146]. Since this breakthrough, the standard CNN has adopted a similar architecture to AlexNet. Multiple convolutional layers are stacked on top of each other, periodically interspersed with pooling layers. Once the output feature map size no longer allows for additional pooling, one or more fully-connected layers are appended before the output layer, as is illustrated in Figure 5.8.

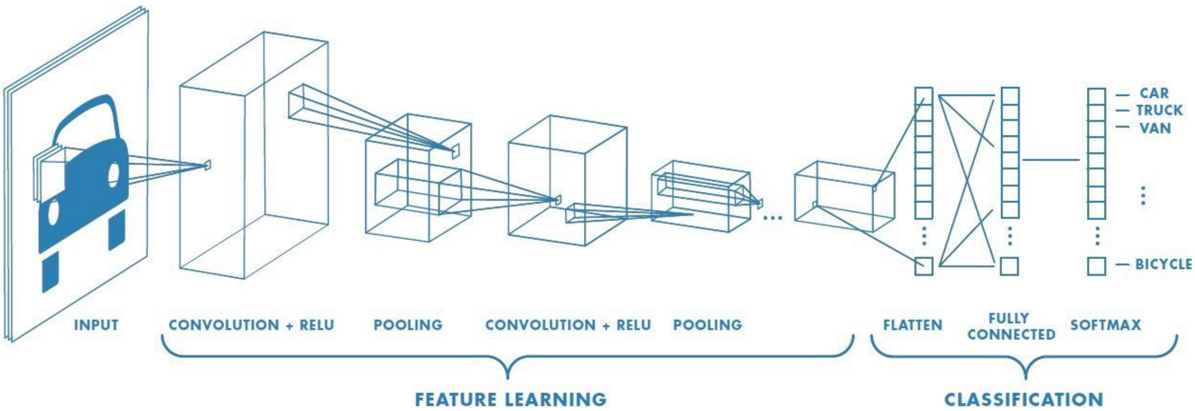


Figure 5.8: Illustration of a typical CNN architecture containing convolutional, pooling, and fully-connected layers before the final output layer. Figure taken from reference [147].

Led primarily by large research teams at the technology giants, improvements upon this standard architecture have since been made. Initially, this process involved the addition of extra convolutional layers to form deeper and deeper networks, as was done by the VGG architecture in 2014 [148]. Another approach was the introduction of the *inception module* within the GoogLeNet [126] architecture, allowing for different feature scales to be considered.

ResNet introduced residual connections in 2016 [149, 150]. By adding connections skipping specific layers, a larger gradient could reach the lower layers of the network, increasing learning. Recently, the inception module and ResNet concepts have been combined [151], and there has been a significant push for efficient rather than just deeper networks [152, 153]. The common repeating layer patterns that form the above networks are shown in Figure 5.9 for reference.

5.3.4 Regularisation

A key challenge when training supervised machine learning models is ensuring that they can generalise to new, previously unseen data, not within the training dataset. With networks typically containing millions of trainable parameters, it can become effortless for them to learn specific features and noise of the training dataset, rather than generalisable features. This unwanted learning, called *overfitting*, is ubiquitous when training CNNs. Methods used within this work to prevent overfitting are outlined below, all of which are commonly referred to as *regularisation* techniques.

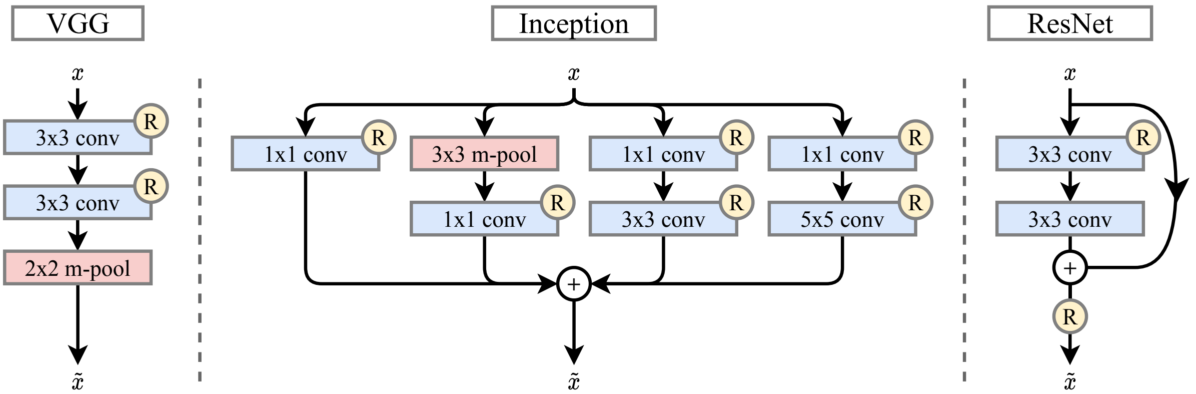


Figure 5.9: Common repeating layer *blocks* used within CNN architectures, taking x as input and producing $\tilde{x}$ as output through operations whose forward pass is shown with the arrows. The blue and red boxes represent convolutional (conv) and max-pooling (m-pool) layers respectively, with the size of the operation shown. The circular yellow R indicates the use of the ReLU activation function.

Stochastic gradient descent

The gradient descent update equation outlined in equation (5.9) updates the network weights at each training iteration using the gradient calculated over the full training dataset. This procedure is called *batch training*. It is instead much more common to calculate an approximation to the complete gradient at each iteration using a *minibatch* of the full dataset. This is done by considering just a subset of the training data with a size commonly referred to as the *batch size* and typically equal to a power of two for computational reasons.

This modification to standard batch training gradient descent is called *stochastic gradient descent* as it introduces stochasticity to the training process, providing two main advantages. Firstly, the computational time of each iteration is significantly reduced, and crucially the memory requirements lowered. Secondly, the addition of minibatch specific noise decreases the chance that the minimisation will get stuck in a local minimum suited to overfitting the training dataset.

Early stopping

Early stopping is another simple regularisation procedure. During training, the training dataset is commonly iterated over multiple times, with each full iteration called an *epoch*. By evaluating the error on an independent *validation* dataset at the end of each

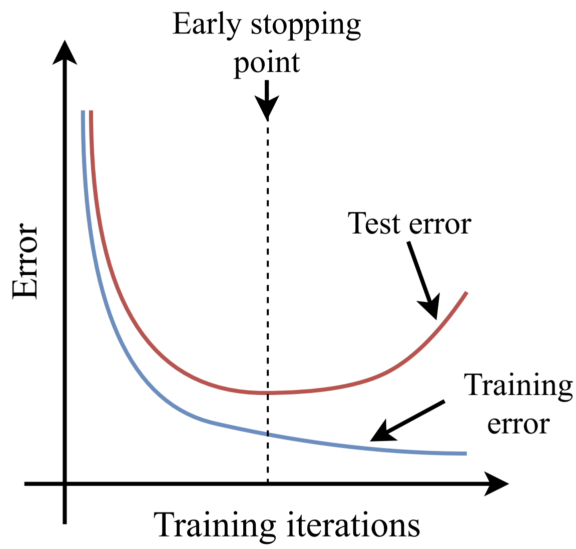


Figure 5.10: Illustration of the early stopping procedure. Initially both the training and test error decrease, but, at some point the test error will start to increase due to overfitting, at this point the training is stopped.

epoch, the point at which overfitting starts to occur can be determined, as illustrated in Figure 5.10. At the determined epoch, the training is stopped to return the best possible generalised model. In practice, it is common only to stop training after n epochs have passed with no validation error improvement.

Batch normalisation

The training of a neural network is found to work best when the inputs of each neuron are centred on zero with respect to the bias of the neuron. This is because large input values can cause saturation of the activation function and subsequent vanishing of the associated gradient, reducing the ability of the network to learn. To counter this, *batch normalisation* introduces layers that standardise their inputs by using both the mean and variance of each minibatch [154]. This modification not only speeds up training by preventing the vanishing of gradients but also reduces overfitting by again using the stochasticity of the minibatch.

Dropout

Dropout is another simple technique to reduce overfitting [155]. At each training iteration, each neuron has a probability p_d, to be *dropped out* and ignored for that iterations

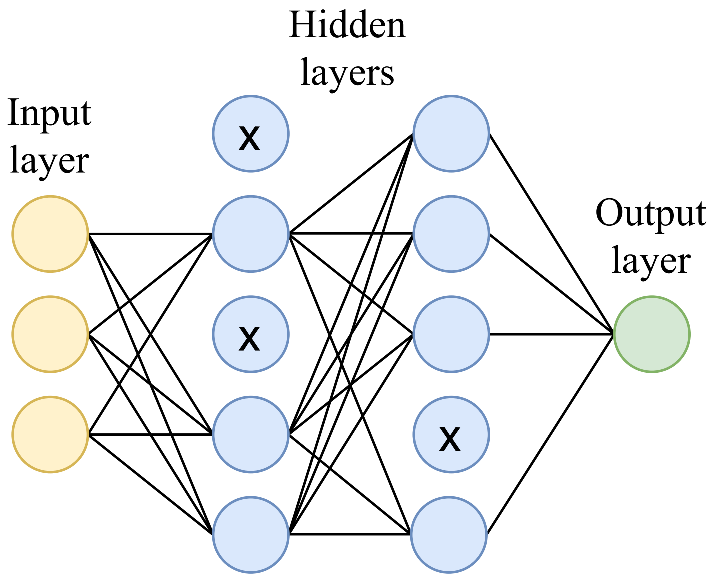


Figure 5.11: Example of dropout applied to the network shown in Figure 5.2. Neurons are randomly *dropped out* and not considered for a training step. This reduces the strong connections between neurons that can lead to overfitting.

calculation. This is illustrated in Figure 5.11. By ignoring a subset of neurons at each iteration, it is difficult for the network to form the particularly strong connections that are usually responsible for overfitting, leading to greater generalisation.

5.4 A baseline implementation for CHIPS

The raw output from a water Cherenkov detector, such as those envisioned by the CHIPS concept, is a simple image of each event where two channels of information are known for each PMT: the number of collected photoelectrons, and the associated hit times. Therefore, it is a natural fit to use CNNs primarily developed for image-based computer vision tasks for CHIPS event analysis.

For this purpose, a Python-based software package named *chipsnet* [156] has been built. By using the high-level *Application Programming Interfaces* (APIs) provided by the Tensorflow framework (version 2.3.0) [157], a full pipeline including data preparation, network training, and performance evaluation has been implemented. In this section, the baseline CNN implementation built into chipsnet is outlined. The specific network implementations described in Section 5.5 share this common baseline with a few specific differences.

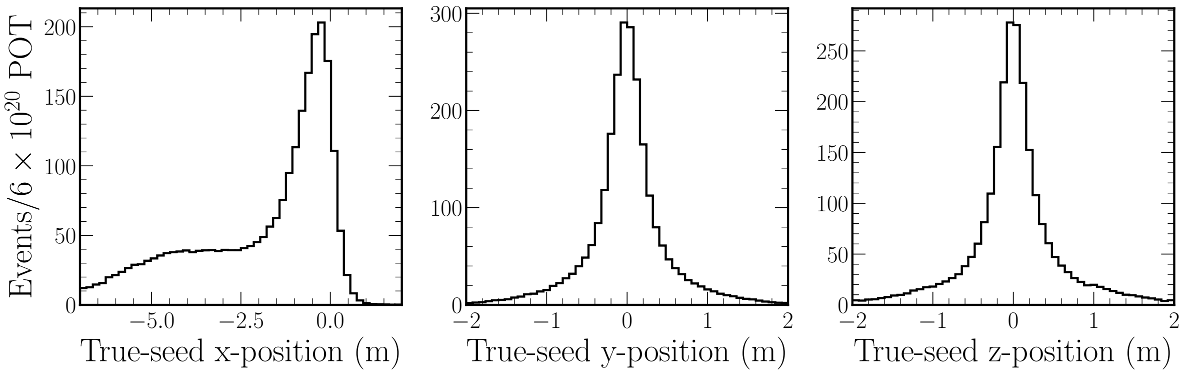


Figure 5.12: Difference between the true and seed estimated vertex position, split by component. The large negative tail for the x component shows the tendency of the seeding procedure to estimate the x component closer to the downstream detector wall than reality.

5.4.1 Baseline inputs

The primary difficulty in the application of CNNs to Chips is determining how to map an event captured on a cylindrical surface to a two-dimensional grid. Furthermore, this must be done in such a way as not to distort the underlying Cherenkov emission topology, which could inhibit network learning. As a solution, this work takes inspiration and then builds upon the ideas outlined in reference [158]. Simply put, an event is mapped onto a two-dimensional grid as though it is viewed from its estimated interaction vertex position. The primary motivation behind this is to remove any detector shape effects and to focus on the underlying event topology and Cherenkov emission profiles.

To estimate each event's interaction vertex position, the top-scoring seed from the seeding procedure introduced in Section 5.2.1 is used. This process, unlike the full likelihood fit, requires no predefined track hypothesis and typically takes under 0.1 seconds per event on a standard batch farm computing node. The difference between the true and seed estimated vertex position for a sample of expected beam events are shown in Figure 5.12. The x component (along the direction of the beam) is commonly estimated closer to the downstream wall of the detector than reality, however, as the y and z components perpendicular to the beam primarily drive event distortions, the impact on event topology is small.

Using θ and ϕ components calculated as viewed from the estimated interaction vertex position facing along the x-axis (downstream), hit PMTs are mapped onto a 64×64 grid. This procedure is used to generate two event *maps*. Firstly, a *hit-charge* map where

Event map	Cap-point	Capped percentage
hit-charge	25 p.e	0.10%
hit-time	120 ns	0.15%
hough-height	3500 p.e	0.23%

Table 5.1: Table showing the input event map cap-points (maximum value of the encoded range) and the associated percentage of bin values that are capped at the maximum 8-bit value of 255 as a consequence. The cap-points are specifically chosen so that the capped percentage is approximately 0.1%, keeping any information loss small.

each grid bin is given by the total collected photoelectrons from all PMTs mapped to that bin. Secondly, a *hit-time* map where each grid bin is given by the first hit time (in nanoseconds) across all PMTs mapped to that bin. Each hit-time map is further corrected so that the first hit time across all bins lies at zero. Note that within this work veto PMTs are ignored for simplicity.

By design, the Hough transform within the seeding procedure uses the estimated interaction vertex position to generate the transform space. Therefore, by re-binning the transform space to a 64×64 grid, a third *hough-height* map is generated for each event. This event map aims to provide a complementary but different representation of the event where Cherenkov rings are instead represented as peaks, allowing for additional discriminating features to be learnt.

All three event maps: hit-charge, hit-time, and hough-height are down-sampled using 8-bit encoding by converting each 32-bit float value to an integer between 0 and 255. Encoding not only significantly reduces storage requirements but also dramatically increases the speed with which data can be loaded during training (which turns out to be the primary bottleneck). For each map type, a range over which to encode from zero up to a *cap-point* is chosen to minimise the number of bin values that are capped at the maximum encoded value of 255. Table 5.1 shows the cap-points and the associated percentage of bin values capped for each map type, while Figure 5.13 shows the distribution of bin values for each event map across the encoded range.

Event maps for example events generated using the above procedure are shown in Figure 5.14 for a CC resonant ν_e event, in Figure 5.15 for a CC DIS ν_μ event, in Figure 5.16 for a NC DIS event, and in Figure 5.17 for a cosmic muon event.

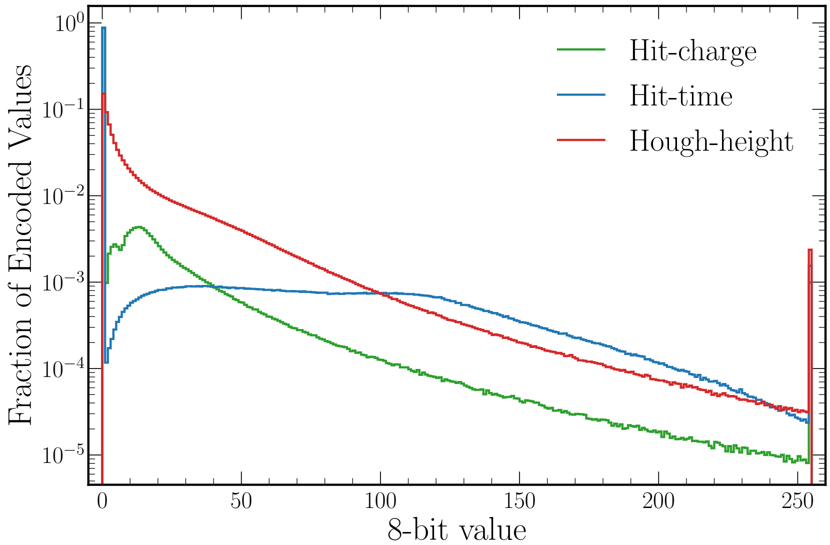


Figure 5.13: The distribution of encoded 8-bit values for the hit-charge, hit-time and hough-height event maps.

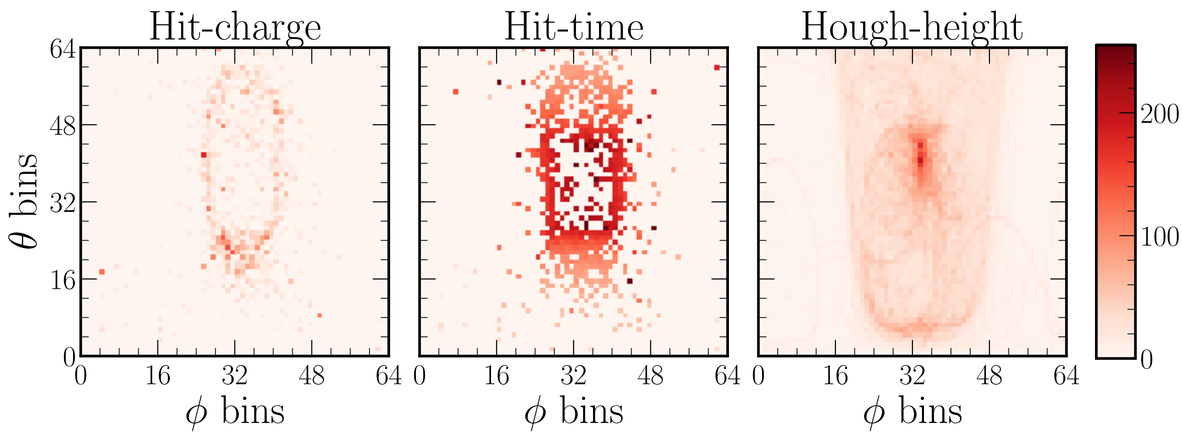


Figure 5.14: Three map representation of a CC resonant ν_e event. Initiated by a ν_e of energy 3.3 GeV the final state particles above the Cherenkov threshold include a e^- of energy 2.8 GeV and a 0.3 GeV π^0.

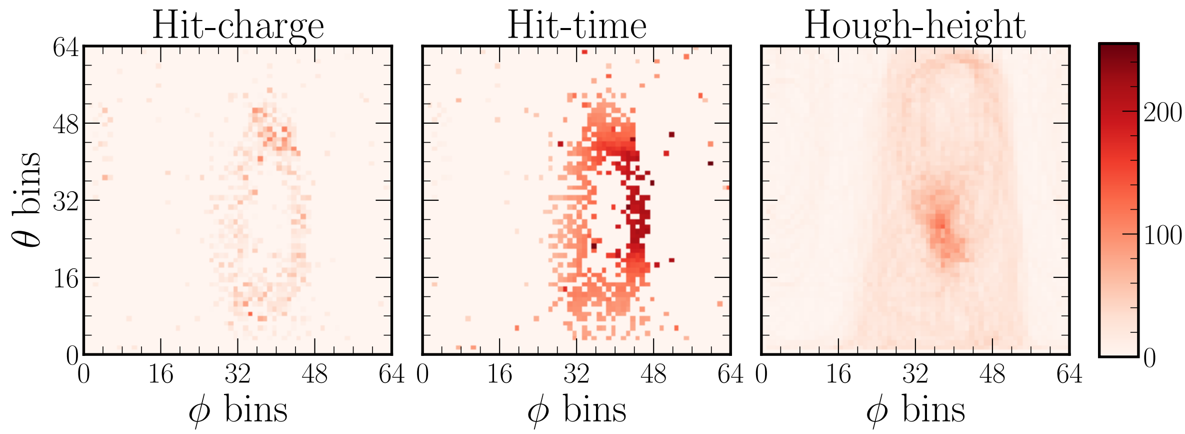


Figure 5.15: Three map representation of a CC DIS ν_μ event. Initiated by a ν_μ of energy 3.5 GeV the final state particles above the Cherenkov threshold include a μ^- of energy 1.9 GeV, a proton of energy2.0 GeV, and a 0.2 GeV π^-.

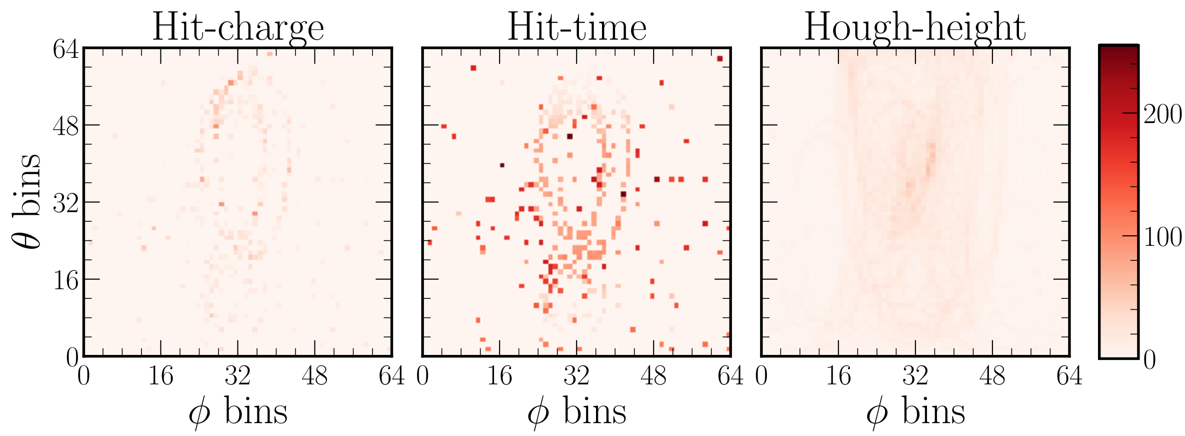


Figure 5.16: Three map representation of a NC DIS event. Initiated by a ν_e of energy 9.3 GeV the final state particles above the Cherenkov threshold include a proton of energy 2.6 GeV and a 2.5 GeV π^-.

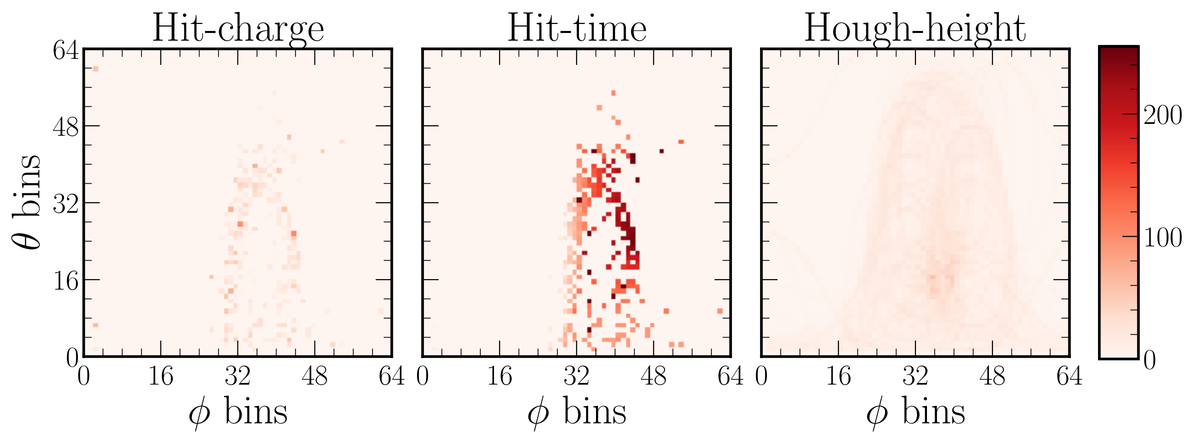


Figure 5.17: Three channel representation of a cosmic muon event, containing a μ^- of energy 2.9 GeV.

5.4.2 Baseline architecture

An illustrative diagram of the baseline chipsnet architecture is shown in Figure 5.18. Based on the VGG network previously mentioned in Section 5.1 and detailed in reference [148] there are a few key differences from the literature defined network:

- Each of the three event maps: hit-charge, hit-time, and hough-height are initially fed into three separate branches. Each branch contains two VGG blocks with two convolutional layers each (four convolutional layers in total). The outputs from each branch are merged using a concatenation layer before being fed to the rest of the network. This configuration allows for event map specific features to be learnt independently before combined features are learnt by the rest of the network.
- Batch normalisation as described in Section 5.3.4 is included before the activation (ReLU) function for every convolutional layer.
- Squeeze-and-excitation units, as detailed in reference [159] are included after the max-pooling operation in all VGG blocks. These units introduce extra parameters to model the interdependencies between output feature maps, allowing the network to learn how to weight each feature map effectively.
- Dropout is included at the end of each VGG block as well as after the final fully-connected layer. Instead of dropping individual kernel elements, the dropout within the VGG blocks drops entire kernels at each training iteration, this is commonly called *two-dimensional spatial dropout*. The dropout after the fully-connected layers is standard, in that it drops out individual fully-connected neurons.

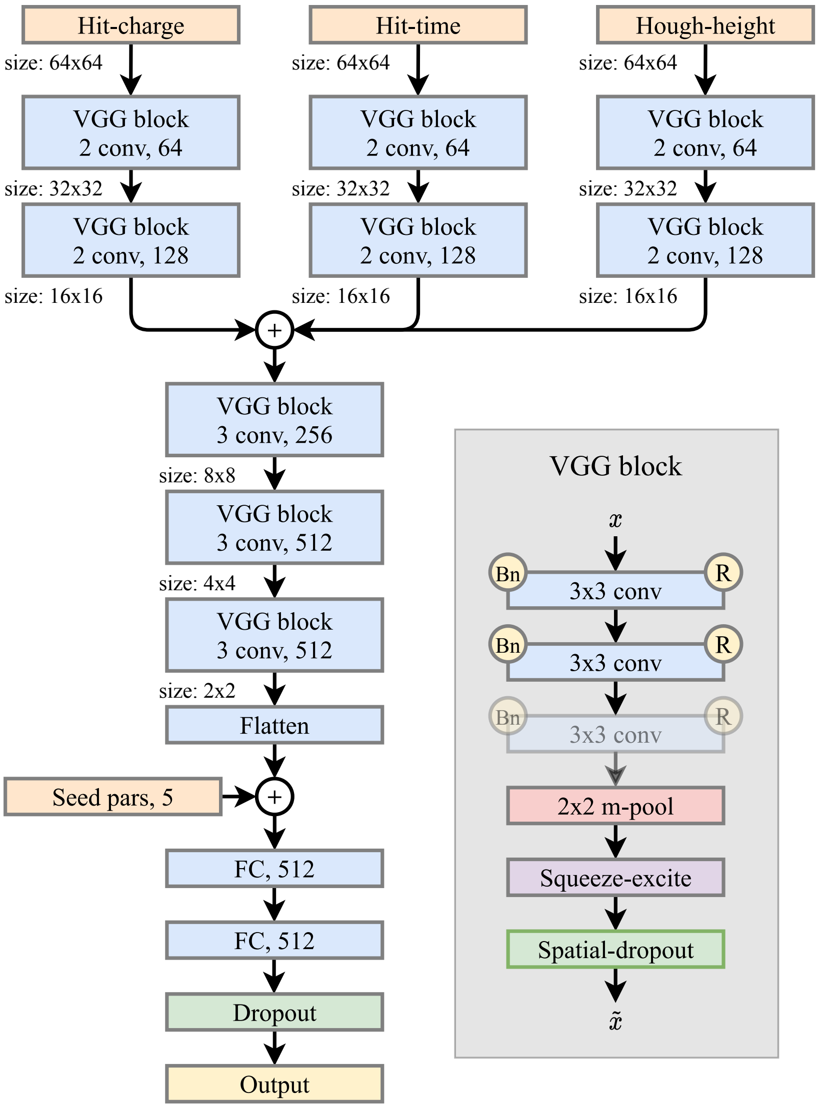


Figure 5.18: Illustrative diagram of the baseline chipsnet architecture. The three input event maps are separately passed through two VGG blocks each before their outputs are combined and passed through a further three VGG blocks together. The flattened VGG blocks outputs are then concatenated with five seed parameters (seed pars) and passed through two fully-connected (FC) layers of 512 neurons each before the output layer. Both the number of convolutional units (1st value) and kernels (2nd value) is shown for each block. The detailed VGG block structure is shown within the grey box. The circular yellow R and Bn indicate the use of the ReLU activation function and batch normalisation, respectively.

- Five output parameters from the seeding process (seed pars) are concatenated with the flattened layer before the fully-connected layers. Included are the three components of the estimated interaction vertex position (s_x, s_y, s_z), and the two components of the estimated track direction (s_θ, s_ϕ). These parameters provide the network with spatial context as to where in the detector the input event maps have been generated and the dominant direction of PMT activity.

The chipsnet baseline architecture is implemented using the Keras API built into Tensorflow [160]. Keras allows for predefined common layers such as a two-dimensional convolution or a max-pooling layer to be easily structured into a full network definition.

5.4.3 Baseline outputs

Many CNN applications are found to benefit from learning multiple tasks at the same time. This is believed to be the case as training with multiple tasks tends to return a network with an improved generalised representation of the inputs, with features learnt for one task improving the performance of another. Additionally, multiple tasks work to prevent any one output from overfitting. Commonly named *multi-task* learning, this methodology is used extensively in this work.

To train a network with multiple tasks (outputs), a loss function E_{tot}, must be defined to combine the individual loss functions for each task E_i. The simplest way to do this is via a linear weighted sum, such that

$$E_{tot} = \sum_{i=1}^{i=N} w_i E_i, \tag{5.20}$$

where N is the number of tasks and w_i are the associated weights. In this work this is referred to as the *simple* multi-task loss.

However, the final network performance can strongly depend on the relative weighting between loss functions, especially when the values returned by each differ by many order of magnitude (common when combining regression and classification tasks). Therefore, finding the optimal w_i weights can be both difficult and time-consuming. Another approach outlined in reference [161] remedies this problem by learning the optimal weighting between loss functions. This is done by introducing an additional trainable parameter σ_i, for each loss function, such that

$$E_{tot} = \sum_{i=1}^{i=N} \frac{1}{2\sigma_i^2} E_i + \log \sigma_i. \tag{5.21}$$

In this work we refer to this as the *learnt* multi-task loss.

The specific number and nature of outputs for the specific networks are detailed in Section 5.5. Although physically motivated to some degree, the exact set of tasks and the loss combination technique used is mainly driven by extensive trial-and-error. The chipsnet software is specifically designed to enable this process by making it easy to configure the network outputs via a simple configuration file.

5.4.4 Baseline training

All networks are trained on an 18 core CPU (36 thread) machine equipped with four NVIDIA GeForce RTX 2080 graphics processing units (GPUs). The Tensorflow dataset API is used to create an efficient input data pipeline where data is loaded on-the-fly at training time. This procedure ensures all CPU threads are utilised loading, decoding, and preprocessing data for the primary GPU based network calculations before being needed, maximising computational efficiency.

During preprocessing, all 8-bit input event map values are converted to 32-bit float values bounded between zero and one. Furthermore, a random factor scaling is applied to each map bin. Generated from a normal distribution centred on one with a standard deviation of σ_r, by fluctuating the bin values the network is forced to focus less on the absolute bin values and more on the underlying event topology. Not only does this process provide valuable regularisation to reduce overfitting, but also makes the trained networks robust to small changes within the input (explored within Section 6.3).

A minibatch training strategy of minibatch size of n_b, using the Adam optimiser [162] ($\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e - 7$) is used. The exact training sample size and composition for each specific network are given in Section 5.5, but for all networks a 95% training to 5% validation data split is employed across the full training sample. Consequently, early stopping is used. The learning rate for each epoch η_e, is set to decrease throughout training according to

$$\eta_e = \frac{\eta_0}{1 + c_d(e - 1)}, \tag{5.22}$$

where η_0 is the initial learning rate, e is the epoch number (starting at one), and c_d is the learning rate decay coefficient.

Therefore, when training each network, there is a list of tunable *hyperparameters*, all of which are optimised using the SHERPA hyperparameter tuning framework [163]. To maximise performance, SHERPA uses a *random search* algorithm to select random configurations of hyperparameters which are then tested by training the network for five epochs on the available training data. Each configuration's performance is assessed by using a metric detailed for each network in Section 5.5. The search space is confined to a specific range or selection of choices for each hyperparameter, with:

- the **initial learning rate $\boldsymbol{\eta_0}$**, in a range from 0.00005 to 0.001;
- the **learning rate decay coefficient $\boldsymbol{c_d}$**, in a range from 0.2 to 0.8;
- the **dropout probability $\boldsymbol{p_d}$**, in a range from 0.0 to 0.5;
- the **random scaling size $\boldsymbol{\sigma_r}$**, in a range from 0.0 to 0.1;
- the **minibatch size $\boldsymbol{n_b}$**, choosing from either 32, 64, 128, or 256; and
- the **multi-task loss combination strategy**, choosing from either simple or learnt.

5.5 Specific implementations for CHIPS

The specific CNN implementations for cosmic rejection, beam classification, and energy estimation are outlined below. It is important to note that the exact configuration of networks outlined here is the result of extensive testing designed to maximise the selection of a pure and efficient sample of appeared CC ν_e beam events whose neutrino energy can also be accurately determined.

As an example of an alternative implementation, if cosmic rejection and beam classification are combined into a single network, both objectives see a reduction in performance. The same is also true if either cosmic rejection or beam classification is combined with neutrino energy estimation. However, specific secondary outputs such as counting the number of primary particles in conjunction with beam classification are seen to improve performance. It is clear, therefore, that the multi-task approach only works for tasks that require a similar learnt representation of the inputs. Put simply; the tasks must be similar.

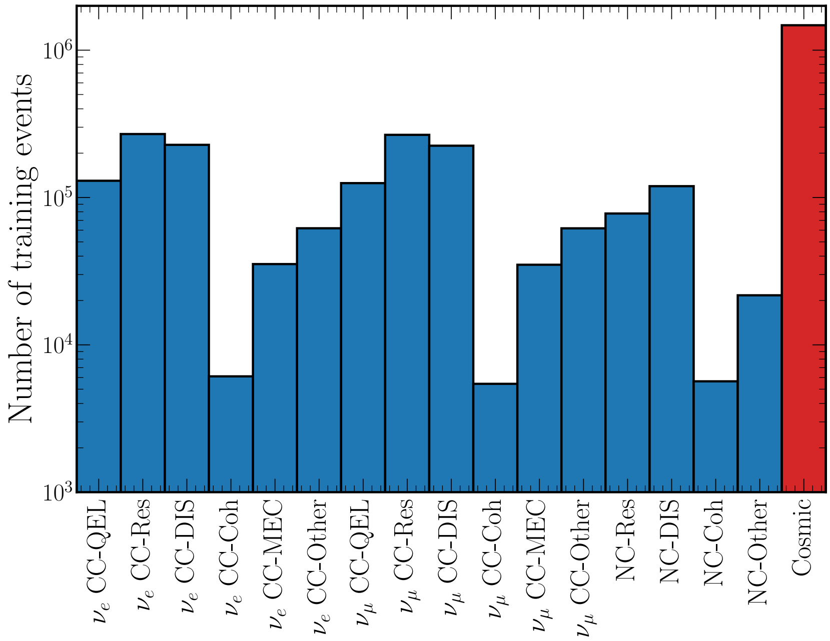


Figure 5.19: Number of training events per category for the cosmic rejection network. All beam event interaction types are shown, however, all are classed as beam events (blue) in training against the cosmic events (red).

5.5.1 Cosmic rejection

The cosmic rejection network aims to prevent the vast cosmic muon background from contaminating the final selected sample of beam events. Therefore, the primary task is a simple binary classification between beam and cosmic events. Additionally, training the network to also separate events where the primary charged lepton escapes the detector volume or not, is found to improve cosmic rejection performance. As a large proportion of cosmic muons are relatively high in energy and, therefore, escape the detector in this fashion, there is motivation as to why this additional task is helpful.

The network is trained on a sample of 3.15 million simulated events produced using the detector simulation and event generation methods outlined in Section 3.3. Roughly $1/3^{rd}$ are ν_μ beam events, $1/3^{rd}$ ν_e beam events, and $1/3^{rd}$ cosmic muon events, the counts of which are shown in Figure 5.19. All beam events (both ν_μ and ν_e) are generated using the expected unoscillated CHIPS-5 ν_μ energy spectrum to closely mimic the dominant ν_μ beam component and appeared ν_e signal. Every event in the sample is used for training with no preselection, as this is found to be the best for cosmic rejection performance.

There are two outputs to the cosmic rejection network:

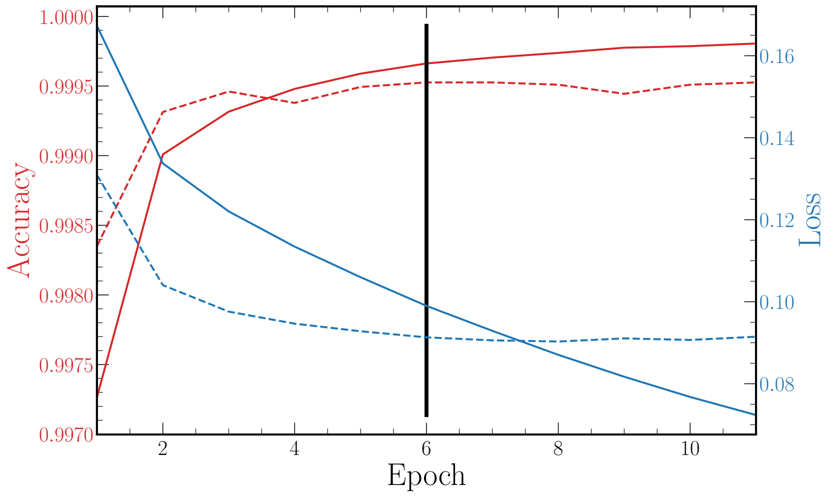


Figure 5.20: Total loss and *cosmic score* accuracy for both the training sample (solid) and validation sample (dashed) throughout training for the cosmic rejection network. The final network weights are taken at epoch six, as shown by the vertical black line.

1. **Cosmic score (1 classification neuron):** Returns a score between zero and one corresponding to whether the event is beam or cosmic like. The binary cross-entropy loss function in equation (5.5) is used for training with a simple multi-task weight of 1.

2. **Escapes score (1 classification neuron):** Returns a score between zero and one corresponding to whether the charged lepton in an event is contained or escapes the detector. The binary cross-entropy loss function in equation (5.5) is used for training with a simple multi-task weight of 1. NC beam events without a charged lepton are masked (do not contribute to the loss) during training for this output.

The network is allowed to train for up to 30 epochs using the SHERPA optimised hyperparameters: $\eta_0 = 0.00005$, $c_d = 0.7$, $p_d = 0.1$, $\sigma_r = 0.02$, and $n_b = 128$, with a simple multi-task loss combination as given in equation (5.20). The *cosmic score* accuracy metric, defined as the fraction of validation sample events that are correctly classified when a cut value of 0.5 on the *cosmic score* output is used to determine the classification of each event, is used for SHERPA optimisation and early stopping. Typically, only 6 epochs are required to reach the maximum validation sample *cosmic score* accuracy, with early stopping halting training after 11 epochs (taking 15 hours), as can be seen in Figure 5.20.

5.5.2 Beam classification

The beam classification network aims to separate beam events by their neutrino and interaction type to primarily select a pure and efficient sample of appeared ν_e events, but also a sample of survived CC ν_μ events. Therefore, the principal task is a categorical classification between CC ν_e, CC ν_μ, and NC events. No attempt is made to separate the appeared CC ν_e and intrinsic beam CC ν_e components as they are impossible to tell apart, except for their distribution in neutrino energy.

Similar to the implementation used by DUNE [131], alongside the core classification additional classification and particle counting tasks outlined below are included to improve performance. Note that the particle counting tasks are not used in this work for anything but increasing the primary classification performance. Future work, however, could exploit any ability to separate exclusive final states, deduced from these particle counts, to reduce both energy resolution and systematic errors. As an example of a method already in use, NOvA use their ability to accurately determine the hadronic energy of an event to split their CC ν_μ sample into populations of different energy resolution. Each population can then be treated separately in the analysis before being combined, increasing overall performance [164].

The network is trained on a sample of 1.67 million simulated events produced using the detector simulation and event generation methods outlined in Section 3.3. Roughly half are ν_μ beam events, with the other half being ν_e beam events, as shown in Figure 5.21. All events (both ν_μ and ν_e) are generated using the expected unoscillated CHIPS-5 ν_μ energy spectrum to closely mimic the dominant ν_μ beam component and appeared ν_e signal. All events are used for training with no preselection as this is found to be the best for beam classification performance, especially NC rejection.

There are nine outputs to the beam classification network:

1. **Combined category (3 classification neurons):** Returns a classification probability score between zero and one for each of CC ν_e, CC ν_μ, and NC (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

2. **CC category (6 classification neurons):** Returns a classification probability score between zero and one for each of CC-QEL, CC-Res, CC-DIS, CC-Coh, CC-MEC, and CC-other (summing to one). The categorical cross-entropy loss function

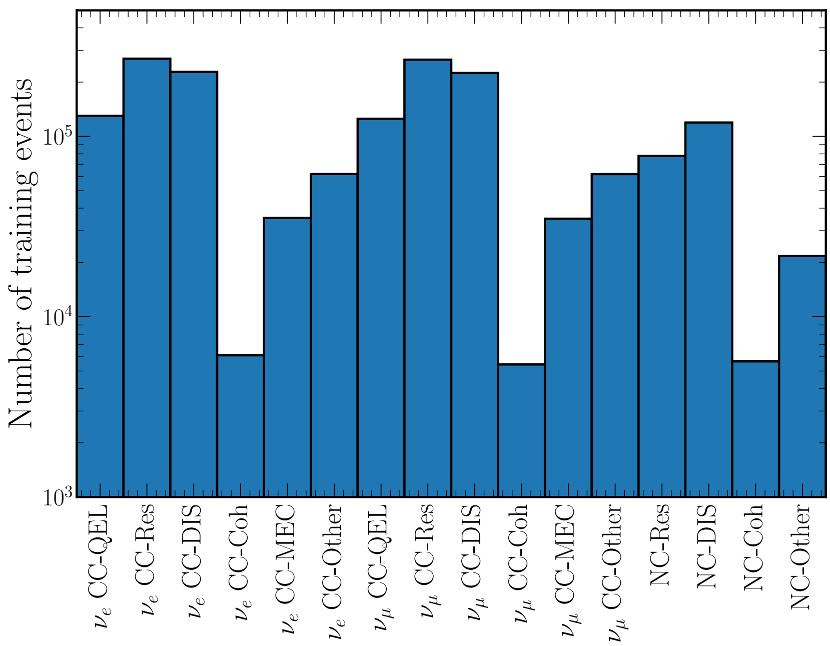


Figure 5.21: Number of training events per category for the beam classification network. All beam event interaction types are shown.

in equation (5.7) is used for training with a simple multi-task weight of 1. NC events are masked (do not contribute to the loss) during training for this output.

3. **NC category (4 classification neurons):** Returns a classification probability score between zero and one for each of NC-Res, NC-DIS, NC-Coh, and NC-other (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1. CC events are masked (do not contribute to the loss) during training for this output.

4. **Electron count (4 classification neurons):** Returns a classification probability score between zero and one for each of 0, 1, 2, and 3+ electrons in the final state (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

5. **Muon count (4 classification neurons):** Returns a classification probability score between zero and one for each of 0, 1, 2, and 3+ muons in the final state (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

6. **Proton count (4 classification neurons):** Returns a classification probability score between zero and one for each of 0, 1, 2, and 3+ protons in the final state

(summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

7. **$\pi^{\pm}$ count (4 classification neurons):** Returns a classification probability score between zero and one for each of 0, 1, 2, and 3+ charged pions in the final state (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

8. **π^{0} count (4 classification neurons):** Returns a classification probability score between zero and one for each of 0, 1, 2, and 3+ neutral pions in the final state (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

9. **Photon count (4 classification neurons):** Returns a classification probability score between zero and one for each of 0, 1, 2, and 3+ photons in the final state (summing to one). The categorical cross-entropy loss function in equation (5.7) is used for training with a simple multi-task weight of 1.

The network is allowed to train for up to 30 epochs using the SHERPA optimised hyperparameters: $\eta_0 = 0.0002$, $c_d = 0.5$, $p_d = 0.1$, $\sigma_r = 0.02$, and $n_b = 128$, with a simple multi-task loss combination as given in equation (5.20). The *combined category* accuracy metric, defined as the fraction of validation sample events that are correctly classified when the highest scoring *combined category* output neuron is used to determine the classification of each event, is used for SHERPA optimisation and early stopping. Typically, only 7 epochs are required to reach the maximum validation sample *combined category* accuracy, with early stopping halting training after 12 epochs (taking 15 hours), as can be seen in Figure 5.22.

5.5.3 Energy estimation

Accurate neutrino energy estimation is accomplished using multiple networks trained on separate samples of ν_e and ν_μ events across multiple CC interaction types. It is found that separation such as this results in greater performance compared to if a single energy estimation network or even separate ν_e and ν_μ networks are trained. This is expected as a single set of network weights is unlikely to be able to capture the specific topological features that contribute to the energy for all types of event.

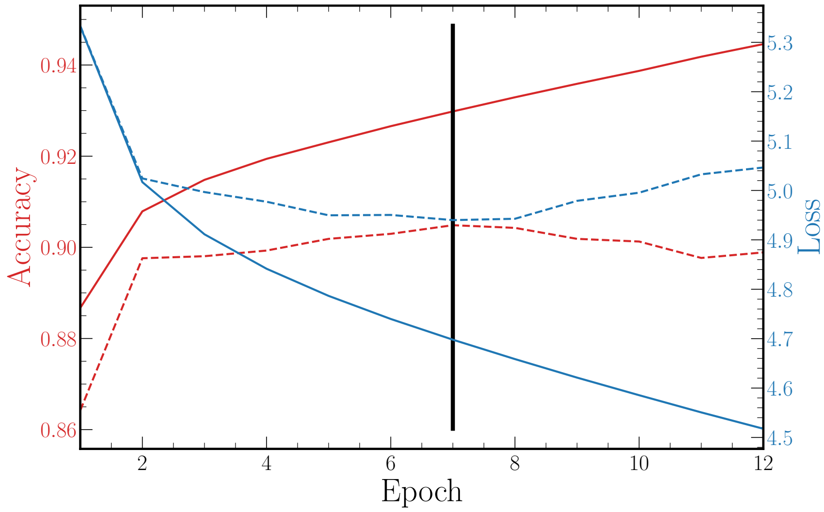


Figure 5.22: Total loss and *combined category* accuracy for both the training sample (solid) and validation sample (dashed) throughout training for the beam classification network. The final network weights are taken at epoch seven, as shown by the vertical black line.

Alongside the primary neutrino energy regression task, additionally learning the primary charged lepton energy and the interaction vertex position and time are found to improve performance. Although this improvement is relatively small for neutrino energy estimation, it dramatically improves primary charged lepton energy estimation when compared to being predicted alone. With two energy tasks, the network is encouraged to learn how the primary charged lepton and neutrino energies are related. As the interaction vertex position within the detector and hence distance from the instrumented wall can impact the number of deposited photoelectrons, there is motivation as to why this additional task is also helpful.

Separate networks are trained for each of CC-QEL (and CC-MEC), CC-Res, and CC-DIS for both ν_e and ν_μ events (6 in total) using 250000 corresponding simulated events each. All events (both ν_μ and ν_e) are produced using the detector simulation and event generation methods outlined in Section 3.3. The expected unoscillated CHIPS-5 ν_μ energy spectrum is used to generate all events to closely mimic the dominant ν_μ beam component and appeared ν_e signal. Only events for which the primary charged lepton is fully contained within the detector volume are used for training, with no additional preselection applied. Note that CC-QEL and CC-MEC energy estimation is combined into a single network as both have incredibly similar final state topologies (a single charged lepton).

There are six outputs to each of the energy estimation networks:

1. **Neutrino energy (1 regression neuron):** Returns the estimated neutrino energy. The mean-squared error loss function in equation (5.8) is used for training.
2. **Charged lepton energy (1 regression neuron):** Returns the estimated primary charged lepton energy. The mean-squared error loss function in equation (5.8) is used for training.
3. **Interaction vertex x-position (1 regression neuron):** Returns the estimated interaction vertex x-position. The mean-squared error loss function in equation (5.8) is used for training.
4. **Interaction vertex y-position (1 regression neuron):** Returns the estimated interaction vertex y-position. The mean-squared error loss function in equation (5.8) is used for training.
5. **Interaction vertex z-position (1 regression neuron):** Returns the estimated interaction vertex z-position. The mean-squared error loss function in equation (5.8) is used for training.
6. **Interaction time (1 regression neuron):** Returns the estimated interaction time relative to the first PMT hit for each event. The mean-squared error loss function in equation (5.8) is used for training.

Each network is allowed to train for up to 30 epochs using the SHERPA optimised hyperparameters: $\eta_0 = 0.0002$, $c_d = 0.5$, $p_d = 0.1$, $\sigma_r = 0.0$, and $n_b = 128$, with a learnt multi-task loss combination as given in equation (5.21). The *neutrino energy* mean absolute error metric, defined as the average difference between the true and estimated neutrino energies across all validation sample events, is used for SHERPA optimisation and early stopping. Early stopping typically halts training after approximately 20 epochs (taking 2 hours). An example of how an energy estimation networks training typically proceeds is given in Figure 5.23.

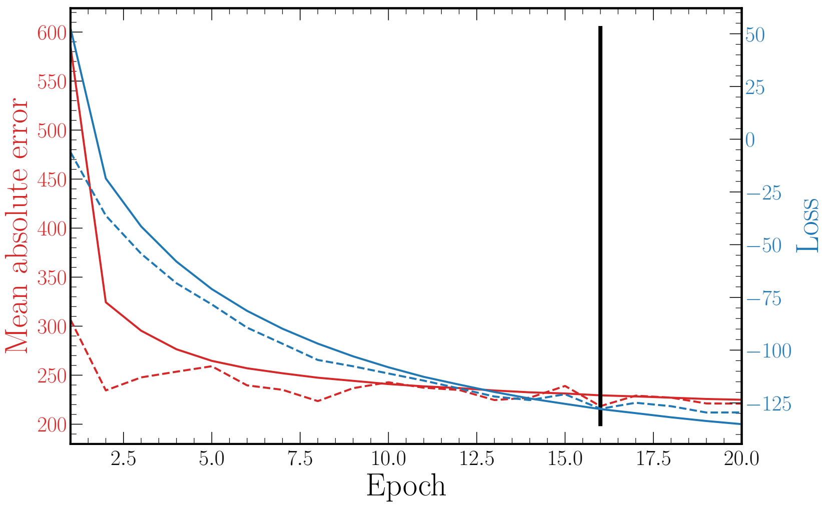


Figure 5.23: Total loss and *neutrino energy* mean absolute error for both the training dataset (solid) and validation dataset (dashed) throughout training for the energy estimation network trained on CC-QEL and CC-MEC ν_e events. The final network weights are taken at epoch sixteen as shown by the vertical black line.

Chapter 6

Network evaluation

This chapter provides a comprehensive evaluation of how the new CNN approach behaves. Four categories of understanding are explored (each a section of this chapter):

1. determining the beam selection and energy estimation performance;
2. explaining the inner workings of the trained networks;
3. studying how robust the network outputs are to changes in the input; and
4. exploring alternative implementations to see which factors drive performance.

Beforehand, a hugely impactful advantage of the CNN approach must be highlighted. Although the time taken to train the CNNs in this work is approximately two days, once trained the time required to calculate all network outputs (inference time) for a single event is on the order of 2 ms. When combined with event seeding and event map generation, the total time taken to fully reconstruct and classify a raw event is less than 0.1 seconds. Compared to the ∼15 minutes required for each event using the standard reconstruction and classification methods (with multiple hypotheses), the difference is stark.

Armed with this incredible speed, the time taken to fully process a large dataset containing millions of events becomes a matter of hours, compared to the many weeks typically required. This change has far-reaching implications for how physics analysis is conducted. By removing the processing bottleneck, larger datasets can be used without worry, new techniques can be tested quickly, and overall analysis turnover increased.

6.1 Performance

Here the performance of the CNN approach is presented and compared to that achieved by the standard CHIPS methods as well as comparable experiments. Primarily this focuses on the principal aim of selecting an efficient and pure appeared CC ν_e signal sample with accurate neutrino energy estimation. However, the survived CC ν_μ selection is also presented for completeness.

6.1.1 Evaluation sample and preselection

An independent sample of events is used to evaluate the combined performance of the trained CNNs. The evaluation sample consists of 400000 beam and 350000 cosmic muon events produced in the same way as the training and validation events, using the detector simulation and event generation methods outlined in Section 3.3. The beam events include the expected ν_μ, $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$ components as well as events generated to mimic the appeared (sometimes referred to as *App*) ν_e component. Only the neutrino mode (forward horn current) of NuMI beam operation is considered here. During the evaluation, the intrinsic neutrino and antineutrino components of the beam are considered the same for simplicity.

All evaluation events are weighted to match the expected spectrum at the CHIPS-5 detector using the unoscillated flux, cross-sections, and oscillation probabilities (derived from the NuFIT oscillation parameters shown in Figure 2.8, assuming the normal hierarchy, and including matter effects). Additional weighting also scales the sample to match data taking in the NuMI beam for a single year, corresponding to 6×10^{20} protons on target (POT). Cosmic muon events are weighted according to the 11.8 kHz expected CHIPS-5 rate outlined in Section 3.3.3. The final weighted spectrum of events is shown in Figure 6.1 with the combination of underlying beam neutrino interaction types shown in Figure 6.2.

Separate from the CNN driven work, a simple preselection is applied to all evaluation sample events. Designed to reject cosmic and NC events while keeping the selection efficiency of CC beam events high, the preselection consists of four simple cuts, shown in Figure 6.3. Firstly, the total number of collected photoelectrons (charge) across all PMTs in the event must be greater than 250. Secondly, the maximum Hough transform space height must be greater than 250 photoelectrons. Thirdly, the seeding procedure

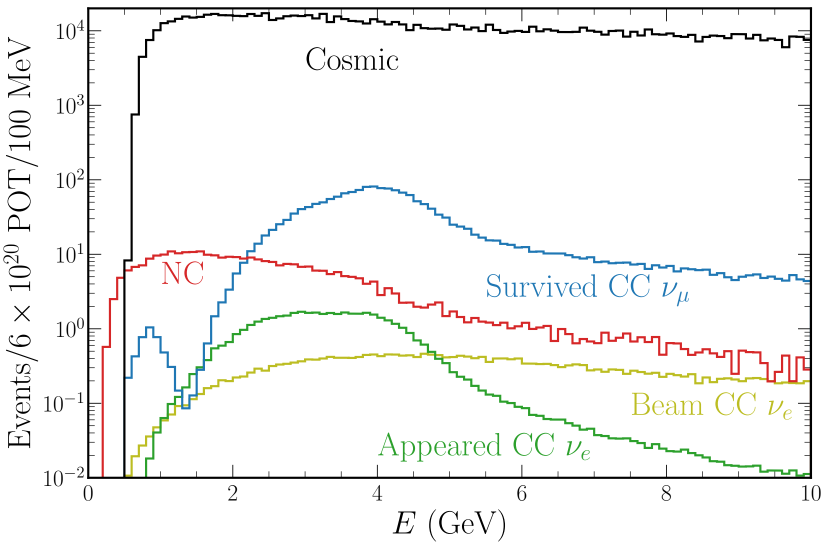


Figure 6.1: Weighted spectrum of events contained within the evaluation sample. The weighting is designed to mimic the expected event spectrum of the CHIPS-5 detector. Beam events are weighted by combining the expected unoscillated flux with cross-sections and standard oscillation probabilities, while cosmic events are weighted using the expected cosmic rate. Shown in blue, green, and olive are the surviving CC ν_μ, appearing CC ν_e, and intrinsic beam CC ν_e spectra respectively, binned in terms of their neutrino energy. Shown in red is the NC event spectra, binned in terms of the energy of the hadronic component (excluding the outgoing neutrino energy) to represent more accurately the energy visible to the detector. Finally, shown in black is the cosmic muon event spectra binned in terms of the muon energy.

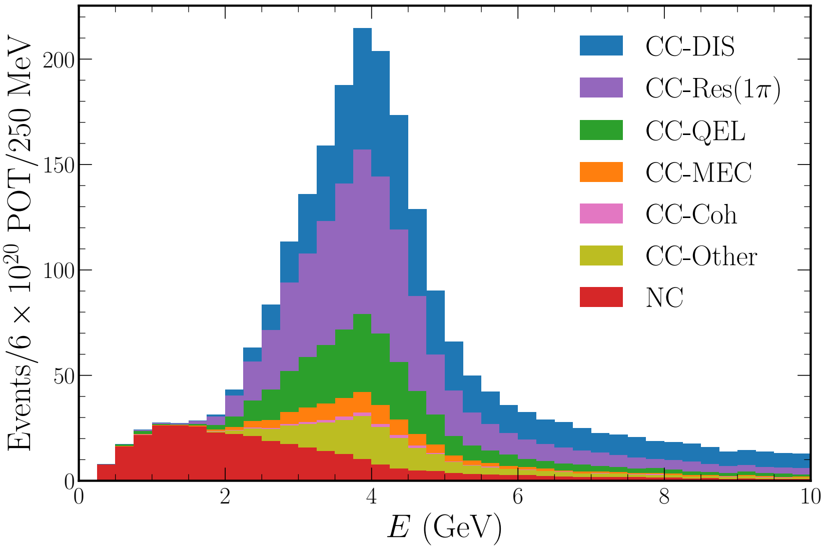


Figure 6.2: Weighted spectrum of ν_μ and ν_e beam events contained within the evaluation sample separated by interaction type. CC events are binned in terms of the true neutrino energy while NC events are binned in terms of the true hadronic component energy.

$\cos(\theta)$ direction must be between ± 0.7. Finally, the seeding procedure ϕ direction must be between ± 1.1 radians. The first two cuts reject low energy events which are usually NC, while the last two reject events whose activity is not along the beamline, typically cosmic.

6.1.2 Cosmic rejection and containment

Cosmic score

The *cosmic score* output from the trained cosmic rejection network shows an excellent separation between beam (output close to zero) and cosmic (output close to one) events, as can be seen in Figure 6.4. Notably, 99.2% of beam events are associated with a score incredibly close to zero (< 0.0001) as is more clearly shown in Figure 6.5. Given this fact, a *cosmic score* of below 0.0001 is chosen to select beam-like events. Out of the total 350000 cosmic events in the evaluation sample, all are rejected by this cut alongside preselection.

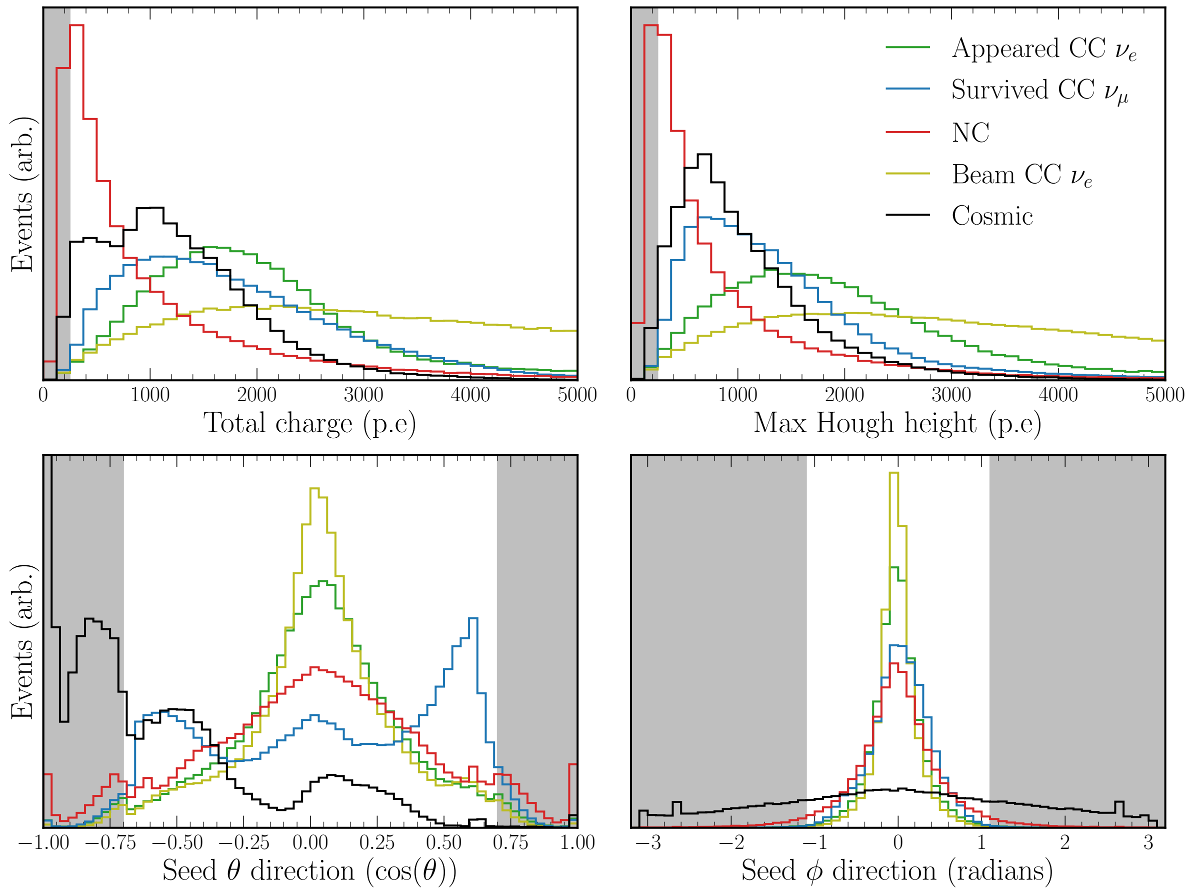


Figure 6.3: Plots showing the four preselection cuts and how they affect the different event categories. The grey regions indicate rejected events.

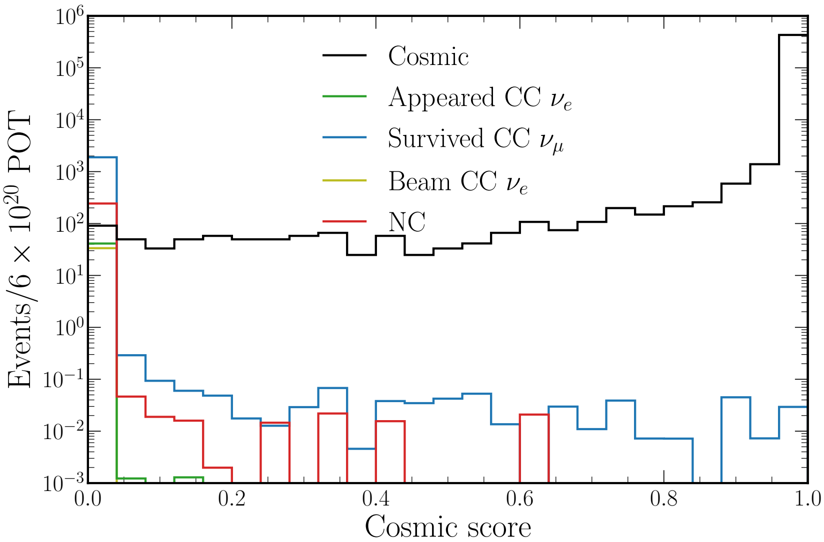


Figure 6.4: Distribution of *cosmic score* output values from the trained cosmic rejection network for the different event categories. A score close to one signifies a cosmic-like event, while a score close to zero corresponds to a beam-like event. Only preselected events are shown to better highlight the events this output aims to classify.

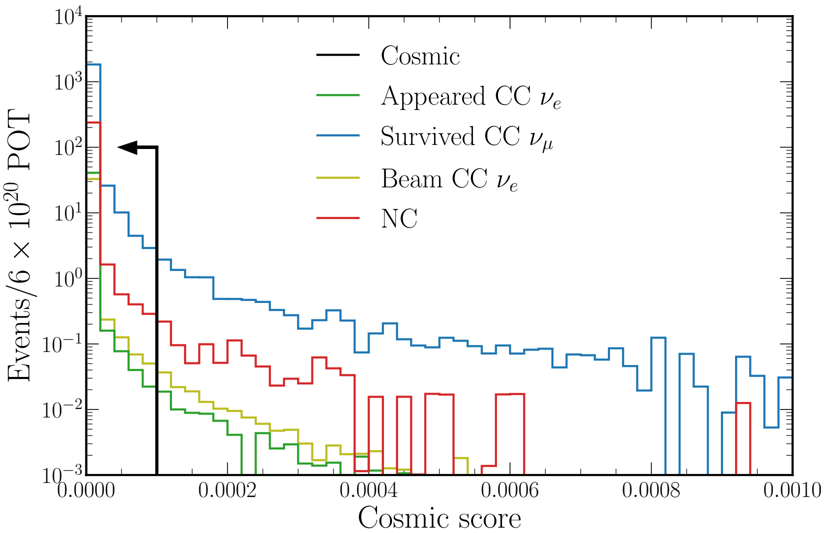


Figure 6.5: Distribution of *cosmic score* output values from the trained cosmic rejection network close to zero for the different event categories. The cosmic rejection cut value is shown at 0.0001 with the arrow indicating selected events. Only preselected events are shown to better highlight the events this output aims to classify. No cosmic events are within the shown range due to the limitations of the evaluation sample.

Escapes score

It is crucial for accurate neutrino energy estimation that the activity of an event is fully contained within the volume of the detector. If charged event particles instead leave the detector and emit Cherenkov radiation not captured by PMTs, it can be incredibly difficult to estimate the resulting missing energy and hence neutrino energy. Within the CHIPS-5 detector, this is particularly important for long track CC ν_μ events for which only 44% of the primary charged muons are fully contained within the detector volume.

Therefore, the second output from the cosmic rejection network, *escapes score* is also used to select events. Although this output only considers the primary charged lepton, instead of all event particles, it still acts as a reasonable proxy for event containment. The distribution of *escapes score* output values for each event category is shown in Figure 6.6.

An *escapes score* value below 0.33 is chosen to select events for which the primary charged lepton is deemed to be fully contained within the detector. This cut value is chosen to maximise the fraction of CC ν_μ events which are correctly classified as having their primary charged lepton contained or not within the detector, leading to $96.8 \pm 0.1\%$ of CC ν_μ events being classified correctly. As expected, the vast majority ($97.2 \pm 0.1\%$) of the short track CC ν_e and NC events are selected by this cut.

For comparison with other experiments, the escapes cut effectively works as a quasi fiducial volume cut, for an energy-dependent region near the downstream wall of the detector. Fiducial volume cuts are common in HEP experiments to remove events whose activity is close to the detector walls and, therefore, can not be reconstructed well. Future work should explore how a fully implemented fiducial cut using the reconstructed interaction vertex position, impacts performance.

Combined rejection

The total number of expected events per year that pass each successive cut (including preselection) for each event category is shown in Table 6.1. Both CC ν_e categories are selected with an efficiency greater than 92%, relative to the total number of expected events, while CC ν_μ events are reduced to a $38.9 \pm 0.1\%$ selection efficiency, mainly by the *escapes score* cut (as desired). Furthermore, NC events are found to be primarily rejected by the preselection, while cosmic events are heavily rejected by both the preselection and *cosmic score* cuts.

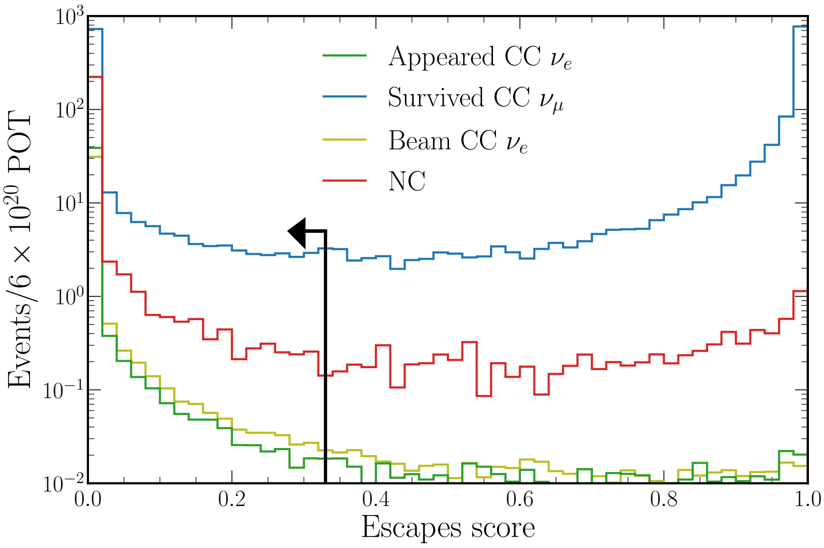


Figure 6.6: Distribution of *escapes score* output values from the trained cosmic rejection network for the different event categories. A score close to one signifies an escaped primary charged lepton like event, while a score close to zero corresponds to a contained primary charged lepton like event. The containment cut value is shown at 0.33 with the arrow indicating the events that are selected.

A future data-driven estimate, using actual CHIPS-5 data collected outside of the NuMI beam spill window, could allow for a much-improved understanding of the cosmic muon background rate. However, for now, an upper limit of < 2 cosmic muon events per year passing all the cuts is determined[1], showing that the cosmic muon rejection works well. Crucially, of all the true cosmic muon events with a *cosmic score* less than 0.9, none would be classified as signal CC ν_e events by the beam classification detailed in Section 6.1.3. Therefore, the expected cosmic muon contamination of both beam selections (CC ν_e and CC ν_μ) is expected to be negligible relative to the selected number of signal events for each and ignored for the rest of this performance evaluation.

6.1.3 Beam classification

The output values from each of the *combined category* neurons of the trained beam classification network give the probability score that an event belongs to the corresponding category. As the neuron output scores collectively sum to one, the highest-scoring neuron can be used to classify events as either CC ν_e, CC ν_μ, or NC. Figure 6.7 shows the resulting classification matrix using this approach.

[1]Using the Neyman confidence interval on a poisson distribution.

Selection	App CC ν_e	CC ν_μ	Beam CC ν_e	NC	Cosmic
Total events	44.17 ± 0.15	2045.9 ± 5.8	35.06 ± 0.08	354.7 ± 2.4	2100000 ± 4200
+ Preselection	41.21 ± 0.14	1889.5 ± 5.6	33.52 ± 0.08	243.2 ± 2.0	430000 ± 1900
+ Cosmic cut	41.10 ± 0.14	1874.4 ± 5.5	33.35 ± 0.08	241.6 ± 2.0	< 2
+ Escapes cut	40.68 ± 0.14	795.7 ± 3.5	32.86 ± 0.08	233.0 ± 2.0	< 2
Cuts Eff	$92.1 \pm 0.1\%$	$38.9 \pm 0.1\%$	$93.7 \pm 0.1\%$	$65.7 \pm 0.3\%$	$< 9.5 \times 10^{-7}$

Table 6.1: The total number of expected (weighted) events and the number that pass successive selection cuts for the different event categories. The preselection, *cosmic score* cut, and *escapes score* cut numbers are shown. The selection efficiency relative to the total number of events after all the cuts have been applied is also shown for each event category. As zero out of the 350000 cosmic events in the evaluation sample are selected after the *cosmic score* cut, an upper limit on the values is instead given. The given errors correspond to the statistical uncertainty only.

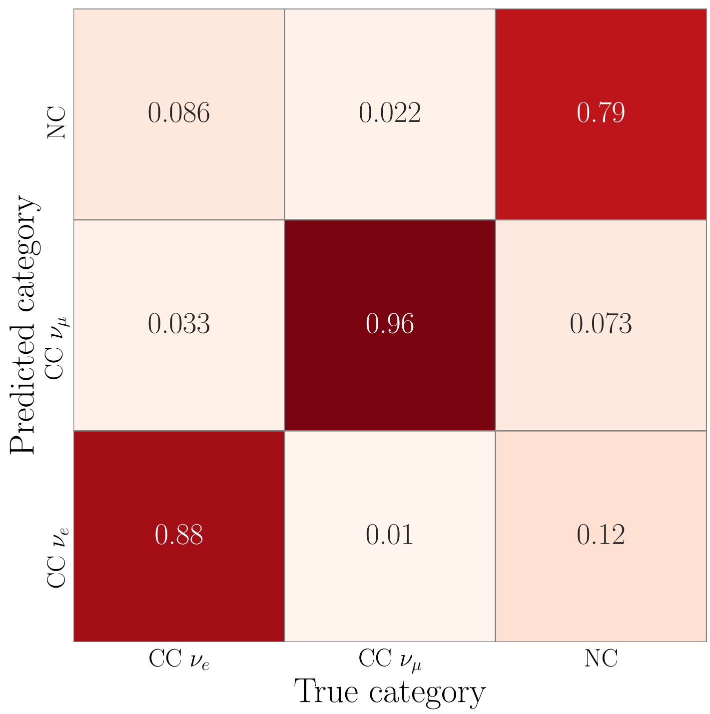


Figure 6.7: Classification matrix for the *combined category* output of the trained beam classification network. Events are simply classified using the categorical score for which they have the highest value. The numbers shown are the fraction of true category events classified into each of the three possible categories.

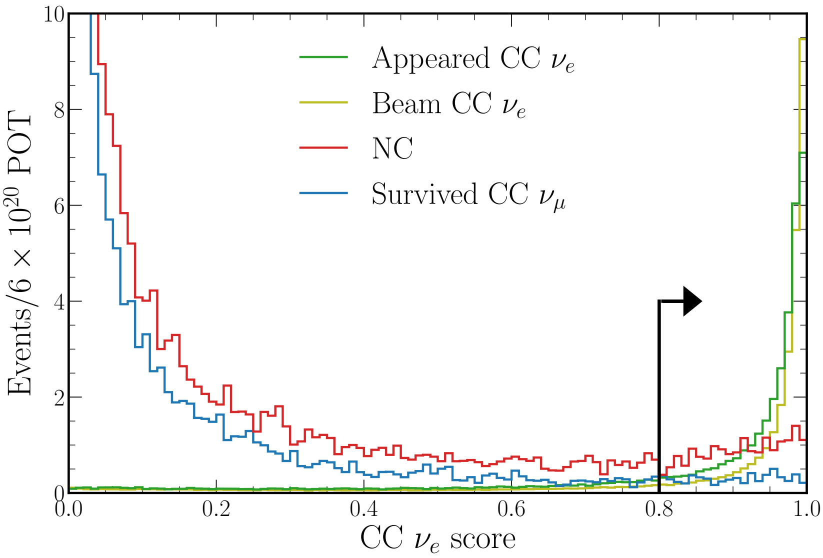


Figure 6.8: Distribution of *combined category* CC ν_e scores from the trained beam classification network for the different event categories. A score close to one signifies a CC ν_e like event. The FOM-ν_e optimised cut value is shown at 0.8 with the arrow indicating the events that are selected. The y-axis has been truncated so that the CC ν_μ and NC components are not fully visible to better show the distribution of signal CC ν_e events.

To assess the beam classification performance more rigorously, a selection score for each of the output categories, found by maximising a figure-of-merit (FOM), is instead calculated. All events with a score above this optimised value are then deemed signal. To minimise the expected measurement statistical error, the value of efficiency $\times$ purity (proportional to the square of $s/\sqrt{s+b}$) is optimised as the FOM [165]. Below, the results for both appeared CC ν_e and survived CC ν_μ selections using this methodology are presented.

CC ν_e selection

The distribution of CC ν_e scores for the different event categories are shown in Figure 6.8. A strong separation between appeared CC ν_e signal and both CC ν_μ and NC background events is achieved. As no attempt is made to separate the appeared CC ν_e signal component from the intrinsic beam CC ν_e background, both are clustered with scores close to one as expected.

The efficiency, purity, and their product (the FOM-ν_e) for CC ν_e events (both appeared and beam) as a function of selecting events above a certain CC ν_e score are shown in

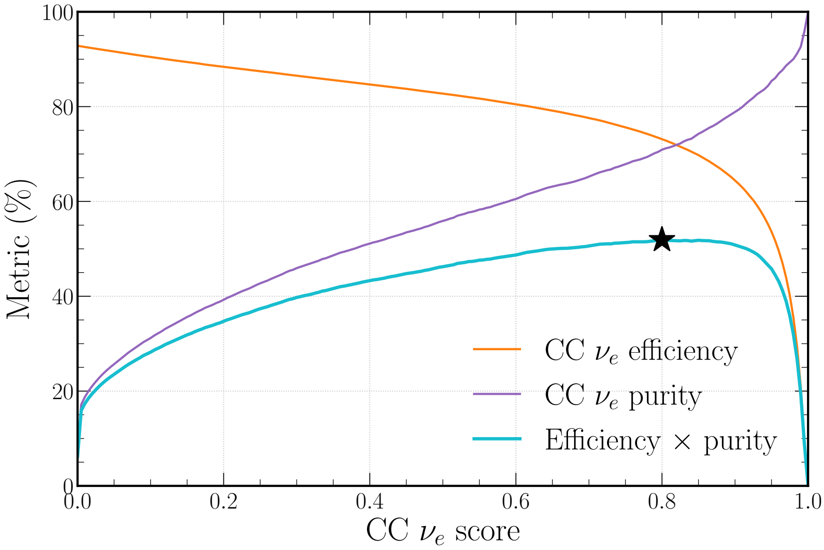


Figure 6.9: CC ν_e efficiency, purity, and efficiency $\times$ purity (FOM-ν_e) curves for different values of CC ν_e score selection. The maximum FOM-ν_e value of 0.519 ± 0.004 is indicated at a CC ν_e score of 0.8 by the star.

Figure 6.9. The FOM-ν_e is optimised by selecting events with a CC ν_e score above 0.8, achieving a value of 0.519 ± 0.004. Note that the FOM-ν_e is optimised considering both the appeared and beam CC ν_e components as signal due to their indistinguishable nature.

The total number of events, those selected by the previously mentioned cuts, and those furthermore selected by the FOM-ν_e optimised selection are shown in Table 6.2 for each event category alongside the corresponding selection efficiencies and appeared CC ν_e signal and combined CC ν_e purities. The purities are defined as the fraction of events within the selection that are true signal events. The final FOM-ν_e selected appeared CC ν_e signal purity of $38.3 \pm 0.3\%$ may appear low, but this is mainly due to the indistinguishable intrinsic beam CC ν_e contamination, note that when both CC ν_e components are considered signal, the selection purity becomes $70.9 \pm 0.6\%$.

The FOM-ν_e optimised CC ν_e selection efficiency, relative to the total number of events as a function of energy for the different event categories, is shown in Figure 6.10 alongside the signal purity. From low neutrino energies, both CC ν_e category selection efficiencies rise to a plateau of approximately 80% beginning at 4 GeV. This is expected as low energy CC ν_e events have less well-defined electron Cherenkov rings, leading to their rejection. Problematically, this turn-on behaviour cuts into the true appeared CC ν_e distribution, especially around the 1.5 GeV oscillation maximum. Future work should

Selection	CC ν_e sig	CC ν_μ bkg	CC ν_e bkg	NC bkg	Purity sig	Purity CC ν_e
Total events	44.17 ± 0.15	2045.9 ± 5.8	35.06 ± 0.08	354.7 ± 2.4	$1.78 \pm 0.02\%$	$3.19 \pm 0.03\%$
+ Cuts	40.68 ± 0.14	795.7 ± 3.5	32.86 ± 0.08	233.0 ± 2.0	$3.69 \pm 0.02\%$	$6.67 \pm 0.03\%$
+ FOM-ν_e	31.27 ± 0.12	6.0 ± 0.3	26.69 ± 0.07	17.8 ± 0.6	$38.3 \pm 0.3\%$	$70.9 \pm 0.6\%$
Cuts Eff	$92.1 \pm 0.1\%$	$38.9 \pm 0.1\%$	$93.7 \pm 0.1\%$	$65.7 \pm 0.3\%$	-	-
FOM-ν_e Eff	$70.8 \pm 0.2\%$	$0.29 \pm 0.02\%$	$76.1 \pm 0.1\%$	$5.0 \pm 0.2\%$	-	-

Table 6.2: Table showing CC ν_e selected event numbers and corresponding efficiencies for the various event categories as well as associated purities. Shown are the total event numbers, those after the preselection, *cosmic score* cut, and *escapes score* cut (Cuts), in addition to the numbers after the FOM-ν_e optimised selection with the efficiencies relative to the total number of events shown for both. Both the appeared signal CC ν_e purity and the joint appeared and beam CC ν_e purity are shown for each selection. The given errors correspond to the statistical uncertainty only.

explore whether a CC ν_e selection cut that varies with energy can lead to a greater proportion of these low energy events being selected.

Due to the abundance of selected intrinsic beam CC ν_e events at higher energies, the appeared CC ν_e purity is observed to peak at approximately 2.5 GeV (reasonably close to the oscillation maximum) before declining. Importantly, within the key signal region from 2 to 4 GeV, the appeared CC ν_e purity is $> 55\%$, larger than the $38.3 \pm 0.3\%$ across the full evaluation sample. The NC efficiency is seen to slowly increase, approaching 15% for hadronic component energies above 5 GeV; this is likely due to misidentification of high energy pions or protons as electrons. Crucially, however, within the key signal region, NC selection efficiency remains low.

The best way to understand the relative performance of the CNN CC ν_e classification is by comparison with the standard event selection presented in Section 5.2.2. The distribution of output scores from both simple neural networks used in the standard selection are shown in Figure 6.11. By optimising the selection values for both networks, 0.91 and 0.78 respectively, to maximise efficiency $\times$ purity, the standard CC ν_e selected sample is found. All events that pass the preselection, not just those shown in Figure 6.11 are used in this optimisation.

A maximum efficiency $\times$ purity of 0.132 ± 0.005 is achieved; only $\sim 25\%$ the value reached by the CNN approach. Both the combined appeared and beam CC ν_e efficiency of $34.7 \pm 0.8\%$ compared to $73.4 \pm 0.2\%$ and purity of $39.3 \pm 1.2\%$ compared to $70.9 \pm 0.6\%$ are considerably lower than that provided by the new CNN classification.

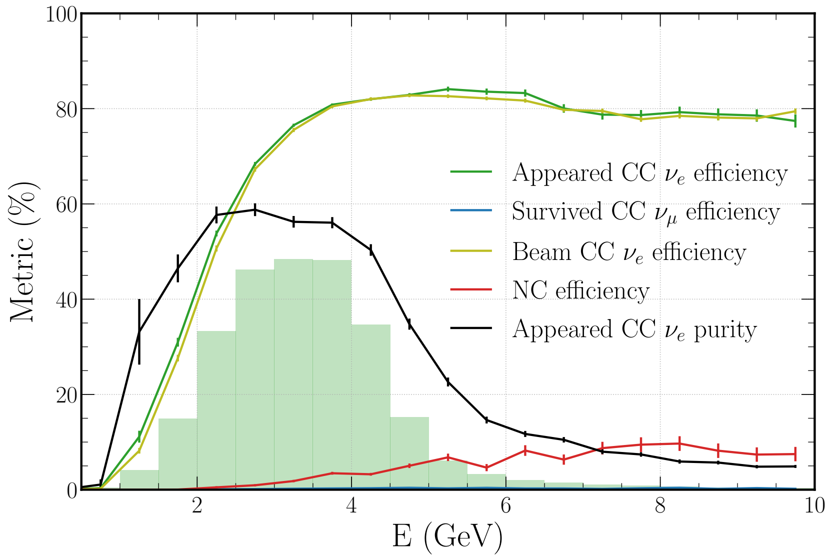


Figure 6.10: FOM-ν_e selection efficiencies relative to the total number of events for the different event categories as well as appeared CC ν_e purity as a function of energy. All CC categories are shown in terms of the true neutrino energy, while NC events are shown in terms of the true hadronic component energy. The survived CC ν_μ efficiency is so low it is barely visible near zero. For reference, the true appeared CC ν_e neutrino energy distribution is shown in green.

Furthermore, the new appeared CC ν_e signal efficiency of $70.8 \pm 0.2\%$ compares well to the 62% and 64% achieved by the NOvA and T2K CC ν_e selections, respectively. However, purity is significantly lower at $38.3 \pm 0.3\%$ compared to the 78% and 80% reached by NOvA and T2K [54, 166]. A large proportion of this difference can be explained by the lower neutrino energies and greater off-axis angles at which these experiments operate. Not only does this increase the proportion of easy to identify CC-QEL events, but it also reduces the indistinguishable beam CC ν_e contamination.

CC ν_μ selection

The distribution of CC ν_μ scores for the different event categories are shown in Figure 6.12. Excellent separation between appeared CC ν_μ signal and both CC ν_e components and NC background is achieved. For high CC ν_μ scores (close to one) the difference between signal and background event counts is approximately three orders of magnitude.

The efficiency, purity, and their product (the FOM-ν_μ) for CC ν_μ events as a function of selecting events above a certain CC ν_μ score are shown in Figure 6.13. FOM-ν_μ is optimised by selecting events with a CC ν_μ score above 0.315, achieving a value of

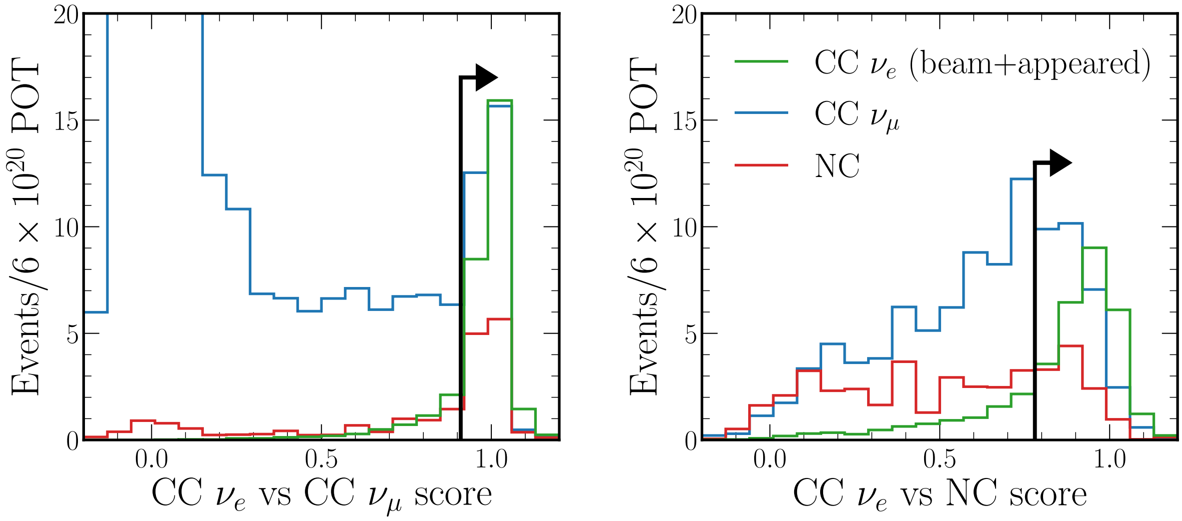


Figure 6.11: Distributions of CC ν_e vs CC ν_μ (left) and CC ν_e vs NC (right) output scores from the two standard event selection neural networks for the different event categories. A score close to one signifies a CC ν_e like event in both cases. Each plot shows events which have passed both the preselection and the optimised cut from the other network. This is done to better show the events which the network in question rejects. Selected CC ν_e events are shown by the arrows. The y-axis for the CC ν_e vs CC ν_μ distributions has been truncated so that the CC ν_μ component is not fully visible to better show the distribution of signal CC ν_e events. Due to the long reconstruction time required, a smaller evaluation sample is used here.

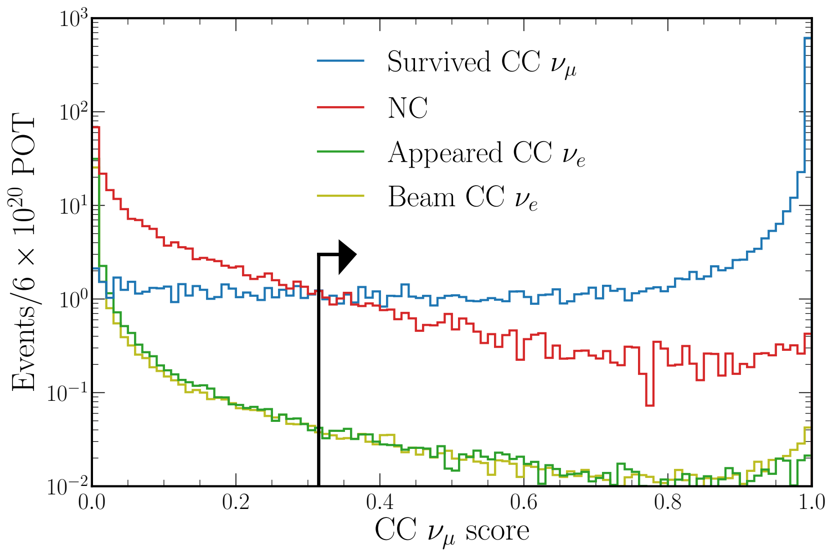


Figure 6.12: Distribution of *combined category* CC ν_μ scores from the trained beam classification network for the different event categories. A score close to one signifies a CC ν_μ like event. The FOM-ν_μ optimised cut value is shown at 0.315 with the arrow indicating the events that are selected.

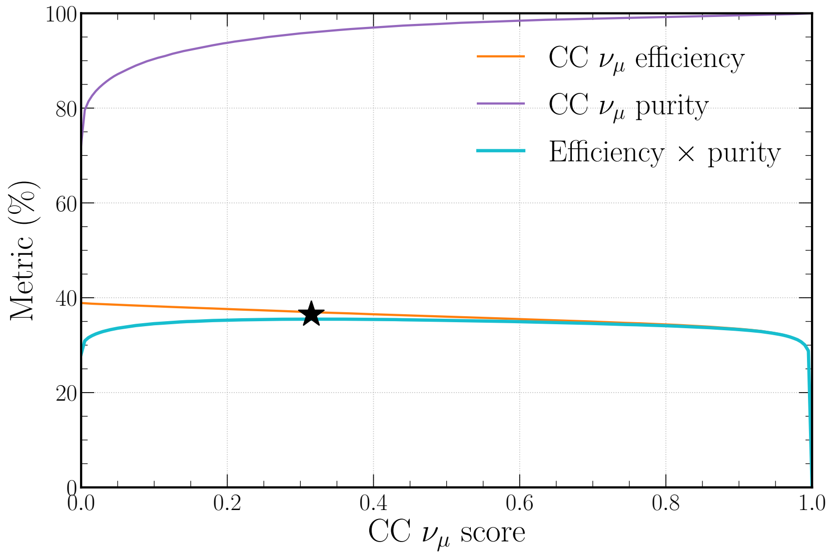


Figure 6.13: CC ν_μ efficiency, purity, and efficiency × purity (FOM-ν_μ) curves for different values of CC ν_μ score selection. The maximum FOM-ν_μ value of 0.365 ± 0.002 is indicated at a CC ν_μ score of 0.315 by the star.

0.365 ± 0.002. The total number of events, those selected by the previously mentioned cuts, and those furthermore selected by the FOM-ν_μ optimised selection are shown in Table 6.3 for each event category alongside the corresponding selection efficiencies and CC ν_μ signal purity.

The signal efficiency of $37.0 \pm 0.1\%$ compares well to the 31% and 36% achieved by the NOvA and T2K CC ν_μ selections, respectively [54, 166]. This is also the case for the signal purity of $96.0 \pm 0.1\%$ compared to the 98.6% and 94% purities of the NOvA and T2K selections. Although the final signal efficiency is low, this is desirable to ensure events are fully contained for energy estimation. When considering just those CC ν_μ events for which the primary charged muon is contained within the detector volume at the truth level, an $87.5 \pm 0.1\%$ selection efficiency is achieved.

The FOM-ν_μ optimised CC ν_μ selection efficiency, relative to the total number of events as a function of energy for the different event categories, is shown in Figure 6.14. Survived CC ν_μ selection efficiency peaks at just below 2 GeV before slowly declining, this is explained by higher energy events being less likely to have their primary charged muon fully contained within the detector. Of interest is the expected dip in the otherwise very high ($> 90\%$) CC ν_μ purity at approximately 1.5 GeV, corresponding to the oscillation maximum (shown in Figure 2.10). As in the CC ν_e selection case, the NC efficiency is

Selection	CC ν_μ sig	App CC ν_e bkg	Beam CC ν_e bkg	NC bkg	Purity sig
Total events	2045.9 ± 5.8	44.17 ± 0.15	35.06 ± 0.08	354.7 ± 2.4	$82.5 \pm 0.2\%$
+ Cuts	795.7 ± 3.5	40.68 ± 0.14	32.86 ± 0.08	233.0 ± 2.0	$72.2 \pm 0.2\%$
+ FOM-ν_μ	756.4 ± 3.4	1.293 ± 0.03	1.315 ± 0.02	29.0 ± 0.7	$96.0 \pm 0.1\%$
Cuts Eff	$38.9 \pm 0.1\%$	$92.1 \pm 0.1\%$	$93.7 \pm 0.1\%$	$65.7 \pm 0.3\%$	-
FOM-ν_μ Eff	$37.0 \pm 0.1\%$	$2.9 \pm 0.1\%$	$3.8 \pm 0.1\%$	$8.2 \pm 0.2\%$	-

Table 6.3: Table showing CC ν_μ selected event numbers and corresponding efficiencies for the various event categories as well as the associated signal purity. Shown are the total event numbers, those after the preselection, *cosmic score* cut, and *escapes score* cut (Cuts), in addition to the numbers after the FOM-ν_μ optimised selection with the efficiencies relative to the total number of events shown for both. The CC ν_μ signal purity is also shown for each selection. The given errors correspond to the statistical uncertainty only.

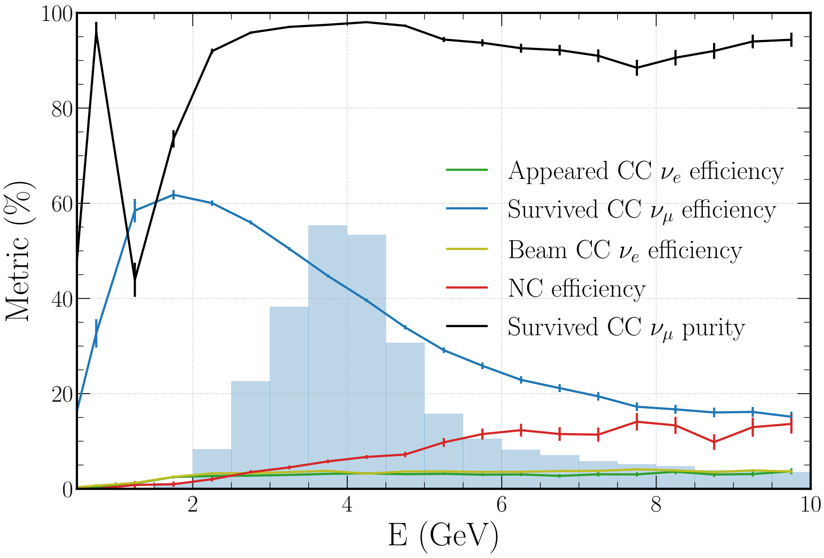


Figure 6.14: FOM-ν_μ selection efficiencies relative to the total number of events for the different event categories as well as the survived CC ν_μ purity as a function of energy. All CC categories are shown in terms of the true neutrino energy, while NC events are shown in terms of the true hadronic component energy. For reference, the true survived CC ν_μ neutrino energy distribution is shown in blue.

seen to rise with energy, again likely due to misidentification of energetic protons and pions.

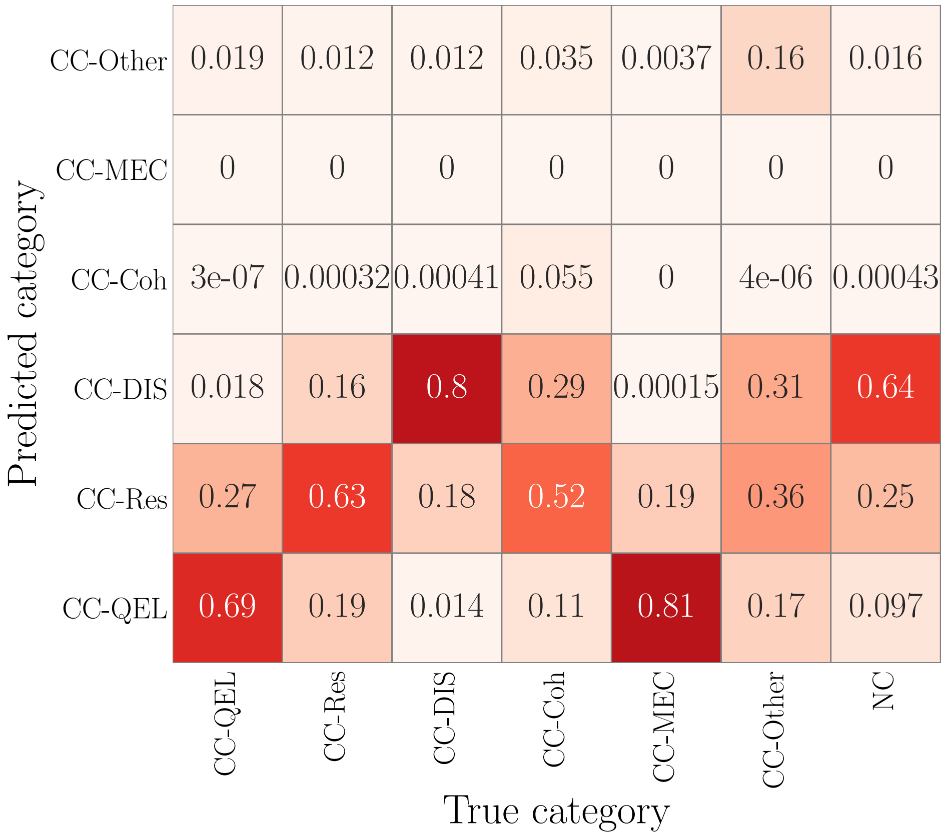


Figure 6.15: Classification matrix for the *CC category* output of the trained beam classification network. Shown are events that have either been selected by the CC ν_e or CC ν_μ selection. Events are simply classified using the categorical score for which they have the highest value. The numbers shown are the fraction of true category events classified into each of the six possible categories.

Interaction type classification

Using the *CC category* output of the trained beam classification network, the CC interaction type (used for energy estimation in Section 6.1.4) for both CC ν_e and CC ν_μ selected events can be determined. As in the *combined category* output case, the highest-scoring neuron can be used for classification, resulting in the matrix shown in Figure 6.15. Note that only events which are selected by either the CC ν_e or CC ν_μ selection are shown.

Reasonable classification accuracy greater than 60% is achieved across the three dominant interaction types, CC-QEL, CC-Res, and CC-DIS. The less common CC-Coh and CC-MEC types are found to be commonly misidentified as CC-Res and CC-QEL respectively, likely due to the imbalanced training dataset and their corresponding

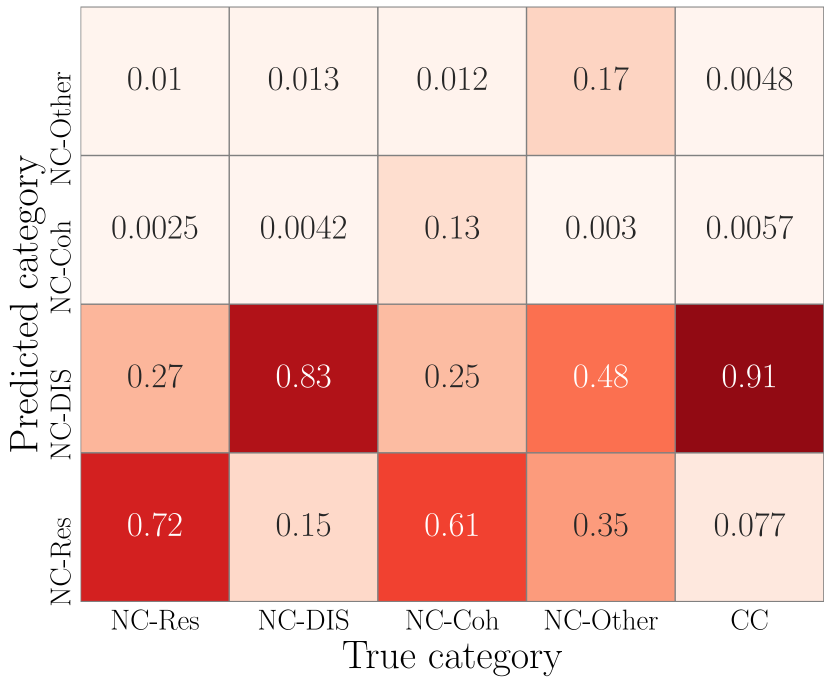


Figure 6.16: Classification matrix for the *NC category* output of the trained beam classification network. Events are simply classified using the categorical score for which they have the highest value. The numbers shown are the fraction of true category events classified into each of the four possible categories.

topological similarities. Background NC events which pass either CC ν_e or CC ν_μ selection (commonly high in energy) are found to be typically classified as CC-DIS; this is expected as they commonly contain multiple energetic particles in the final state.

For completeness, the *NC category* classification matrix is shown in Figure 6.16 for events that are neither classified as CC ν_e or CC ν_μ. As in the CC case, the dominant interaction types NC-Res and NC-DIS are classified well, with NC-Coh events typically being classified as NC-Res. Of the CC events that are not selected, the vast majority are classified as NC-DIS events. Similarly to before, this is likely due to multiple energetic particles in the final state, for which the beam classification network determined no clear charged lepton.

6.1.4 Energy and vertex estimation

By using the CC interaction type classification just presented, the differences between CC interaction types can be exploited to improve neutrino energy, charged lepton energy, and interaction vertex position and time estimation. For events classified as either CC ν_e

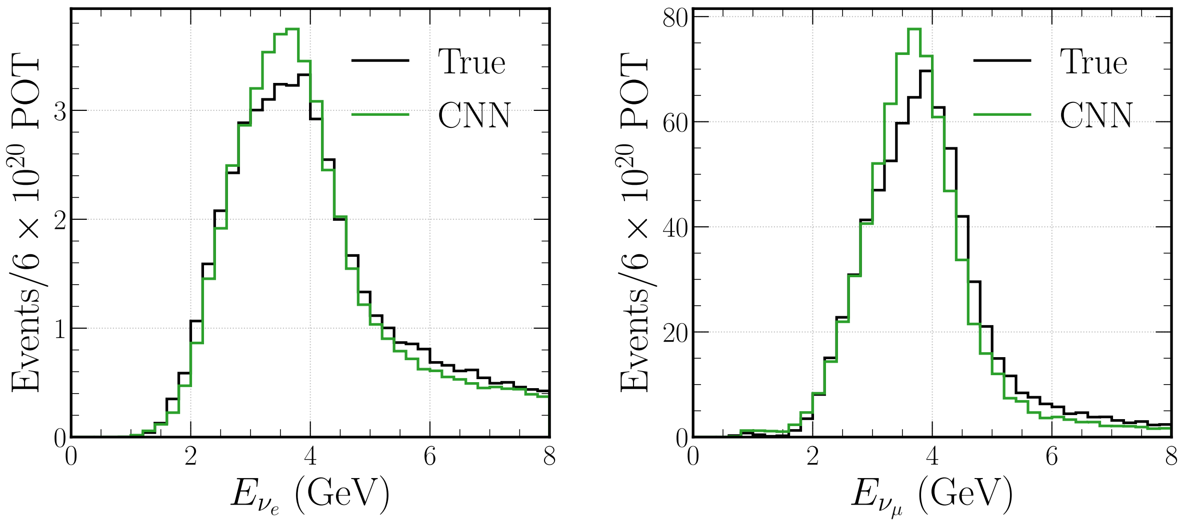


Figure 6.17: Distributions of true and CNN estimated neutrino energy for CC ν_e (left) and CC ν_μ (right) beam events. Only true CC ν_e and CC ν_μ events that are also selected by the CC ν_e and CC ν_μ selections respectively are shown.

or CC ν_μ with an associated *CC category* interaction type, the corresponding bespoke trained network outlined in Section 5.5.3 is used for estimation.

Only three networks for each neutrino type are trained, one for each of the dominant interaction types CC-QEL (and CC-MEC), CC-Res, and CC-DIS. For events not classified by the *CC category* output as one of these categories, such as CC-Coh or CC-Other, the CC-Res network is used as it is the most topologically similar interaction type.

Energy estimation

The distributions of CNN estimated (*neutrino energy* output) and true ν_e and ν_μ neutrino energies for true CC ν_e and CC ν_μ events respectively that are also selected by their corresponding CC selection are shown in Figure 6.17. The CNN estimated distributions match the truth well across the full range of neutrino energies expected within Chips-5, except in the peak regions where the truth distribution shape is not fully captured.

As the training samples contain a spectrum of events typical of the beam, it is important to check that the CNN neutrino energy estimation is not simply predicting an energy close to the expected peak beam energy. The probability of a CNN estimated neutrino energy given a true neutrino energy is shown in Figure 6.18 for both CC ν_e and CC ν_μ events. CNN estimated energy is found to be roughly equivalent to the true energy across the full range of expected Chips-5 beam energies, proving the desired response.

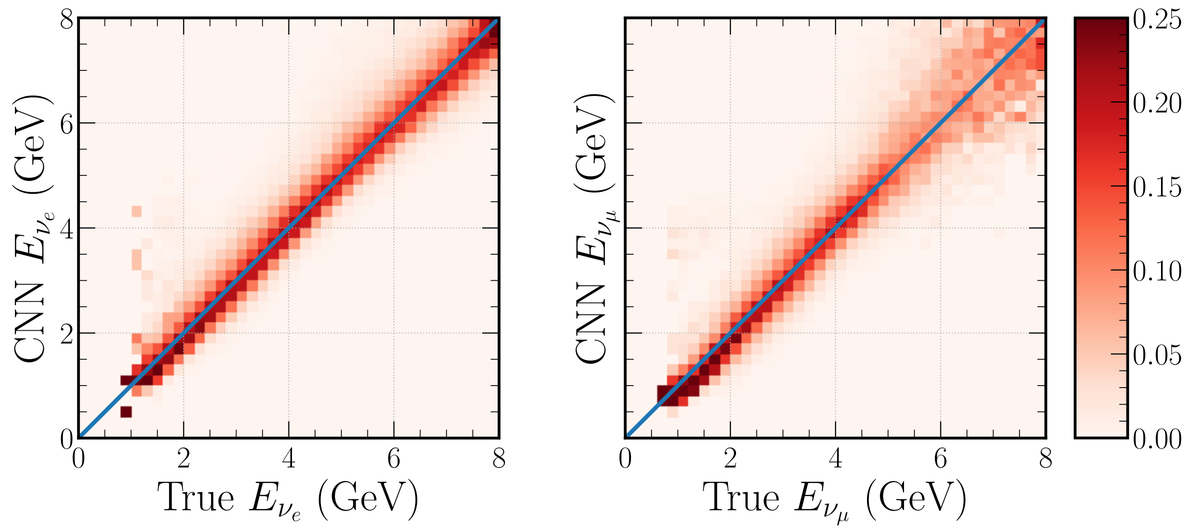


Figure 6.18: Probability of CNN estimated neutrino energy given a true neutrino energy for CC ν_e (left) and CC ν_μ (right) beam events, with their equality shown in blue. Only true CC ν_e and CC ν_μ events that are also selected by the CC ν_e and CC ν_μ selections respectively are shown.

Event type	All	QEL component	Res component	DIS component
CC ν_e	$24.0 \pm 0.3\%$	$16.4 \pm 0.4\%$	$24.3 \pm 0.2\%$	$31.9 \pm 0.2\%$
CC ν_μ	$29.4 \pm 0.4\%$	$14.1 \pm 0.3\%$	$27.2 \pm 0.3\%$	$34.9 \pm 0.3\%$

Table 6.4: Summary of CC ν_e and CC ν_μ neutrino energy FWHM values. Shown for each sample are the FWHM values for all selected signal events and the three dominant interaction type components, QEL, Res, and DIS. The FWHM values are calculated from the fractional energy difference distributions shown in Figure 6.19 and Figure 6.20, and given as a fractional percentage. The given errors correspond to the statistical uncertainty only.

To fully understand CNN energy estimation performance, histograms of ratios of differences between CNN estimated (reco) and true neutrino energy to true neutrino energy for both true CC ν_e and CC ν_μ beam events that are also selected by their full corresponding CC selection (including preselection, escapes, and cosmic cuts) are shown in Figure 6.19. Similar distributions splitting the signal components by interaction type are shown in Figure 6.20. Furthermore, the Full Width Half Maximum (FWHM) values derived from these plots for both CC ν_e and CC ν_μ events are shown in Table 6.4, given as a fractional percentage.

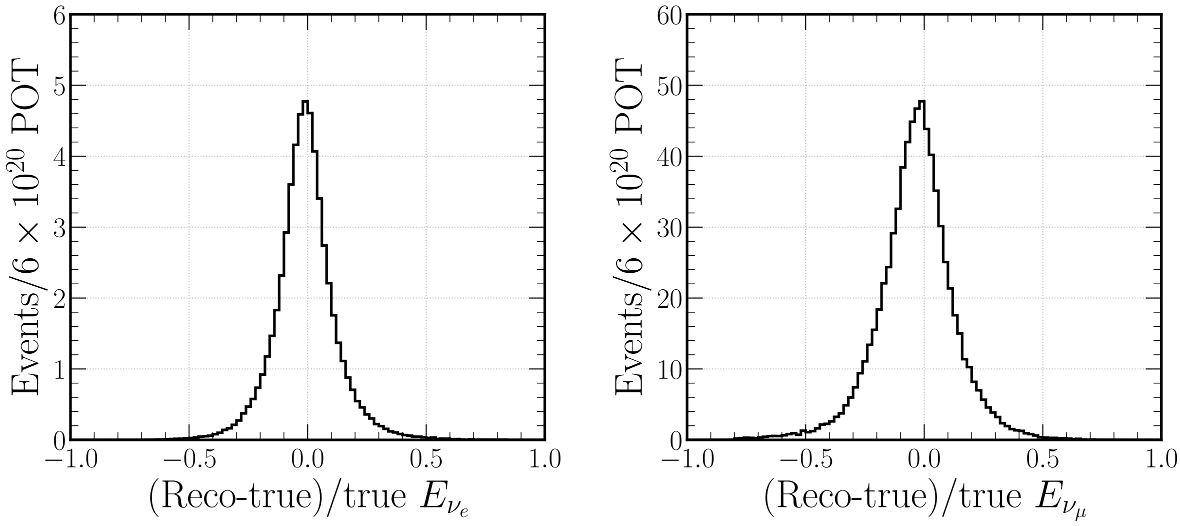


Figure 6.19: Distributions of (reco-true)/true neutrino energies for both selected CC ν_e (left) and CC ν_μ (right) signal beam events.

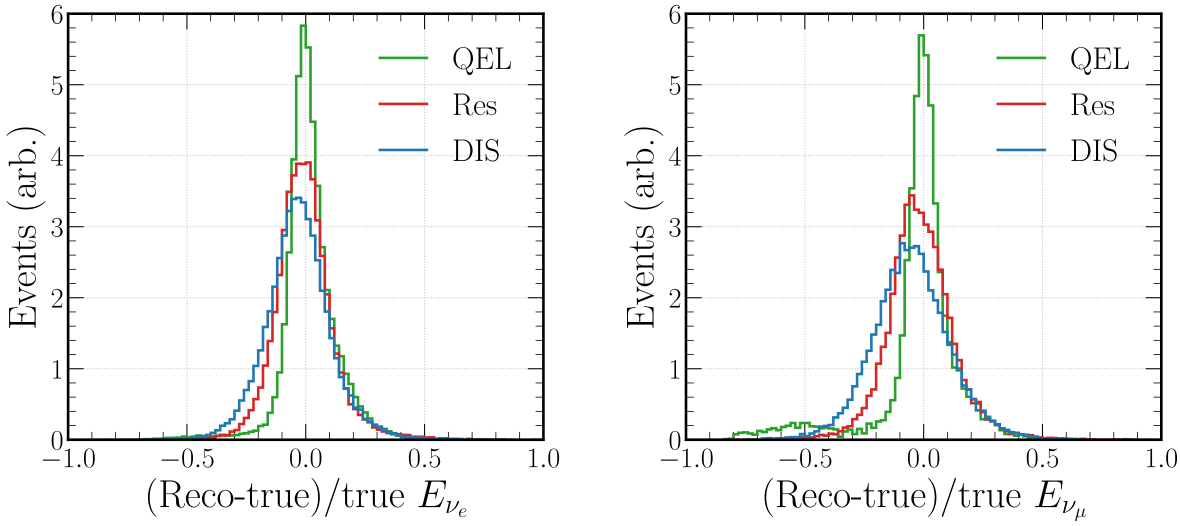


Figure 6.20: Distributions of (reco-true)/true neutrino energies for both selected CC ν_e (left) and CC ν_μ (right) signal beam events by interaction type. The relative number of events between interaction types has been scaled for clearer comparison.

The interaction type FWHM values follow the expected pattern, with the simple to reconstruct single charged lepton QEL interactions achieving a smaller value than multi-particle DIS events. Furthermore, when the approximate resolution is derived from the FWHM values[2] the resolutions of $10.2 \pm 0.2\%$ and $12.5 \pm 0.2\%$ for CC ν_e and CC ν_μ respectively are comparable to the resolutions obtained by NOvA of 10.7% and 9.1% [54].

The energy dependence of the energy resolution is explored by finding the means and FWHM values of the (reco-true)/true signal neutrino energy distributions shown in Figure 6.20 in 1 GeV wide true neutrino energy bins. These values are shown in Figure 6.21 and Figure 6.22 for CC ν_e and CC ν_μ events respectively, split by interaction type.

Reasonably significant bias in the means is observed with respect to the true neutrino energy for both CC ν_e and CC ν_μ events. This is particularly true of DIS events, but still significant for Res and QEL events. As the energy spectrum of events used for training peaks, this is expected. Any future energy estimation work should consider using a flat energy spectrum for training, as done in reference [128].

The FWHM is seen to decrease with true neutrino energy for all interaction types, except for energies above the flux peak. Again, this is most likely due to the spectrum of events used in the training sample. A contributing factor, however, will be due to the inability of PMTs to distinguish between numbers of incident photons at higher counts, more likely at higher energies.

As in the CC ν_e selection case, the best way to understand the relative performance of the energy estimation is by comparison with the standard CHIPS reconstruction presented in Section 5.2.1. Although the standard reconstruction does not attempt to estimate the neutrino energy, the energy of the primary charged lepton in each CC event is predicted. The value can be compared to the *charged lepton energy* output of the energy estimation CNNs. Histograms of ratios of differences between CNN estimated (reco) and true charged lepton energy to true charged lepton energy for both CC ν_e and CC ν_μ beam QEL events are shown in Figure 6.23.

A significant improvement is made using the new CNN approach. FWHM lepton energy values 30% and 39% the size of the standard reconstruction values for CC ν_e and CC ν_μ QEL events respectively is achieved, at $10.0 \pm 0.1\%$ and $9.0 \pm 0.1\%$, in their fractional percentage form. When the approximate resolution is found from the CC ν_e

[2] The approximate resolution, given by the standard deviation σ is found by dividing the FWHM by $2\sqrt{2 \ln 2} \approx 2.355$.

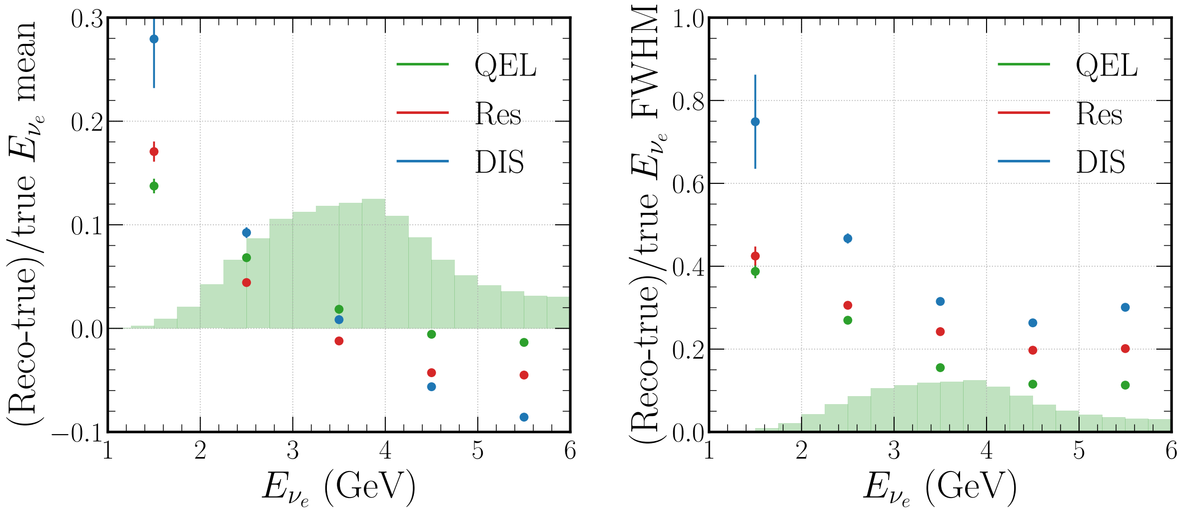


Figure 6.21: Means (left) and FWHM (right) values from distributions of (reco-true)/true neutrino energy for CC ν_e events across a range of 1 GeV wide true neutrino energy bins and split by interaction type. The true distribution of CC ν_e events is shown in green.

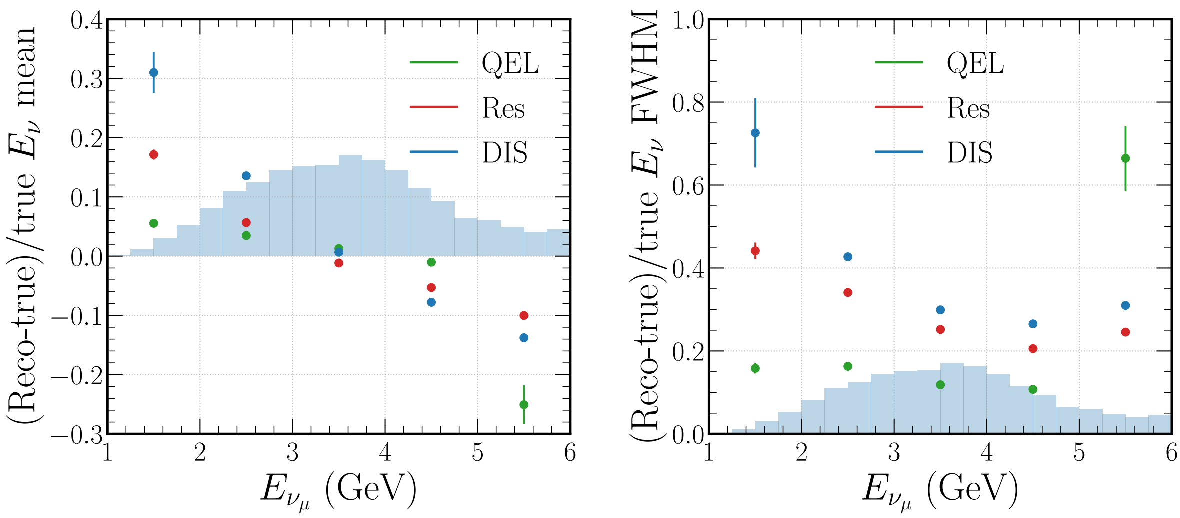


Figure 6.22: Means (left) and FWHM (right) values from distributions of (reco-true)/true neutrino energy for CC ν_μ events across a range of 1 GeV wide true neutrino energy bins and split by interaction type. The true distribution of CC ν_μ events is shown in blue.

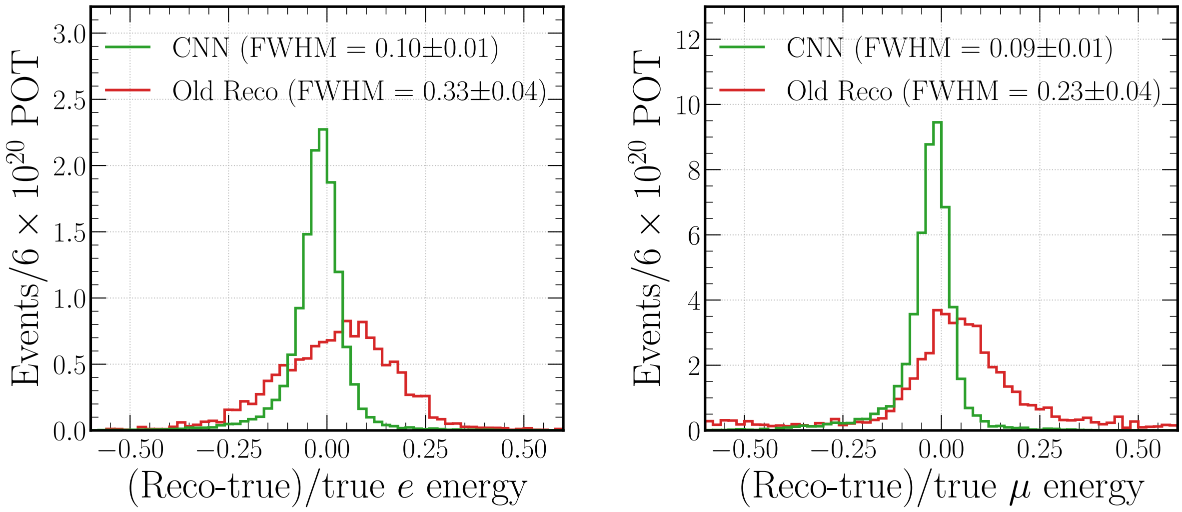


Figure 6.23: Distributions of (reco-true)/true primary charged lepton energies for both CC ν_e (left) and CC ν_μ (right) beam events for both the new CNN approach and standard (old) methods. Only QEL events are shown for clearer comparison with the standard reconstruction methods. The FWHM values of each distribution are indicated in the legends.

FWHM value ($4.2 \pm 0.1\%$), it compares well to the $\sim 2.5\%$ CC QEL charged lepton energy resolution reached by the Super-Kamiokande fiTQun algorithm [137]. Impressive, given the significant differences in detector design.

Interaction vertex estimation

The *interaction vertex x-position*, *interaction vertex y-position*, *interaction vertex z-position*, and *interaction time* outputs from the energy estimation networks can also be used. Although not employed in this work, future analyses may require accurate fiducial volume cuts or separation of events in time, therefore, strong performance is desirable. The CNN estimated (reco) minus truth distributions are shown for selected signal CC ν_e and CC ν_μ QEL events in Figure 6.24 and Figure 6.25, respectively, with the standard reconstruction method distributions shown for comparison.

Comparable resolutions are achieved for CC ν_e events, while CC ν_μ events display considerable improvements in interaction vertex z-position and time prediction. The Super-Kamiokande fiTQun algorithm reaches a resolution on their interaction vertex reco minus true distributions of 20 cm for CC ν_e and 15 cm for CC ν_μ events [137], compared to the approximately 50 cm and 70 cm from this work. Although these are large absolute

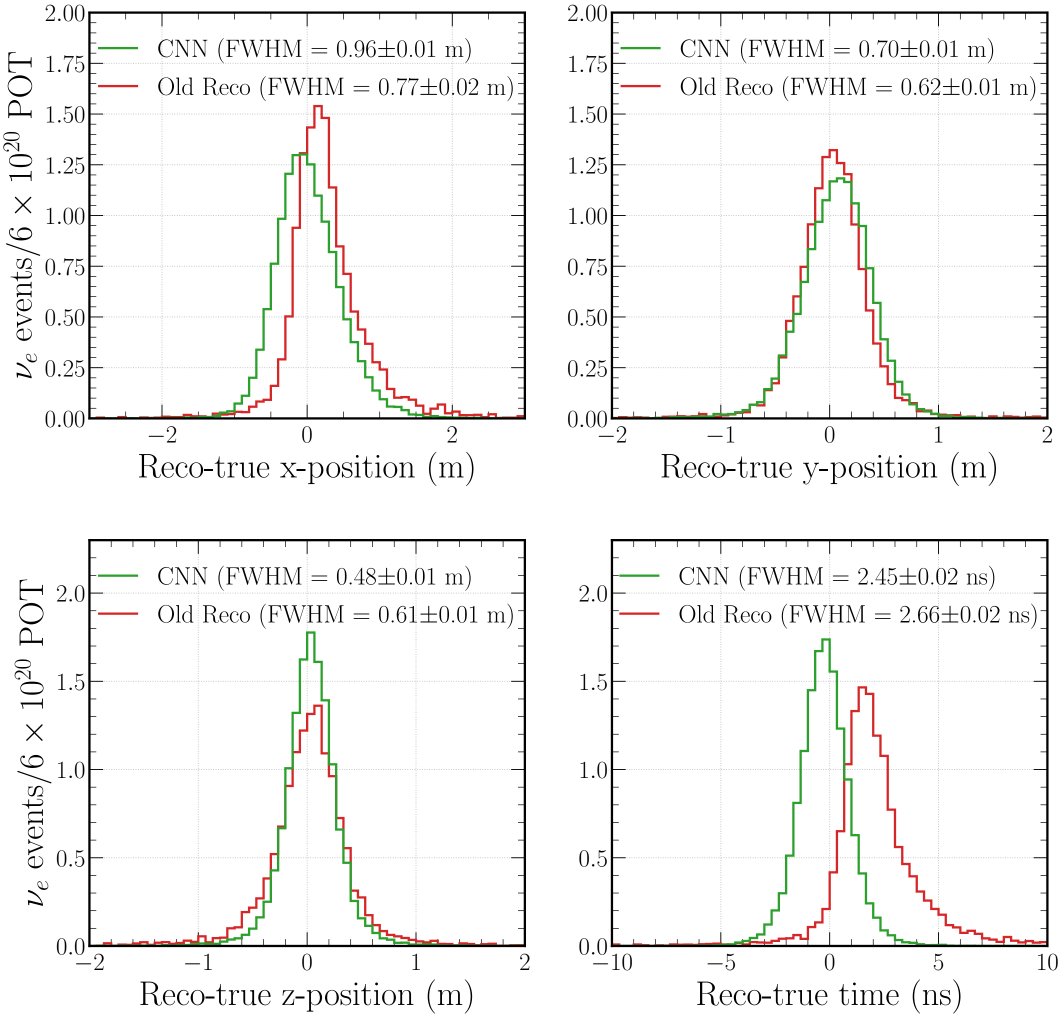


Figure 6.24: Reco-true distributions for the interaction vertex position components and time for CC ν_e QEL events. Both the distributions for the new CNN approach and standard (old) reconstruction methods are shown. The FWHM values of each distribution are indicated in the legends.

differences, relative to the detector size, they still allow for sufficient localisation of interaction vertices.

The charged lepton energy, interaction vertex position, and interaction vertex time resolutions also display the clear advantage of the CNN approach. Long tails and biased means are common in the distributions associated with the standard reconstruction methods when compared to the generally symmetric distributions of the CNN approach. This difference, which is particularly stark for non-QEL event types, highlights how both the limited inputs of the standard reconstruction and the need for a predefined hypothesis can easily bias the outputs and not generalise well to all event types.

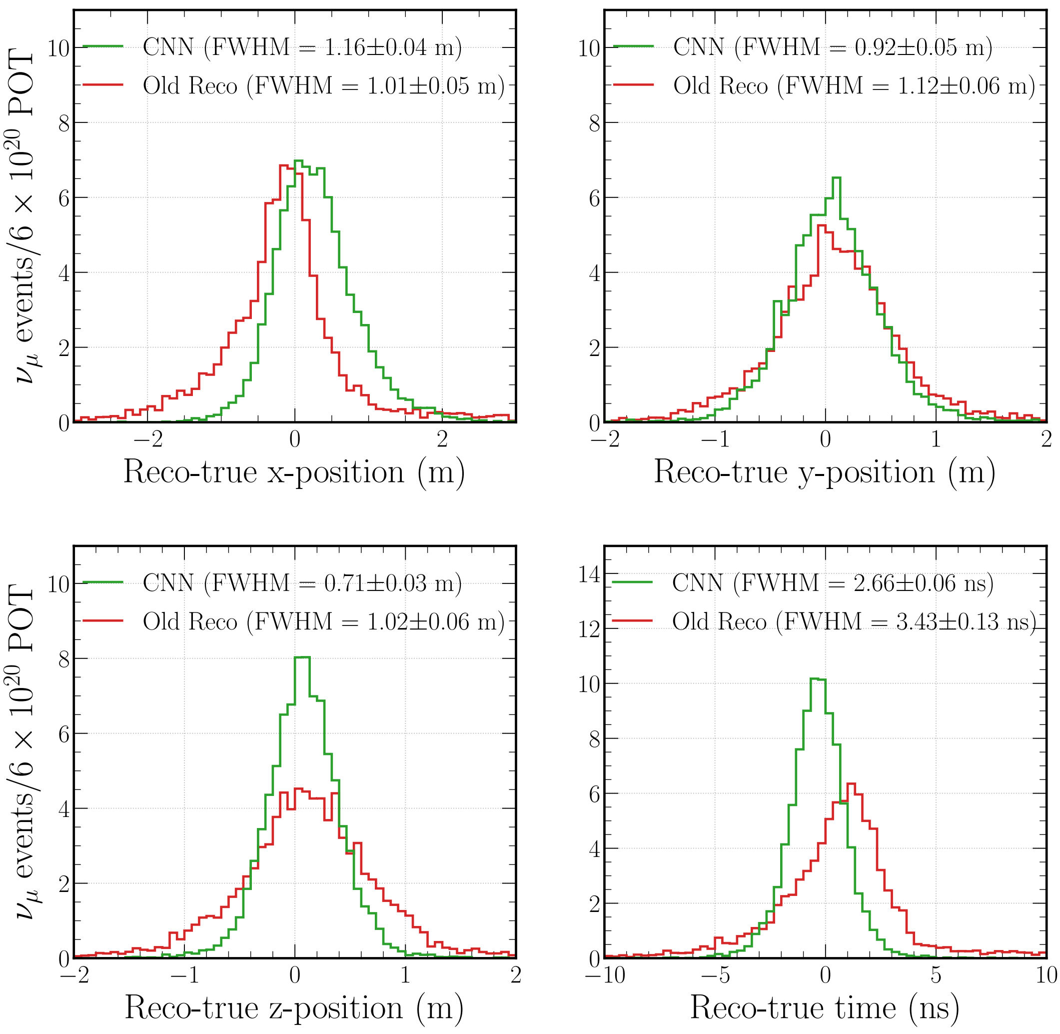


Figure 6.25: Reco-true distributions for the interaction vertex position components and time for CC ν_μ QEL events. Both the distributions for the new CNN approach and standard (old) reconstruction methods are shown. The FWHM values of each distribution are indicated in the legends.

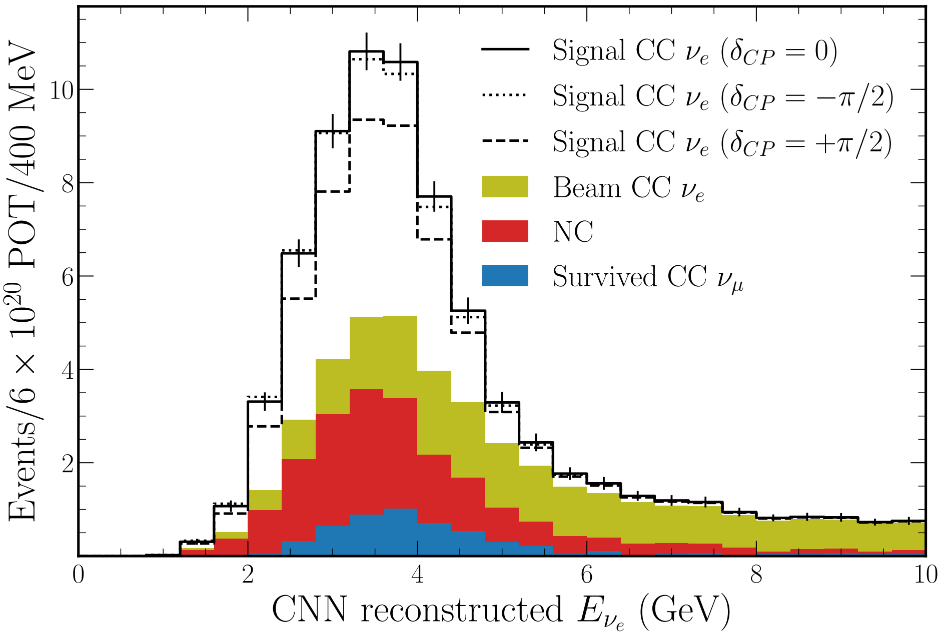


Figure 6.26: Distribution of CNN reconstructed ν_e energies for CC ν_e selected events. The appeared CC ν_e signal component as well as the intrinsic beam CC ν_e, NC and survived CC ν_μ background components are shown stacked to generate the full distribution. Expected totals are show in each bin for $\delta_{CP} = -\pi/2$, 0, and $+\pi/2$.

6.1.5 Combined performance

By combining CC ν_e and CC ν_μ selections with neutrino energy estimation, the final selected spectrum of events within CHIPS-5 running for a year can be estimated. These are shown in Figure 6.26 and Figure 6.27 for CC ν_e and CC ν_μ selections respectively. Although a detailed CHIPS-5 sensitivity analysis is not included in this work, the expected curves for different values of δ_{CP} are shown in the CC ν_e case for interest.

6.2 Explainability

A common and justified concern with CNNs is their tendency to be used as a black box (inputs in, outputs out) with no understanding of their inner working. For detailed physics analyses, this can have significant confidence implications for the final results. Although difficult quantitatively, qualitative assessments of the trained networks can go a long way to proving they behave as desired. Here, a collection of studies, which aim to explain the inner workings of the trained CNNs presented in this work are described.

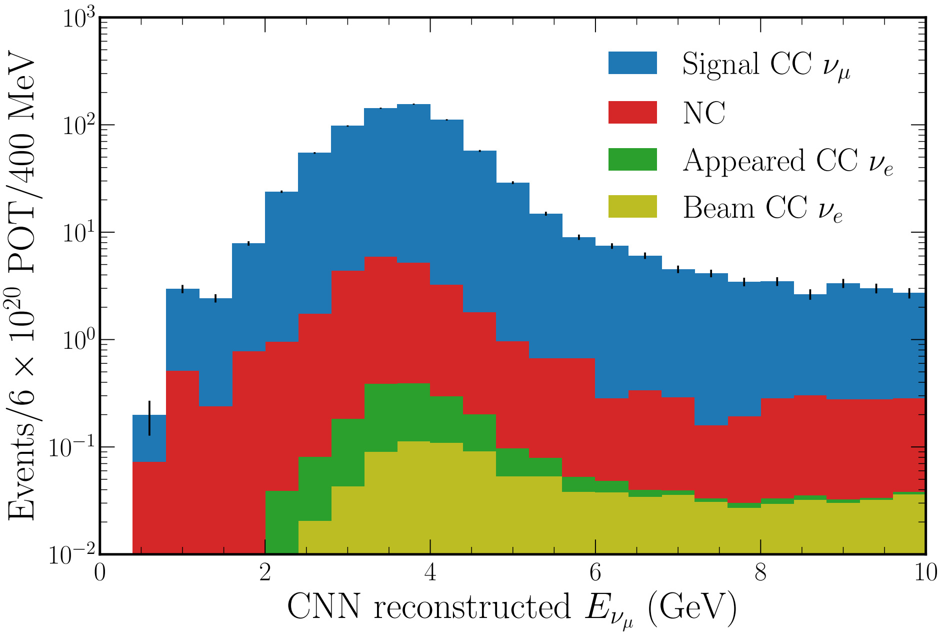


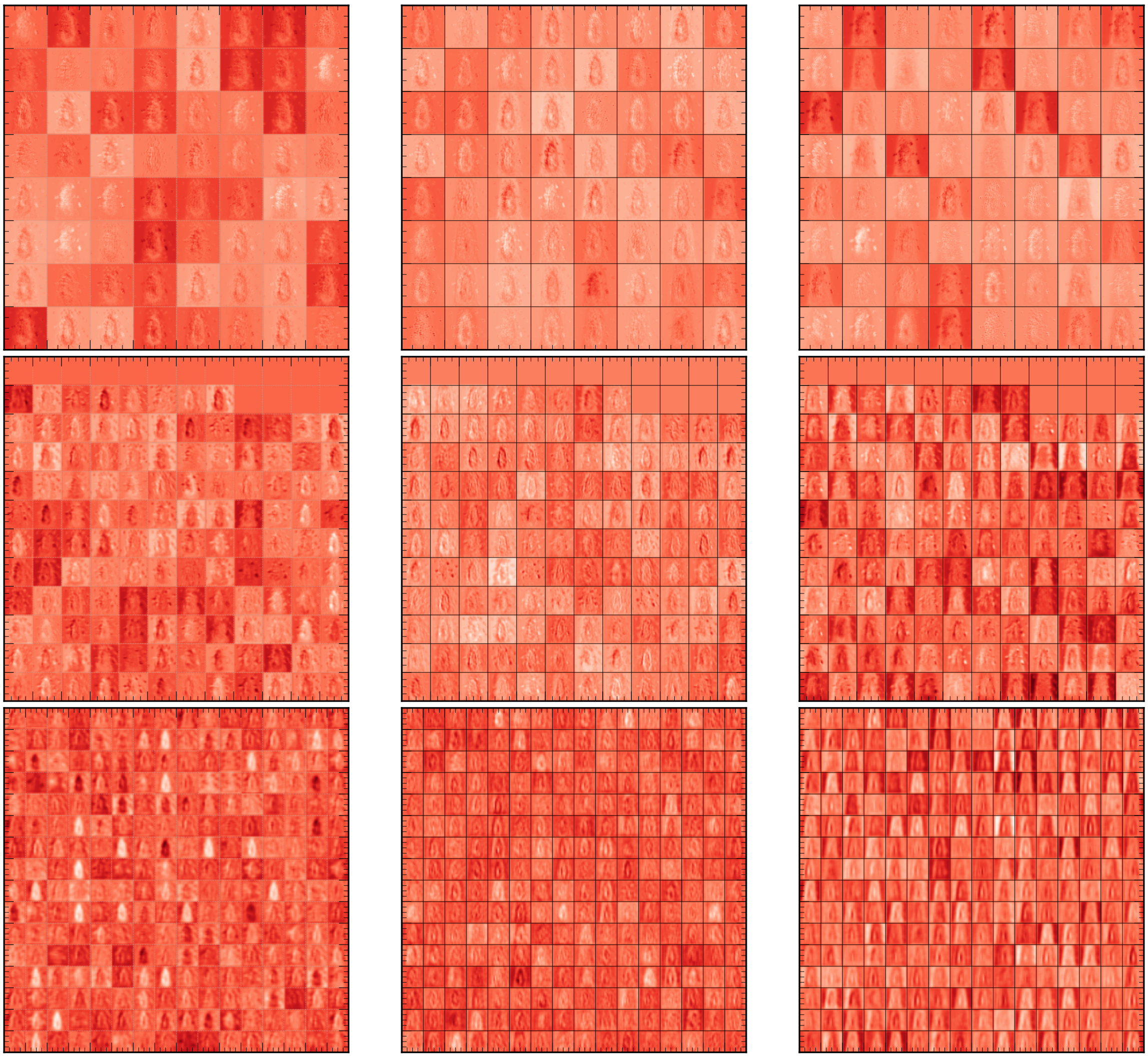
Figure 6.27: Distribution of CNN reconstructed ν_μ energies for CC ν_μ selected events. The survived CC ν_μ signal component as well as the appeared CC ν_e, intrinsic beam CC ν_e, and NC background components are shown stacked to generate the full distribution.

6.2.1 Feature map visualisation

Visualisations of the output feature maps from the first, second, and third VGG blocks for each of the trained networks (cosmic rejection, beam classification, and energy estimation) are shown in Figure 6.28, using the event shown in Figure 6.29 as input. Learnt Cherenkov ring features are observed: ring edges, ring holes, outlying hits, Hough peaks, and a myriad of combinations are seen. Furthermore, there are clear differences between the specific networks, proving each network learns those features found to be important for its tasks.

6.2.2 t-SNE visualisation

Another technique to analyse trained CNNs is t-Distributed Stochastic Neighbour Embedding (t-SNE) [167]. The t-SNE procedure is an unsupervised learning algorithm to visualise the learnt high-dimensional feature-space of a trained network in a lower number of dimensions. It accomplishes this by clustering events with similar features nearby in two-dimensional space and separating events with dissimilar features. Here, the outputs



(a) Cosmic rejection **(b)** Beam classification **(c)** Energy estimation

Figure 6.28: Visualisations of the feature map outputs, using as input the event shown in Figure 6.29. Shown are all the activated feature map outputs from the first (top), the second (middle), and the third (bottom) VGG blocks for each of the trained network types, with each small box representing an individual feature map. A chipsnet architecture with only a single branch and all three input event maps stacked into a single three-channel input image is used for simplicity.

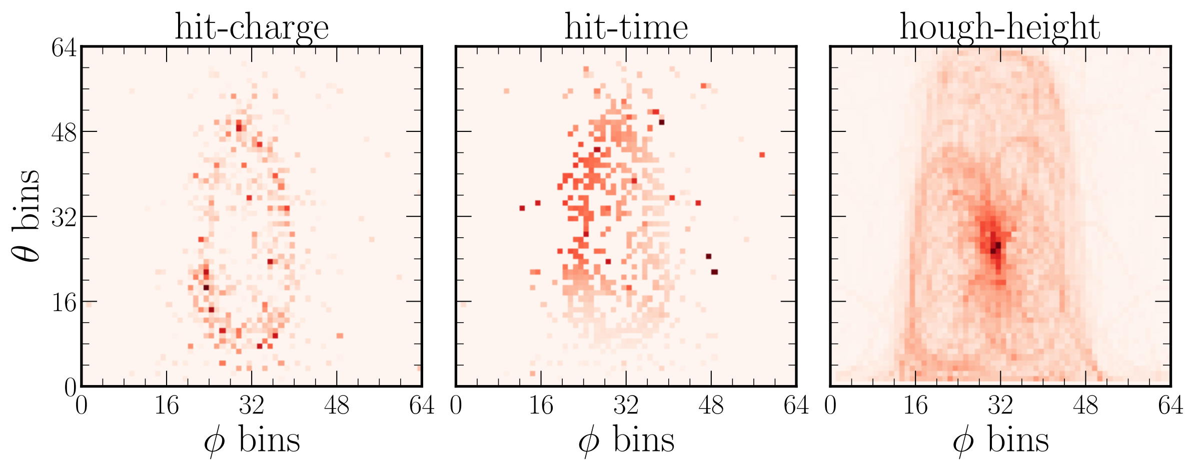


Figure 6.29: Three map representation of a CC quasi-elastic ν_e event. Initiated by a ν_e of energy 2.4 GeV with a final state e^- of energy 1.6 GeV.

from the last fully connected layer before the output layer (with 512 dimensions) are used as input, as they provide the final representation of the learnt network features.

Visualisations of the t-SNE algorithm outputs, when applied to both the trained cosmic rejection and beam classification networks, are shown in Figure 6.30 and Figure 6.31, respectively. The very strong cosmic-like to beam-like separation presented in Section 6.1.2 is clear from the cosmic rejection visualisation. Conversely, for the beam classification network, the separation is weaker, with major overlap between categories, especially for CC ν_e and NC events.

For the beam classification network, three events, labelled in the t-SNE space of Figure 6.31, are shown in Figure 6.32. Each event is highly representative of its class, achieving a high respective *combined category* score. Both the CC ν_e and NC events are typical of that expected. However, the CC ν_μ event contains a primary charged lepton that escapes the detector volume, identified by the central peak. This topology suggests that strongly classified CC ν_μ events can be identified by this 'escaping' feature rather than the shape of the muon ring. Future work, therefore, should explore using only fully contained events during beam classification training.

6.3 Robustness

Recent CNN research has focused on another concern; they tend to not generalise well under distributional changes within the input data [168]. As the CNNs presented in this

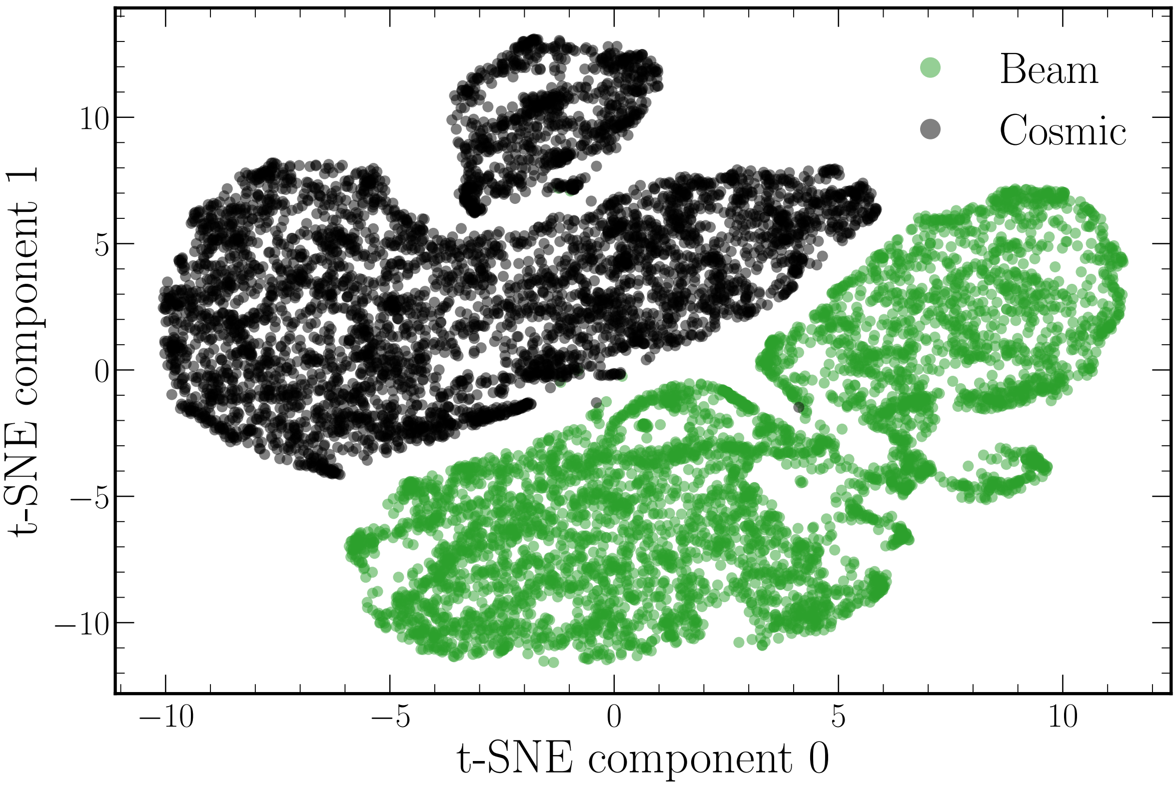


Figure 6.30: Two dimensional probability space of beam and cosmic events generated using the t-SNE procedure on the final fully-connected layer of the trained cosmic rejection network.

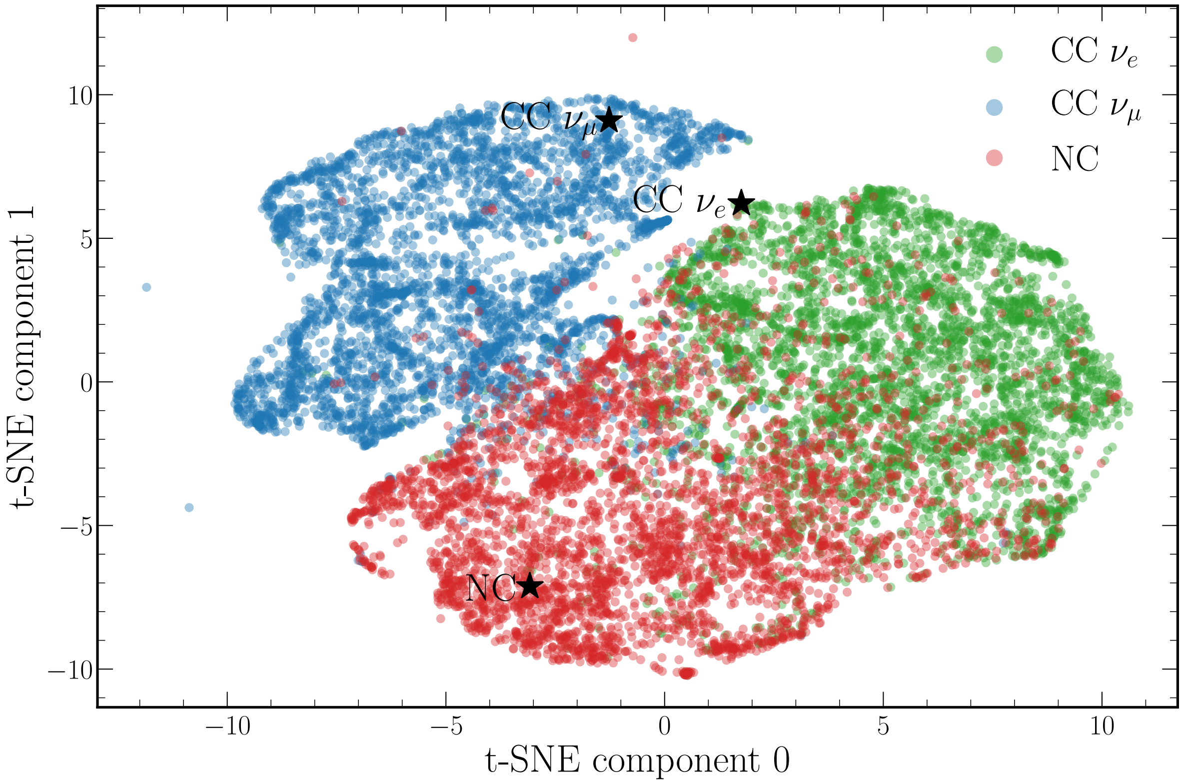


Figure 6.31: Two dimensional probability space of different beam events generated using the t-SNE procedure on the final fully-connected layer of the trained beam classification network. Three events, one highly CC ν_e like, one highly CC ν_μ like, and one highly NC like are highlighted and shown in Figure 6.32.

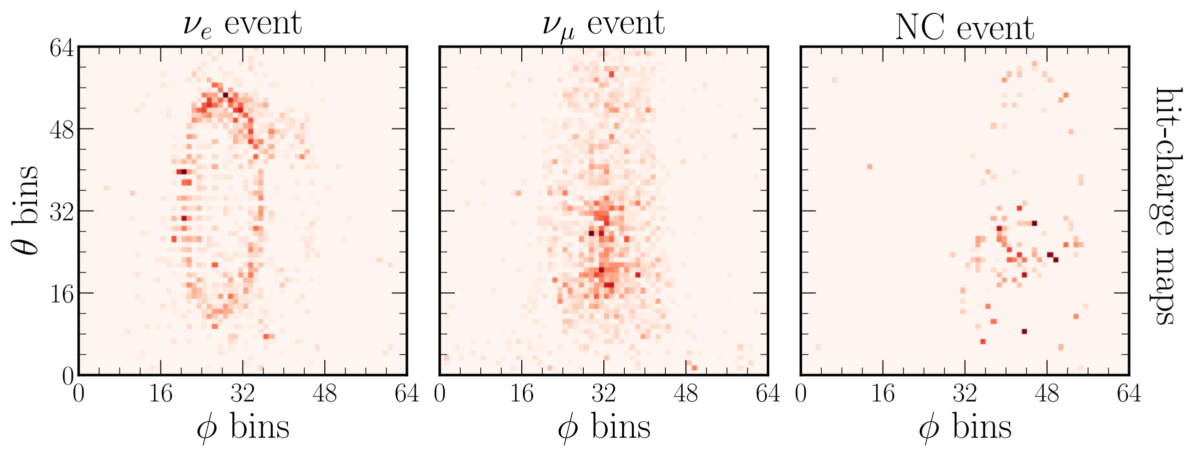


Figure 6.32: Hit-charge maps of the highly CC ν_e like, CC ν_μ like, and NC like events from Figure 6.31.

work are trained and evaluated on simulated Monte Carlo events, this effect is of particular relevance. If the neutrino events used differ from those measured by the real CHIPS-5 detector, the network outputs may not be reliable. An effective calibration procedure and simulation improvements can ensure any discrepancy is minimised; however, a small distributional difference is inevitable.

A selection of studies are presented here to help prove the robustness of the trained CNNs to such changes within the input. Broadly, the CNN inputs can be characterised as being dependent on three sets of PMT information: the hit times, the hit charges, and the hit positions. An accurate PMT position is deemed unimportant to the trained networks as the 64×64 input grid roughly corresponds to bins of size 2.5 m in θ by 2.0 m in ϕ within the CHIPS-5 detector. As this binning is much larger than the actual distance between PMTs, resolving individual PMTs becomes impossible.

Consequently, changes in only the PMT hit times and hit charges are considered in three discrete studies: the smearing of hit times, the smearing and shifting of hit charges, and the addition of random noise. In reality, detector data would be influenced by a convolution of these effects; however, here, they remain separate to develop a detailed understanding of how each individually affects the CNN output. Future work should explore how a combination of such changes affects performance.

Five classification performance metrics are used for comparison during this section and the next (Section 6.4):

1. **Max FOM:** The optimised figure-of-merit (efficiency $\times$ purity) value for CC ν_e events. As the FOM is proportional to the square of $s/\sqrt{s+b}$ which increases linearly with detector exposure, an improvement in the FOM value decreases the exposure time required to reach the same physics sensitivity.

2. **High Score Eff (Pur):** The CC ν_e selection efficiency (purity) using the simple highest-scoring output neuron classification methodology. The simple classification strategy is used here instead of the FOM selection as it is less susceptible to trading off between efficiency and purity, making for easier comparison.

3. **ROC Integral:** The area under the Receiver Operating Characteristic (ROC) curve; a standard tool for classification performance comparison. The curve corresponds to the signal CC ν_e efficiency plotted against the background efficiency as the CC ν_e score cut value is varied. A curve which reaches closer to the top-left (high signal

efficiency, low background efficiency) signifies a stronger classification performance. An example ROC curve is shown for the smearing of hit times in Figure 6.34.

4. **PR Integral:** The area under the Precision-Recall (PR) curve; another standard tool for classification performance comparison. The curve corresponds to the signal CC ν_e purity plotted against the signal efficiency as the CC ν_e score cut value is varied. A curve closer to the top-right (high signal purity, high signal efficiency) signifies a stronger classification performance. For imbalanced class frequencies such as those expected within the CHIPS-5 detector, the PR curve is seen as a more reliable indicator of performance than the ROC curve [169]. An example PR curve is shown for the smearing of hit times in Figure 6.34.

6.3.1 Time smearing

The smearing of input hit times is found to produce a minimal reduction in output performance. For every event, each hit-time bin for which there is an entry not equal to zero is smeared using an absolute time randomly generated using a normal distribution with a mean of zero and a standard deviation of σ nanoseconds. The cases when $\sigma = 0$ ns (no smearing), $\sigma = 2$ ns, and $\sigma = 5$ ns are considered. As the HZC PMTs used predominantly within CHIPS-5 have a photon hit time resolution of $\sim$2 ns (already modelled in the simulation) and other timing error effects (such as cable length calibration) are expected to remain < 2 ns, these values represent a realistic range of time smearing values that could be seen within CHIPS-5.

The actual smearing values used are scaled to the zero to one input range, and any post smearing out-of-range values are clipped to the range boundaries. For each case of σ, the resulting efficiency, purity, and their product (the FOM) for CC ν_e events as a function of selecting events above a particular CC ν_e score is shown in Figure 6.33. The classification performance metrics are presented in Table 6.5 with the ROC and PR curves shown in Figure 6.34. Cosmic rejection and energy estimation performance are not presented, as the resulting output changes are negligible.

With an effective calibration, a realistic discrepancy in hit times of less than 2 ns should be expected. Given this, it is promising to observe that for the smearing of $\sigma = 2$ ns the beam classification performance change is minimal. At $\sigma = 5$ ns more significant degradation starts to occur; however, there is no dramatic fall-off in performance.

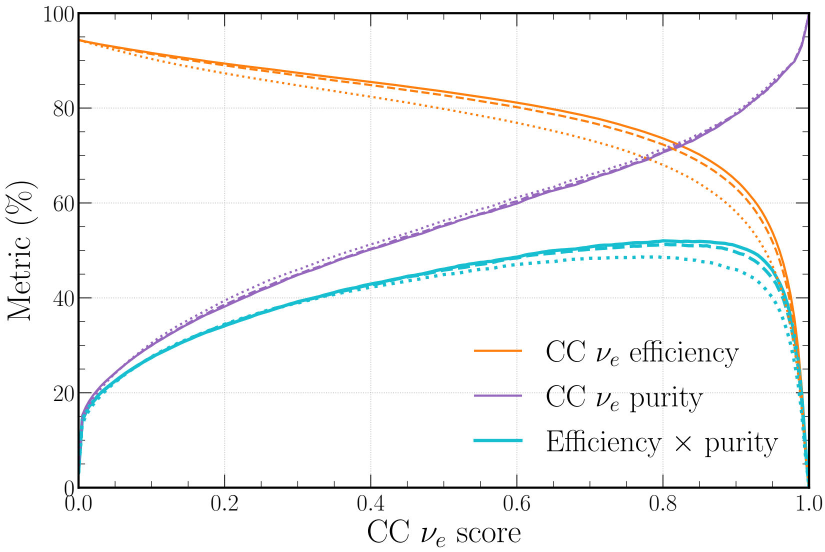


Figure 6.33: CC ν_e efficiency, purity, and efficiency $\times$ purity at different values of CC ν_e score selection for different levels of hit-time smearing. The $\sigma = 0$ ns curves are shown by the solid lines, $\sigma = 2$ ns curves by the dashed lines, and $\sigma = 5$ ns curves by the dotted lines.

Metric	$\sigma = 0$ ns	$\sigma = 2$ ns	$\sigma = 5$ ns
Max FOM	$\mathbf{0.520 \pm 0.004}$	0.513 ± 0.004	0.487 ± 0.004
High Score Eff	$\mathbf{0.835 \pm 0.001}$	0.827 ± 0.001	0.798 ± 0.001
High Score Pur	0.552 ± 0.004	0.555 ± 0.004	$\mathbf{0.563 \pm 0.004}$
ROC Integral	$\mathbf{0.828 \pm 0.015}$	0.828 ± 0.015	0.826 ± 0.015
PR Integral	$\mathbf{0.756 \pm 0.014}$	0.751 ± 0.014	0.730 ± 0.013

Table 6.5: Classification performance metrics for different levels of hit-time smearing. The highest scoring values for each metric are indicated in bold. The ROC and PR integrals are taken from under the curves shown in Figure 6.34. The given errors correspond to the statistical uncertainty only.

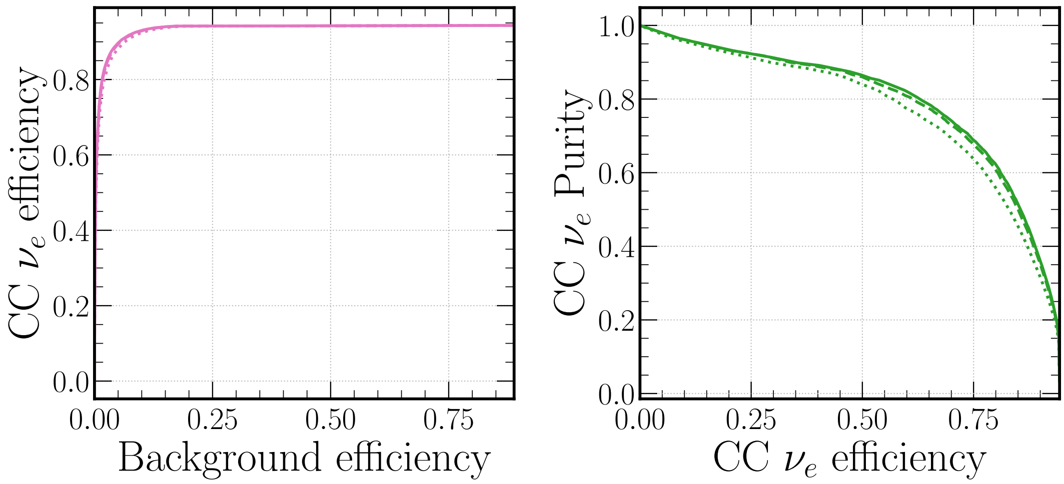


Figure 6.34: ROC (left) and PR (right) curves for different levels of hit-time smearing. The $\sigma = 0$ ns curves are shown by the solid lines, $\sigma = 2$ ns curves by the dashed lines, and $\sigma = 5$ ns curves by the dotted lines.

Interestingly, the *High Score Pur* is seen to increase with greater hit-time smearing, indicating that CC ν_e signal events are proportionally more robust to hit-time input smearing than background events.

6.3.2 Charge smearing and shifting

The smearing and shifting of input hit charges behave as expected and produce no significant reduction in output performance. For every event, each hit-charge and hough-height bin is scaled by a factor randomly drawn from a normal distribution with a mean of μ and a standard deviation of σ. This methodology differs from the absolute hit-time smearing above by modifying bins proportional to their charge, instead of using an absolute value. Any post modification bin values outside the zero to one input range are clipped to the range boundaries.

Alongside the default case when $\mu = 1.0$ and $\sigma = 0.0$, two smearing and two shifting cases are considered: $\mu = 1.0, \sigma = 0.2$ and $\mu = 1.0, \sigma = 0.4$ for smearing; and $\mu = 1.2, \sigma = 0.2$ and $\mu = 1.4, \sigma = 0.2$ for shifting. Note that a standard deviation of $\sigma = 0.2$ is used to introduce some width to the distribution from which the shift value is sampled for the shifting cases. With an effective PMT calibration, it is realistic to assume differences on the order of a few per cent for the hit-charge and hough-height inputs [170]. Therefore, the cases outlined above represent smearing and shifting cases

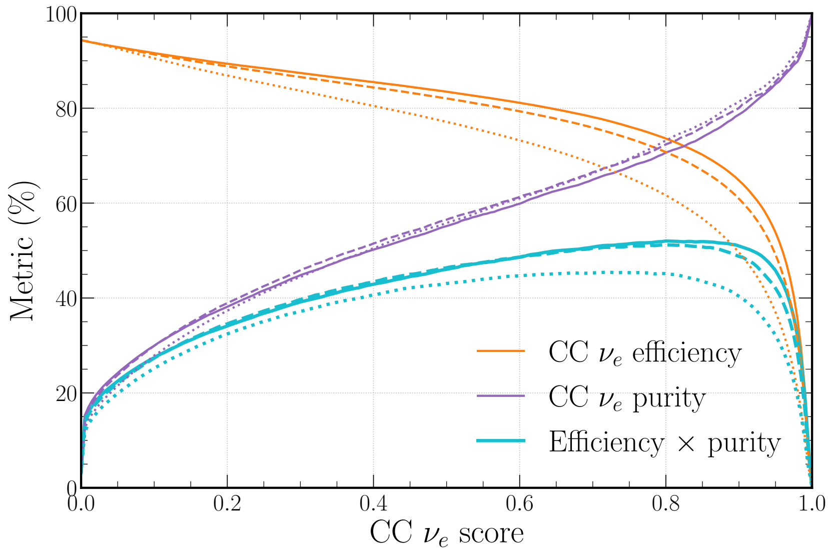


Figure 6.35: CC ν_e efficiency, purity, and efficiency $\times$ purity curves at different values of CC ν_e score selection for different levels of hit-charge smearing where $\mu = 1.0$. The $\sigma = 0.0$ curves are shown by the solid lines, $\sigma = 0.2$ curves by the dashed lines, and $\sigma = 0.4$ curves by the dotted lines.

Scaling (μ, σ)	$(1.0, 0.0)$	$(1.0, 0.2)$	$(1.0, 0.4)$	$(1.2, 0.2)$	$(1.4, 0.2)$
Max FOM	$\mathbf{0.520 \pm 0.004}$	0.512 ± 0.004	0.454 ± 0.004	0.518 ± 0.004	0.498 ± 0.004
High Score Eff	$\mathbf{0.835 \pm 0.001}$	0.820 ± 0.001	0.772 ± 0.001	0.829 ± 0.001	0.820 ± 0.001
High Score Pur	0.552 ± 0.004	$\mathbf{0.565 \pm 0.004}$	0.558 ± 0.004	0.549 ± 0.004	0.533 ± 0.004
ROC Integral	$\mathbf{0.828 \pm 0.015}$	0.827 ± 0.015	0.825 ± 0.015	0.827 ± 0.015	0.825 ± 0.015
PR Integral	$\mathbf{0.756 \pm 0.014}$	0.751 ± 0.014	0.707 ± 0.013	0.750 ± 0.014	0.756 ± 0.013

Table 6.6: Classification performance metrics for different levels of hit-charge smearing and shifting. The highest scoring values for each metric are indicated in bold. The given errors correspond to the statistical uncertainty only.

much larger than expected within CHIPS-5. However, these values are presented here to demonstrate when charge smearing and shifting start to impact performance.

For each smearing case, the resulting efficiency, purity, and their product (the FOM) for CC ν_e events as a function of selecting events above a particular CC ν_e score is shown in Figure 6.35. The equivalent plot for each shifting case is shown in Figure 6.36. The classification performance metrics are presented in Table 6.6, with distributions of (reco-true)/true CC QEL ν_e energies for each case also shown in Figure 6.37. Cosmic rejection performance is not presented as the resulting output changes are negligible.

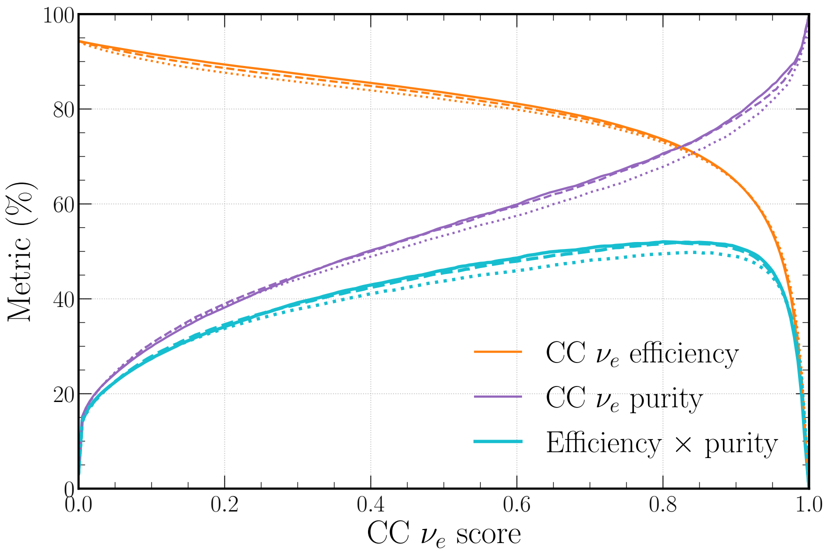


Figure 6.36: CC ν_e efficiency, purity, and efficiency × purity curves at different values of CC ν_e score selection for different levels of hit-charge shifting. The $\mu = 1.0$ curves are shown by the solid lines, $\mu = 1.2$ curves by the dashed lines, and $\mu = 1.4$ curves by the dotted lines.

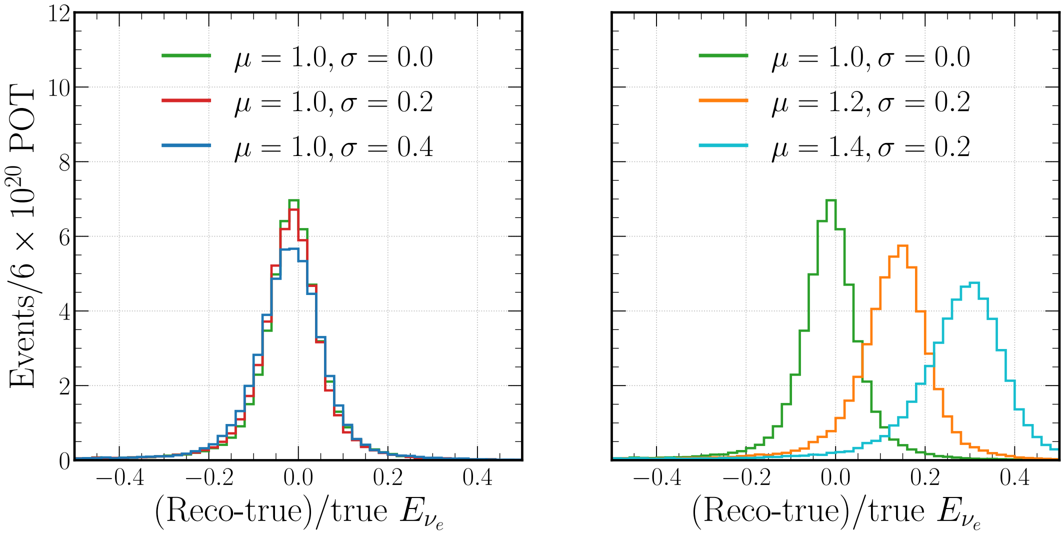


Figure 6.37: Distributions of (reco-true)/true CC QEL ν_e energies for different levels of hit-charge smearing and shifting. Shown are the different smearing levels (left) and the different shifting levels (right).

The extensive input modifications made here (40% shifts, for example) still produce only relatively small reductions in beam classification performance (up to a $\sim 4\%$ reduction in maximum FOM value); therefore, a negligible effect can be assumed. Interestingly, smearing of $\sigma = 0.4$ is seen to impact performance to a greater degree than shifting with $\mu = 1.4$. This behaviour can be explained by the overall relative event topology remaining consistent under a shift compared to when smeared.

For neutrino energy estimation, the impact of smearing is small, but shifting is seen to produce significant changes. This is understandable when assuming that the energy estimation (even for a CNN) relies principally on the counting of the deposited charge for each event, as is the case for most experimental particle physics predictions. Therefore, any systematic shift in the measured input data is always expected to bias the output prediction heavily. However, for realistic input differences on the order of a few per cent, only small output changes should be expected here.

6.3.3 Random noise

Random PMT noise added to the input is found to have a negligible impact on output performance. For every event hit-charge bin, a normal distribution with a mean of zero and a standard deviation of μ photoelectrons is randomly sampled. If a value greater than that corresponding to a single photoelectron is returned, the bin charge is incremented by the returned value. Furthermore, for each modification made, the corresponding hit-time bin is updated by randomly sampling a uniform time distribution and choosing the earliest of the return value and the already set bin time. Alongside the default case with no noise (0.0%), values of μ are chosen here so that either 2.3% or 6.7% of bins are modified by random noise in each event.

Testing has shown that at room temperature, the Nikhef Pom HZC PMTs produce a dark noise rate of $\sim$1 KHz. Therefore, only $\sim 0.02\%$ of event map bins are expected to be affected by PMT noise throughout a typical 100 ns event. Therefore, the cases outlined above represent PMT noise rates more than two orders of magnitude greater than expected within Chips-5. However, these values are presented here in order to demonstrate when random noise starts to impact performance.

For each case, the resulting efficiency, purity, and their product (the FOM) for CC ν_e events as a function of selecting events above a particular CC ν_e score is shown in Figure 6.38. The classification performance metrics are presented in Table 6.7. Cosmic

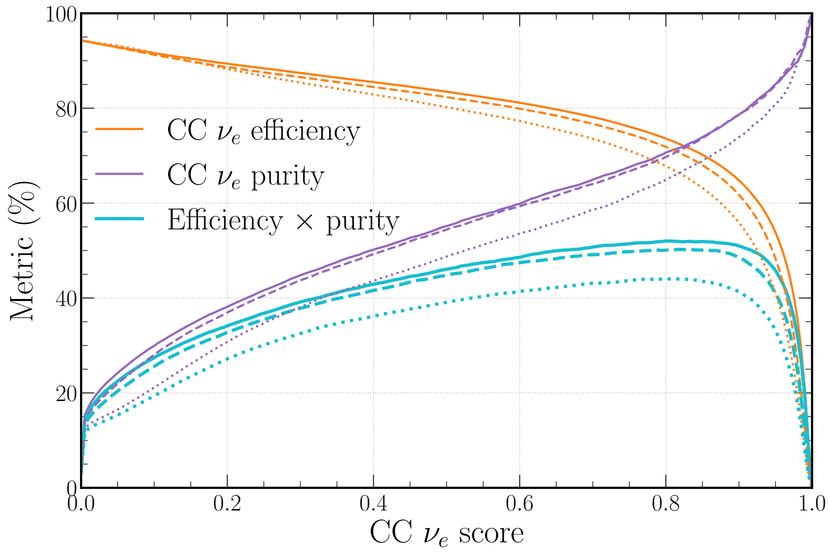


Figure 6.38: CC ν_e efficiency, purity, and efficiency × purity curves at different values of CC ν_e score selection for different levels of random noise. The 0.0% curves are shown by the solid lines, 2.3% curves by the dashed lines, and 6.7% curves by the dotted lines.

Metric	0.0%	2.3%	6.7%
Max FOM	$\mathbf{0.520 \pm 0.004}$	0.503 ± 0.004	0.440 ± 0.003
High Score Eff	$\mathbf{0.835 \pm 0.001}$	0.823 ± 0.001	0.802 ± 0.001
High Score Pur	$\mathbf{0.552 \pm 0.004}$	0.544 ± 0.004	0.487 ± 0.004
ROC Integral	$\mathbf{0.828 \pm 0.015}$	0.827 ± 0.015	0.822 ± 0.015
PR Integral	$\mathbf{0.756 \pm 0.014}$	0.740 ± 0.013	0.681 ± 0.012

Table 6.7: Classification performance metrics for different levels of random noise. The highest scoring values for each metric are indicated in bold. The given errors correspond to the statistical uncertainty only.

rejection and energy estimation performance are not presented as the resulting output changes are negligible. Excellent beam classification robustness is observed, with the small reduction in performance seen for the 2.3% case indicating any output changes with realistic PMT noise levels would be negligible.

6.4 Alternative implementations

Here a sample of alternative (but ultimately unsuccessful) Chips CNN implementations are presented and compared to the final implementation used. Those chosen highlight interesting factors that are found to drive (or not drive) performance. Only the beam classification performance, specifically the primary CC ν_e selection is considered here for comparison; however, these findings also translate to cosmic rejection and energy estimation performance.

6.4.1 Alternative inputs

How the raw PMT hits of an event are represented is found to impact performance significantly. Three different representations are considered. Firstly, the (successful) *vertex view*, where event maps are generated in θ and ϕ as viewed from the seed estimated interaction vertex position. Secondly, the *origin raw view*, where event maps are generated in θ and ϕ as viewed from the centre of the detector (the origin). Finally, the *origin iso view*, where event maps are generated using a PMT position parametrisation in X_+ and X_- as viewed from the centre of the detector.

The origin iso view follows the parametrisation from reference [171], used for exploratory Super-Kamiokande (and Hyper-Kamiokande) CNN studies. This mapping attempts to equally distribute PMT density across the whole two-dimensional input representation. The PMT positions in cylindrical coordinates (ρ, ϕ, z) are mapped to two dimensions, X_+ and X_- using

$$X_\pm = \begin{cases} 1 - \chi_\mp & (z \geq 0) \\ \chi_\pm & (z < 0), \end{cases} \tag{6.1}$$

where

$$\chi_\pm = W(\rho, z)\frac{\pi \pm \phi}{2\pi}, \tag{6.2}$$

and

$$W(\rho, z) = \sqrt{\frac{\rho^2 - 2R|z| + RH}{R^2 + RH}}, \tag{6.3}$$

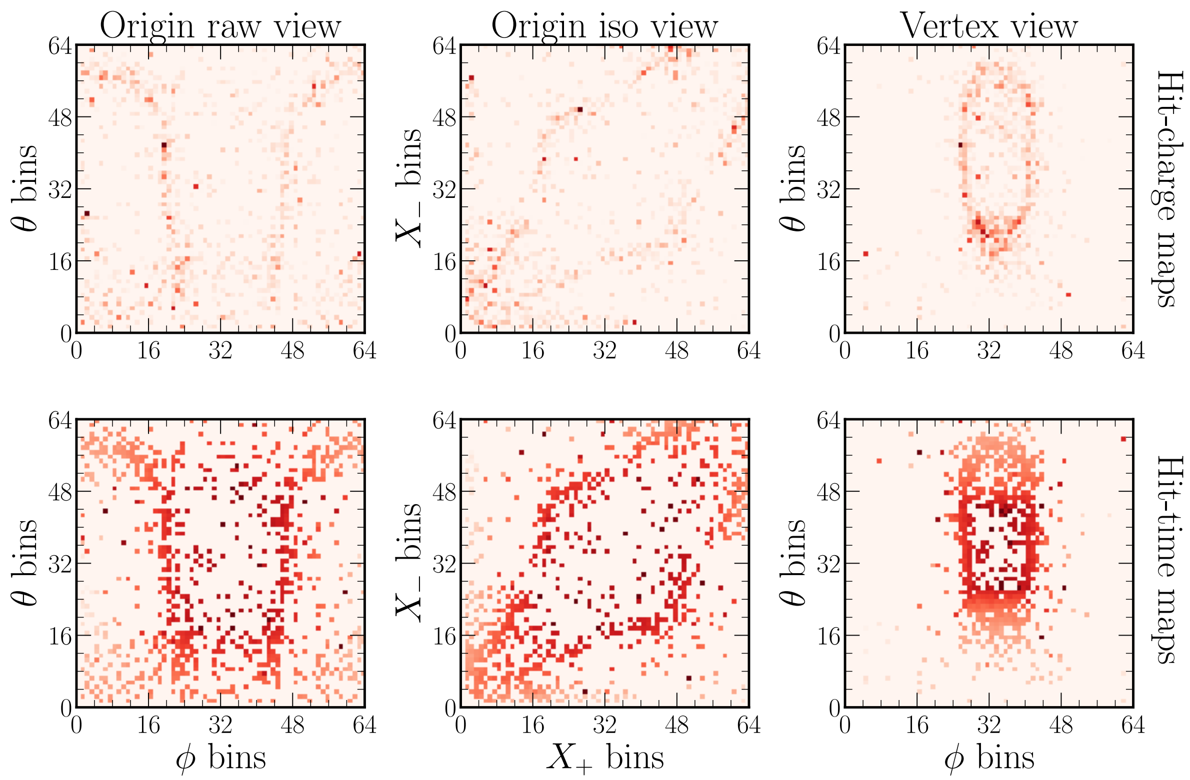


Figure 6.39: Example CC resonant ν_e event shown for each of the three input representations, with the hit-charge (top) and hit-time (bottom) event maps shown for each. The event is initiated by a ν_e of energy 3.3 GeV with the final state particles above the Cherenkov threshold including a e^- of energy 2.8 GeV and a 0.3 GeV π^0.

with R and H being the radius and height of the cylindrical detector respectively.

The same example CC resonant ν_e event for each representation is shown in Figure 6.39 for reference. The example event highlights the advantages of representing the event as viewed from its estimated interaction vertex position. The emitted Cherenkov radiation and resulting ring are viewed without detector or representation parametrisation distortions, showing their true physical topology. This clarity allows the CNN to solely learn the underlying topological differences between event types instead of also having to untangle distortions.

A beam classification network is trained for each representation. Only the hit-charge and hit-time event maps are used as input as the hough-height map can only be generated from the seed estimated interaction vertex position. For each representation the resulting efficiency, purity, and their product (the FOM) for CC ν_e events as a function of selecting events above a particular CC ν_e score is shown in Figure 6.40. Performance comparison metrics are also presented in Table 6.8.

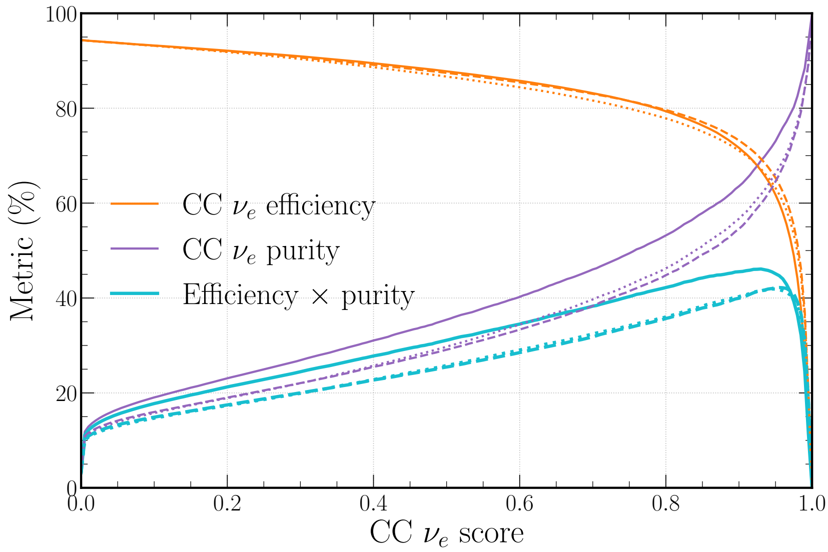


Figure 6.40: CC ν_e efficiency, purity, and efficiency $\times$ purity curves for different values of CC ν_e score selection for the three input representations. The *vertex view* curves are shown by the solid lines, the *origin raw view* curves by the dashed lines, and the *origin iso view* curves by the dotted lines.

Metric	Vertex View	Origin Raw View	Origin Iso View
Max FOM	$\mathbf{0.461 \pm 0.004}$	0.422 ± 0.003	0.419 ± 0.003
High Score Eff	$\mathbf{0.878 \pm 0.001}$	0.874 ± 0.001	0.867 ± 0.001
High Score Pur	$\mathbf{0.354 \pm 0.002}$	0.291 ± 0.002	0.298 ± 0.002
ROC Integral	$\mathbf{0.825 \pm 0.015}$	0.822 ± 0.015	0.821 ± 0.015
PR Integral	$\mathbf{0.707 \pm 0.013}$	0.675 ± 0.012	0.670 ± 0.012

Table 6.8: Classification performance metrics for the three input representations. The highest scoring values for each metric are indicated in bold. The given errors correspond to the statistical uncertainty only.

A significant reduction in performance is observed when using either origin view representation, quantifying the qualitative description using the example event given above. Interestingly, both origin view representations result in a very similar performance, showing that a uniform PMT distribution across the input representation does not provide an improvement. Future work should consider other methods to reduce distortions, such as the smearing of hits across nearby bins to reduce isolated peaks and an improved interaction vertex position estimation procedure.

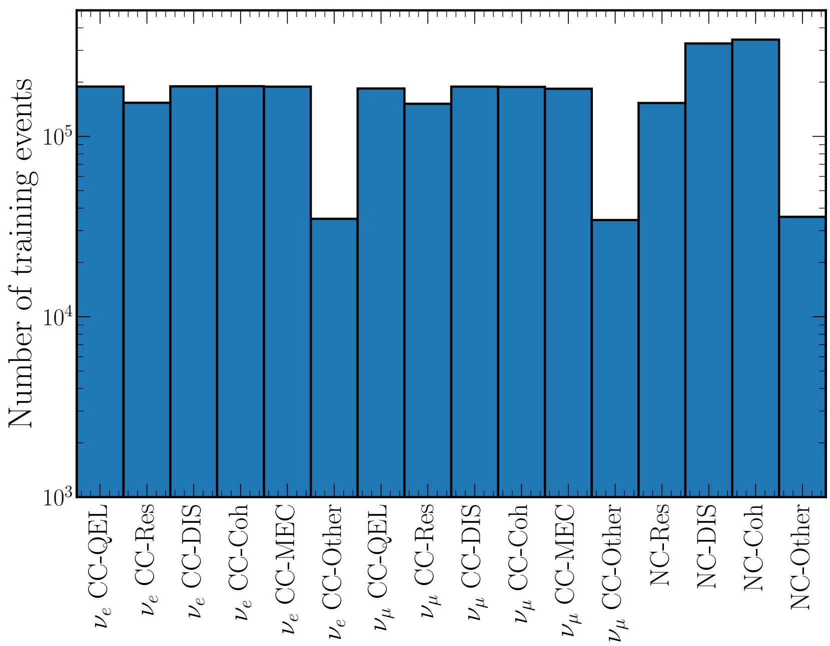


Figure 6.41: The number of training events per category for the uniform beam classification training sample, with roughly equal numbers of CC ν_e, CC ν_μ, and NC training events. All beam event interaction types are shown.

6.4.2 Alternative training samples

The relative proportions of training sample interaction types should match the expected sample as closely as possible. Three different training samples are considered. Firstly, the (successful) *flux* sample representative of the expected spectrum. Secondly, a *uniform* sample, using a roughly similar number of events per interaction type as shown in Figure 6.41. Finally, a *both* sample using an equivalent combination of both.

A beam classification network is trained using each of the samples with an equal number of events. For each sample the resulting efficiency, purity, and their product (the FOM) for CC ν_e events as a function of selecting events above a particular CC ν_e score is shown in Figure 6.42. Performance comparison metrics are also presented in Table 6.9.

Using the uniform training sample is found to reduce beam classification performance significantly. Given that the *both* sample performance lies between the uniform and flux cases, it can be assumed that as the training sample tends towards the expected distribution of events, the performance increases. This finding highlights a key feature of CNNs; they play the game of statistics and only statistics, making them less smart than they may first appear.

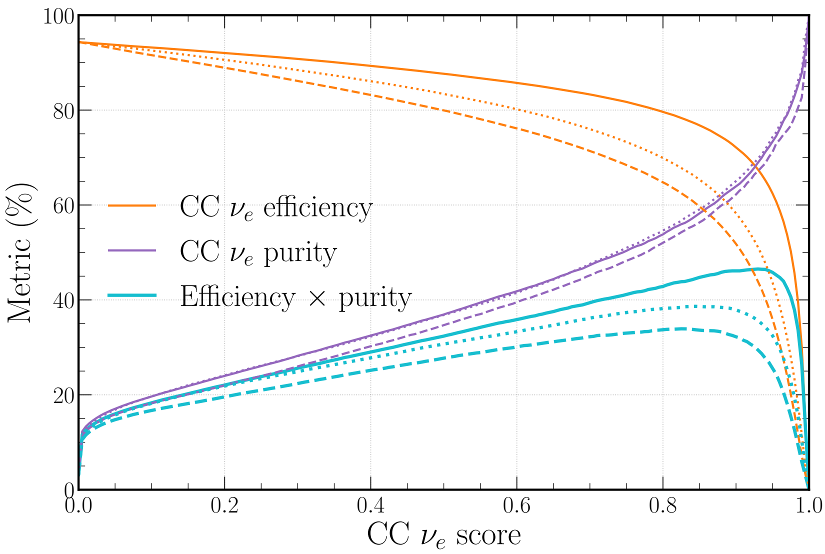


Figure 6.42: CC ν_e efficiency, purity, and efficiency × purity curves for different values of CC ν_e score selection for the different training samples. The *flux* sample curves are shown by the solid lines, the *uniform* sample curves by the dashed lines, and the *both* curves by the dotted lines.

Metric	Flux	Uniform	Both
Maximum FOM	**0.465 ± 0.004**	0.339 ± 0.003	0.386 ± 0.003
High Score Eff	**0.877 ± 0.001**	0.799 ± 0.001	0.834 ± 0.001
High Score Pur	**0.369 ± 0.003**	0.347 ± 0.003	0.368 ± 0.003
ROC Integral	**0.826 ± 0.015**	0.815 ± 0.015	0.820 ± 0.015
PR Integral	**0.712 ± 0.013**	0.568 ± 0.011	0.634 ± 0.012

Table 6.9: Classification performance metrics for the different training samples. The highest scoring values for each metric are indicated in bold. The given errors correspond to the statistical uncertainty only.

In the beam classification case, the CNN learns to minimise the categorical cross-entropy loss function, maximising the fraction of total training events that are classified correctly. An easy way to do this is to become heavily biased by the averages of the sample used during training, such that the final performance does not generalise well to a different evaluation sample.

For the flux training sample case, less common event classes are ignored by the network during training, as solely focusing on the dominant classes is found to be the easiest way to minimise the loss. The effect of this can be seen in the interaction type classification matrices in Section 6.1.3 for CC-Coh and CC-MEC events. Future work should consider the use of a different loss function to promote learning for the less common interaction types and to maximise the physics sensitivity rather than majority accuracy.

6.4.3 Alternative architectures

The choice of CNN architecture is found to have a minimal impact on performance. Four different network architectures are considered: the default VGG based chipsnet architecture, an Inception (GoogLeNet) based architecture [126], a ResNet based architecture [150], and an Inception-ResNet based architecture [151]. Each is modified to fit the standard chipsnet pattern with separate initial branches for each event map followed by a series of combined layers. For a fair comparison, each network is scaled so that they all have approximately the same number of trainable parameters as the default VGG based architecture.

A beam classification network is trained using each network architecture with individually optimised hyperparameters (using SHERPA). For each network architecture, the classification performance metrics are presented in Table 6.10 along with the number of trainable parameters (Number of pars) and the time taken in milliseconds to complete a single training iteration step (Iter time).

Only small performance differences are observed across the different CNN architectures, with the simplest and least modern, the VGG architecture, proving the highest performing. Furthermore, the VGG architecture provides the fastest training step iteration time by a significant margin, leading to drastically lower overall training times.

The other architectures are primarily the result of advances aimed at improving the classification of real-world objects, such as cars, trees, and dogs, with image sizes

Metric	VGG	Inception	ResNet	Inception-ResNet
Number of pars	17,225,296	16,893,216	**16,526,288**	17,145,238
Iter time (ms)	$\mathbf{88 \pm 3}$	192 ± 8	122 ± 5	209 ± 10
Maximum FOM	$\mathbf{0.465 \pm 0.004}$	0.459 ± 0.004	0.445 ± 0.004	0.445 ± 0.004
High Score Eff	$\mathbf{0.877 \pm 0.001}$	0.870 ± 0.001	0.870 ± 0.001	0.874 ± 0.001
High Score Pur	0.369 ± 0.003	0.373 ± 0.003	$\mathbf{0.374 \pm 0.003}$	0.349 ± 0.003
ROC Integral	$\mathbf{0.826 \pm 0.015}$	0.825 ± 0.015	0.824 ± 0.015	0.824 ± 0.015
PR Integral	$\mathbf{0.712 \pm 0.013}$	0.706 ± 0.013	0.688 ± 0.012	0.699 ± 0.012

Table 6.10: Classification performance metrics for the different network architectures. The highest scoring values for each metric are indicated in bold. The given errors correspond to the statistical uncertainty only.

typically 256×256 pixels or greater. Therefore, it is likely that the task of learning subtle changes in Cherenkov ring topologies, with a relatively limited number of classes, on small 64×64 images, does not require the advanced feature representation they provide. Using the increased complexity models instead introduces additional bias, acting to decrease performance instead of providing additional discriminating power.

Chapter 7

Summary and conclusion

This thesis presented a range of work for the CHIPS neutrino detector R&D project. CHIPS puts forward a novel water Cherenkov based concept to counter the vast expense, increased complexity, and long construction time expected from future long-baseline neutrino oscillation experiments. As detailed, this feat is possible via a series of steps: deploying detector modules in bodies of water on the Earth's surface rather than deep underground; using commercially available rather than bespoke components wherever possible; and optimising photocathode coverage to study accelerator beam neutrinos exclusively. These steps reduce the total cost per kt of sensitive mass to between \$200k-\$300k and should allow for megaton scale detectors to become a reality.

Moreover, CHIPS detector modules are expected to be relatively easy to build, quick to deploy, and can be upgraded once operational, making them a much more attractive proposition when resources are constrained. It is hoped that future CHIPS detectors will be able to help answer some of the key unsolved questions of neutrino physics, such as the mass hierarchy ambiguity and the search for CP violation in neutrino oscillations.

During the summer of 2019 a CHIPS prototype, CHIPS-5 was deployed into the Wentworth 2W disused mine pit in northern Minnesota, USA. Although CHIPS-5 is not yet fully proven and future plans are still in flux, the amount of knowledge gained during its construction, deployment, and initial commissioning can not be overstated. Alongside proving that a large CHIPS detector module can be constructed and deployed, CHIPS-5 acted as a brilliant testbed for the development of the CHIPS data acquisition system.

Notably, the use of commercially available single-board Beaglebone machines and the open-source Elasticsearch monitoring solution were found to be highly successful. Within both future CHIPS detectors, and hopefully the broader experimental particle physics field, bespoke components should continue to be phased out, with commercially available or open-source components used instead. Not only does this drastically reduce the implementation effort required by collaborators and cost, but leads to a much improved final result due to the pooling of resources.

As is the case within the world around us, future CHIPS concept detectors will also make ever-increasing use of modern deep learning techniques. This comes as a direct result of the dramatic improvement in CHIPS-5 reconstruction and classification performance brought about by the principal work presented in this thesis. Three forms of a Convolutional Neural Network have been trained to reject cosmic muon events, classify beam events, and estimate neutrino energies, all using only the raw detector event as input. This new approach replaces the standard likelihood-based reconstruction and simple neural network classification, greatly increasing generalisability and processing speed.

With the primary goal of selecting an efficient and pure appeared CC ν_e sample for which the neutrino energy can be accurately determined, the new CNN-based approach is found to provide excellent performance. In some cases, the performance is comparable with similar experiments, impressive given the significant differences in detector design. The vast cosmic muon background of CHIPS-5 is found to be accepted by only a factor of $< 9.5 \times 10^{-7}$ (without the help of a veto), equivalent to < 2 cosmic muon events contaminating the beam sample per year, of which none are expected to be classified as CC ν_e events.

Furthermore, the key performance metrics for both the CC ν_e and CC ν_μ beam selections are summarised in Table 7.1, with a comparison of the metrics available for the old likelihood-based approach given in Table 7.2. The new CNN approach is found to improve the primary CC ν_e selection signal efficiency by an impressive $212 \pm 5\%$, and the corresponding signal purity by $180 \pm 6\%$. Energy estimation is also significantly improved with the approximate energy resolution for CC ν_e QEL event electrons $30 \pm 4\%$ that provided by the old likelihood-based method.

Not only are the trained CNNs found to provide excellent performance, but some insight into their inner workings was achieved, and their outputs are found to be robust to a sample of tested distributional changes in the input. These findings go some way to answering the common and justified concern that they are too often used as a black box (input in, outputs out). Cherenkov ring and Hough peak features are extracted from

Selection	Signal Efficiency	Signal Purity	Approximate ν Energy Resolution
CC ν_e	$73.4 \pm 0.2\%$	$70.9 \pm 0.6\%$	$10.2 \pm 0.2\%$
CC ν_μ	$37.0 \pm 0.1\%$	$96.0 \pm 0.1\%$	$12.5 \pm 0.2\%$

Table 7.1: The key performance metrics for both the CC ν_e and CC ν_μ beam selections using the new CNN-based approach. The signal efficiency relative to the total number of expected events, the signal purity defined as the fraction of selected events which are signal, and the approximate signal neutrino energy resolution. The values considering both the appeared and beam CC ν_e components as signal are given for the CC ν_e selection.

Metric	Likelihood-based (old)	CNN-based (new)	Percentage Change
Max FOM-ν_e	0.132 ± 0.005	0.519 ± 0.004	$393 \pm 15\%$
CC ν_e Efficiency	$34.7 \pm 0.8\%$	$73.4 \pm 0.2\%$	$212 \pm 5\%$
CC ν_e Purity	$39.3 \pm 1.2\%$	$70.9 \pm 0.6\%$	$180 \pm 6\%$
QEL ν_e lepton energy res	$14.1 \pm 1.7\%$	$4.2 \pm 0.1\%$	$30 \pm 4\%$
QEL ν_μ lepton energy res	$9.9 \pm 1.6\%$	$3.8 \pm 0.1\%$	$39 \pm 6\%$

Table 7.2: Comparison of performance metrics between the old likelihood-based approach and the new CNN-based approach with the percentage change given for reference. The maximum FOM value metric alongside the signal efficiency and purity are given for the CC ν_e beam selection, as well as the QEL event approximate lepton energy resolutions for both CC ν_e and CC ν_μ selected signal events.

the input images, resulting in a learnt representation of the inputs seen to have strong discriminating power between categories when visualised using the t-SNE technique. Additionally, realistic modifications to the input hit times and charges and the addition of random noise are all found to have a negligible effect on output performance.

It is sincerely hoped that other water Cherenkov neutrino experiments will take inspiration from and then build upon the work presented in this thesis for their own Convolutional Neural Network implementations. Although the results presented in this work are incredibly compelling, there are still clear avenues for exploration and improvement. These are all principally related to the critical performance drivers outlined within this thesis.

Firstly, generating the input event maps to focus on the underlying Cherenkov profiles is incredibly important; therefore, any methodology to remove distortions further or more accurately determine the interaction vertex position will be beneficial. Secondly, the distribution of events (in energy or type) used within the training sample heavily impacts performance; thus, a comprehensive study of this behaviour could optimise the sample used. Finally, multi-task learning clearly shows promise, with further trial-and-error or a more generalised approach likely to uncover additional valuable tasks.

Likely, the true potential of these methods is just beginning to be realised. As is always the case, only time will tell.

Bibliography

[1] W. Pauli, "Letter to the radioactive ladies and gentlemen", Letter to the physical society of Tübingen (1930).

[2] J. Chadwick, "The intensity distribution in the magnetic spectrum of beta particles from radium (B+C)", Verh. Phys. Gesell. **16**, 383–391 (1914).

[3] E. Fermi, "An attempt of a theory of beta radiation", Zeitschrift für Physik **88**, 161–177 (1934).

[4] H. Bethe and R. Peierls, "The 'Neutrino'", Nature **133**, 532–532 (1934).

[5] C. Cowan, F. Reines, F. Harrison, H. Kruse, and A. McGuire, "Detection of the free neutrino: a confirmation", Science **124**, 103–104 (1956).

[6] G. Danby, J. Gaillard, A. Konstantin, L. Lederman, B. Nari, M. Schwartz, and J. Steinberger, "Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos", Physical Review Letters **9**, 36–44 (1962).

[7] G. Arnison et al. (UA1 Collaboration), "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider", Physics Letters B **126**, 398–410 (1983).

[8] G. Arnison et al., "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV", Physics Letters B **122**, 103–116 (1983).

[9] G. Abrams, C. Adolphsen, D. Averill, J. Ballam, B. C. Barish, T. Barklow, B. Barnett, J. Bartelt, S. Bethke, D. Blockus, et al., "Measurements of z-boson resonance parameters in e+ e- annihilation", Physical Review Letters **63**, 2173 (1989).

[10] The ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration, the OPAL Collaboration, the SLD Collaboration, the LEP Electroweak Working Group, the SLD electroweak, heavy flavour groups, "Precision electroweak measurements on the Z resonance", Physics Reports **427**, 257–454 (2006).

[11] M. L. Perl et al., “Evidence for anomalous lepton production in $e^+ - e^-$ annihilation”, Physical Review Letters **35**, 1489–1492 (1975).

[12] K. Kodama et al., “Observation of tau neutrino interactions”, Physics Letters B **504**, 218–224 (2001).

[13] P. Zyla et al. (Particle Data Group Collaboration), “Review of particle physics”, Progress of Theoretical and Experimental Physics **2020**, 083C01 (2020).

[14] R. Davis Jr, D. Harmer, and K. Hoffman, “Search for neutrinos from the sun”, Physical Review Letters **20**, 1205 (1968).

[15] J. Bahcall, N. Bahcall, and G. Shaviv, “Present status of the theoretical predictions for the Cl-36 solar neutrino experiment”, Physical Review Letters **20**, 1209 (1968).

[16] K. Hirata et al. (Kamiokande-II Collaboration), “Observation of B-8 solar neutrinos in the Kamiokande-II detector”, Physical Review Letters **63**, 16 (1989).

[17] A. Abazov et al. (SAGE Collaboration), “Search for neutrinos from the sun using the reaction Ga-71 (ν_e e^-) Ge-71”, Physical Review Letters **67**, 3332 (1991).

[18] P. Anselmann et al. (GALLEX Collaboration), “GALLEX results from the first 30 solar neutrino runs”, Physics Letters B **327**, 377–385 (1994).

[19] K. Hirata et al. (Kamiokande-II Collaboration), “Experimental study of the atmospheric neutrino flux”, Physics Letters B **205**, 416 (1988).

[20] R. Becker-Szendy et al., “Electron-and muon-neutrino content of the atmospheric flux”, Physical Review D **46**, 3720 (1992).

[21] T. Kajita et al. (Super-Kamiokande Collaboration), “Atmospheric neutrino results from Super-Kamiokande and Kamiokande - Evidence for ν_μ oscillations”, Nuclear Physics B Proceedings Supplements **77**, 123–132 (1999).

[22] Q. R. Ahmad et al. (SNO Collaboration), “Measurement of day and night neutrino energy spectra at SNO and constraints on neutrino mixing parameters”, Physical Review Letters **89**, 011302 (2002).

[23] Wikipedia contributors, *Standard Model — Wikipedia, The Free Encyclopedia*, `https://en.wikipedia.org/w/index.php?title=Standard_Model&oldid=970721485`.

[24] N. Aghanim et al. (Planck Collaboration), “Planck 2018 results. VI. Cosmological parameters”, (2018), arXiv:`1807.06209 [astro-ph.CO]`.

[25] M. Aker et al. (KATRIN Collaboration), "Improved upper limit on the neutrino mass from a direct kinematic method by KATRIN", Physical Review Letters **123**, 10.1103/physrevlett.123.221802 (2019).

[26] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles", Progress of Theoretical Physics **28**, 870–880 (1962).

[27] B. Pontecorvo, "Neutrino experiments and the problem of conservation of leptonic charge", Soviet Physics JETP **26**, 984–988 (1968).

[28] V. Gribov and B. Pontecorvo, "Neutrino astronomy and lepton charge", Physics Letters B **28**, 493–496 (1969).

[29] L. Wolfenstein, "Neutrino oscillations in matter", Physical Review D **17**, 2369–2374 (1978).

[30] S. Mikheyev and A. Smirnov, "Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos", Soviet Journal of Nuclear Physics **42**, 913–917 (1985).

[31] T. Katori, "Meson exchange current (MEC) models in neutrino interaction generators", AIP Conference Proceedings **1663**, 030001 (2015), arXiv:1304.6014 [nucl-th].

[32] C. Andreopoulos et al., "The GENIE neutrino Monte Carlo generator", Nucl. Instrum. Meth. A **614**, 87–104 (2010), arXiv:0905.2517 [hep-ph].

[33] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher, G. Perdue, and J. Yarba, "The GENIE neutrino Monte Carlo generator: physics and user manual", (2015), arXiv:1510.05494 [hep-ph].

[34] J. A. Formaggio and G. P. Zeller, "From eV to EeV: neutrino cross sections across energy scales", Reviews of Modern Physics **84**, 1307–1341 (2012).

[35] I. Esteban, M. Gonzalez-Garcia, A. Hernandez-Cabezudo, M. Maltoni, and T. Schwetz, "Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ_{23}, δ_{CP}, and the mass ordering", Journal of High Energy Physics **2019**, 1–35 (2019).

[36] I. Esteban, M. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, "The fate of hints: updated global analysis of three-flavor neutrino oscillations", (2020), arXiv:2007.14792 [hep-ph].

[37] B. T. Cleveland, T. Daily, J. Raymond Davis, J. R. Distel, K. Lande, C. K. Lee, P. S. Wildenhain, and J. Ullman, “Measurement of the solar electron neutrino flux with the Homestake chlorine detector”, The Astrophysical Journal **496**, 505–526 (1998).

[38] F. Kaether, W. Hampel, G. Heusser, J. Kiko, and T. Kirsten, “Reanalysis of the GALLEX solar neutrino flux and source experiments”, Physics Letters B **685**, 47–54 (2010).

[39] J. N. Abdurashitov et al. (SAGE Collaboration), “Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002–2007 data-taking period”, Physical Review C **80**, 015807 (2009).

[40] J. Maneira et al. (SNO Collaboration), “Combined analysis of all three phases of solar neutrino data from the Sudbury neutrino observatory”, in 13th ICATPP Conference on Astroparticle, Particle, Space Physics and Detectors for Physics Applications (2012), pp. 360–366.

[41] J. Hosaka et al. (The Super-Kamiokande Collaboration), “Solar neutrino measurements in Super-Kamiokande-I”, Physical Review D **73**, 112001 (2006).

[42] J. P. Cravens et al. (The Super-Kamiokande Collaboration), “Solar neutrino measurements in Super-Kamiokande-II”, Physical Review D **78**, 032002 (2008).

[43] K. Abe et al. (The Super-Kamiokande Collaboration), “Solar neutrino results in Super-Kamiokande-III”, Physical Review D **83**, 052010 (2011).

[44] Y. Nakano et al., “Solar neutrino results from Super-Kamiokande”, Proceedings of Science **ICHEP2016**, 462 (2017).

[45] G. Bellini et al. (Borexino Collaboration), “Precision measurement of the ^{7}Be solar neutrino interaction rate in Borexino”, Physical Review Letters **107**, 141302 (2011).

[46] G. Bellini et al. (Borexino Collaboration), “Measurement of the solar ^{8}B neutrino rate with a liquid scintillator target and 3 MeV energy threshold in the Borexino detector”, Physical Review D **82**, 033006 (2010).

[47] G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, B. Caccianiga, L. Cadonati, F. Calaprice, A. Caminata, P. Cavalcante, et al., “Neutrinos from the primary proton–proton fusion process in the sun”, Nature **512**, 383–386 (2014).

[48] A. Gando et al. (The KamLAND Collaboration), “Constraints on θ_{13} from a three-flavor oscillation analysis of reactor antineutrinos at KamLAND”, Physical Review D **83**, 052002 (2011).

[49] A. Karle, J. Ahrens, J. Bahcall, X. Bai, T. Becka, K.-H. Becker, D. Besson, D. Berley, E. Bernardini, D. Bertrand, et al., “IceCube – the next generation neutrino telescope at the south pole”, Nuclear Physics B-Proceedings Supplements **118**, 388–395 (2003).

[50] M. G. Aartsen et al. (IceCube Collaboration), “Determining neutrino oscillation parameters from atmospheric muon neutrino disappearance with three years of IceCube DeepCore data”, Physical Review D **91**, 072004 (2015).

[51] K. Abe et al. (Super-Kamiokande Collaboration), “Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV”, Physical Review D **97**, 072001 (2018).

[52] P. Adamson et al. (MINOS Collaboration), “Measurement of neutrino and antineutrino oscillations using beam and atmospheric data in MINOS”, Physical Review Letters **110**, 251801 (2013).

[53] P. Adamson et al. (MINOS Collaboration), “Electron neutrino and antineutrino appearance in the full MINOS data sample”, Physical Review Letters **110**, 171801 (2013).

[54] M. A. Acero et al. (NOvA Collaboration), “First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by NOvA”, Physical Review Letters **123**, 151803 (2019).

[55] A. Himmel, *New oscillation results from the NOvA experiment*, `https://zenodo.org/record/4142045`, Talk given at the XXIX International Conference on Neutrino Physics and Astrophysics, Chicago, USA, July 2020.

[56] P. Dunne, *Latest neutrino oscillation results from T2K*, `https://zenodo.org/record/4154355`, Talk given at the XXIX International Conference on Neutrino Physics and Astrophysics, Chicago, USA, July 2020.

[57] F. An et al. (Daya Bay Collaboration), “Observation of electron-antineutrino disappearance at Daya Bay”, Physical Review Letters **108**, 171803 (2012).

[58] F. P. An et al., “Improved measurement of the reactor antineutrino flux and spectrum at Daya Bay”, Chinese Physics C **41**, 013002 (2017).

[59] J. Ahn et al. (RENO Collaboration), “Observation of reactor electron antineutrinos disappearance in the RENO experiment”, Physical Review Letters **108**, 191802 (2012).

[60] G. Bak et al. (RENO Collaboration), "Measurement of reactor antineutrino oscillation amplitude and frequency at RENO", Physical Review Letters **121**, 201801 (2018).

[61] Y. Abe et al. (Double Chooz Collaboration), "Reactor electron antineutrino disappearance in the Double Chooz experiment", Physical Review D **86**, 052008 (2012), arXiv:1207.6632 [hep-ex].

[62] K. Abe et al. (T2K Collaboration), "Evidence of electron neutrino appearance in a muon neutrino beam", Physical Review D **88**, 032002 (2013).

[63] P. Adamson and other (NOvA Collaboration), "First measurement of electron neutrino appearance in NOvA", Physical Review Letters **116**, 151806 (2016).

[64] F. An, G. An, Q. An, V. Antonelli, E. Baussan, J. Beacom, L. Bezrukov, S. Blyth, R. Brugnera, M. B. Avanzini, and et al., "Neutrino physics with JUNO", Journal of Physics G: Nuclear and Particle Physics **43**, 030401 (2016).

[65] A. de Gouvea et al., "Neutrinos", (2013), arXiv:1310.4340 [hep-ex].

[66] K. Abe and other (T2K Collaboration), "Search for CP violation in neutrino and antineutrino oscillations by the T2K experiment with 2.2×10^{21} protons on target", Physical Review Letters **121**, 171802 (2018).

[67] R. Acciarri et al., "Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) conceptual design report volume 1: the LBNF and DUNE projects", (2016), arXiv:1601.05471 [physics.ins-det].

[68] B. Abi et al. (DUNE Collaboration), "Deep Underground Neutrino Experiment (DUNE), far detector technical design report, Volume II DUNE physics", (2020), arXiv:2002.03005 [hep-ex].

[69] U.S. Department Of Energy Office Of Project Management Oversight And Assessments, *Independent cost review of the Long Baseline Neutrino Facility/Deep Underground Neutrino Experiment project at Fermi National Accelerator Laboratory*, Aug. 2015.

[70] American Institute of Physics, *FY21 budget request: DOE office of science*, https://www.aip.org/fyi/2020/fy21-budget-request-doe-office-science, Mar. 2020.

[71] Hyper-Kamiokande Proto-Collaboration, "Hyper-Kamiokande design report", (2018), arXiv:1805.04163 [physics.ins-det].

[72] Hyper-Kamiokande Working Group, "A long baseline neutrino oscillation experiment using J-PARC neutrino beam and Hyper-Kamiokande", (2014), arXiv:1412.4673 [physics.ins-det].

[73] Nature, *Japan will build the worlds largest neutrino detector*, https://www.nature.com/articles/d41586-019-03874-w, Dec. 2019.

[74] P. Adamson et al. (CHIPS Collaboration), "CHerenkov detectors In mine PitS (CHIPS) letter of intent to FNAL", (2013), arXiv:1307.5918 [physics.ins-det].

[75] A. Perch, "Construction of the CHIPS-M prototype and simulations of a 10 kiloton module", (2015), arXiv:1505.00042 [physics.ins-det].

[76] M. Pfützner, "Prototype detection unit for the CHIPS experiment", Journal of Physics: Conference Series **888**, 012059 (2017).

[77] M. Pfützner, "Sensitivity study and first prototype tests for the CHIPS neutrino detector R&D program", https://discovery.ucl.ac.uk/id/eprint/10052874/1/Pfutzner_thesis.pdf, PhD thesis (University College London, July 2018).

[78] F. Amat et al. (CHIPS Collaboration), "Measuring the attenuation length of water in the CHIPS-M water Cherenkov detector", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **844**, 108–115 (2017).

[79] M. Campbell, "Measuring neutrino oscillations in the NOvA and CHIPS Detectors", https://discovery.ucl.ac.uk/id/eprint/10097512/1/Campbell_10097512_thesis_sig-removed.pdf, PhD thesis (University College London, Aug. 2020).

[80] E. Adde Chase, "Sensitivity determination in the CHIPS neutrino detector", https://scholarworks.wm.edu/honorstheses/927, MA thesis (College of William and Mary, 2016).

[81] K. Lang, "Prospects for CHIPS (R&D of water Cherenkov detectors in mine pits)", (2015), arXiv:1504.08330 [physics.ins-det].

[82] D. van Eijk, "Electronics and DAQ for the CHIPS experiment", (2018), arXiv:1805.12206 [physics.ins-det].

[83] P. Adamson et al., "The NuMI neutrino beam", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **806**, 279–306 (2016).

[84] K. S. McFarland (MINERvA), "MINERvA: a dedicated neutrino scattering experiment at NuMI", Nuclear Physics B Proceedings Supplements **159**, 107–112 (2006), arXiv:physics/0605088.

[85] A. J. Perch, "Three-flavour neutrino oscillations with MINOS and CHIPS", https://discovery.ucl.ac.uk/id/eprint/1551615/1/Perch_ID_thesis.pdf, PhD thesis (University College London, Jan. 2017).

[86] K. Hanson and O. Tarasova, "Design and production of the IceCube digital optical module", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **567**, Proceedings of the 4th International Conference on New Developments in Photodetection, 214–217 (2006).

[87] R. Bruijn and D. van Eijk, "The KM3NeT multi-PMT digital optical module", PoS **ICRC2015**, 1157 (2016).

[88] *HZC PMT Datasheet*, http://www.hzcphotonics.com/products/ProductManual.pdf.

[89] U. Katz, K. Consortium, et al., "Status of the KM3NeT project", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **602**, 40–46 (2009).

[90] A. Martinez et al., "Letter of intent for KM3NeT 2.0", Journal of Physics G: Nuclear and Particle Physics **43**, 084001 (2016).

[91] *Hamamatsu R6091 PMT Datasheet*, https://www.hamamatsu.com/resources/pdf/etd/R6091_TPMH1114E.pdf.

[92] R. Arnold, C. Augier, A. Bakalyarov, J. Baker, A. Barabash, P. Bernaudin, M. Bouchel, V. Brudanin, A. Caffrey, J. Cailleret, et al., "Technical design and performance of the nemo 3 detector", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **536**, 79–122 (2005).

[93] *WCSim git repository*, https://github.com/WCSim/WCSim.

[94] S. Agostinelli et al., "Geant4 – a simulation toolkit", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250–303 (2003).

[95] J. Allison et al., "Geant4 developments and applications", IEEE Transactions on Nuclear Science **53**, 270–278 (2006).

[96] J. Allison et al., "Recent developments in Geant4", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **835**, 186–225 (2016).

[97] A. Blake, S. Germani, Y. B. Pan, A. J. Perch, M. M. Pfützner, J. Thomas, and L. H. Whitehead (CHIPS Collaboration), "CHIPS event reconstruction and design optimisation", (2016), arXiv:1612.04604 [physics.ins-det].

[98] *Photomultiplier Tubes - Basics and Applications*, https://www.hamamatsu.com/resources/pdf/etd/PMT_handbook_v3aE.pdf.

[99] *WCSim: How to add your own photodetector tutorial*, https://github.com/WCSim/WCSim/wiki/Tutorial:-How-to-add-your-own-photdetector.

[100] C. Hagmann, D. Lange, J. Verbeke, and D. Wright, "Cosmic-ray shower library (CRY)", (2012).

[101] C. Hagmann, D. Lange, and D. Wright, "Monte Carlo simulation of proton-induced cosmic-ray cascades in the atmosphere", (2012).

[102] S. Klimushin, E. Bugaev, and I. A. Sokalski, "Precise parametrizations of muon energy losses in water", in 27th international cosmic ray conference, Vol. 3 (June 2001), p. 1009, arXiv:hep-ph/0106010.

[103] *chipsgen git repository*, https://gitlab.com/chipsneutrino/chips-gen.

[104] S. V. Cao, H. Junting, L. Karol, and N. Federico, "Cosmic ray muon rates in water Cherenkov detectors at shallow depths", (2013).

[105] E. V. Bugaev, A. Misaki, V. A. Naumov, T. Sinegovskaya, S. Sinegovsky, and N. Takahashi, "Atmospheric muon flux at sea level, underground, and underwater", Physical Review D **58**, 054001 (1998), arXiv:hep-ph/9803488.

[106] M. Lipiski, T. Wostowski, J. Serrano, and P. Alvarez, "White Rabbit: a PTP application for robust sub-nanosecond synchronization", in 2011 ieee international symposium on precision clock synchronization for measurement, control and communication (2011), pp. 25–30.

[107] *WR switch open hardware repository*, https://ohwr.org/projects/white-rabbit/wiki/switch.

[108] *WR chromium switch webpage*, https://redmine.nikhef.nl/et/projects/chromium/wiki/Backplane-20SFP.

[109] *WR-LEN webpage*, https://sevensols.es/index.php/index/timing-products/wr-len.

[110] J. D. Cockcroft and E. T. Walton, “Experiments with high velocity positive ions.(i) further developments in the method of obtaining high velocity positive ions”, Proceedings of the royal society of London. Series A, containing papers of a mathematical and physical character **136**, 619–630 (1932).

[111] S. Biagi, T. Chiarusi, P. Piattelli, and D. Real (KM3NeT), “The data acquisition system of the KM3NeT detector”, PoS **ICRC2015**, 1172 (2016).

[112] *Beaglebone webpage*, `https://beagleboard.org/bone`.

[113] S. Aiello et al. (KM3NeT), “KM3NeT front-end and readout electronics system: hardware, firmware and software”, J. Astron. Telesc. Instrum. Syst. **5**, 046001 (2019), arXiv:`1907.06453 [astro-ph.IM]`.

[114] *microDAQ git repository*, `https://github.com/WIPACrepo/microdaq`.

[115] *chipsdaq git repository*, `https://gitlab.com/chipsneutrino/chips-daq`.

[116] *BOOST.Asio webpage*, `https://www.boost.org/doc/libs/1_66_0/doc/html/boost_asio.html`.

[117] *Elasticsearch webpage*, `https://www.elastic.co/elasticsearch`.

[118] *Elastic stack webpage*, `https://www.elastic.co/elastic-stack`.

[119] *Kibana webpage*, `https://www.elastic.co/kibana`.

[120] *Kibana webpage*, `https://elastalert.readthedocs.io/en/latest/elastalert.html`.

[121] K. Fukushima and S. Miyake, “Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition”, in Competition and cooperation in neural nets (1982), pp. 267–285.

[122] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, `http://www.deeplearningbook.org` (MIT Press, 2016).

[123] F. Psihas, M. Groh, C. Tunnell, and K. Warburton, “A Review on Machine Learning for Neutrino Experiments”, (2020), arXiv:`2008.01242 [physics.comp-ph]`.

[124] R. Perrault, Y. Shoham, E. Brynjolfsson, J. Clark, J. Etchemendy, B. Grosz, T. Lyons, J. Manyika, S. Mishra, and J. C. Niebles, “The AI index 2019 annual report”, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA (2019).

[125] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, and P. Vahle (NOvA Collaboration), "A convolutional neural network neutrino event classifier", Journal of Instrumentation **11**, P09001–P09001 (2016).

[126] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition (2015), pp. 1–9.

[127] F. Psihas, E. Niner, M. Groh, R. Murphy, A. Aurisano, A. Himmel, K. Lang, M. D. Messier, A. Radovic, and A. Sousa (NOvA Collaboration), "Context-enriched identification of particles with a convolutional network for neutrino events", Physical Review D **100**, `10.1103/physrevd.100.073005` (2019).

[128] P. Baldi, J. Bian, L. Hertel, and L. Li (NOvA Collaboration), "Improved energy reconstruction in NOvA with regression convolutional neural networks", Physical Review D **99**, `10.1103/physrevd.99.012011` (2019).

[129] R. Acciarri et al., "Design and construction of the MicroBooNE detector", Journal of Instrumentation **12**, P02017 (2017).

[130] R. Acciarri et al. (MicroBooNE Collaboration), "Convolutional neural networks applied to neutrino Events in a liquid argon time projection chamber", Journal of Instrumentation **12**, P03011–P03011 (2017).

[131] B. Abi et al. (DUNE Collaboration), "Neutrino interaction classification with a convolutional neural network in the DUNE far detector", (2020), arXiv:`2006.15052 [physics.ins-det]`.

[132] E. Racah, S. Ko, P. Sadowski, W. Bhimji, C. Tull, S.-Y. Oh, P. Baldi, and Prabhat (Daya Bay Collaboration), "Revealing fundamental physics from the Daya Bay neutrino experiment using deep neural networks", 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), `10.1109/icmla.2016.0160` (2016).

[133] S. Aiello et al. (KM3NeT Collaboration), "Event reconstruction for KM3NeT/ORCA using convolutional neural networks", (2020), arXiv:`2004.08254 [astro-ph.IM]`.

[134] A. Abhishek, W. Fedorko, P. de Perio, N. Prouse, and J. Z. Ding, "Variational autoencoders for generative modelling of water Cherenkov detectors", (2019), arXiv:`1911.02369 [physics.ins-det]`.

[135] R. Patterson, E. Laird, Y. Liu, P. Meyers, I. Stancu, and H. Tanaka (MiniBooNE Collaboration), "The extended-track event reconstruction for MiniBooNE", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **608**, 206–224 (2009).

[136] A. Hocker et al., "TMVA - Toolkit for multivariate data analysis", (2007), arXiv:physics/0703039.

[137] M. Jiang et al. (Super-kamiokande Collaboration), "Atmospheric neutrino oscillation analysis with improved event reconstruction in Super-Kamiokande IV", Progress of Theoretical and Experimental Physics **2019**, 053F01, 10.1093/ptep/ptz015 (2019), eprint: https://academic.oup.com/ptep/article-pdf/2019/5/053F01/28638877/ptz015.pdf.

[138] A. D. Missert et al. (T2K Collaboration), "Improving the T2K oscillation analysis with fiTQun: a new maximum-likelihood event reconstruction for Super-Kamiokande", JPhCS **888**, 012066 (2017).

[139] *chips-reco git repository*, https://gitlab.com/chipsneutrino/chips-reco.

[140] J. Illingworth and J. Kittler, "A survey of the hough transform", Computer vision, graphics, and image processing **44**, 87–116 (1988).

[141] R. Brun and F. Rademakers, "ROOT – an object oriented data analysis framework", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389**, New Computing Techniques in Physics Research V, 81–86 (1997).

[142] P. Werbos, "Beyond regression: new tools for prediction and analysis in the behavioral sciences", PhD thesis (Harvard University, 1974).

[143] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists", Physics Reports **810**, 1–124 (2019).

[144] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning", Nature **521**, 436–44 (2015).

[145] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database", in 2009 IEEE conference on computer vision and pattern recognition (Ieee, 2009), pp. 248–255.

[146] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in Advances in Neural Information Processing Systems (2012), pp. 1097–1105.

[147] *A Comprehensive Guide to Convolutional Neural Networks*, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[148] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, (2014), arXiv:1409.1556 [cs.CV].

[149] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, in Proceedings of the IEEE conference on computer vision and pattern recognition (2016), pp. 770–778.

[150] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks”, in European Conference on Computer Vision (Springer, 2016), pp. 630–645, arXiv:1603.05027 [cs.CV].

[151] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the impact of residual connections on learning”, (2016), arXiv:1602.07261 [cs.CV].

[152] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: inverted residuals and linear bottlenecks”, in Proceedings of the IEEE conference on computer vision and pattern recognition (2018), pp. 4510–4520.

[153] M. Tan and Q. V. Le, “EfficientNet: rethinking model scaling for convolutional neural networks”, (2019), arXiv:1905.11946 [cs.LG].

[154] S. Ioffe and C. Szegedy, “Batch Normalization: accelerating deep network training by reducing internal covariate shift”, in Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML’15 (2015), pp. 448–456, arXiv:1502.03167 [cs.LG].

[155] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors”, (2012), arXiv:1207.0580 [cs.NE].

[156] *chipsnet git repository*, https://gitlab.com/chipsneutrino/chips-net.

[157] Martn Abadi et al., *TensorFlow: large-scale machine learning on heterogeneous systems*, Software available from https://www.tensorflow.org/, 2015.

[158] T. Theodore, “Particle identification in Cherenkov detectors using convolutional neural networks”, https://particlephysics.ca/wp/wp-content/uploads/CERN_summer_student_report_2016_Theodore_Tomalty.pdf, MA thesis (University of Toronto, 2016).

[159] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks”, in Proceedings of the IEEE conference on computer vision and pattern recognition (2018), pp. 7132–7141.

[160] F. Chollet et al., *Keras*, `https://keras.io`, 2015.

[161] A. Kendall, Y. Gal, and R. Cipolla, “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics”, in Proceedings of the IEEE conference on computer vision and pattern recognition (2018), pp. 7482–7491.

[162] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization”, (2014), arXiv:`1412.6980 [cs.LG]`.

[163] L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, “Sherpa: robust hyperparameter optimization for machine learning”, (2020), arXiv:`2005.04048 [cs.LG]`.

[164] M. Acero et al. (NOvA), “New constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in the NOvA experiment”, Physical Review D **98**, 032012 (2018), arXiv:`1806.00096 [hep-ex]`.

[165] B. List, “Why and when to optimize efficiency times purity”, ETH Zurich internal note (2002).

[166] K. Abe et al. (T2K Collaboration), “Measurements of neutrino oscillation in appearance and disappearance channels by the T2K experiment with 6.6×10^{20} protons on target”, Phys. Rev. D **91**, 072010 (2015).

[167] L. v. d. Maaten and G. Hinton, “Visualizing data using t-SNE”, Journal of Machine Learning Research **9**, 2579–2605 (2008).

[168] J. Djolonga, J. Yung, M. Tschannen, R. Romijnders, L. Beyer, A. Kolesnikov, J. Puigcerver, M. Minderer, A. D’Amour, D. Moldovan, S. Gelly, N. Houlsby, X. Zhai, and M. Lucic, “On robustness and transferability of convolutional neural networks”, (2020), arXiv:`2007.08558 [cs.CV]`.

[169] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets”, PloS one **10**, e0118432 (2015).

[170] K. Abe, Y. Hayato, T. Iida, K. Iyogi, J. Kameda, Y. Kishimoto, Y. Koshio, L. Marti, M. Miura, S. Moriyama, et al., “Calibration of the super-kamiokande detector”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **737**, 253–272 (2014).

[171] L. Berns, "New ideas for ring formation and reconstruction in water-Cherenkov detectors", Journal of Physics: Conference Series **1468**, 012165 (2020).

List of figures

List of tables