

UNIVERSITY OF COPENHAGEN

MASTER THESIS

---

**Determining Detector Efficiency And Timing  
Through Muon Tomography And Simulation  
Calibration Using Graph Neural Networks On  
IceCube Data**

---

*Author:*  
Simon C. DEBES

*Supervisor:*  
Troels C. PETERSEN

*A thesis submitted in fulfillment of the requirements  
for the degree of Master of Science  
in the*

**Experimental Particle Physics  
The Niels Bohr Institute**

September 5, 2023

*"The best that most of us can hope to achieve in physics is simply to misunderstand at a deeper level."*

Wolfgang Pauli

UNIVERSITY OF COPENHAGEN

## *Abstract*

Faculty of Science  
The Niels Bohr Institute

Master of Science

### **Determining Detector Efficiency And Timing Through Muon Tomography And Simulation Calibration Using Graph Neural Networks On IceCube Data**

by Simon C. DEBES

Some physical effects cause measurements (pulses) in the detector to be split needlessly, and a pulse merging algorithm is developed. The algorithm, when applied to both MC and data improves the similarity of the charge distributions in HLC hits and the number of pulses per DOM. Since DynEdge trains on pulses, the pulse-level data/MC agreement improvement could improve DynEdge's performance while slightly decreasing the size of the data.

By geometrically reconstructing a photon vector for each DOM, the efficiency of the entire detector is determined to be  $0.22 \exp -\frac{x}{44m} + 0.020$  for data and  $0.45 \exp -\frac{x}{42m} + 0.025$  for MC, where  $x$  is the distance the photon propagates, to great statistical precision. The plot of efficiency as a function of the zenith angle of the photon is shown to have a different shape in data than it does in MC, and the efficiency is found to be non-uniform in azimuth. The efficiency also varies greatly with the  $z$ -coordinate, seemingly in accordance with previously determined values of absorption lengths in the ice as a function of  $z$ .

The reconstructed photon vectors are also used to compute the time at which the muon stops ( $t_A$ ) from all pulses. How much  $t_A$  varies, indicates the resolution of the timing in the detector, and how much the photons scatter in the ice. The resolution is measured for different distances to the DOMs and is fitted with the linear function  $0.75 \frac{\text{ns}}{\text{m}} \cdot x + 50\text{ns}$  in data, and  $0.65 \frac{\text{ns}}{\text{m}} \cdot x + 48\text{ns}$  where  $x$  is the propagation distance of the photon. The offset corresponds to the resolution excluding the effects of the ice. Lastly, the resolution is calculated for MC using the true values and the reconstructed values. The difference between the two makes up the resolution due to the imperfect resolution.

## *Acknowledgements*

Thank you to my supervisor Troels Petersen for being an amazing presence in the office. You are always genuinely eager to discuss physics, and I always leave your meetings with rekindled inspiration and motivation. Coupled with your encouraging and positive spirit, I could not have asked for a better supervisor.

Thank you to my fellow master's students, Peter, Andreas, Morten, Moust, Frederik, and Hermansen whom I've had the pleasure of sharing an office with. We have shared a lot of breakfasts, dinners, and even some trips, and without you, this year surely would have been a lot less enjoyable.

I would also like to thank Rasmus Ørsøe and Andreas Søgaard who have been a massive help in offering brilliant insights on the weekly GraphNeT meetings.

Lastly, I would like to thank Sofie and my parents for reminding me there are things more important than muons.

# Contents

<b>Abstract</b>	ii
<b>Acknowledgements</b>	iii
<b>1 Introduction</b>	<b>1</b>
<b>2 Particle Physics</b>	<b>3</b>
2.1 The standard model . . . . .	3
2.2 The Neutrino . . . . .	4
2.3 Neutrino Oscillation . . . . .	5
2.4 Cherenkov Radiation . . . . .	6
2.5 Interactions . . . . .	8
2.6 The Muon . . . . .	10
2.6.1 Stopped Muons . . . . .	11
2.7 The Origins Of The Particles . . . . .	11
<b>3 The IceCube Neutrino Observatory</b>	<b>13</b>
3.1 The Detector . . . . .	13
3.1.1 DeepCore . . . . .	13
3.1.2 IceTop . . . . .	14
3.1.3 Challenges of the Detector . . . . .	15
3.2 Digital Optical Modules . . . . .	16
3.2.1 Photomultiplier Tubes . . . . .	17
3.2.2 Digitizers . . . . .	17
3.3 From Photon to Hit . . . . .	18
3.4 Ice Properties . . . . .	19
3.4.1 Dust . . . . .	22
<b>4 Data</b>	<b>24</b>
4.1 Features . . . . .	24
4.2 Monte Carlo Simulation . . . . .	25
4.2.1 Reconstruction . . . . .	26
4.3 Noise . . . . .	26
4.4 Pre-Pulses, Afterpulses, Late Pulses and Pulsesplitting . . . . .	26
4.4.1 Late Pulses . . . . .	27
4.4.2 Afterpulses . . . . .	27
4.4.3 Pre-Pulses . . . . .	27
4.5 IceCube Data Selection . . . . .	28
4.5.1 Local Coincidence Hits . . . . .	28
4.5.2 IceCube Neutrino Selection Levels . . . . .	28

<b>5 Machine Learning</b>	<b>30</b>
5.1 Basics of Machine Learning . . . . .	30
5.1.1 Supervised Learning . . . . .	30
5.1.2 Loss Functions . . . . .	31
5.1.3 Gradient Descent . . . . .	33
5.1.4 Model Convergence . . . . .	34
5.2 Neural Networks . . . . .	35
5.3 Graph Neural Networks . . . . .	37
5.4 DynEdge . . . . .	38
<b>6 Analysis</b>	<b>40</b>
6.1 Overview . . . . .	40
6.2 Data Selection . . . . .	41
6.2.1 Event Type Multiclass Classification . . . . .	41
6.2.2 Stopped Muon Classifier . . . . .	41
6.2.3 Reweighting Data . . . . .	43
6.3 Reconstruction . . . . .	44
6.3.1 Muon Azimuth Reconstruction . . . . .	44
6.3.2 Muon Zenith Reconstruction . . . . .	46
6.3.3 Muon Stopping Point Reconstruction . . . . .	48
6.4 Pulse merging . . . . .	51
6.4.1 The Effect of Merging on Charge Distribution . . . . .	52
6.4.2 Pulses Per DOM . . . . .	52
6.5 Photon Reconstruction . . . . .	54
6.6 Efficiency Analysis . . . . .	56
6.6.1 Photon Zenith Angle . . . . .	57
6.6.2 Photon z Coordinate . . . . .	58
6.6.3 Photon Azimuth Angle . . . . .	61
6.7 Hit Timing . . . . .	63
6.7.1 Pulse selection . . . . .	64
6.7.2 Resolution Of The Timing . . . . .	66
<b>7 Conclusion And Outlook</b>	<b>70</b>
7.1 Conclusion . . . . .	70
7.2 Outlook . . . . .	71
7.2.1 Pulsemerging . . . . .	72
7.2.2 Photon Vector Reconstruction . . . . .	73
7.2.3 Efficiency . . . . .	73
7.2.4 Scattering . . . . .	74
7.2.5 Hit Timing . . . . .	74
<b>A Cut-Plots</b>	<b>75</b>
A.1 Systematic Monte Carlo Sets . . . . .	77

# List of Figures

1.1	Illustration showing the different areas that IceCube are studying. Illustration from [1]. . . . .	1
2.1	The particles of the standard model and their properties. Illustration from [5] . . . . .	3
2.2	A plot showing the probability of a muon neutrino having becoming a tau neutrino upon reaching IceCube, as a function of the zenith angle and the energy of the particle. [11] . . . . .	7
2.3	On the left, a charged particle moving slower than the phase velocity of light. On the right, a charged particle moving faster than the phase velocity of light, causing the wavefronts to bunch up. . . . .	8
2.4	A plot showing the Feynman diagrams of the neutrino interactions. [15] . . . . .	9
2.5	Each dot is a DOM. The size indicates the measured charge, and the color indicates the time, where red dots are early hits, and green dots are later hits. [16] . . . . .	9
2.6	Two plots showing the cross-section of each of the three charged current interactions as a function of energy. The left image is for neutrinos and the right image is for antineutrinos. Illustration from [17] . . . . .	10
2.7	The stopping power of a muon as a function of muon momentum in copper. Illustration from [18] . . . . .	10
2.8	A figure showing the neutrino flux as a function of the energy. . . . .	12
2.9	Illustration of the subatomic particles that are created in the air shower produced when incoming cosmic rays decay in the atmosphere. . . . .	12
3.1	Schematic showing the architecture of the IceCube detector. . . . .	14
3.2	Figure showing the DOM distribution in the detector as well as how the absorption varies with depth. Illustration from [22]. . . . .	15
3.3	The schematics of a digital optical module. Illustration from [24]. . . . .	16
3.4	Schematic showing the composition of a PMT. Illustration from [27]. . . . .	17
3.5	Images showing how the three digitizers respond to the same signal, and in the bottom image, the resulting pulses. Illustration from [30] . . . . .	18
3.6	Single photon electron (SPE) template. Illustration from [26]. . . . .	19
3.7	Schematic showing how a photon becomes a hit. Illustration from [25] . . . . .	20
3.8	Plots showing the effective scattering length and the absorption coefficient as a function of depth in the ice. Illustration from [23]. . . . .	20
3.9	Plot showing the temperature taken inside the AMANDA boreholes, as a function of the depth. Illustration from [33]. . . . .	21
3.10	Plots of different simulations of the angular acceptance rate of the DOMs as a function of the zenith angle. Illustration from [36]. . . . .	21
3.11	The ratio between the charge measured in data and the charge simulated in MC, as a function of the azimuth angle between emitter and receiver. Illustration from [37]. . . . .	22
3.12	The top image shows how photons propagate from a light source without accounting for birefringence, and the bottom image shows the propagation while accounting for birefringence. Illustration from [38]. . . . .	23

4.1	Plot showing a histogram of time between hits in HitSpool and the distributions of the causes [47]. . . . .	27
4.2	This image shows the ratio between the charge measured in data and the charge simulated in MC, as a function of the azimuth angle between emitter and receiver. Illustration from [30]. . . . .	29
5.1	MAE and MSE. . . . .	31
5.2	Binary cross entropy loss. . . . .	32
5.3	A figure showing an example of the parameter space of a neural network simplified to just three dimensions. Illustration from [51] . . . . .	33
5.4	A figure showing the separation of the two classes, blue and red. The black line represents a robust model that has converged, but not overfitted, and the green line represents an overfitted model. Illustration from [53] . . . . .	35
5.5	Figure showing how the loss of the test set and validation set evolves through iterations. Illustration from [54] . . . . .	35
5.6	A neuron that takes two inputs and returns one output. Illustration from [55] . . . . .	36
5.7	The two activation functions, sigmoid and RELU. . . . .	36
5.8	Neural network consisting of three layers of neurons. Illustration from [55] . . . . .	37
5.9	The architecture of DynEdge. Illustration from [2]. . . . .	39
6.1	Decision tree showing the selection process. First, the events that are predicted to be muons are selected, and then the muons that are predicted to stop within the detector are selected. . . . .	41
6.2	The distribution of the probability of being stopped for each muon. Stopped muons are in orange and through muons are in orange. . . . .	42
6.3	The ROC curve of the predictions of the muons being stopped muons on the test set. . . . .	43
6.4	The four figures show the distributions of the four features that we are using to weight the events before and after weighting. . . . .	43
6.5	This plot shows the residuals of the predictions of the azimuth angle for the validation set of the MuonGun stopped muons. . . . .	45
6.6	The distributions of the true values of azimuth in blue, and the predicted values of azimuth in orange. . . . .	45
6.7	The residuals of the predictions of the zenith angle for the validation set of the MuonGun stopped muons. . . . .	46
6.8	The distributions of the true values of zenith in blue, and the predicted values of zenith in orange. . . . .	47
6.9	The residuals of the predictions of the three stopping coordinates for the validation set of the MuonGun stopped muons. . . . .	48
6.10	The distribution of the distance between the predicted stopping point and the true stopping point for all stopped muons. . . . .	49
6.11	The distributions of the x-coordinates of the test set, and their corresponding predictions. . . . .	49
6.12	The distributions of the y-coordinates of the test set, and their corresponding predictions. . . . .	49
6.13	The distributions of the z-coordinates of the test set, and their corresponding predictions. . . . .	50
6.14	Visualization of three pulses recorded during an event before and after merging. . . . .	51
6.15	Charge distribution of HLC and SLC separately in MC and data, before and after merging. The plots in the bottom row are the ratios between MC and data before and after merging. . . . .	52

6.16 Charge distribution of HLC hits in MC and data before and after merging. The plots in the bottom row are the ratios between the distributions. . . . .	53
6.17 The number of pulses per DOM in MC and data before and after merging pulses. . . . .	53
6.18 Illustration of the points and vectors needed for photon reconstruction. . . . .	54
6.19 The top plot shows histograms of the photon distance to all DOMs and activated DOMs respectively, for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot. . . . .	58
6.20 The top plot shows histograms of the photon distance to all DOMs and activated DOMs respectively, for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot. . . . .	59
6.21 The top plot shows histograms of the photon $z$ coordinate for all DOMs and activated DOMs respectively, for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot. . . . .	60
6.22 The efficiency of the detector in data as a function of the photon $z$ coordinate. . . . .	60
6.23 The top plot shows histograms of the number of DOMs and activated DOMs for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot. . . . .	61
6.24 The ratio of recorded charge for data vs. MC and data vs. truth, as a function of the azimuth angle. . . . .	62
6.25 Figure showing the muon trajectory in the detector and the emitted Cherenkov radiation. . . . .	63
6.26 Histogram of the pulse series, where the median is marked in red, and the 800ns acceptance interval is shown by the grey dotted lines. The mean of the pulses within the interval is marked by the black dotted line. . . . .	64
6.27 Fabricated example a pulse series containing two peaks, where the median of the series is located in between the peak, such that no signal is within the acceptance window. . . . .	65
6.28 Histograms of the mean $t_A$ for data, MC, and truth. . . . .	65
6.29 Histograms showing the distributions of all pulses and those that are in the allowed time interval for data . . . . .	65
6.30 Histograms showing $t_A$ for MC, data, and truth when applying no restriction, and selecting photons traveling less than 100m respectively. . . . .	66
6.31 The distribution of $t_A$ for photons traveling between 0m and 5m, fitted with a CrystalBall function. . . . .	67
6.32 The distribution of $t_A$ for photons traveling between 50m and 55m, fitted with a CrystalBall function. . . . .	67
6.33 The distribution of $t_A$ for photons traveling between 95m and 100m, fitted with a CrystalBall function. . . . .	67
6.34 Plot of the resolution of the timing for data, MC, and MC truth, as a function of photon travel distance. . . . .	68
6.35 Plot of the resolution of the timing for data, MC, and MC truth, as a function of photon travel distance. The effects of the reconstruction are also plotted. . . . .	69
6.36 Plot of the resolution of the timing for data, MC, and MC truth, as a function of photon travel distance. The effects of the ice and the reconstruction are also plotted. . . . .	69
7.1 2D histogram of the time difference and charge ratio between pulses in a DOM during an event for data and MC. . . . .	72

7.2 Plot showing the charges and timings of pulses in a DOM. The mean, median, and time interval are shown by the vertical line . . . . .	74
A.1 The charge distribution of the first two pulses in a DOM, divided by the time difference between the two pulses for data . . . . .	75
A.2 The charge distribution of the first two pulses in a DOM, divided by the time difference between the two pulses for MC . . . . .	75
A.3 The charge distribution for pulse series, divided into how many pulses are in the pulse series. . . . .	76
A.4 The distribution of the number of pulses, number of strings, number of DOMs activated, and the average number of pulses per DOM in MC and data. . . . .	76

# List of Tables

4.1	The features of the event file.	24
4.2	The features of the pulsemmap.	25
6.1	The number of events before and after both selections.	42
6.2	The features of the pulsemmap after photon reconstruction.	57
7.1	The number of events before and after both selections.	70
7.2	The resolutions of all the reconstructed variables.	70
7.3	The efficiency as a function of photon travel distance for data, MC, and truth respectively.	71
7.4	The resolution as a function of photon travel distance for data, MC, and truth respectively.	71
A.1	Table showing the different systematic sets, and which parameters have been changed. Table from [48]	77

## Chapter 1

# Introduction

Neutrinos are very light fundamental particles that only very rarely interact with anything. Their low interaction rate makes them very hard to detect but also makes them more interesting to study. The traditional method of astronomy relies on the observation of photons that are easily absorbed or bent by electromagnetic or gravitational fields. Neutrinos on the other hand do not interact with anything until we detect them here on Earth. Therefore they will travel in a straight line from the point of creation to the point of observation, which cannot be guaranteed by photons, allowing for more accurate source detection. This is the premise of **multimessenger astronomy**, allowing us to observe events through different methods that offer different benefits. Neutrinos in themselves also have some very interesting and poorly understood properties such as neutrino oscillation and neutrino mass ordering, the study of which could lead to physics beyond the standard model.

The IceCube collaboration lowered thousands of detectors into a cubic kilometer of ice at the south pole to detect these weakly interacting particles for the study of the concepts mentioned above, along with many others detailed in figure 1.1.

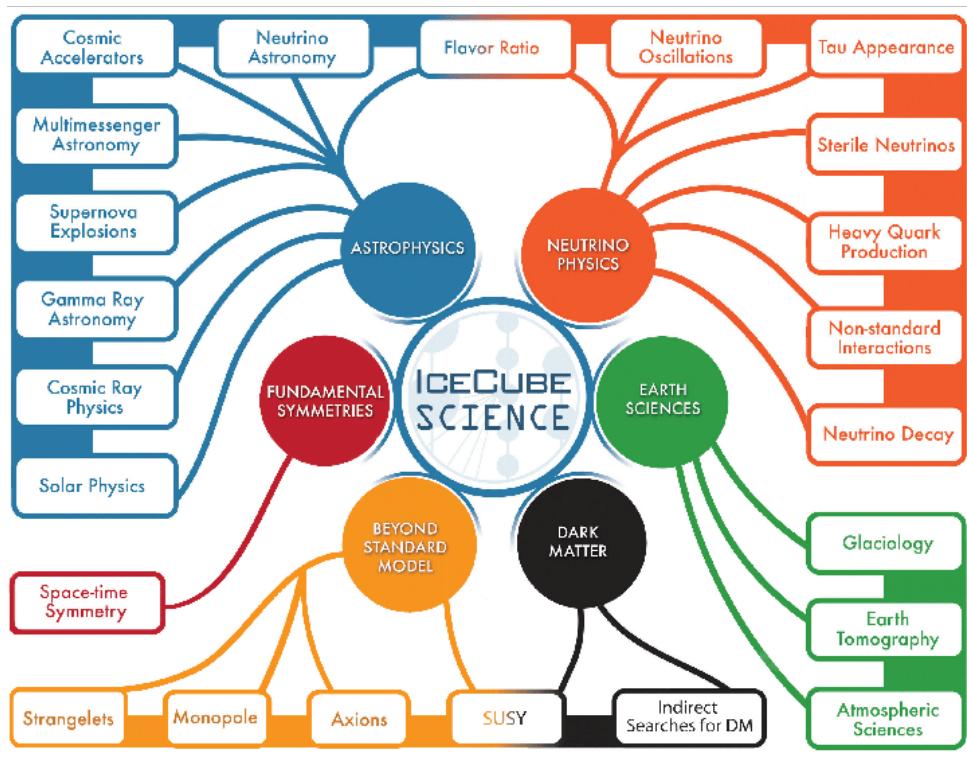


FIGURE 1.1: Illustration showing the different areas that IceCube are studying.  
Illustration from [1].

In all these endeavors, they use recorded data from the detector alongside events simulated using **Monte Carlo (MC)** techniques. IceCube bases their reconstruction on simulations, meaning that the reconstruction is only as good as the simulation. The goal of this thesis is to improve on the systematic errors of both the simulations, by investigating the disagreement between data and MC, and the measurement of data, through an investigation of the timing resolution of data events and MC events. These simulations depend on many parameters, such as how photons propagate in the ice, and the synchronization of the many modules to maintain consistent timing, and it is these parameters we aim to understand better in this thesis. Such an improvement would result in better performance for all reconstruction and analyses in IceCube.

The current algorithm used for event reconstruction at IceCube, **RETRO**, is relatively slow. In recent years it has been proven that machine learning methods can improve reconstruction speeds by orders of magnitude, while also improving performance considerably [2]. **DynEdge** of the **GraphNeT** [3] python library has proven to be one of the top models available regarding neutrino telescope event reconstruction, and it is therefore used here to identify and reconstruct stopped muons using machine learning at higher statistics than previously seen, to achieve optimal resolution.

Muons are very well suited for calibration because they are so abundant compared to the elusive neutrinos. Stopped muons are even better because we know where they stop, from which we can infer their energy more accurately. They also have lower energy than muons that make it through the detector, and therefore lose energy at a more constant rate, and thus behave more predictably, making them easier to reconstruct.

With this large sample of stopped muons I will begin my analysis of three parts. The first part, **Pulsemerging**, attempts to highlight where data and MC are different and introduces my algorithm **PulseMerger**. PulseMerger combines measurements (pulses) where there is reason to believe a pulse was mistakenly split, and the signal is represented more accurately as one pulse.

The second part reconstructs the trajectories of a large amount of recorded photons and uses them to analyze the efficiency of the detector as a function of the angles, depth, and distance to the DOM.

The third part uses the reconstructed photons to study their scattering and how well the detectors are synchronized based on the timing of the photons. The resolution of the detector will be modeled and the contributions from different components are identified. How the photons propagate in ice also reveals properties of the ice, which will also improve simulations.

In figure 1.1, this thesis would fall under cosmic ray physics, glaciology, and Earth tomography. Nevertheless, the result of this work would improve every other analysis, as it could improve the simulations in IceCube.

## Chapter 2

# Particle Physics

### 2.1 The standard model

The standard model (SM) is currently the best singular model physics has to offer to explain the fundamental particles and forces of the universe [4]. It explains and describes matter, antimatter, and their interactions through the electromagnetic force, **the weak nuclear force**, and **the strong nuclear force**. Simply put, the strong force is the force that keeps atoms together, and the weak force is the force that breaks them apart or changes them. Notably the model is incomplete as it is not compatible with gravity and multiple other phenomena, such as neutrinos having mass. Studying neutrinos could therefore be the key to a more complete model of the universe.

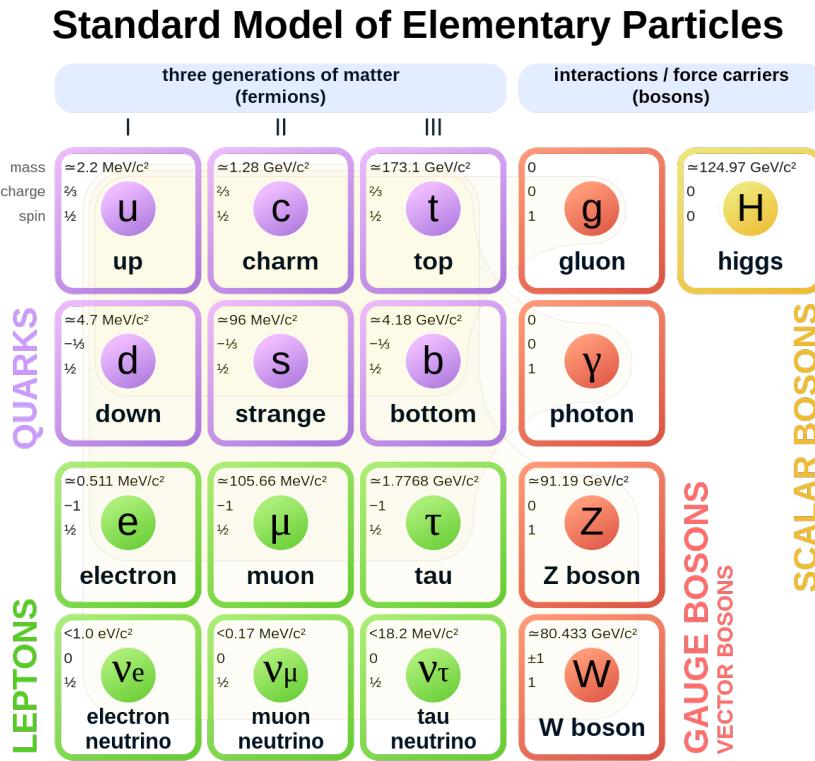


FIGURE 2.1: The particles of the standard model and their properties. Illustration from [5]

The elementary particles of the standard model are split into two groups; the fermions and bosons. The **bosons** are integer-spin particles that obey Bose-Einstein statistics. The gauge bosons are the mediators of the fundamental forces covered by the standard model. The gluon

mediates the strong force, the W and Z bosons mediate the weak force, and the photon mediates the electromagnetic force. Lastly, the Higgs boson is what gives all the other massive particles their mass [6].

The **fermions**, all spin 1/2 particles that obey Fermi-Dirac statistics, are further split into two groups; the **quarks** and the **leptons**. The quarks, unable to exist in singular form, make up baryons (quark-trios) like protons and neutrons, and mesons (quark-pairs) like pions and kaons. Particles that consist of at least two quarks are also called **hadrons**.

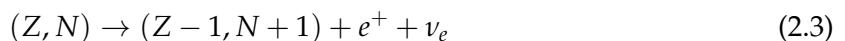
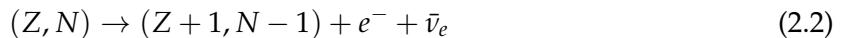
The leptons consist of the three charged particles, electron, muon, and tauon, and their respective neutrino counterpart (electron-neutrino, muon-neutrino, tau-neutrino).

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} \quad (2.1)$$

There are three generations (or flavors) of fermions, and their masses increase with each generation. Due to the law of conservation of lepton numbers, each charged lepton must decay into its corresponding neutrino.

## 2.2 The Neutrino

The neutrino was first postulated by physicist Wolfgang Pauli in 1930 [7] when looking at  $\beta$ -decays [4]. A  $\beta$ -decay is a process where a neutron in a nucleus decays into a proton, or a proton decays into a neutron.



where  $(Z, N)$  is a nucleus with proton number  $Z$  and nucleon number  $N$ . Due to conservation of charge, this decay demands the emission of a  $\beta$ -particle, which can be either an electron ( $e^-$ ) or a positron ( $e^+$ ). Initially,  $\beta$ -decay processes were believed to be two-body problems. Since only singular electrons would be emitted from the nucleus, it was expected that they would assume specific discrete energies, but that was not what was observed. Instead, the energies of the emitted electrons were continuous. In an attempt to explain this phenomenon, Pauli postulated that  $\beta$ -decays were not a two-body problem, but a three-body problem. If the process produced another particle alongside the  $\beta$ -particle, and the energy would be shared between the two produced particles, it would explain the continuous energy distribution. As the two-body problem satisfied charge conservation already the proposed particle would need to be neutral, hence the name *neutrino* which means "little neutral one" in Italian.

In the absence of neutrinos in eq. (2.2) all the energy lost by the nucleus

$$E_e = \Delta Mc^2 = [M(Z, N) - M(Z + 1, N - 1)] c^2 \quad (2.4)$$

would go to the emitted electron. Since the antineutrino is there as well, the energy of the electron is only bounded by

$$m_e c^2 \leq E_e \leq (\Delta M - m_{\nu_e}) c^2. \quad (2.5)$$

Since the two leptons share the energy in any fraction, the upper bound on the electron neutrino mass also can be estimated experimentally in this way, by isolating  $m_{\nu_e}$  when  $E_e = \max(E_e)$ .

In 1956 Clyde L. Cowan and Frederick Reines published *Detection of the Free Neutrino: a Confirmation* [8], wherein they described the first experimental observation of neutrinos. They used radioactive fission products from a reactor, that were likely to experience  $\beta$ -decay, producing

many antineutrinos. These antineutrinos would interact with protons, in inverse  $\beta$ -decay processes, producing positrons, which upon annihilation with an electron, would produce gamma rays that were amplified by a liquid scintillator, and detected by a photomultiplier tube.

In the late 1960s, after the discovery of electron neutrinos, Raymond Davis and John Bahcall conducted the Homestake experiment [9], named after the mine in South Dakota in which it took place. A large tank was placed deep below the Earth's surface to detect the flux of electron neutrinos from the Sun. The tank contained Chlorine-37, which becomes Argon-37, as electron neutrinos interact with it. When the experiment was done, they determined the amount of Argon-37 in the tank and found it was 1/3 of the expected amount based on neutrino flux calculations in the Sun. This problem was called the solar neutrino problem and was explained many years later by the inception of the idea of **neutrino oscillation**, where neutrinos can change flavors as they propagate. The Sun almost only produces the electron flavor neutrino and upon reaching the Chlorine tanks, two-thirds of the neutrinos had changed flavor to muon and tau neutrinos.

As the neutrino has no charge, it cannot interact electromagnetically and only interacts through the weak force. A criterion for the ability to oscillate is that there exist mass states of different masses, meaning neutrinos cannot have zero mass as predicted by the standard model. At lengths greater than the scale of an atom (femtometers), the strength of the weak force diminishes drastically, unlike the electromagnetic force, meaning that the weak-force cross-section is much lower than the electromagnetic cross-section. Therefore neutrinos interact much less often than their charged cousins, so much that even though about 100 trillion neutrinos pass through your body every second, only around 275 neutrinos are detected in the cubic kilometer of ice in IceCube per day [10]. Therefore a neutrino detector needs to be very big, as the detection rate scales with interaction volume. The low detection rate has also earned the neutrino the nickname "the ghost particle".

## 2.3 Neutrino Oscillation

Neutrino oscillation is the phenomenon where a neutrino changes into another flavor of neutrinos after traveling for a long enough distance, i.e. it oscillates between neutrino states. This cannot occur without **neutrino mixing**, which is the assumption that the three flavor states of neutrinos  $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$  do not have a certain mass, but exist as linear combinations of three mass eigenstates,  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ , with definite mass  $m_1$ ,  $m_2$ , and  $m_3$  [11]. The oscillation then occurs between these mass states, and because oscillation has been observed, the mass of the three states must be different and therefore non-zero [4]. The relationship between the flavor states and mass states is governed by the unitary mixing matrix  $U$ , called the **PMNS matrix**<sup>1</sup> [12].

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = U \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} \quad (2.6)$$

Each flavor state,  $\alpha = e, \mu, \tau$ , can be expressed by its mass state components,  $i = 1, 2, 3$ , through the PMNS matrix elements

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle. \quad (2.7)$$

Each element of the matrix  $U_{\alpha i}$  represents the amplitude of the mass state  $i$ , in terms of the flavor state  $\alpha$  and  $|U_{\alpha i}|^2$  describes the fraction of  $\nu_\alpha$  being in the mass state  $\nu_i$ .  $U$  can be

---

<sup>1</sup>Named after Pontecorvo, Maki, Nakagawa, and Sakata.

expressed as a product of three separate matrices that dictate the oscillation between the three possible combinations of the three states.

$$U = U_{23}U_{13}U_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{i\delta_{CP}} \\ 0 & 1 & 0 \\ s_{13}e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.8)$$

$$= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}e^{i\delta_{CP}} & c_{23}c_{13} \end{bmatrix}$$

where  $\sin \theta_{ij}$  and  $\cos \theta_{ij}$  is shortened to  $s_{ij}$  and  $c_{ij}$  respectively,  $\theta$  is the mixing angle, and  $\delta_{CP}$  is a charge-parity violating phase.

For simplicity, let's consider the mixing between 1 and 2, governed by  $U_{12}$ , where we will leave out the third row.

$$\begin{bmatrix} \nu_\alpha \\ \nu_\beta \end{bmatrix} = \begin{bmatrix} c_{12} & s_{12} \\ -s_{12} & c_{12} \end{bmatrix} \begin{bmatrix} \nu_i \\ \nu_j \end{bmatrix} \quad (2.9)$$

Then the two flavor states become

$$\nu_\alpha = \nu_i \cos \theta_{ij} + \nu_j \sin \theta_{ij} \quad (2.10)$$

$$\nu_\beta = -\nu_i \sin \theta_{ij} + \nu_j \cos \theta_{ij}. \quad (2.11)$$

The probability of finding  $\nu_\alpha$  as  $\nu_\beta$  is given as

$$P(\nu_\alpha \rightarrow \nu_\beta) \approx \sin^2(2\theta) \sin^2\left(\frac{L}{L_0}\right) \quad (2.12)$$

where  $L$  is the distance traveled, and  $L_0$  is the oscillation length defined as

$$L_0 = \frac{4E\hbar c}{(m_j^2 - m_i^2)c^4} \quad (2.13)$$

and usually ranges from  $\mathcal{O}(10^2)\text{m}$  to  $\mathcal{O}(10^5)\text{m}$  [4]. Here  $E$  is the energy of the particle, and  $\hbar$  is the reduced Planck constant. Equations 2.12 and 2.13 show that the probability depends on the distance traveled,  $L$ , the masses of  $\nu_i$  and  $\nu_j$  and the energy of the particle. The oscillation length depends on the mass difference of  $\nu_i$  and  $\nu_j$ , and if they were equal, the denominator would be 0, and the length would be undefined, showing that neutrinos must have mass.

Figure 2.2 shows the probability of a muon neutrino produced in the atmosphere being detected as a tau neutrino upon reaching IceCube. The probability depends on the energy, and the zenith angle of the neutrino, as it dictates the distance from the outer atmosphere to the detector. The structure of the plot is periodical because the muon neutrino will oscillate between the states.

It is known that  $\nu_1$  and  $\nu_2$  have similar but not identical masses and that  $\nu_3$  either has a much larger mass (called normal mass ordering, NMO) or a much smaller mass (called inverted mass ordering, IMO). These orderings are also studied at IceCube.

## 2.4 Cherenkov Radiation

The IceCube detector records data by measuring light, but as we have established, neutrinos do not interact electromagnetically and are thus incapable of emitting light. What IceCube actually

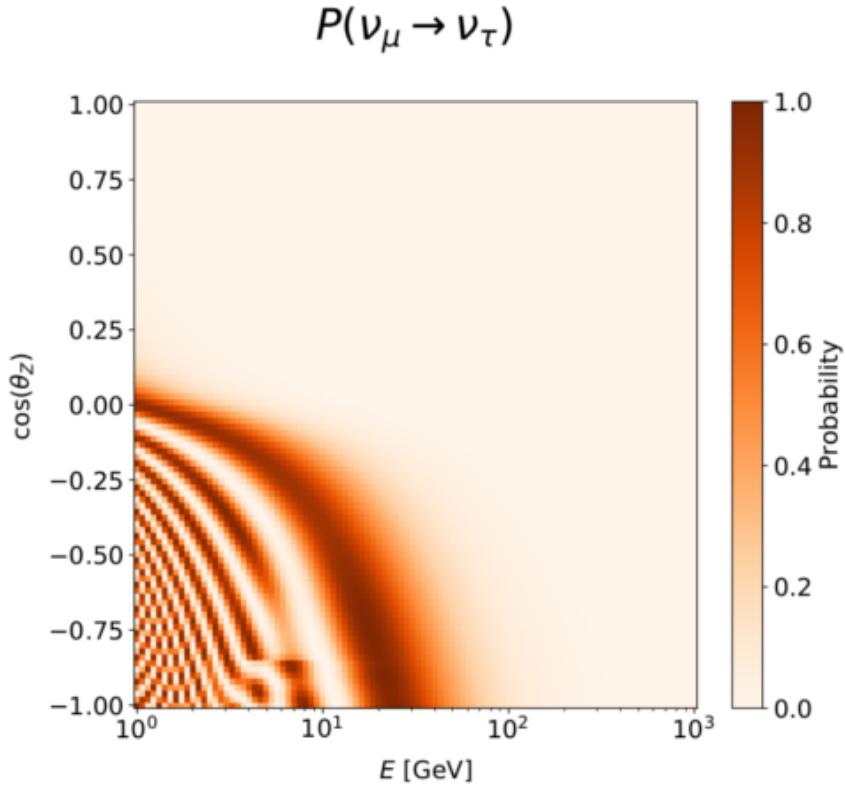


FIGURE 2.2: A plot showing the probability of a muon neutrino having becoming a tau neutrino upon reaching IceCube, as a function of the zenith angle and the energy of the particle. [11]

records is the **Cherenkov radiation** from the charged secondary particles the neutrinos decay into. Therefore neutrinos are only measured indirectly after they decay. Muons on the other hand are charged particles that emit Cherenkov radiation.

It is commonly known that nothing can move faster than the speed of light in a vacuum,  $c = 299,792,458 \text{ m/s}$ . If light has to travel through some optical medium such as water or the ice at the south pole, it will decrease the phase velocity of the light. The phase velocity decreases because photons, being electromagnetic fields, interact with the electromagnetic fields of the charged particles in the ice. These particles absorb the photons and re-emit them in the same direction. This process does not change the direction of the light but introduces a slight delay. The ratio between  $c$  and the velocity of the light in the medium is called the medium's **refractive index**,  $\eta$ . The refractive index generally increases with the density of the medium, as there will be more charged particles to interact with. Cherenkov light is a continuous spectrum, but peaks around 420nm [13], producing the characteristic blue color. The refractive index of ice varies slightly with the depth of the ice, and the wavelength of the light. A study in IceCube was done to describe the refractive index more accurately [14], but the error was deemed minimal, so they decided to use a fixed value of  $\eta = 1.32$ , and so will we in this thesis.

With this value, we can calculate the velocity of light in the ice to be

$$v_\gamma = \frac{1}{\eta}c = \frac{1}{1.32}c \approx \frac{3}{4}c. \quad (2.14)$$

As the charged particle travels through the medium, it will polarize the medium, exciting the electrons in the atoms. As the particle travels away, the electrons fall back into lower energy states and emit photons. If the charged particle causing this polarization is moving slower

than the phase velocity of light, these light waves emitted from the medium will interfere destructively, which is why we don't see it. If, on the other hand, the particle moves faster than the phase velocity of light, the waves will interfere constructively, and form a wavefront. The polarization occurs isotropically around the charged particle, such that light is emitted in all directions. If the source moves faster than the emitted waves, they will bunch up in front of the source, creating a cone of light like in figure 2.3.

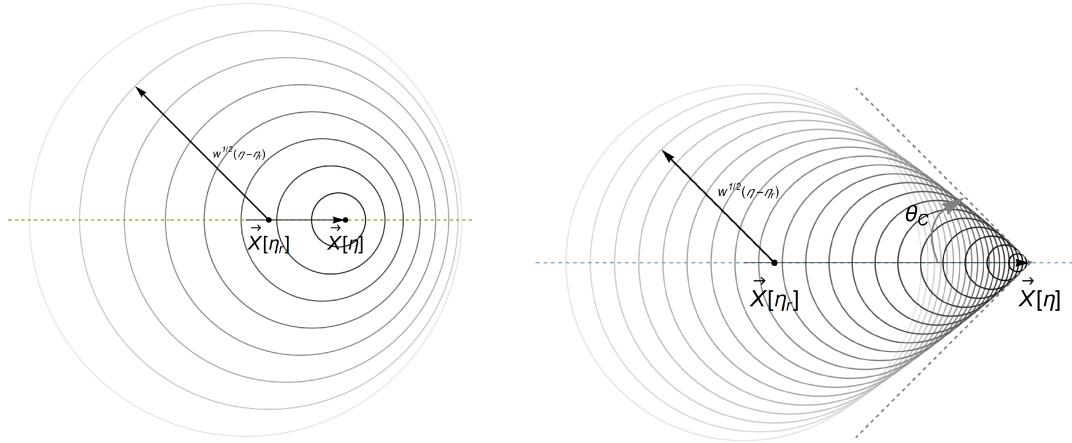


FIGURE 2.3: On the left, a charged particle moving slower than the phase velocity of light. On the right, a charged particle moving faster than the phase velocity of light, causing the wavefronts to bunch up.

The Cherenkov angle can be calculated directly from only the refractive index and the velocity of the particle

$$\theta_C = \arccos\left(\frac{1}{\beta\eta}\right) \quad (2.15)$$

where  $\beta$  is the velocity relative to  $c$ ,  $\beta = v/c$ . For ultra-relativistic muons  $\beta \approx 1$ , so eq. 2.15 becomes

$$\theta_C = \arccos\left(\frac{1}{\eta}\right) = \arccos\left(\frac{1}{1.32}\right) = 40.75^\circ, \quad (2.16)$$

which is the value used in this thesis.

## 2.5 Interactions

In order for neutrinos to be detected they must decay into other charged particles, that are able to produce Cherenkov radiation. In IceCube, we observe this through four separate interactions that are either charged current (CC) interactions (mediated by the  $W^\pm$  boson) or neutral current interactions (NC) (mediated by the  $Z^0$  boson).

In figure 2.4 we see that:

- a) NC interaction does not change the flavor of the neutrino and only produces a cascade.
- b) The CC interaction between an electron neutrino and a nucleon results in both electromagnetic and hadronic cascades.
- c) The CC muon neutrino interaction is the only one that produces a muon, which can travel far, in addition to the hadronic cascade.

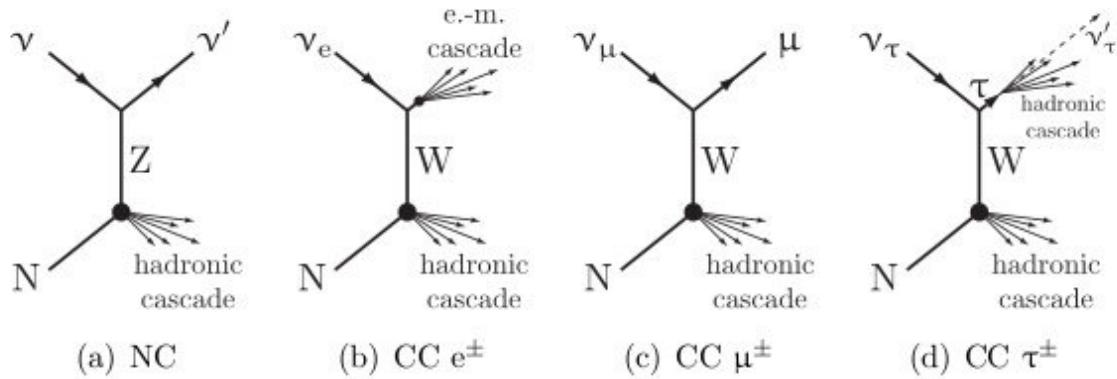


FIGURE 2.4: A plot showing the Feynman diagrams of the neutrino interactions.  
[15]

- d) The tau neutrino also produces its corresponding lepton, but with the short lifetime of tau, it quickly decays. Most often it will produce hadronic and electromagnetic showers, but sometimes it will also produce a muon and thus resemble (c).

Figure 2.6 shows that the rates of these interactions depend on the energy of the particle. Interactions below 1 GeV are dominated by the quasi-elastic scattering (QE), where a neutrino interacts with a nucleus in the ice, and a nucleon is scattered off. Higher energies (10+ GeV) are dominated by the deep inelastic scattering (DIS) where once again a nucleon is hit, but with enough energy to split the nucleon and produce a cascade of hadrons. Between 1 and 10 GeV, we find the resonant production (RES), where the neutrino excites the nucleon to a resonant state, which then decays to create a hadronic cascade.

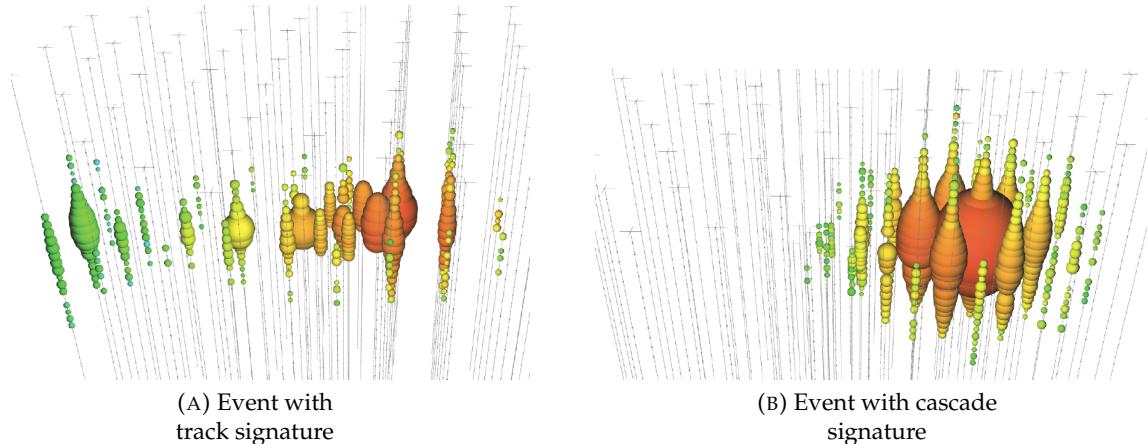


FIGURE 2.5: Each dot is a DOM. The size indicates the measured charge, and the color indicates the time, where red dots are early hits, and green dots are later hits. [16]

All non-noise events in IceCube are classified as either **track** or **cascade**. A cascade event will activate a lot of detectors around the interaction vertex, where a track will form a long trail of activated DOMs. Cascade events feature interactions with hadronic cascades, where tracks will indicate the trajectory of a far-reaching muon, as muons are the only charged particles to travel far in the detector. Figure 2.5 shows the detector signatures of the two types of neutrino events. As track events trigger DOMs along its path, and cascade events only trigger DOMs around the interaction point, it is easier to fit a vector to the long arrow-like track than to the spherical cascade.

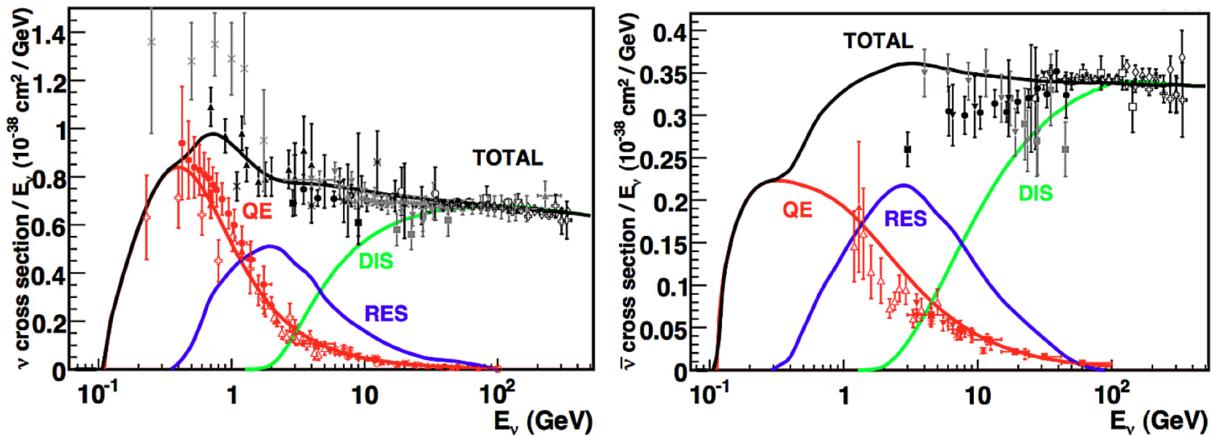


FIGURE 2.6: Two plots showing the cross-section of each of the three charged current interactions as a function of energy. The left image is for neutrinos and the right image is for antineutrinos. Illustration from [17]

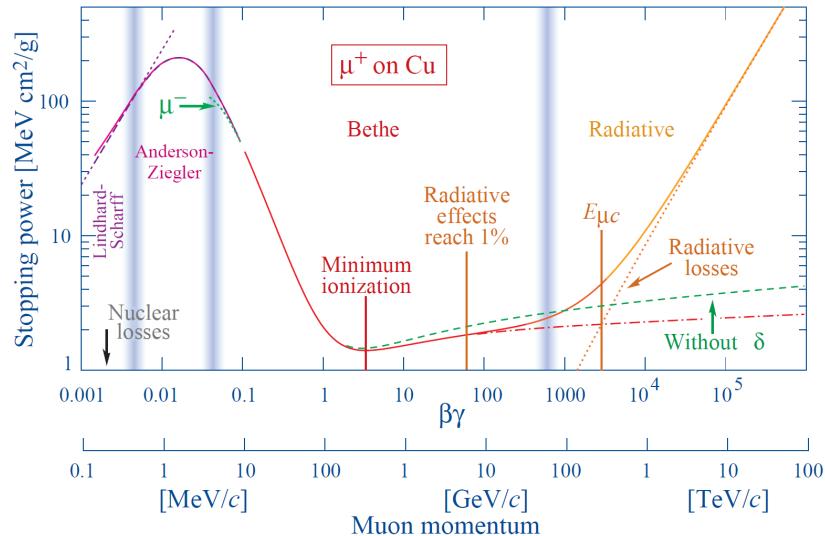


FIGURE 2.7: The stopping power of a muon as a function of muon momentum in copper. Illustration from [18]

Figure 2.6 shows that for low-energy neutrinos, all three charged interactions need to be considered, but for higher energies, only DIS is observed. This is true for both neutrinos and antineutrinos although their corresponding shapes are different in the figure.

## 2.6 The Muon

The lifetime of the muon is 2.2  $\mu$ s. Light can travel 660m during the lifetime of the muon, which is far below the distance from the outer layer of the atmosphere to IceCube. Cosmic and atmospheric muons will therefore only reach IceCube if they are moving fast enough to be relativistic, because relativistic particles experience time-dilation, where time is perceived as passing slower than for an observer at rest. This mechanic allows muons to travel far enough to be detected in IceCube. As a result, all muons detected in IceCube are going at approximately the speed of light. The electron has a very low mass and is very easily stopped, so it does not travel far. The tauon, while much heavier, has a very short lifetime, and does not travel far

either. The muon, being from the middle generation, has a mass and lifetime that allows it to travel far. Therefore it is also the particle that is recorded by IceCube by far the most.

2.7 shows the stopping power of muons as a function of the momentum of the muon in copper, but the curve is similar in ice, and the same principles apply. The stopping power of a particle is defined as the amount of energy the particle loses per distance traveled. The stopping power between the energies 100 MeV and 100 GeV is low, and for 100+ GeV it rapidly increases. This means that high-energy muons will lose their energy relatively quickly, lessening the difference in propagation length for high and low-energy muons. The similar propagation lengths mean that we see a broader spectrum of energies in muons that stop within the detector. It also causes a broader energy spectrum for any angle interval. This is beneficial for IceCube because the wider representation of energies and angles in stopped muons gives us a better foundation for simulating particles in all combinations of angles and energies.

### 2.6.1 Stopped Muons

IceCube is a neutrino observatory and muons are not the primary area of interest and are not studied very much in IceCube. Nevertheless, because neutrinos are observed so rarely compared to muons, the high statistics make muons more suited for calibrating the simulations. The subset of muons that decay within the detector is even better suited. Because we can fairly accurately deduce where they stop in the detector, and we know from figure 2.7 how muons lose energy, we can more precisely infer the energy of the muon. This is the reason the main analysis presented in this thesis is regarding stopped muons.

The stopped muons are obviously also the ones with the lowest energy, otherwise, they would not stop. The low-energy muons are more predictable because they emit light at a constant rate and intensity, unlike high-energy muons, and can therefore be reconstructed more accurately.

Muons are also used to determine the resolution with which IceCube can determine the angles of particles. The Moon blocks cosmic rays, and its position can therefore be determined by finding the missing circle of muons in the sky. Because we know exactly the position and size of the Moon, it is easy to compare results.

## 2.7 The Origins Of The Particles

The neutrinos detected at IceCube have multiple sources. The most interesting are active galactic nuclei, black holes, supernovae, and other cosmic events because they can be used to enlighten us about these mysterious phenomena.

Unfortunately, IceCube only detects neutrinos down to about 1 GeV, and as figure 2.8 shows, from around 0.1 GeV to 10 TeV, the flux of atmospheric neutrinos caused by cosmic rays dominates other sources.

Cosmic rays are 99% hydrogen and helium nuclei, but can also be heavier nuclei up to uranium [19]. Upon reaching Earth, the cosmic rays interact with nuclei of atoms in the atmosphere to produce a shower of different particles, as shown in figure 2.9, but mainly pions ( $\pi$ ). These pions quickly decay to produce muons and neutrinos that can reach the surface, whereas the rest of the products of the shower do not.

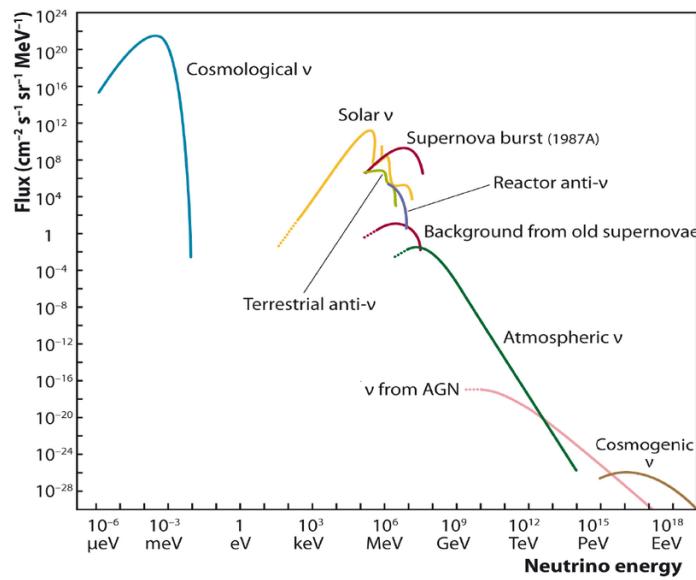


FIGURE 2.8: A figure showing the neutrino flux as a function of the energy.

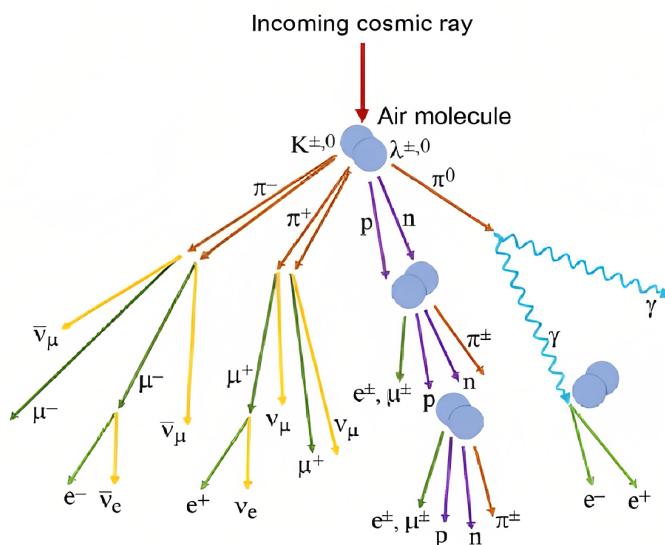


FIGURE 2.9: Illustration of the subatomic particles that are created in the air shower produced when incoming cosmic rays decay in the atmosphere.

## Chapter 3

# The IceCube Neutrino Observatory

In the 1960s it was proposed to build neutrino detectors with extremely large detection volumes to make up for the low interaction rate of neutrinos. The proposed media were freshwater (BAIKAL in Lake Baikal), seawater (DUMAND off the coast of Hawaii), and ice (AMANDA in Antarctica) [20]. As this was uncharted territory, it was hard to know what to expect, and it was not certain which medium was best suited, so all were tested. The two detectors in water proved to be problematic, due to currents, animals, and changing conditions, whereas the detector in the ice proved to be a better fit than anticipated. Especially the clarity of the ice was an unexpected, but welcome perk. Water has a higher absorption rate, but a lower scattering rate than ice, so ice is not definitively better, and more neutrino telescopes in water are being planned.

The **Antarctic Muon And Neutrino Detector Array (AMANDA)** was built into the antarctic ice in the early to mid-90s and served as proof of concept for IceCube before they merged under the name of the latter in 2005 in conjunction with the beginning of the construction of IceCube. AMANDA was turned off in 2009, and IceCube completed construction on December 17, 2010.

This section explains the architecture of the detector, the more intricate details of how the machinery works, and the properties of the ice that encapsulates the detector.

### 3.1 The Detector

The IceCube detector is one cubic kilometer of instrumented ice located below the Amundsen-Scott South Pole Station in Antarctica, making it the largest neutrino telescope in the world. 5160 Digital Optical Modules (DOMs) distributed onto 86 strings have been put in the ice at depths between 1450m to 2450m. Figure 3.1 shows the almost hexagonal layout of the detector, where the strings are placed in a triangular lattice with approximately 125m (some up to 157m and down to 97m) in between. The detector has three parts, the IceCube main array, DeepCore, and IceTop. The DOMs on a given string in the IceCube main array are separated by around 17m vertically. The detector is officially sensitive to particles of energies between 10 GeV to 1 EeV [21], but can also detect particles slightly outside this interval, but with decreased accuracy.

#### 3.1.1 DeepCore

The original IceCube array detects particles in the 150 GeV - 1 EeV energy range, but figure 2.8 showed that there are interesting neutrinos in the sub-GeV energy regime. This motivated the addition to IceCube called **DeepCore**, which lowered the lower boundary on detectable particle energy 10 GeV [21]. It was named DeepCore because extra DOMs were added to the deepest part of the center of the detector. Figure 3.2 shows that Deepcore was placed just below the dust layer, where the absorption and scattering is the lowest, to ensure the maximum amount of signal is recorded. It also shows that DeepCore consists of 8 newer DeepCore strings and 7 of the older IceCube strings.

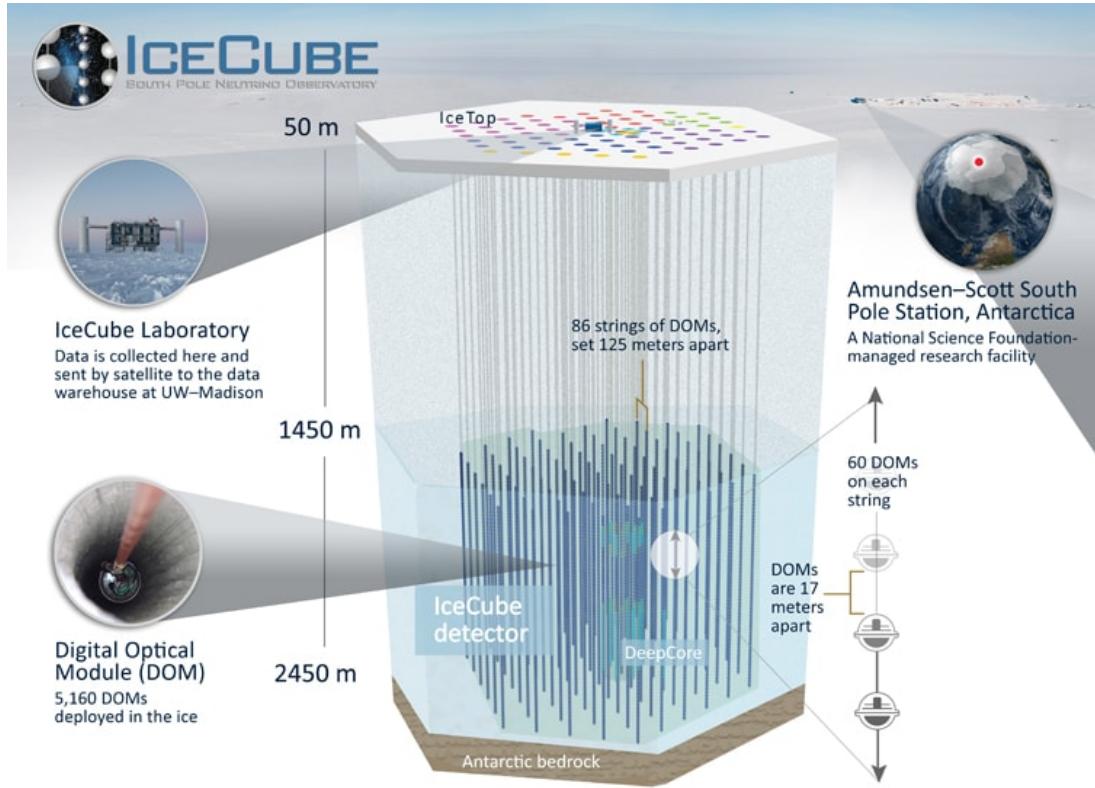


FIGURE 3.1: Schematic showing the architecture of the IceCube detector.

6 of the 8 new strings called HQE (High Quantum Efficiency) DeepCore Strings are equipped with newer hardware, that has a 35% higher quantum efficiency, while the last 2 strings (called DeepCore Infill Strings) have the original hardware. Quantum efficiency is defined as the ratio of emitted electrons to absorbed photons of the photocathode. A higher quantum efficiency therefore means that they are more sensitive to light, allowing them to detect lower energy photons, and therefore lower energy neutrinos. The new strings are split into two parts by the dust layer, where DeepCore is the part below, at depths of 2100m to 2450m, and the part above is called the veto cap. The veto cap allows for vetoing atmospheric particles [21]. The 50 DOMs per string in DeepCore, on top of being more sensitive to light, are only 7m apart vertically and 75m horizontally, and the 10 DOMs per string in the veto cap have the same horizontal distance, because they are on the same strings, but are 10m apart vertically.

With the higher quantum efficiency and the 5 times higher DOM density than the rest of the array, DeepCore is more sensitive to signal, decreasing the lower energy boundary for detection [22]. The horizontal axis in figure 2.2 shows that oscillations occur at energies below 100GeV. Most of the structure in the figure occurs at energies below 10 GeV, so while DeepCore has been important for oscillation analysis, another upgrade of the sensitivity of IceCube would allow for the lower energy interval to be explored.

### 3.1.2 IceTop

As mentioned in section 2.7, a lot of neutrinos and muons are created when hydrogen and helium nuclei collide with the atmosphere. When looking for extra terrestrial neutrino sources, it is necessary to filter out these atmospheric neutrinos, which is the purpose of IceTop. IceTop is a cosmic ray air shower array located on the surface of the ice. It consists of 162 tanks of ice containing two DOMs each [23]. As these atmospheric neutrinos are created in showers, they

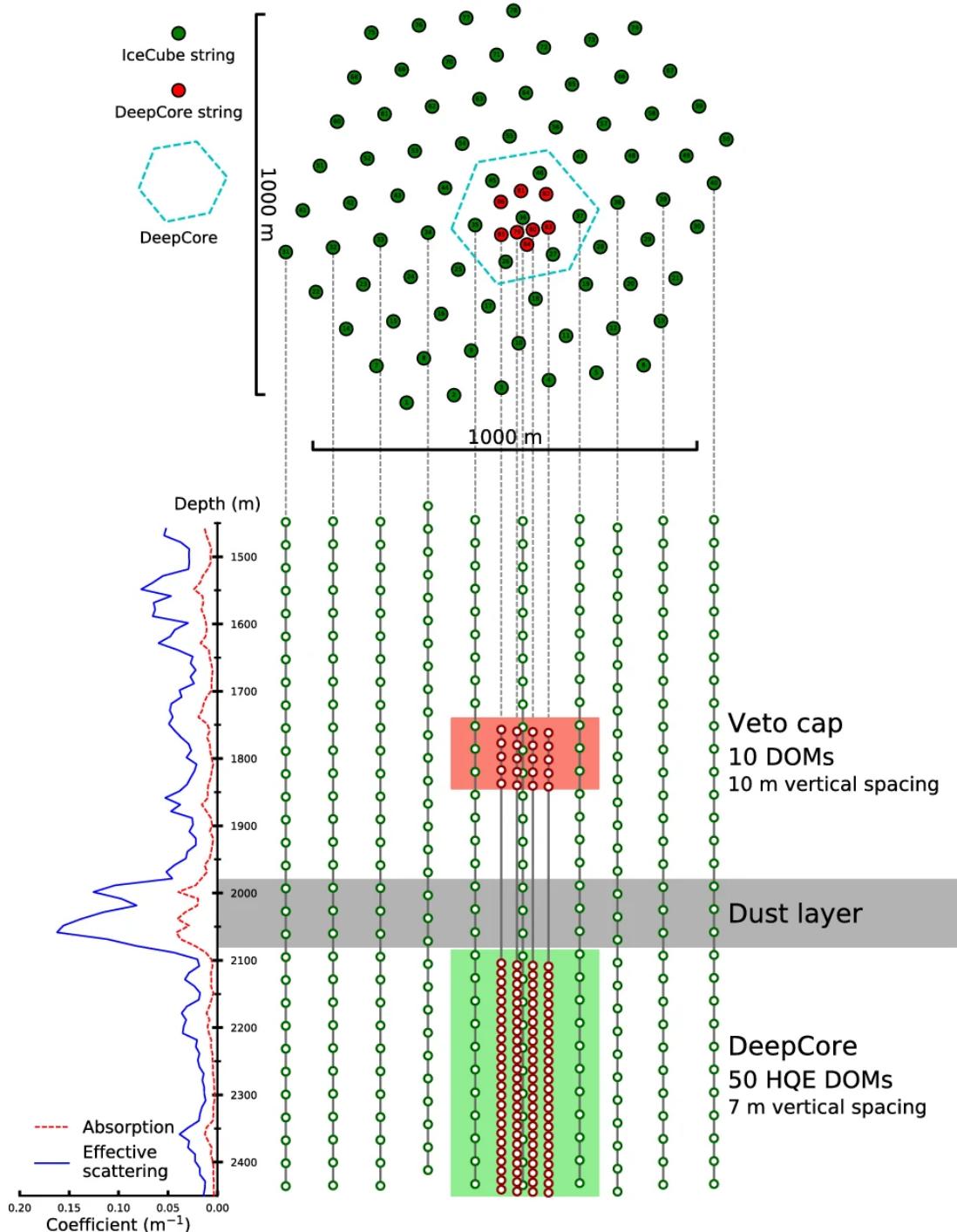


FIGURE 3.2: Figure showing the DOM distribution in the detector as well as how the absorption varies with depth. Illustration from [22].

usually come in a larger group of charged particles, unlike cosmic neutrinos. If IceTop detects a shower in conjunction with a neutrino of the same angle, it is probably atmospheric.

### 3.1.3 Challenges of the Detector

While the detector is as good as we could have hoped for in many ways, there are a few problems arising from the fact that the detectors have to endure extreme conditions being frozen deep in the ice. Out of the 5160 DOMS, some so-called "dead DOMs" have stopped working,

but are still in the ice, due to the large amount of resources it would take to replace them. Because the DOMs are inaccessible after deployment, it is not possible to alter the hardware, it is only possible to make corrections in the software.

The construction of neutrino telescopes on such a large scale was still not an easy ordeal, even after building AMANDA, meaning that some unforeseen obstacles were inevitable. One problem that in hindsight could have been avoided, is caused by the hexagonal structure of the IceCube Array along with the triangular grid of strings. This constellation of neatly ordered rows of strings means that there are also rows of absence of strings, (so-called corridors). Sometimes muons will enter the detector volume through these corridors, with adequate distance to the nearest DOM so as to not be detected, until it gets close enough to a DOM to be detected. In this case, the first measurement of this muon will occur somewhere in the center of the detector. A muon appearing in the middle of the detector is the hallmark of the decay of a neutrino which births a muon, and these corridor-muons will usually be classified as neutrinos.

## 3.2 Digital Optical Modules

The measurement in IceCube is done by the DOMs shown in figure 3.3, which are glass spheres containing photomultiplier tubes, that are connected to the surface by transfer cables. The 13mm thick glass is designed with a refractive index close to that of the ice, allowing photons to pass through seamlessly while resisting the pressures at great depths. The DOMs are filled with dry nitrogen, meaning the gas has been stripped of water molecules, at a pressure of around half the atmospheric pressure. The DOMs are fitted with pressure sensors [24] to monitor pressure, and flasher boards to calibrate PMT responses. These flasher boards will emit light, and based on the nearby recording times, the distance between DOMs can be calculated [25], to calibrate the DOMs. The glass surrounding the detector is designed to have a low impurity, so it contains as little as possible radioactive material, such that particles emitted from the glass is a negligible source of noise. The glass is also lined with a gel to minimize photon scattering and fitted with a metal mesh to shield it from Earth's magnetic field.

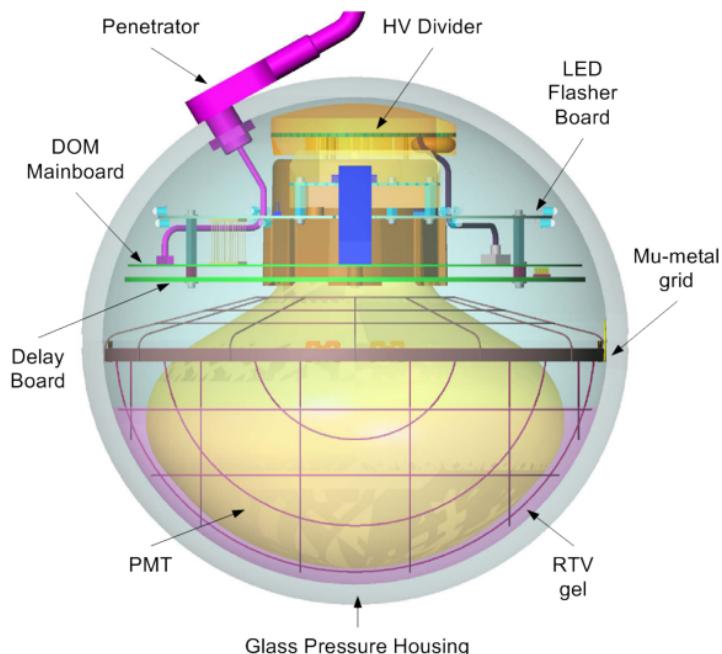


FIGURE 3.3: The schematics of a digital optical module. Illustration from [24].

### 3.2.1 Photomultiplier Tubes

There are two types of **photomultiplier tubes (PMTs)** used in IceCube. The original DOMs deployed in the first round in 2010 were fitted with Hamamatsu Photonics R7801-02, and when DeepCore was created in 2016, the newer HQE model Hamamatsu R7081MOD [26] [27] was used. Within the PMT, the photons initially hit a photocathode that uses the photoelectric effect to emit photoelectrons towards the 10 serialized dynodes in the PMT [28], as shown in figure 3.4. The dynodes are charged metal plates brimming with electrons, such that a very small charge is able to knock electrons off one plate, sending them towards the next plate, which in turn knocks off a larger amount of electrons, creating an amplifying chain reaction. These electrons produce a charge after the final amplification, which is measured as the strength of the signal. The higher the energy of the photon, the more electrons will be knocked off initially, and therefore the energy of the photon will be directly proportional to the output charge. There is some uncertainty in the photon-energy-to-output-charge ratio due to the variance in how many electrons are knocked off during each step, as it depends on multiple factors, such as the angle of the incident photon.

The PMT in the DOMs is located on the underside, which favors the observation of up-going particles more than the down-going. This design choice also acts as a noise filter, as neutrinos are the only particles we observe going through Earth. Newer proposed DOMs called D-eggs for a future upgrade of IceCube have both an upwards and a downwards-facing PMT [29].

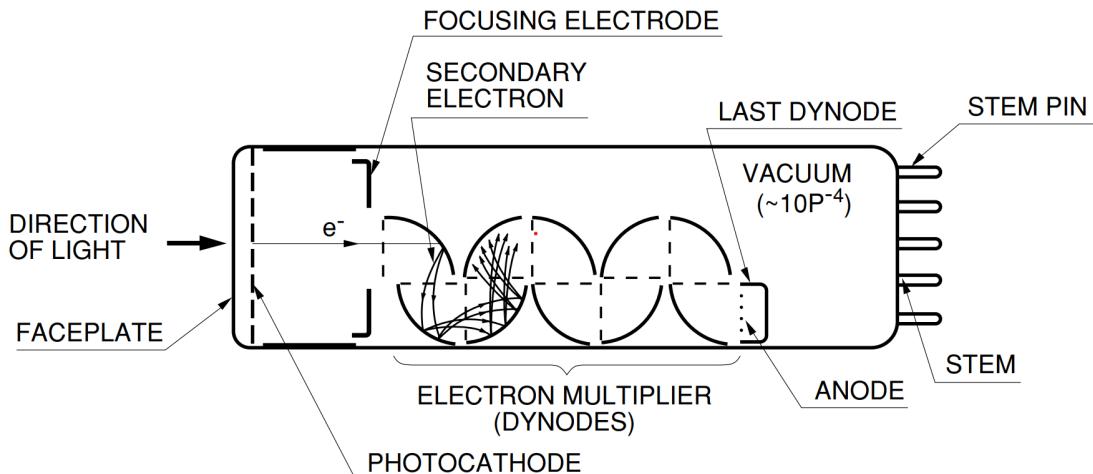


FIGURE 3.4: Schematic showing the composition of a PMT. Illustration from [27].

### 3.2.2 Digitizers

The charge produced from the PMT is transformed into digital waveforms through a digitizer. Each DOM uses two types of digitizers. One has a higher sample rate in a short time window, and the other has a lower sample rate in a longer time window. The **Analog Transient Waveform Digitizer (ATWD)** takes 128 samples with 3.3 ns in between. For longer signals, a second digitizer with a lower rate is used. The **Fast Analog to Digital Converter (FADC)** takes 256 samples with 25ns in between resulting in a period of 6400ns, which is chosen as it is longer than any physical signal is expected to last. Each DOM has two ATWD chips, that take turns measuring to prevent downtime due to the export of data, such that the second chip can measure while the first reads out, and vice versa [26].

Figure 3.5 shows how the digitizers respond to a signal. The two ATWDs record with a fine resolution in a window too small to capture the entire signal, whereas the FADC captures the

entire signal, but with a lower resolution. The bottom plot shows how the final pulsemmap looks after the waveforms have been processed to produce pulses with timestamps. Although it is not shown here, the charge of the reconstructed pulse is also saved.

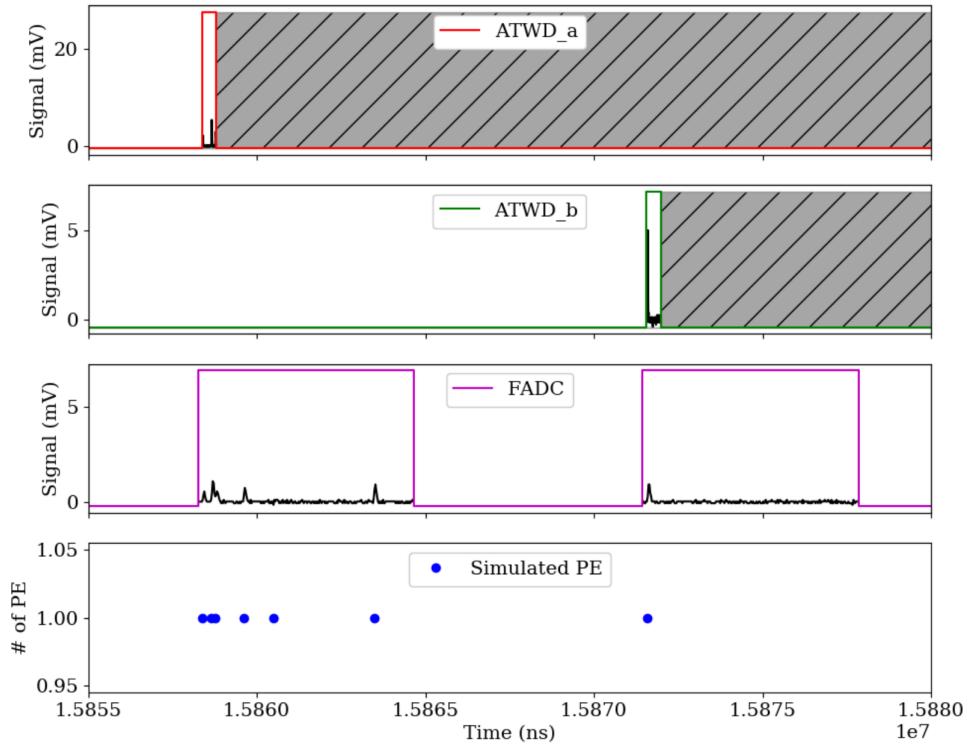


FIGURE 3.5: Images showing how the three digitizers respond to the same signal, and in the bottom image, the resulting pulses. Illustration from [30]

### 3.3 From Photon to Hit

When the PMT in the DOM receives a Cherenkov photon from a particle, it starts the domino effect described in section (3.2.1), producing photoelectrons. If an incident photon has an energy corresponding to a charge greater than the onboard discriminator trigger threshold (approximately 1.2mV or 0.23PE<sup>1</sup> [26]), it produces a digital waveform representing the measurement. The waveform seen in figure 3.5 shows the charges recorded on a time axis. From the waveform, a list of pulses with discrete times and charges are inferred through a linear algebra based fitting algorithm called **WaveDeform** [31] that fits a linear combination of single photoelectron basis functions (or SPE pulse templates) to the waveform, where the SPE pulse templates emulate the response of a single photoelectron hitting the PMT. Such a template is shown in figure 3.6.

Thus each pulse is reconstructed separately. Two pulses arriving in quick succession might be digitized as one, although measures are in place to reduce this effect [26]. The SPE distribution is found by averaging measurements of 118 PMTs in an above-ground freezer meant to emulate the conditions in the ice.

Figure 3.7 shows the process of converting recorded photons to a pulsemmap. As soon as the discrimination threshold is surpassed, the timestamp is saved and a signal is sent to neighboring DOMs, to see if they also recorded a signal. The signal then goes through the digitizers to

<sup>1</sup>1 PE or 1 photoelectron is a unit equal to the charge of 1 electron.

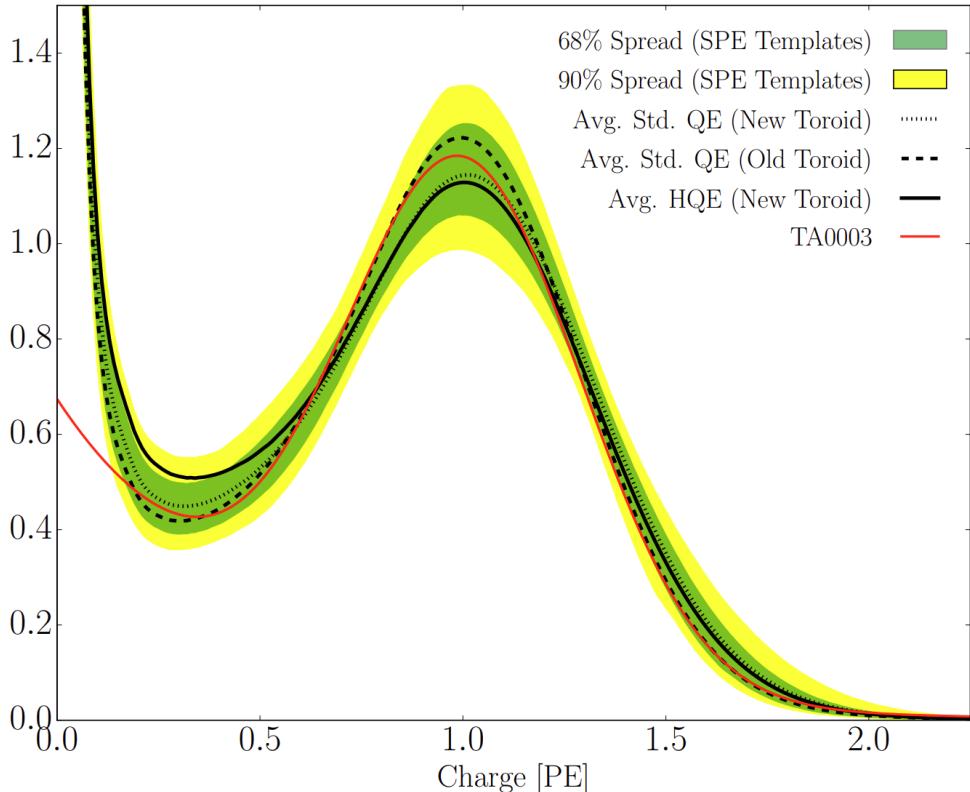


FIGURE 3.6: Single photon electron (SPE) template. Illustration from [26].

produce the digital waveform, and checks if any neighboring DOMs recorded coincidence hits, which decides the resolution the waveform data is saved with before being sent to the surface.

### 3.4 Ice Properties

The Antarctic ice is the product of millions of years of compounding layers of snow. As snow is covered by newer layers of snow, it is compressed by the weight, turning it into ice [32]. Since the snow properties depend on multiple parameters that shift in time, such as sun exposure and atmospheric composition, each layer of ice will also have different properties. Furthermore, the ice experiences greater pressure, the more ice accumulates on top, leading to the deeper ice being more compressed, also changing its properties. Figure 3.8 shows that the effective scattering length of photons and the absorption length vary greatly with depth. It is also worth noting that the two graphs are not linear, meaning that it is not as simple as the ice simply becoming clearer with depth, although they do trend downward. Part of the success of IceCube can be attributed to the very high clarity of the deep glacial ice, which causes the high absorption lengths, meaning photons can travel far before being absorbed.

Figure 3.9 shows how the temperature varies with depth in the ice [33]. The measurement was done using thermistors<sup>2</sup> in AMANDA boreholes. As IceCube spans the depths 1450m - 2450m, the temperature spans from around  $-42^{\circ}$  to  $-20^{\circ}$ , which affects the rate of noise created in the DOMs.

The ice in Antarctica is constantly shifting and the bottom layers are moving faster than the top layers, causing the angle of the strings to change by less than 0.01 degrees per year [23], so it can be safely neglected.

<sup>2</sup>A portmanteau of "thermal" and "resistor". A type of resistor that is especially sensitive to temperature.

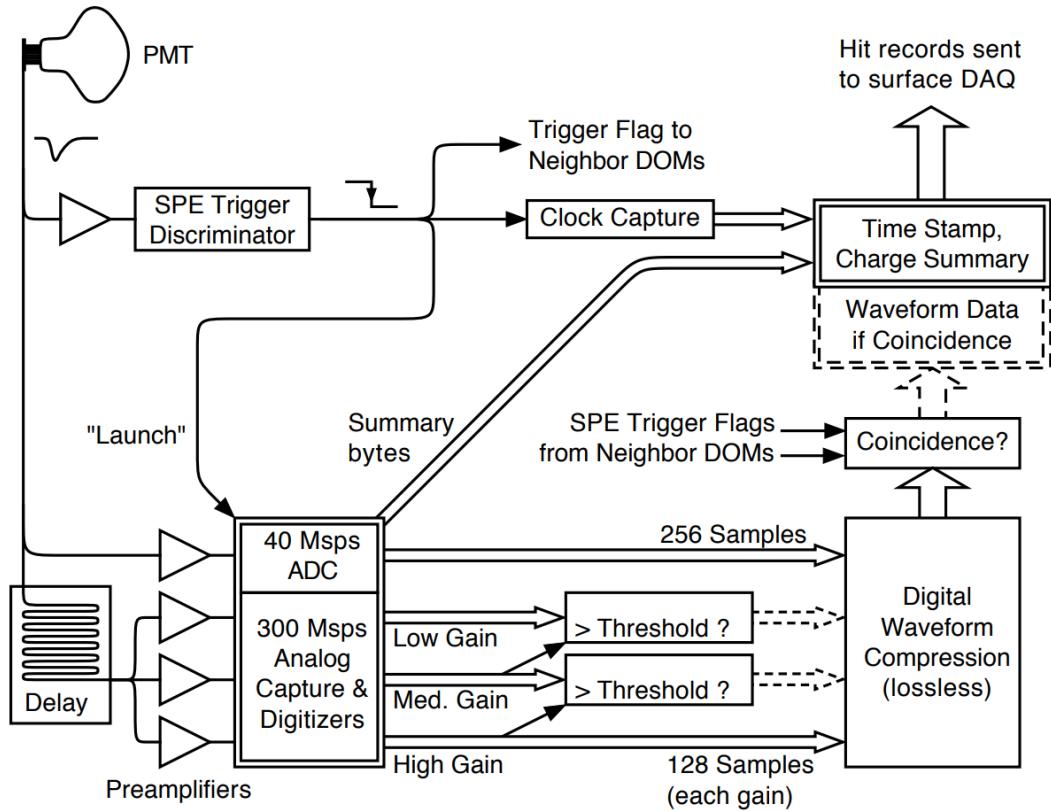


FIGURE 3.7: Schematic showing how a photon becomes a hit. Illustration from [25]

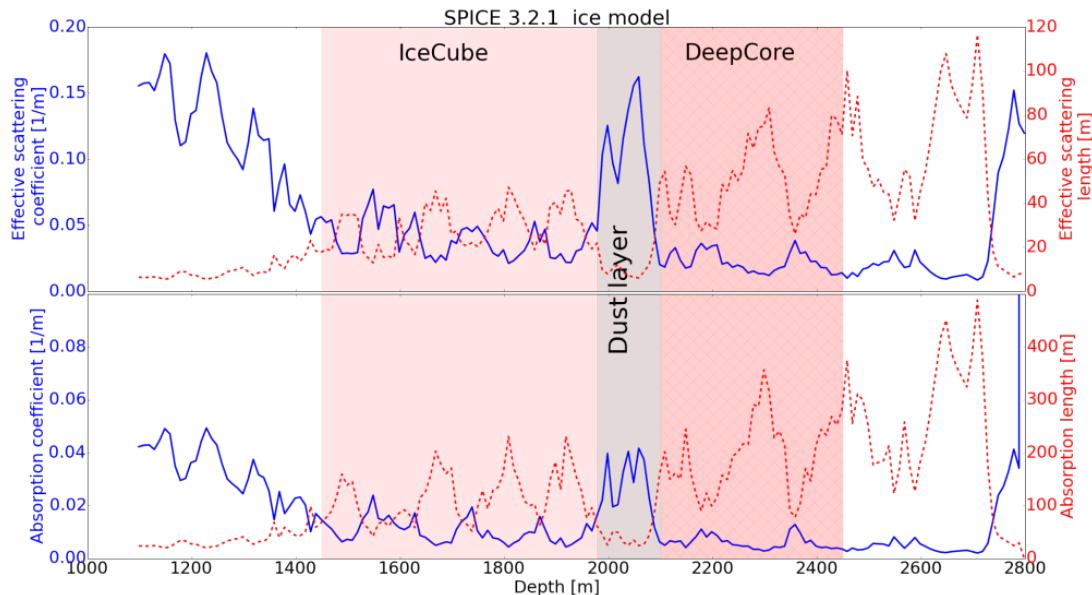


FIGURE 3.8: Plots showing the effective scattering length and the absorption coefficient as a function of depth in the ice. Illustration from [23].

In order to place the DOMs in the ice, it was necessary to drill holes in the ice that they could lower the strings into. This was done using a hot water drill, to melt a 60cm diameter hole down to depths of 2450m [34]. After lowering down the strings, the thawed ice refroze, creating cylinders of the so-called "hole ice", with different optical properties than the surrounding ice.

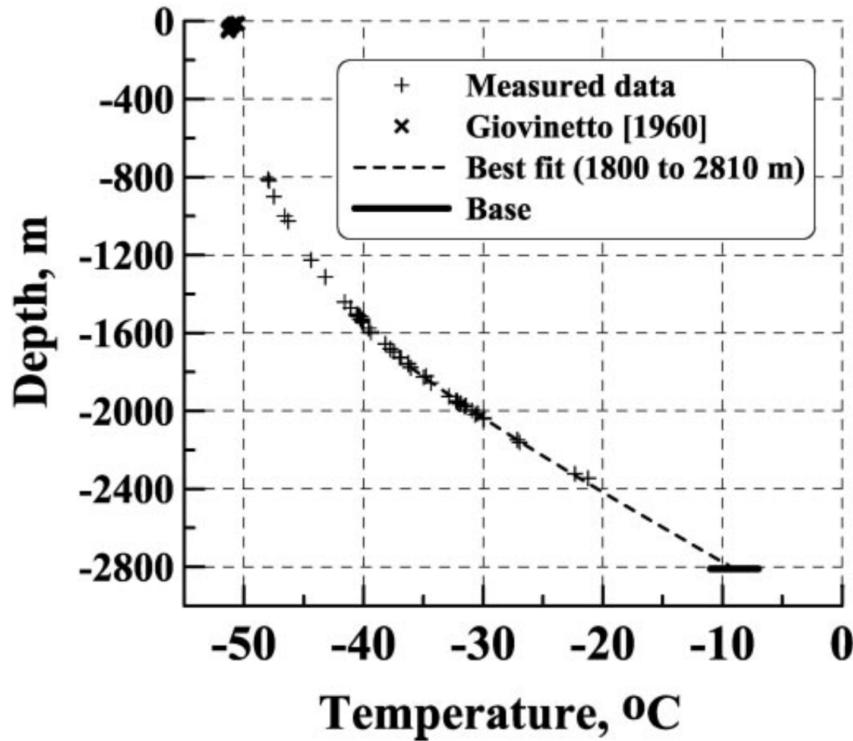


FIGURE 3.9: Plot showing the temperature taken inside the AMANDA boreholes, as a function of the depth. Illustration from [33].

The scattering of photons is dominated by the bubbles in shallower ice, but the effect vanishes at depths below 1350m as the higher pressure turns the bubbles to non-scattering air hydrates, which have refractive indexes very similar to the ice [35].

The exact properties of the hole ice are not very well understood. Figure 3.10 shows calculations for the angular acceptance rate of photons, as a function of the cosine to the zenith angle. The different simulations are very similar until the photons start pointing straight up toward the surface, along the column of hole ice, where the uncertainty is very high.

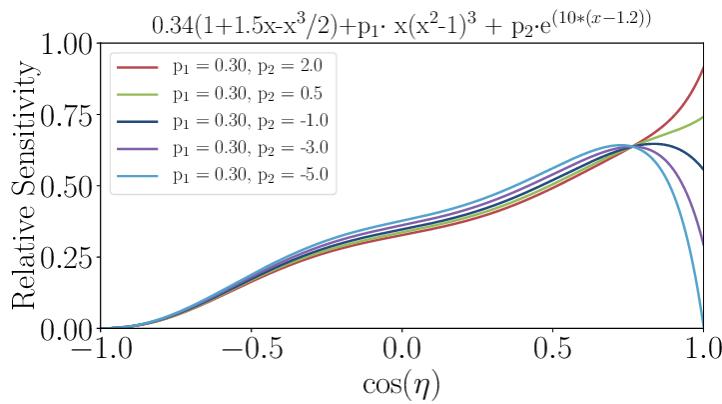


FIGURE 3.10: Plots of different simulations of the angular acceptance rate of the DOMs as a function of the zenith angle. Illustration from [36].

Recently it was discovered through experiments that the attenuation of the light in the ice is anisotropic, and aligned with the local flow of the ice [37]. Using the flashers in the DOMs, light was measured by the DOMs with similar depth in the surrounding strings. As shown in figure 3.11, it was discovered that charge was measured at almost twice the rate from photons

along the flow axis, than the orthogonal tilt axis. This data was collected by using the flashers in the DOM to emit light that was then recorded by the DOMs on the surrounding strings, approximately 125m away.

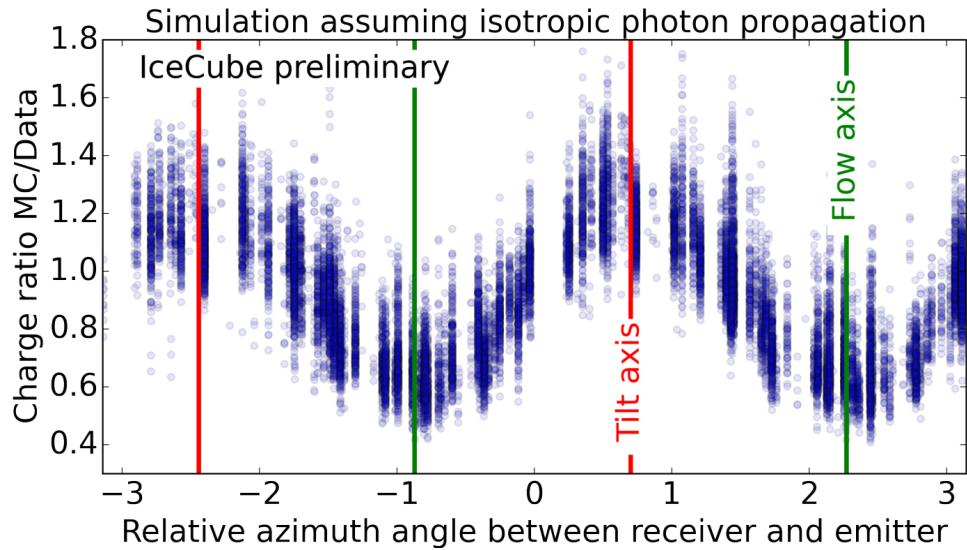


FIGURE 3.11: The ratio between the charge measured in data and the charge simulated in MC, as a function of the azimuth angle between emitter and receiver. Illustration from [37].

The anisotropy in the MC/data ratio arises from the fact that the effect is not yet included in the simulation, so the charge distribution is isotropic in azimuth for MC. Although the cause of the anisotropy is still not certain, the effect is now being included in future simulations. [23] [38].

This phenomenon is believed to be caused by birefringence in the ice, meaning that the refractive index depends on the direction of propagation of the light. The birefringence also causes a bending of light towards the preferred axis, which is illustrated in figure 3.12.

### 3.4.1 Dust

There are impurities such as dust in all layers of the ice, but most significantly in the layer between depths of approximately 1980m to 2080m, which is called the dust layer. Each layer of the ice is affected by the climate of the era, and a lot of dust accumulated in Antarctica during a particularly dry and dusty period. Photons travel easily in pure ice but will scatter or be absorbed upon hitting an impurity such as dust [35]. It is visible in figure 3.2 that the absorption and scattering of the ice increase significantly in that layer. While it makes the signal less clear in the detector, it is also an obvious calibration source. If a model of absorption, scattering, or DOM efficiency does not indicate the presence of the dust layer, something is wrong.

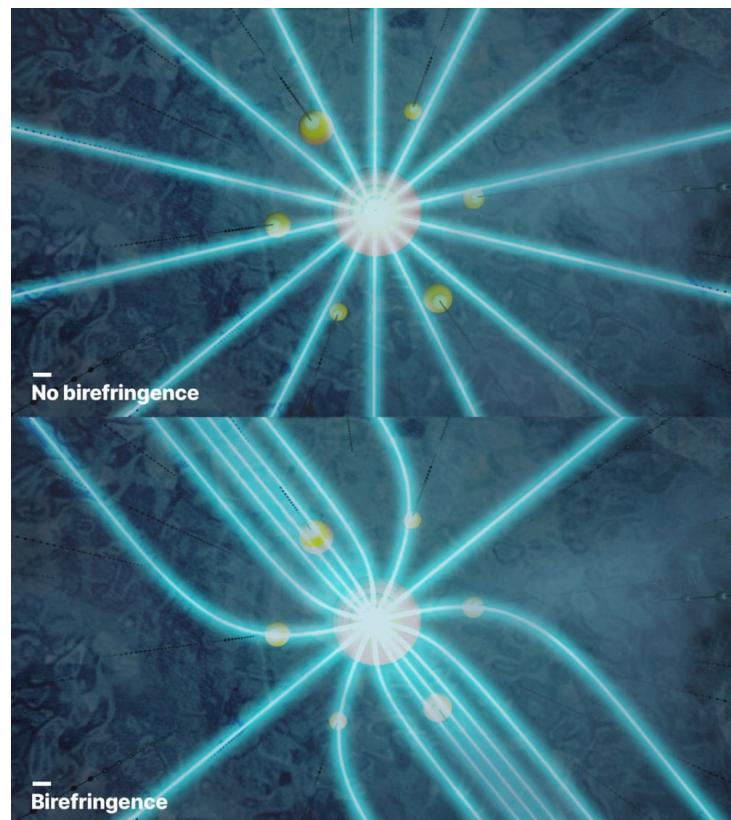


FIGURE 3.12: The top image shows how photons propagate from a light source without accounting for birefringence, and the bottom image shows the propagation while accounting for birefringence. Illustration from [38].

## Chapter 4

# Data

Of the total set of data recorded by the IceCube detector between 2011 and 2021, only 1% (corresponding to 92.4 million events) has been unblinded. The 1% was chosen as all subruns between 2011 and 2021 with an ID ending in 00, to ensure no systematic error related to the concurrence of the subruns. This sample is referred to as the **burnsample** and is the data used in this analysis.

The Monte Carlo events used were a sample of MuonGun events from the IceCube OscNext group and contained the features in table 4.1, as predicted by former Master's student Leon Bozianu [39] also using a DynEdge model trained on stopped muons.

### 4.1 Features

One set of data comes in two parts; the event file and the pulsemapper. The event file has one row per event containing some features describing the event. These files can have more than 50 columns, where most are of no importance to the analyses in this thesis, like predictions on elasticity, or a boolean indicating if the veto cap was activated. In this thesis, only the handful of features in table 4.1 from the event file were used.

Feature name	Description
zenith_pred	Prediction of the zenith angle of the particle
azimuth_pred	Prediction of the azimuth angle of the particle
position_x_pred	Prediction of the x-coordinate of the particle decay position
position_y_pred	Prediction of the y-coordinate of the particle decay position
position_z_pred	Prediction of the z-coordinate of the particle decay position

TABLE 4.1: The features of the event file.

All features ending in \_pred are predicted values, which are predicted using GNNs. The MC set contains the true values as well as the predictions.

The other half is the pulsemapper. The pulsemapper contains the features shown in table 4.2 for each pulse in the set. All of these except QE and pmt\_area were used in this thesis.

Feature name	Description
dom_x	x-coordinate of the triggered DOM
dom_y	y-coordinate of the triggered DOM
dom_z	z-coordinate of the triggered DOM
dom_charge	Charge of the pulse
dom_time	Time from readout start to trigger
QE	Quantum efficiency of the DOM
width	Resolution of the pulse (1.0 if HLC, 8.0 if SLC)
pmt_area	The area of the PMT

TABLE 4.2: The features of the pulsemmap.

## 4.2 Monte Carlo Simulation

Much of the data analysis in IceCube is based on knowledge gained through the simulation of physical events. These simulations take all relevant information into account, such as ice properties and DOM conditions, which is why it is important to know these well. There are many different physical phenomena that constitute the data recorded at IceCube, and multiple algorithms are needed to simulate them.

- **CORSIKA** [40]

CORSIKA (Cosmic Ray Simulations for Kaskade) simulates the entire cosmic ray shower of particles illustrated in figure 2.9, even though only the muons are used. This is to achieve more realistic energy distributions of the muons.

- **MuonGun** [41] [42]

The MuonGun algorithm is derived from CORSIKA algorithm when it became necessary to simulate singular muons instead of full showers. MuonGun simulates single muons as if they were shot into the detector with a gun, and only contain events that pass through DeepCore. Because it is only a fraction of a CORSIKA simulation, it is also much faster and more customizable. The muons are simulated one by one, neglecting the phenomenon of bunched muons, where muons arrive in groups, which accounts for around 10% of the muons detected in real data.

- **GENIE** [43] [44]

GENIE simulates neutrinos and the neutrino-nuclei interactions mentioned in section 2.5 in the ice, and the resulting interaction products in the energy range MeV to PeV.

- **VUVUZELA** [45]

Vuvuzela is the aptly named simulation of random coincident noise, i.e. noise that by chance satisfies the local coincidence filter.

The different algorithms above cover different phenomena that occur simultaneously in the detector, so to create a realistic MC set that is similar to data, a combination of them all is needed.

### 4.2.1 Reconstruction

The method used by IceCube to predict the variables of data events, such as energy and angles, is the reconstruction algorithm called **RETRO** [46]. RETRO simulates photons as if they were emitted in all directions from the DOMs, and absorbed by the ice. Then it constructs tables for each z-layer of the ice, mapping a survivability factor to all areas of the detector based on how many photons were absorbed in that area (with some normalization). Then it reverses the photons, hence the name, such that they originate in the ice and are absorbed by the DOM. The probability of a photon being detected is then based on the survivability of the photon, which is based on where it was created, and in which direction the detecting DOM is.

The expected amount of photons and their timing are calculated and compared to an event hypothesis, to produce a likelihood that the hypothesis is true. An iterative process modifies the parameters of the hypothesis to maximize the likelihood.

RETRO has its shortcomings, as first of all, multiple physical effects are ignored for simplicity. RETRO ignores the azimuthally dependent ice anisotropy, and since it simulates each DOM response as if the other DOMs were not there, reflection and absorption from the surrounding DOMs and strings are ignored. Secondly, RETRO is relatively slow and takes 5-40s to reconstruct a single event.

## 4.3 Noise

At approximately 350Hz **dark noise** [26] makes up a majority of the noise hits, and is significantly more than the rate of cosmic ray muons, which is recorded at a rate around 5Hz to 25Hz depending on the depth. Dark noise, as the name suggests, are hits from photoelectrons that are not caused by photons originating outside the DOM. Dark noise has multiple sources, such as electronic noise, luminescence of the glass in the PMT, radioactive decay in the glass emitting charged particles that produce Cherenkov radiation, field emissions in the PMT, and thermionic emission. The radioactive decays constitute the majority of the dark noise because the glass spheres of the DOMs contain the radioactive element potassium, which causes around 100 beta decays per DOM per second. The glass of the DOM has therefore been treated to minimize the potassium content [25].

**Thermal noise** arises from the high sensitivity of the dynodes of the PMTs, meaning that small fluctuations in the thermal energy of the PMT are enough to spontaneously emit an electron, which can then trigger the PMT. The thermal noise is very sensitive to thermal fluctuations, and figure 3.9 showed that the temperature of the DOMs varies by around 20°C.

Uncorrelated noise covers electronic noise and radioactive decay inside the PMT and the glass DOM sphere. The correlated noise might be from scintillation in the glass of the DOM caused by the radioactive decay. Figure 4.1 shows a histogram of the time between all hits in HitSpool, which is a log of all hits in the detector. The figure also shows that the uncorrelated noise is Poissonian, the correlated noise follows a log-normal distribution, and the afterpulses are Gaussian.

## 4.4 Pre-Pulses, Afterpulses, Late Pulses and Pulsesplitting

There are numerous effects caused by the hardware and the digitization process that result in either hits being early or late or being incorrectly merged or split. The fitting of the SPE template to the data can fit one SPE distribution to a signal that was actually caused by two separate signals, or the other way around, one large signal can incorrectly be identified as two. While these effects are known by IceCube, and their algorithms take them into account, it is

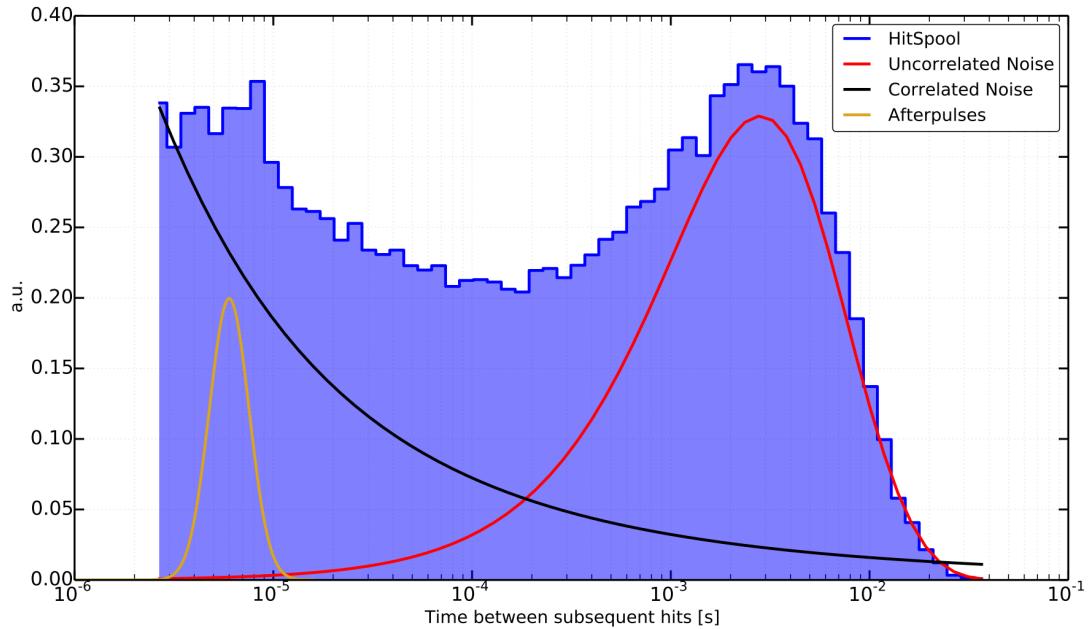


FIGURE 4.1: Plot showing a histogram of time between hits in HitSpool and the distributions of the causes [47].

still a source of error, and elimination of the effects would improve the agreement between MC and data.

#### 4.4.1 Late Pulses

A photoelectron hitting the first dynode of the PMT can scatter off, and then be redirected back into the dynode, producing an additional pulse from the same photoelectron [26]. The backscattering can either be elastic or inelastic, resulting in the late pulse having the same or less energy. The time between a pulse and its corresponding late pulse has been measured to be upwards of 70ns.

#### 4.4.2 Afterpulses

When the free electrons produced in the PMT gain enough charge, they can ionize the residual gas in the PMT. The positively charged atoms in the gas will be accelerated towards the photocathode, creating a second pulse. These pulses have been measured to arrive 300ns to 11000ns after the initial pulse, with a peak around approximately 600ns [26].

#### 4.4.3 Pre-Pulses

Figure 3.4 showed the process of photons becoming electrical signal, starting with the photon hitting the photocathode. Occasionally photons will sneak past the photocathode in the PMT and interact with nothing before hitting the first dynode of the PMT. Because it skipped the initial amplification stage of the photocathode, the resulting charge will be 10 to 20 times lower than an SPE pulse, and the timing of the hit will be sooner. As there usually are a lot of photons, most will hit the photocathode and produce an additional signal after the photon that snuck by the photocathode. These prepulses have been measured to arrive approximately 30ns before the signal from the photocathode, and at a rate less than 1% of the SPE rate [26].

## 4.5 IceCube Data Selection

The detector triggers at a rate of around 2.5 kHz, and only around 275 neutrinos are detected per day, meaning that the vast majority of triggers are caused by noise and muons. To save on the relatively expensive computation cost, certain selections are employed. IceCube records around 1 terabyte of raw data every day, but after these selections are made, only around 100 gigabytes are sent to the mainland via satellite [10].

### 4.5.1 Local Coincidence Hits

Because there is such a low signal-to-noise ratio in IceCube, data is only saved if the measurement fulfills some predetermined requirements. Firstly, the measured charge has to be above a threshold of around 0.23PE [26]. Secondly, the hit has to have a local coincidence with another hit. That means that another hit has to be recorded in a neighboring<sup>1</sup> DOM. If this coincidence happens within a time window of  $\pm 1 \mu\text{s}$ , all coincident hits are labeled as **Hard Local Coincidence (HLC)** hits [21]. All coincident hits within  $\pm 10 \mu\text{s}$  are labeled **Soft Local Coincidence (SLC)** hits.

HLC hits have their full waveform information from both the FADC and ATWD recorded, and SLC hits only have limited waveform information from the three bins around the peak in the FADC waveform saved. Given the limited output, and that it is only information from the FADC, which has a lower sample rate than the ATWD, the signal of SLC hits is poorer than HLC hits because there is a lower probability that they contain signal of physical origin.

There are different triggers of different strictness used for different purposes. High-energy events, are expected to trigger more DOMs than low-energy events and can tolerate a stricter cleaning.

For high energy, the **single multiplicity trigger (SMT)** SMT8 demands 8 HLC hits within  $5 \mu\text{s}$ . If that is satisfied, the window is widened to  $10 \mu\text{s}$  before and after the initial trigger, where SLC hits are also included. As low-energy neutrinos activate fewer DOMs, SMT8 is too strict, and for the low-energy neutrinos detected in the denser DeepCore, SMT3 is used, where only 3 HLC hits in DeepCore within  $2.5 \mu\text{s}$  are required [48].

The **Seeded Radius Time (SRT) cleaning** [24] algorithm starts with the HLC hits, and includes SLC hits if they are within some radius and time, that corresponds with what could be expected of causally connected hits. If accepted, the SLC hits become the new seeds, and the algorithm checks for SLC hits within some time and distance of the newly accepted SLC hits. It is an iterative process, that runs until there are no more suitable hits.

### 4.5.2 IceCube Neutrino Selection Levels

After it has been determined which hits are included in the events, IceCube uses a 7-step event selection process to go from the raw data to a sample that is mostly neutrinos. Each level aims to remove more and more muons and noise events and utilizes the filters mentioned above. The first two levels of cleaning are applied before the data is sent from Antarctica to The University of Wisconsin, so level 2 data is the rawest data available. As this thesis concerns itself with muons, level 2 data is used. The bonus in using the least cleaned data available is that it removes the computation time of the last five levels, saving time.

Figure 4.2 shows the rate of data for each particle type through each level of selection. The data used in this thesis is from OscNext, which uses the DeepCore filter in level 2, which selects events with more than three SMT3 hits. The remaining levels employ both simple cuts and the use of BDTs<sup>2</sup> (Boosted Decision Trees). Level 6 events are reconstructed, the event type

<sup>1</sup>The neighborhood of a DOM is the two DOMs above and the two below on the same string.

<sup>2</sup>More specifically LightGBM.

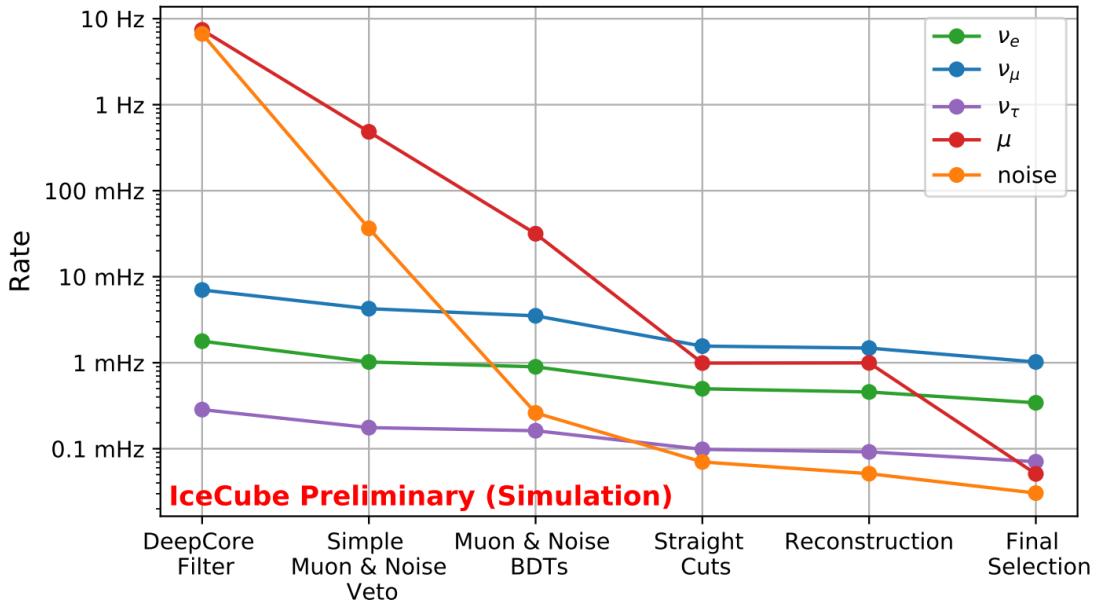


FIGURE 4.2: This image shows the ratio between the charge measured in data and the charge simulated in MC, as a function of the azimuth angle between emitter and receiver. Illustration from [30].

is predicted, and those unlikely to be neutrinos are removed. This selection manages to lower the rate of muons and noise by 6 orders of magnitude, while the rate of neutrinos decreases by less than 1.

The selection process could be replaced by a GNN predicting on level 2 data, which would be magnitudes faster. The "noise" mentioned here are events that do not contain signal from a particle, and are not to be confused with the noise hits also present in muon and neutrino events. These noise hits can also be identified and removed to improve the quality of the events. This can be done either by classification of pulses being noise or signal using a GNN or by calculating the timing of the hits, to determine if it is within a reasonable time from the moment the photons would have been emitted.

## Chapter 5

# Machine Learning

In recent years, machine learning (ML) has proved itself an extremely useful tool in particle physics due to its ability to process large amounts of data quickly and accurately. Not only is ML often orders of magnitudes faster than competing methods, but it can also yield better results [22].

The current reconstruction algorithm IceCube uses, **RETRO** [46], is based on likelihood tables, and is relatively slow, given the vast amount of data continuously being produced. One reconstruction with RETRO takes around 5-40 seconds, whereas the novel GNN is capable of making thousands of reconstructions per second, while also being more accurate. The speed increase associated with using GNNs instead of RETRO would also allow for IceCube to run the reconstructions directly on the south pole, instead of only sending level 2 data to the mainland. The increased efficiency would also decrease the consumption of energy, which would be useful in Antarctica.

The python package **GraphNeT**<sup>1</sup> [2] [3] [22] offers a GNN specifically designed for IceCube, called **DynEdge**, which was used in this thesis for the data selection and reconstruction. Before we get to DynEdge, we will go through an introduction to machine learning.

## 5.1 Basics of Machine Learning

Machine learning is a term used for multiple types of algorithms, that fall into three categories supervised learning, unsupervised learning, and reinforcement learning. This thesis only uses supervised learning, and will not touch on the two other categories.

### 5.1.1 Supervised Learning

Supervised learning has two tasks: classification and regression. Classification is the prediction of class, which is a categorical value such as "neutrino" or "muon". If there are only two classes, the task is known as a binary classification, and if there are more, it is known as a multiclass classification. Regression predicts a continuous variable e.g. a decimal value for the prediction of the energy of a particle. To be clear, a classifier will predict the probability of being a certain class, meaning that the output is a continuous variable between 0 and 1.

Supervised learning gets its name from the fact that the model needs to be supplied with examples where the target is known, and has its performance evaluated throughout the training to know what parameters to change and when to stop. The idea is that the model is given a large amount of example data where the target is known and that it can extract and remember patterns and characteristics. When the model is presented with new data, where the target is not known, it can compare it to its large memory of examples to estimate what the prediction should be.

---

<sup>1</sup>Graph neural networks for Neutrino Telescopes.

### 5.1.2 Loss Functions

The model evaluates the quality of its performance by comparing its predictions on the validation set to the true values of the validation set. Since the validation set is not used in training, it serves as an unbiased sample to evaluate performance on. At the end of each epoch of training the model will make predictions for the validation data. We then provide the model with a metric called the **loss function** on which to compare the predictions to the true values, to produce some value characterizing the performance, called the **loss**. The name alludes to the idea that the higher the value the loss has, the worse the prediction is. The loss function is often denoted  $\mathcal{L}(y, \hat{y})$ , where  $y$  is the true value, and  $\hat{y}$  is the prediction of the value of  $y$ . In general, the loss function is a function that takes a value and the prediction of that value, compares them, and then returns a value that represents how different the two are. The larger difference the higher the output of the loss function is. The difference between true value and prediction can then be scaled or transformed, which is useful in some cases. The simplest example of a loss function for regression is the mean absolute error (MAE)

$$\mathcal{L}(y, \hat{y}) = \frac{1}{N} \sum_i^N |y_i - \hat{y}_i| \quad (5.1)$$

which produces a higher value, the further apart the prediction is from its true value. In some cases, a similar function, mean squared error (MSE)

$$\mathcal{L}(y, \hat{y}) = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2 \quad (5.2)$$

is more suited. MAE and MSE are very similar, but figure 5.1 shows that for values below 1, squaring decreases the loss, and for values above 1, squaring increases the loss. MSE therefore is more forgiving if the prediction is in the right ballpark, but much stricter on outliers.

It is important to choose the right loss function for the specific task at hand, as results can vary greatly, and some loss functions that perform well on one problem, may not on another. Loss functions are usually split into the same two categories: classification and regression because the tasks produce different types of output.

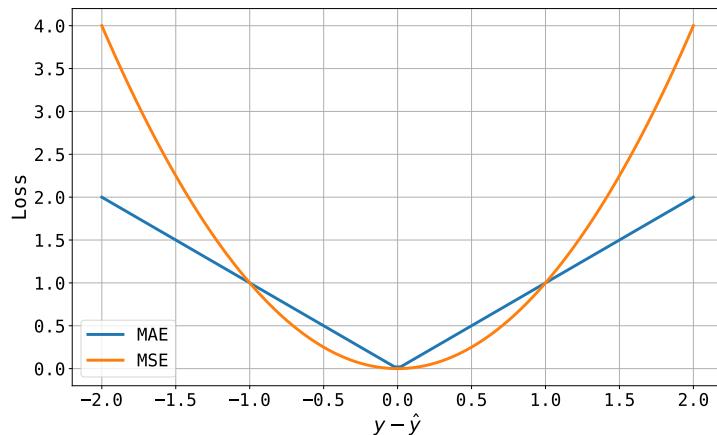


FIGURE 5.1: MAE and MSE.

For classification tasks cross-entropy loss functions are very popular, and also the ones used in this thesis. In the case of two classes, we can assign them the labels 0 and 1, and the **binary cross entropy loss function** is defined as

$$\mathcal{L}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(\hat{y}) + (1 - y_i) \cdot \log(1 - \hat{y}) \quad (5.3)$$

where  $\hat{y}$  is the predicted probability of  $y_i$  of having the label 1, meaning that  $\hat{y}$  must assume values between 0 and 1. Figure 5.2 shows the loss as a function of the prediction. Obviously, the function gives a different loss depending on the true label, but the two are mirror images and behave equally for the two labels. The loss is zero for a correct prediction, and then rises slowly but increases rapidly if the model is fairly confident of the incorrect label.

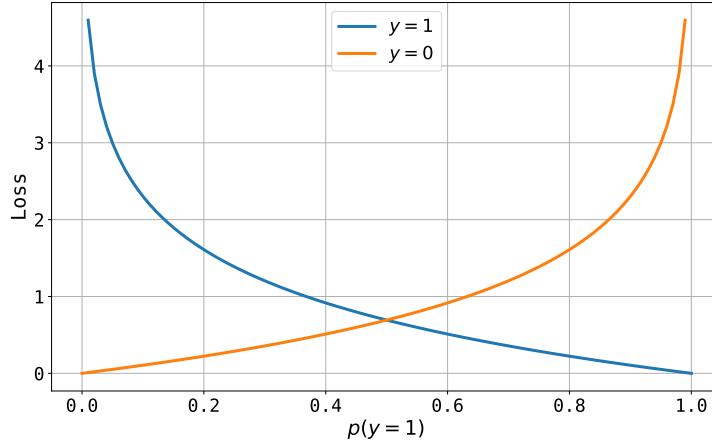


FIGURE 5.2: Binary cross entropy loss.

Regarding the prediction of angles, the choice of loss function is not trivial and must be chosen carefully. In the prediction of azimuth the MAE or MSE would be poor choices as the two angles  $v$  and  $v + 2\pi$  are equal on the unit circle but not on a number line, and if a prediction is "off" by  $2\pi$ , it would produce the largest loss possible. Former NBI master student Rasmus Ørsøe found in his master's thesis [49] that the 2-dimensional **von Mises-Fisher (vMF)** loss function yielded the best results in azimuth prediction for DynEdge, so that is what is used in this thesis. The von Mises-Fisher distribution (also known as a directional Gaussian, due to their similarity) has the probability density function

$$p_n(\mathbf{x}|\mathbf{u}, \kappa) = C_n(\kappa) \exp(\kappa \mathbf{u} \cdot \mathbf{x}) = C_n \exp(\kappa \cos \Delta\phi) \quad (5.4)$$

where  $\mathbf{u}$  and  $\mathbf{x}$  are n-dimensional unit vectors for the direction and predicted direction, respectively, and  $\kappa$  is analogous to  $\sigma^{-2}$  found in the Gaussian.  $\Delta\phi = y - \hat{y}$  is the difference between the true angle and the predicted angle, and  $C_n$  is the normalization constant defined as

$$C_n(\kappa) = \frac{\kappa^{\frac{n}{2}-1}}{(2\pi)^{\frac{n}{2}} I_{\frac{n}{2}-1}(\kappa)} \quad (5.5)$$

where  $I_v$  is the modified Bessel function [50] of the first kind at the order  $v = \frac{n}{2} - 1$ . In our case, we use the 2D distribution, where  $n = 2$ , and  $v = 0$ , so eq. 5.5 becomes

$$C_2(\kappa) = \frac{1}{2\pi I_0(\kappa)} \quad (5.6)$$

and 5.4 becomes

$$p_2(\bar{x}|\bar{u}, \kappa) = C_2 \exp(\kappa \cos \Delta\phi). \quad (5.7)$$

The distribution becomes a loss function we can minimize when we take the negative logarithm of the function

$$\mathcal{L}(\mathbf{u}, \mathbf{x}) = -\ln(p_n(\mathbf{x}|\mathbf{u}, \kappa)) = -\ln(C_n) - \kappa \cos \Delta\phi. \quad (5.8)$$

Because  $\kappa$  is analogous to  $\sigma^{-2}$  from the Gaussian, it is the width of the distribution. We can then make the model output the value of  $\kappa$  along with each prediction, giving us the uncertainty on the prediction, which usually is not available for other loss functions.

When using the 2D vMF loss function, the target vector is constructed as

$$\mathbf{v}_{target} = (\cos(\phi), \sin(\phi)) \quad (5.9)$$

such that the model predicts a vector of cosine and sine to the angle. The predicted angle can then be recovered through

$$\hat{\phi} = \arctan \frac{\cos \hat{\phi}}{\sin \hat{\phi}}. \quad (5.10)$$

### 5.1.3 Gradient Descent

A machine learning model can have millions of learnable parameters<sup>2</sup>, creating a very high-dimensional parameter landscape. The goal is to find the parameters that minimize the loss in this landscape. The loss function is a function of the input variable  $y$ , and the predicted value of  $y$ ,  $\hat{y}$ . As the prediction depends on the parameters of the model,  $\mathbf{p}$ , it can be written as a function of the parameters  $\hat{y}(\mathbf{p})$ .

At the end of each epoch, the gradient of the loss function  $\frac{\partial \mathcal{L}(\mathbf{p})}{\partial \mathbf{p}}$  with respect to the parameters,  $\mathbf{p}$ , is calculated, which reveals where the slope of the parameter landscape is the greatest. This process is called **backpropagation**. By taking a step in the opposite direction (down the slope), the loss should be lower after the next iteration. Figure 5.3 shows the iterative path of the model from the initial parameters to the global minimum of the parameter space, by going down in the direction where the slope of the landscape is the greatest.

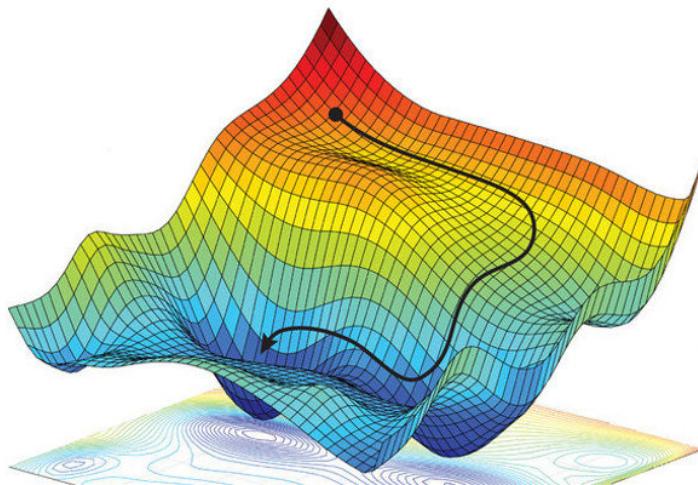


FIGURE 5.3: A figure showing an example of the parameter space of a neural network simplified to just three dimensions. Illustration from [51]

There are multiple layers in a deep neural network, where each node in a layer can be viewed as a function since it takes some values and returns some altered values. The output of one layer is the input of the next and so on, which is why we can find the gradient of the loss

<sup>2</sup>As opposed to hyperparameters.

using the chain rule

$$(f(g(x)))' = f'(g(x)) \cdot g'(x) \quad (5.11)$$

which is a common method of finding the derivative of a function of a function.

The gradient only reveals the slope in the current point in the parameter space occupied by the model, and the model blindly takes a step in that direction with no knowledge of where the global minimum is. The hope is that with enough iterations, this algorithm leads to the global minimum. The **learning rate** hyperparameter determines how large of a step is taken each time. This parameter must be carefully selected, as minima might be small, and learning rates too large could skip the minimum repeatedly, despite finding the right region. Similarly, the learning rate should not be too small either, as the parameter landscape can have multiple local minima, where the model might get stuck if it does not take steps large enough to overcome the barrier to the global minimum. A higher learning rate also means that the model learns faster, and fewer iterations are needed. The models used in this thesis employ a piecewise linear learning rate scheduler that changes the step size throughout the training. The default learning rate is defined as  $10^{-4}$ , and the default number of epochs to train for is defined as 50. The learning rate will then increase linearly from  $10^{-6}$  to  $10^{-4}$ , for the first half of the epoch budget of 50, and then decrease linearly from  $10^{-4}$  to  $10^{-6}$ . This trick benefits both from the increased speed of the large step size but finishes with a smaller step size which is more easily able to find smaller minima in the loss landscape. Peter Andresen also showed in his thesis [52] that this variable learning rate increased the resolution of the angular reconstruction compared to constant learning rates.

### 5.1.4 Model Convergence

Commonly, a data set is split into three parts, a training set, a validation set, and a test set, in approximately a 60% - 20% - 20% split. The model trains on the training set, but in order to fairly judge its performance, a separate set of unseen data (the validation set) is needed. At the end of each iteration of training (called an epoch) the model will predict labels for the validation set, and then compare the predictions to the labels of the validation set, to see how it performs. The use of a validation set prevents **overfitting**. Overfitting is when a model trains too much on a data set and starts to pick up on not only the signal in the data, but also the noise. The black line in figure 5.4 represents a model that has not overfitted, and the green line represents an overfitted model. The overfitted model has also learned the noise in the training sample, and therefore separates the training data perfectly, at the cost of generalizing poorly, and will perform worse on a separate sample of unseen data. This can also be seen in figure 5.5, where training loss continuously decreases, but at some point the validation loss starts to increase again. At the minimum of the validation loss, the model is said to have converged, and any training beyond that leads to overfitting.

It is necessary to train the model enough that it converges towards to best parameters, yet not so much it overfits. This is done by monitoring the model performance on the validation set. Usually, a procedure known as **early stopping** requiring a tolerance parameter  $\mathcal{T}$  is used, where a model will train until the predictions have not improved for  $\mathcal{T}$  epochs. The model will then revert back to the state where the validation loss was the lowest, effectively unlearning the last  $\mathcal{T}$  epochs.

We have the true value of all parameters in simulated events, but unfortunately, nature does not provide us labels on its particles. Therefore we can only test how close the prediction is to the true value in MC, and from that infer the uncertainty in the predictions on data, and assume they are equal to or greater than the uncertainty on the predictions on MC. The better the MC/data agreement is, the lower this uncertainty becomes.

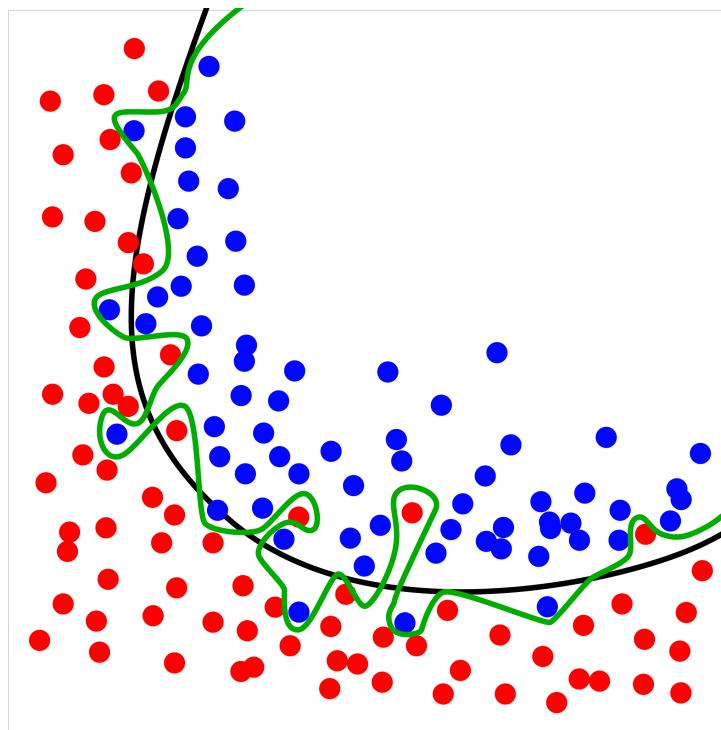


FIGURE 5.4: A figure showing the separation of the two classes, blue and red. The black line represents a robust model that has converged, but not overfitted, and the green line represents an overfitted model. Illustration from [53]

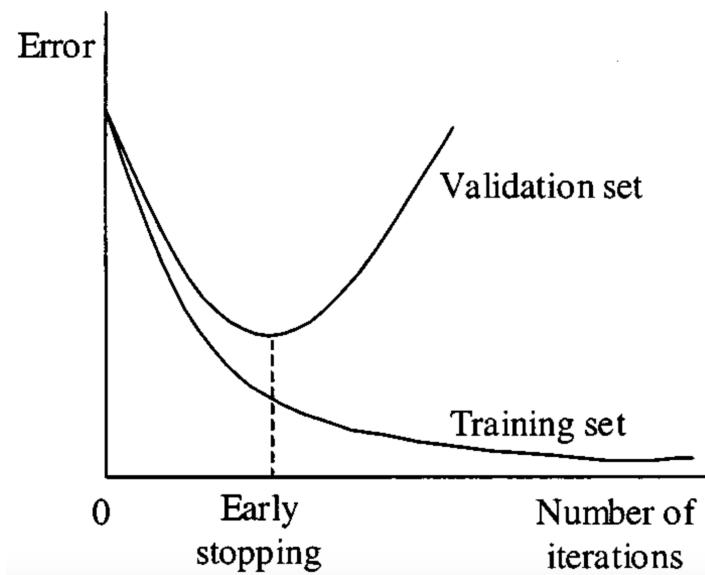


FIGURE 5.5: Figure showing how the loss of the test set and validation set evolves through iterations. Illustration from [54]

## 5.2 Neural Networks

Neural networks are a type of ML algorithm that takes inspiration from the architecture of the brain. In the brain, many cerebral neurons are connected to form a network where they send electric pulses to each other in order to process information and make decisions. This concept is very similar to the artificial neural network (ANN), which is usually shortened to neural

network (NN). A neural network consists of a collection of artificial neurons (here shortened to neurons), named after the cerebral neurons of the brain. A neuron is a function that takes a vector  $\mathbf{x}$  of  $N$  values with and applies the weights  $\mathbf{w}$  and a bias  $b$ , before taking the sum, and puts them through an **activation function**  $\phi$

$$y_i = \phi \left( \sum_{j=0}^{N-1} w_{ij} x_j + b \right). \quad (5.12)$$

If  $b$  is prepended to the vector  $\mathbf{w}$ , and 1 is prepended to the vector  $\mathbf{x}$ , the first term  $w_{i0}x_0$  becomes  $b \cdot 1 = b$ , and eq. 5.12 can be simplified to

$$y_i = \phi \left( \sum_{j=0}^N w_{ij} x_j \right). \quad (5.13)$$

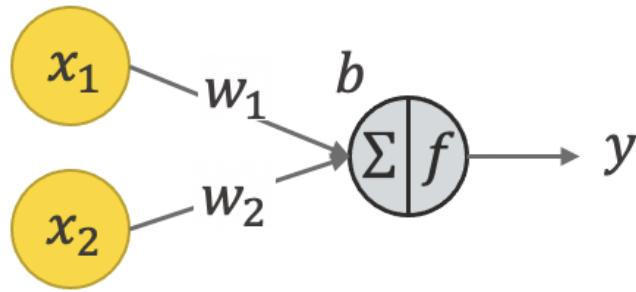


FIGURE 5.6: A neuron that takes two inputs and returns one output. Illustration from [55]

Activation functions ensure nonlinearity in the network. Without activation functions, the output of the network would just be linear combinations of linear combinations, which is still a linear combination, removing the need for multiple layers. Most problems that necessitate neural networks do not have a linear solution, so we use an activation function to add complexity. Further complexity can be added by introducing more neurons, serially or in parallel, to create a neural network.

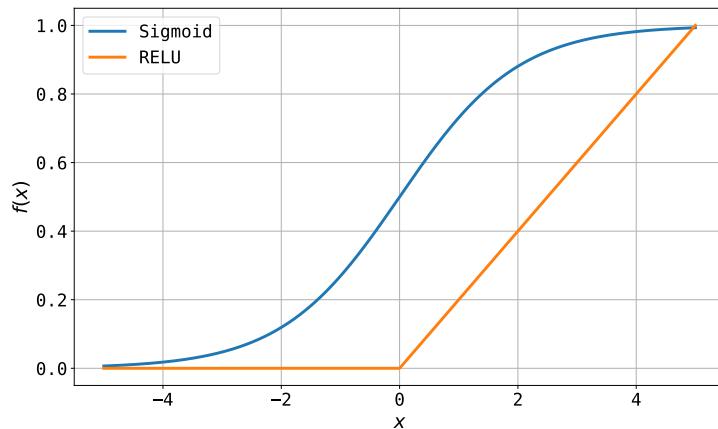


FIGURE 5.7: The two activation functions, sigmoid and RELU.

Figure 5.7 shows how an activation function might look. The Rectified Linear Unit (ReLU) is a simple function  $f(x) = \max(0, x)$  that returns  $x$  if  $x$  is larger than 0, and returns 0 otherwise. The sigmoid, defined as  $f(x) = (1 + e^{-x})^{-1}$ , has the advantage that all outputs between layers assume a value between 0 and 1. Otherwise, the values can quickly increase through the layers and cause an overflow, due to computers often using 32 or 64-bit floats, which can only represent numbers up to  $2^{32} - 1$  and  $2^{64} - 1$  respectively.

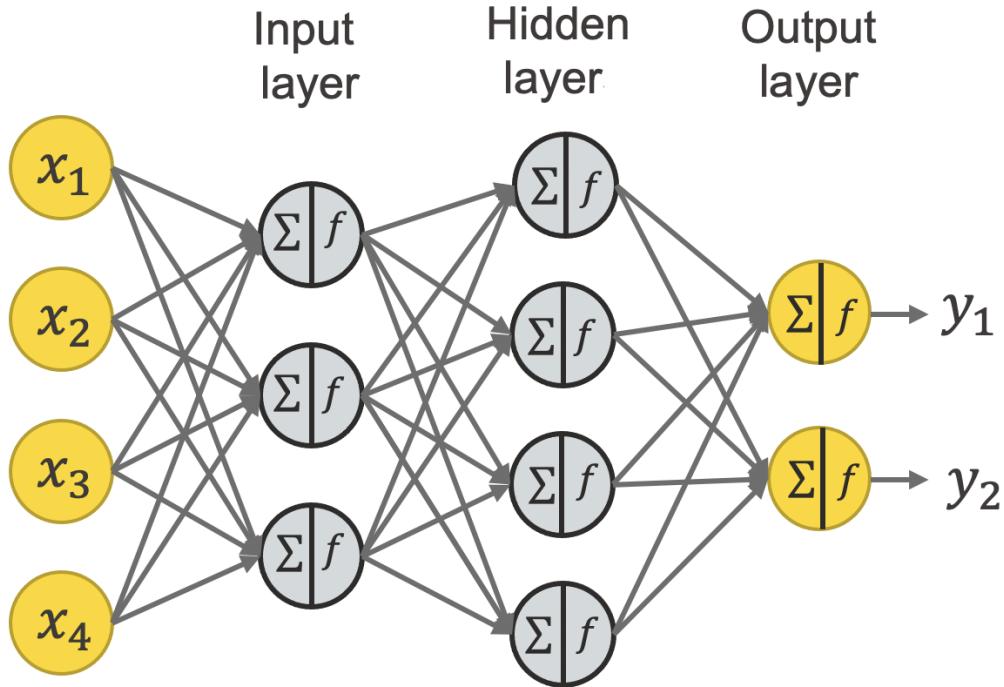


FIGURE 5.8: Neural network consisting of three layers of neurons. Illustration from [55]

Figure 5.8 shows a neural network with three layers, where the output of one layer becomes the input of the next. This is also a deep neural network because there is a layer that is not the input or output layer. Deep neural networks can become quite complex, often having millions of trainable parameters, and are therefore also capable of solving very complex tasks.

The neuron is also referred to as a perceptron, and as the term "neural networks" covers many types of models with the same basic idea, but with slight differences, this type of network is usually referred to as a **Multi-Layer Perceptron (MLP)**.

### 5.3 Graph Neural Networks

It's important to use the right type of ML model for your problem, and for problems with complex geometry, Graph Neural Networks (GNNs) are well suited. A GNN is a type of neural network that takes graphs as input. The concept of graphs is borrowed from the mathematical field of graph theory, where a graph is defined as a set of nodes and edges. The nodes can be thought of as points and edges are connections between the points. A GNN uses attributed graphs, which are graphs where the nodes have attributed node features, and the edges have edge features.

GNNs are essentially generalized **Convolutional Neural Networks (CNN)** which have been used in particle physics for a while. CNNs take data in the form of a matrix<sup>3</sup> and convolves over subgrids and produce some aggregate values. The CNN can be thought of as a special case of the GNN, where the input matrices are just neatly ordered graphs, where each node is an entry in the matrix, and the edges contain no information. The GNN also uses a convolutional operator, which can be thought of as a function that takes node and edge features as input, and returns some aggregate values. GNNs can thus unlike CNNs extract information from the relative positions of the nodes. This is a particularly important feature for IceCube as the geometry of the detector is irregular and contains a lot of information. The nodes and edges of a GNN have no spacial restriction, inviting us to just attribute the nodes to the coordinates of the triggered DOMs.

## 5.4 DynEdge

The model used in this thesis is called DynEdge, which is a GNN created using the GraphNeT [2] framework, which is built upon the PyTorch Geometric library [56]. In this model, each activated DOM is treated as a node in the graph. Each node contains the positional information ( $x, y, z$ ) along with the other features from the pulsemapper 4.2. Each node is connected to its 8 nearest neighbors<sup>4</sup> by edges. There is also some information in knowing the positions of unactivated DOMs, but they were excluded, resulting in significantly lower computation time [49]. Since only activated DOMs are included, the size of the input graph varies from event to event, which would be a problem for a CNN, but not for the GNN. DynEdge's ability to process inhomogeneous data with irregular geometry makes it extremely suited for IceCube.

DynEdge uses a convolutional operator called **EdgeConv** [57]. For every node  $n_j$  with node features  $x_j$ , EdgeConv convolves over all nodes in the neighborhood

$$\mathbf{x}_j = \sum_{i=1}^{N_{neighbors}} \text{MLP}(x_j, x_j - x_i) \quad (5.14)$$

such that the convolved node features  $\mathbf{x}_j$  are a sum of the outputs of the MLP, which takes the node features of  $n_j$ , and all node features of the nodes in the neighborhood of  $n_j$ . The neighborhood can be defined in multiple ways, but in this model, it is the 8 nodes with the shortest Euclidean distance.

In this case, each node has 5 node features which are the number of pulses in the graph and the node homophily ratio [58] of the  $x, y, z, t$  coordinates of each DOM, where the node homophily ratio is defined as the ratio of connected node pairs that have the same label as the node in question

$$\beta = \frac{1}{N} \sum_{i=1}^{N_{nodes}} \frac{\text{Number of } n_i\text{'s neighbors who have the same label as } n_i}{\text{Number of } n_i\text{'s neighbors}}. \quad (5.15)$$

The homophily ratio indicates how many of the pulses in the neighborhood have the same  $x, y, z$  coordinates, i.e. originate from the same DOM, and how many have the same  $t$ , i.e. hit at the same time.

The input graph undergoes four serialized edge convolutions, as seen in figure 5.9. Each edge convolution starts with a K nearest neighbors (KNN) algorithm to calculate the neighborhood of each node. In the first convolution, it is the Euclidean distance, but as the node features are convolved, the nodes are connected in some other fashion, that is increasingly abstract for each edge convolution. Thus the model is not constrained by our idea of distance having to be

<sup>3</sup>Usually images which are just matrices of RGB values.

<sup>4</sup>Given the node has at least 8 neighbors.

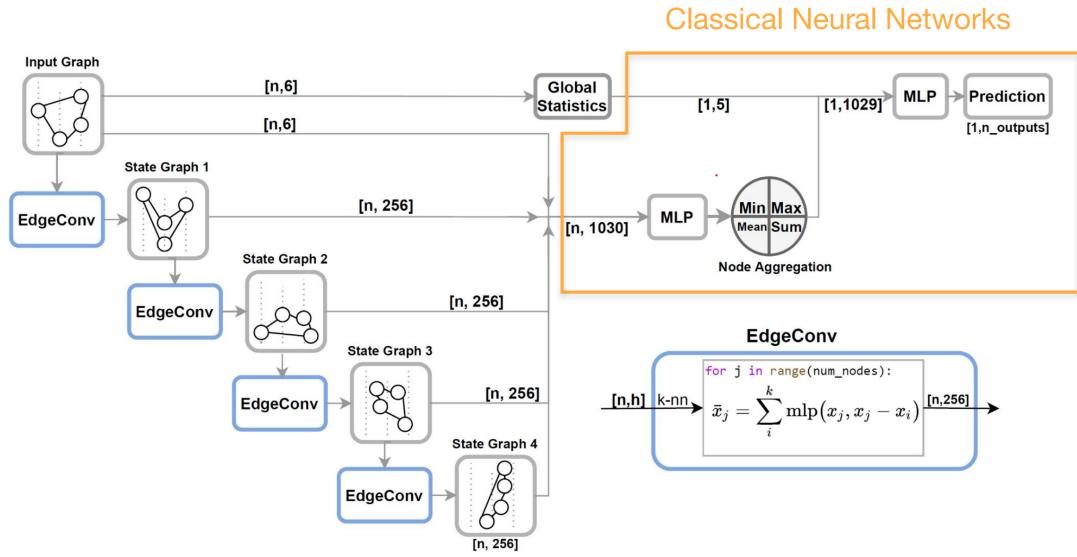


FIGURE 5.9: The architecture of DynEdge. Illustration from [2].

euclidean or even spacial, and can find some other metric where it makes more sense to group the nodes. This concept is illustrated in figure 5.9, where the input graph resembles how the DOMs are ordered in the ice, but after the first edge convolution,  $x, y, z$  become the convolved features  $\tilde{x}, \tilde{y}, \tilde{z}$ , that no longer directly represents the nodes spatial position.

Figure 5.9 shows how the  $[n, 6]$  features from the input graph along with the  $[n, 256]$  features from each subsequent layer are concatenated to form one  $[n, 1030]$  array, which is then fed into a two-layer perceptron. The MLP outputs an  $[n, 256]$  array, which then undergoes node aggregation, and becomes the 4 summary variables **min**, **max**, **mean**, **sum** of the  $n$  rows, producing a  $[1, 4 \cdot 256]$ , which is then concatenated onto to the initial unconvolved 5 global statistics, to form a  $[1, 1029]$  array. The aggregation takes the data from a variable size  $[n, 256]$ , depending on the number of nodes, to the fixed  $[1, 1029]$  shape, allowing the model to work on inputs of any initial size. This aggregated array then is fed into a last MLP, producing the desired prediction.

Because DynEdge takes the pulsemmap as input, it is especially important that the pulsemmaps are similar in data and MC.

## Chapter 6

# Analysis

### 6.1 Overview

The analysis chapter starts with an examination of how the data used in the analysis was selected, and how weights were assigned to the MC set. Then begins the analysis of this thesis, which is divided into three parts.

- **Pulse merging** (section 6.4)

In section 4.4, it was explained that there are numerous physical phenomena that produce excessive pulses, that we are not interested in. Some effects split pulses, such that the total charge is divided between two pulses, and some effects that produce a pulse are just noise, and should be removed. Therefore I developed a pulse merging algorithm, **PulseMerger**, to identify pulses that were likely to be erroneous and merge them with other pulses. Data and MC distributions are compared before and after the algorithm is applied.

- **Efficiency** (section 6.6)

As mentioned in section 3.4 there are many properties of the DOMs and the ice that are poorly understood and information gained on these properties would allow for better simulations, which would improve results across all efforts of the IceCube collaboration. The efficiency of the detector is analyzed by reconstructing the trajectories of the emitted Cherenkov photons, and seeing how many DOMs are triggered and how many are not. There we also see how the efficiency depends on the depth, distance, and the photon angle.

- **Hit timing** (section 6.7)

By reconstructing the photons, we can also calculate the amount of time they travel and where. Thus for each photon that is recorded, we can calculate at what time the corresponding muon would stop. If we do this for all activated DOMs, we can study how well the DOMs keep time compared to each other, as they are supposed to be synchronized. Furthermore, it is possible to calculate the expected time of the photons, and thus classify hits as noise or early hits if they are early, and as scattered photons, late pulses, or noise if they are late. This also allows for the study of how much photons scatter in the ice.

## 6.2 Data Selection

For reasons explained in chapter 1 and section 2.6, the analyses in this chapter are on stopped muons. For Monte Carlo, a set of MuonGun events from the OscNext group is used, and for data, the burnsample is used. The burnsample is just raw data with no selection, so it was necessary to identify the stopped muons. These were found by using DynEdge to first identify the muons, and then divide them into stopped and through as illustrated in figure 6.1.

### 6.2.1 Event Type Multiclass Classification

A multiclass classification DynEdge model was trained to predict the event type of all events in the burnsample to be one of either noise, neutrino, or muon, by former fellow master's student at NBI, Peter Andresen, and is further discussed in his thesis [52]. This classifier was trained on 1.65m MC events. To make a realistic combination of events, all types of events were represented in even amounts. The set was 1/3 noise from Vuvuzela, 1/3 muons from MuonGun and 1/3 neutrinos from Genie. Of the neutrinos, 1/3 was electron neutrinos, 1/3 was muon neutrinos, and 1/3 was tau neutrinos. Neutrinos and antineutrinos behave almost identically and are therefore not distinguished between. From the 62 million events, the 34 million events with a predicted probability of being a muon greater than 0.9 were selected. GNNs are extremely good at distinguishing between noise events and particle events, so the amount of noise in the selection is negligible. Neutrinos and muons can be a bit harder to distinguish between, but because there are so many more muons in the data to begin with, the selection does not have to be very strict for the sample to be very clean.

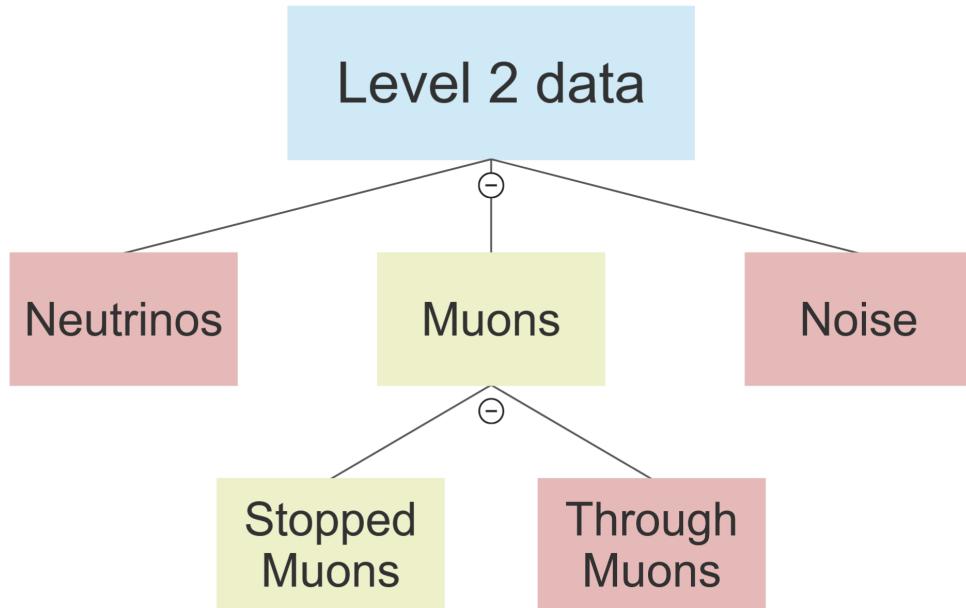


FIGURE 6.1: Decision tree showing the selection process. First, the events that are predicted to be muons are selected, and then the muons that are predicted to stop within the detector are selected.

### 6.2.2 Stopped Muon Classifier

A binary stopped muon classifier was trained on 1 million simulated MuonGun muons, where 500.000 were stopped, and 500.000 went through, to predict if the muons from the burnsample were stopped or through muons. All muons with a prediction of being stopped greater than 0.90 were used in the stopped muon sample.

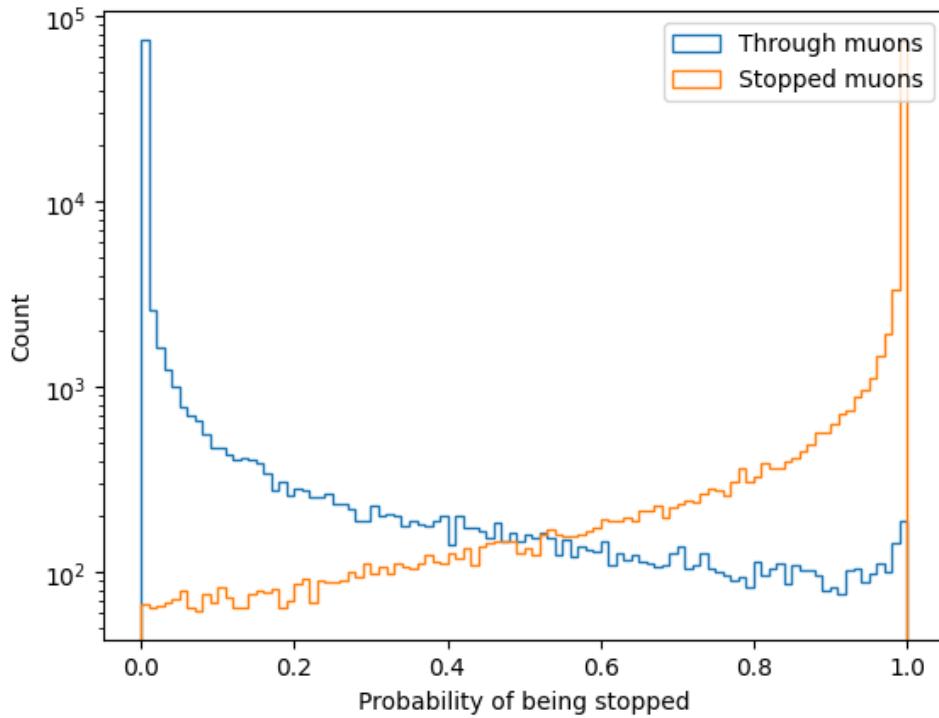


FIGURE 6.2: The distribution of the probability of being stopped for each muon.  
Stopped muons are in orange and through muons are in blue.

Figure 6.2 contains two histograms; one showing the distribution of the predicted probability of being stopped for the muons that were simulated as being through in blue, and the distribution of the predicted probability of being stopped for the muons that were simulated as being stopped in orange. Considering the vertical axis is logarithmic, it already shows great separation of the two classes.

Figure 6.3 shows the **receiver operator characteristic (ROC) curve** of the predictions on the test set. As we have false positives on the horizontal axis and true positives on the vertical, a perfect classification would have a true positive rate of 1 and a false positive rate of 0, making the ROC curve hug the upper left corner. This would also result in the **area under the curve (AUC)** being 1. Assuming an even split of both labels and enough statistics, a random classification of all events would result in a line going across the diagonal of the plot, resulting in an AUC of 0.5. The AUC of the stopped muon classifier being 0.991, means that it performed very well on the test set.

Table 6.1 shows the number of events before and after selecting the muons and stopped muons.

Selection	Number of events	AUC
None	61,739,862	N/A
Muons	34,033,998	0.994
Stopped muons	14,594,307	0.991

TABLE 6.1: The number of events before and after both selections.

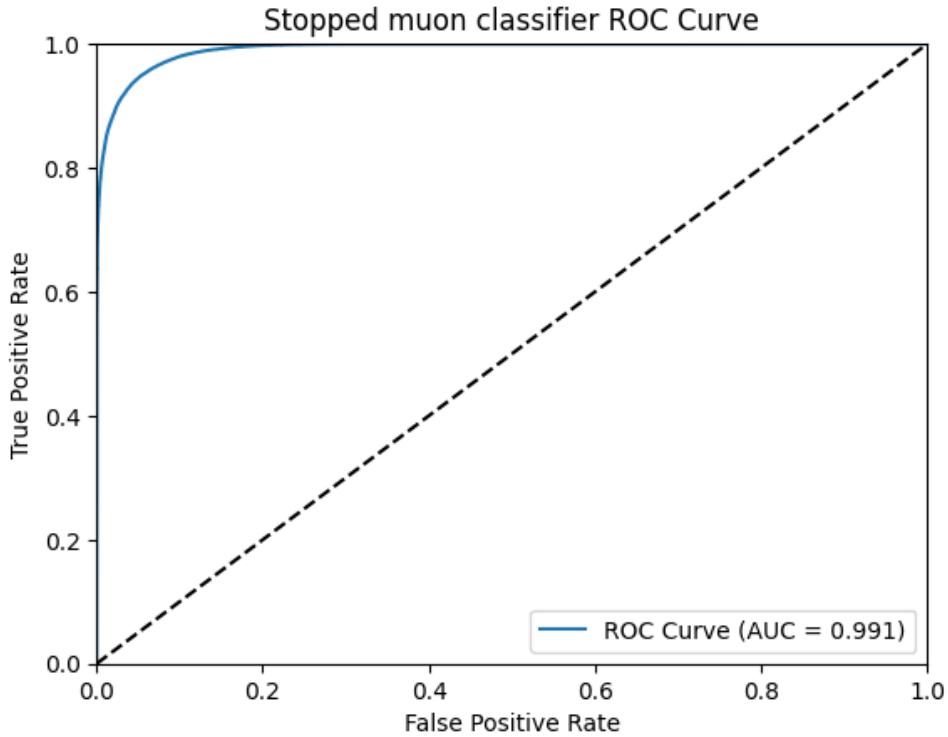


FIGURE 6.3: The ROC curve of the predictions of the muons being stopped muons on the test set.

### 6.2.3 Reweighting Data

When comparing data to simulation, it is common practice to apply weights to the simulations when plotting. This is to account for different effects, such as scaling the distributions to match in volume, and difference in statistics. Each event is given one value such that in a histogram, that event counts not necessarily as 1, but as  $w$ , the weight. If done right, this will result in the different distributions matching better when comparing data to simulation.

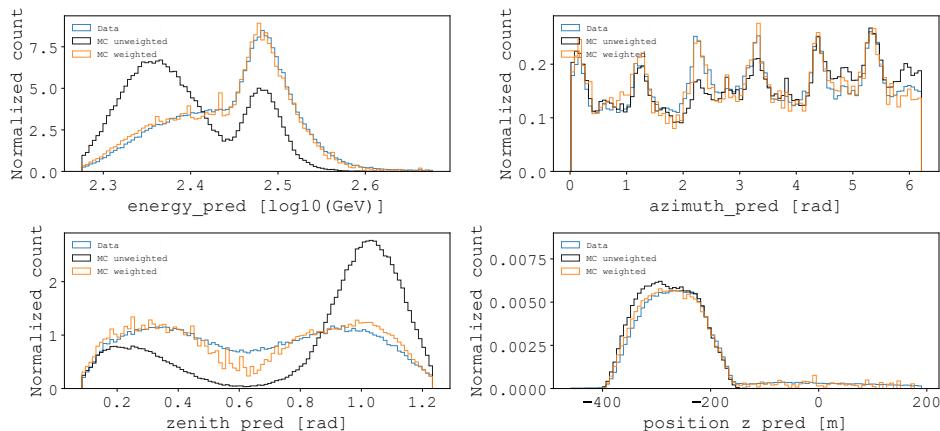


FIGURE 6.4: The four figures show the distributions of the four features that we are using to weight the events before and after weighting.

To assign each event with a weight that improves data similarity across multiple features, the function `GBReweigher` from the python library called `hep_ml` [59] was used. `GBReweigher`

is a reweighter algorithm that utilizes an ensemble of regression trees to minimize the difference between data and simulation across a set of features. The features used to reweight on were `azimuth_pred`, `zenith_pred`, `energy_pred`, `position_z_pred`, where `position_z` refers to the z-coordinate of the stopping point of the muon.

Figure 6.4 shows that while the distributions match poorly initially, they are close to identical after the application of the weights. The coordinate system used at IceCube has defined the point ( $x = 0, y = 0, z = 0$ ) to be somewhere in the middle of the detector, so  $z = 0$  corresponds to a depth of around 2000m below the surface.

## 6.3 Reconstruction

Now we have identified and selected the stopped muons we need, but we still don't know anything about them. For the analysis in section 6.6 and 6.7, we need to know the zenith and azimuth angle of incidence along with the coordinates of the stopping point in the ice. These values are predicted using DynEdge regression models. All regression models were trained on a sample of 1m MuonGun stopped muons. ROC curves can only be used for classification tasks, so to determine performance on the regression tasks, we will inspect residual plots i.e. histograms of the difference between the predicted values of the target variable and the true values of the target variables. Note that the test sample is necessarily MC events and that we can only expect the predictions on data to be of equal or worse quality. When doing calculations based on the predicted values of e.g. the angle of a particle, there will be an uncertainty on the result stemming from the error of the prediction. We will later see how this error can be estimated. Each of the three reconstructions took less than an hour for the 14.6m events.

### 6.3.1 Muon Azimuth Reconstruction

A DynEdge model was trained to predict the azimuth angle of the muon. Figure 6.5 shows the plot of the residuals of the azimuth prediction. The distribution of the residuals has a mean very close to 0, indicating that the model does not tend to predict significantly over or under the true value. The residuals have a standard deviation of 0.211rad or  $12^\circ$ . Since the values of the angles exist on a circle, the residuals have been transformed such that residuals below  $\pi$  have  $2\pi$  added, and residuals above  $\pi$  have  $2\pi$  subtracted. This ensures that residuals close to  $2\pi$  are transformed into residuals close to 0.

Figure 6.6 shows the distribution of the actual angles and the predicted angles. Both histograms have large variations from bin to bin, and no interval of angles are significantly different, i.e. there does not seem to be a specific angle interval the model seems to have more trouble predicting than others. The peaks and valleys arise from the corridors mentioned in section 3.1.3

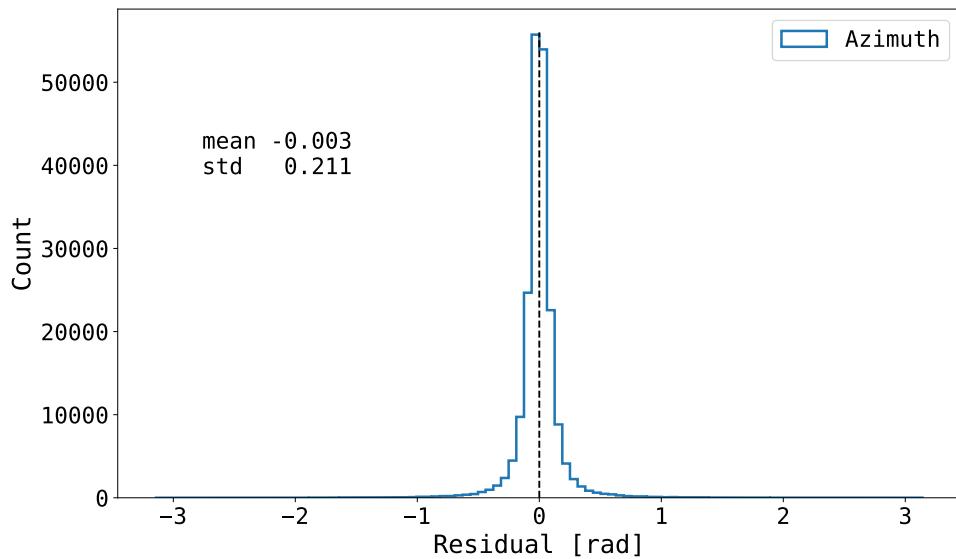


FIGURE 6.5: This plot shows the residuals of the predictions of the azimuth angle for the validation set of the MuonGun stopped muons.

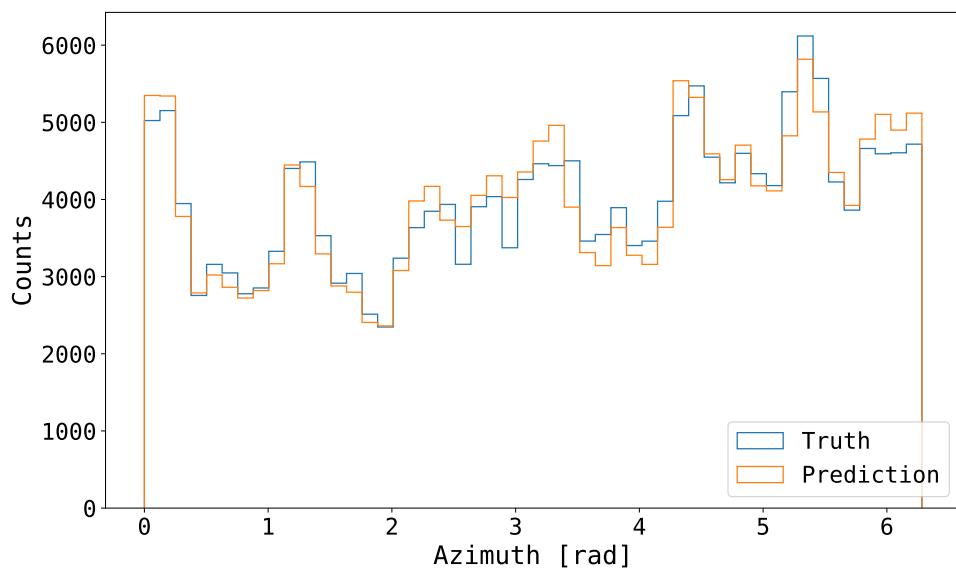


FIGURE 6.6: The distributions of the true values of azimuth in blue, and the predicted values of azimuth in orange.

### 6.3.2 Muon Zenith Reconstruction

Another DynEdge model was trained to predict the zenith angle of the muon. The residual plot of the predictions of the zenith angle is shown in figure 6.7. The residuals here are quite low compared to that of the azimuth reconstruction, being just 0.054 rad or  $3.1^\circ$ , but the interval that the values span is also a lot smaller.

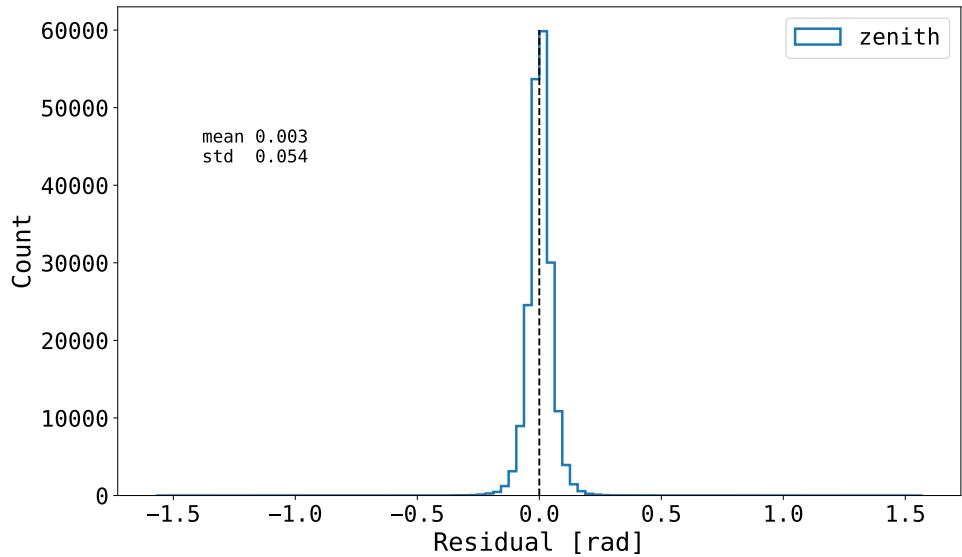


FIGURE 6.7: The residuals of the predictions of the zenith angle for the validation set of the MuonGun stopped muons.

While the allowed values for the zenith angle are  $[0, \pi]$ , we only observe muons in roughly the interval  $[0, \pi/2]$  as shown in figure 6.8, because muons are easily stopped by the Earth and only muons from above will make it to the detector. Figure 6.8 also shows that the distributions of the true values and the predicted values look very similar, with a slight over-prediction where the counts are the highest in truth, and under-prediction where the counts are lowest in truth. ML models sometimes tend to predict close to the most common value if uncertain, because then the expected loss will be lowest. In the same vein, it will avoid predictions in areas with low counts, to avoid high loss if wrong.

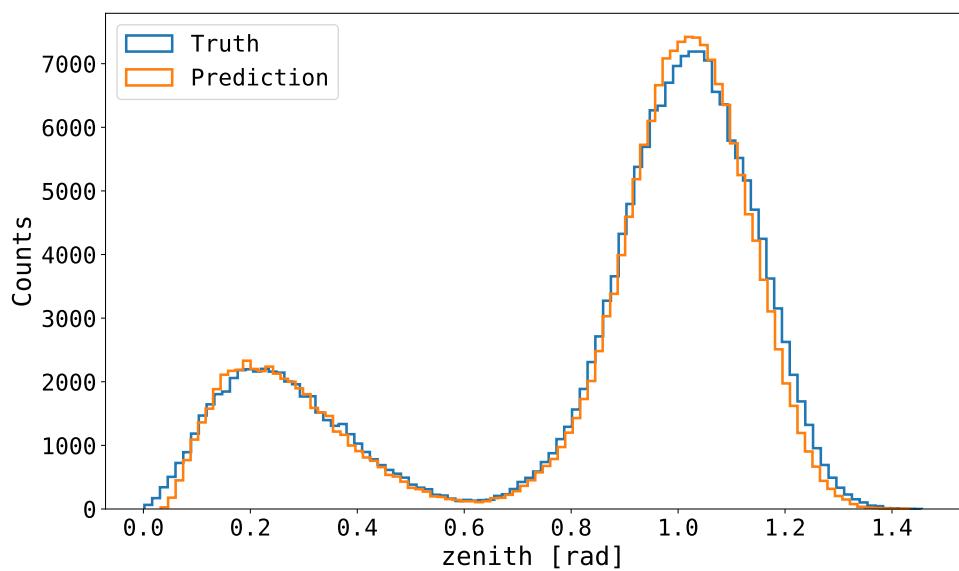


FIGURE 6.8: The distributions of the true values of zenith in blue, and the predicted values of zenith in orange.

### 6.3.3 Muon Stopping Point Reconstruction

Another DynEdge regression model was trained on the same 1m MuonGun sample of exclusively stopped muons, to predict the three coordinates  $x, y, z$  of the stopping point. Figure 6.9 shows the three residual plots on top of each other, along with the color-coded mean and standard deviation of the distributions. The three distributions have similar shapes and standard deviations. The position\_x variable has a mean very close to 0m, while position\_y is slightly further from 0, and position\_z is slightly further still. The offset is clear in these plots, but in such a large detector, where the DOMs are up to 17m apart, and strings are about 125m apart, 3m is not that much.

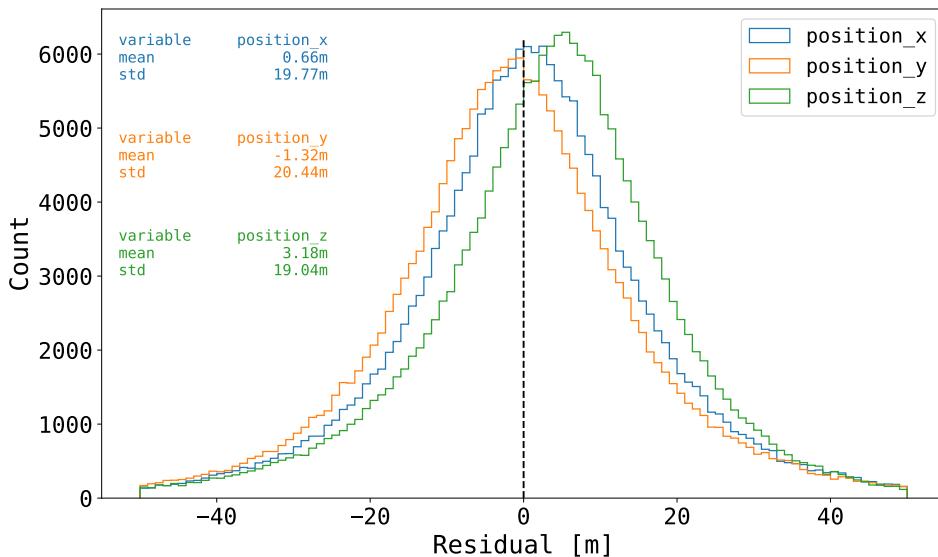


FIGURE 6.9: The residuals of the predictions of the three stopping coordinates for the validation set of the MuonGun stopped muons.

The three distributions in figure 6.9 show the three coordinates independently, making it hard to determine the quality of the stopping point prediction for each event. Figure 6.10 on the other hand, shows a histogram of the distance between the predicted stopping point and the true stopping point for all the events. There we see that the mean of the distribution is 26.8m, the median is 21.8m, and the standard deviation is 21.5m. Again, these numbers are not that large compared to the size of the detector.

Figures 6.11, 6.12, and 6.13 show the distribution of the  $x, y$ , and  $z$  coordinates and their prediction. These prediction distributions match their counterparts decently, with no massive discrepancies. Only the  $z$ -coordinate has a large bump in the lower part of the detector in the reconstruction compared to the true values.

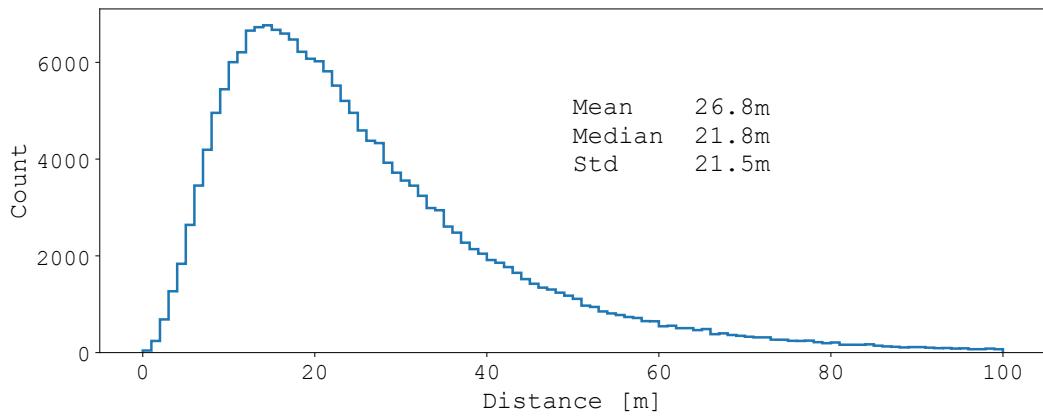


FIGURE 6.10: The distribution of the distance between the predicted stopping point and the true stopping point for all stopped muons.

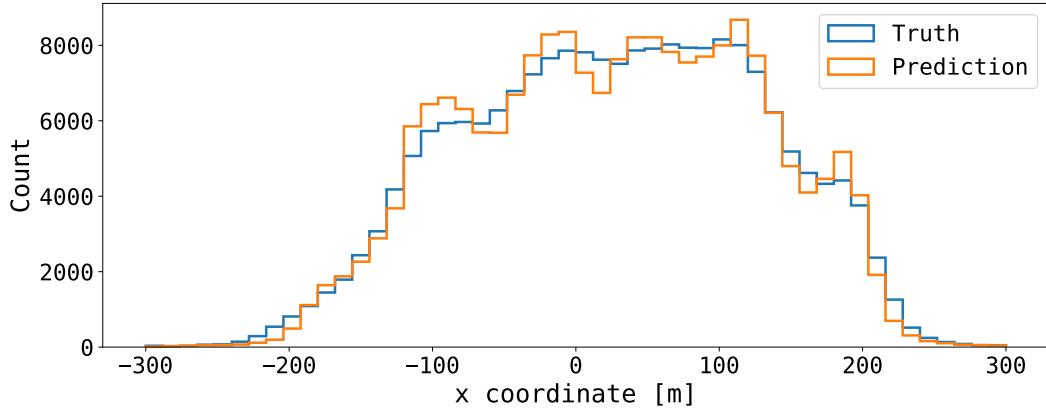


FIGURE 6.11: The distributions of the x-coordinates of the test set, and their corresponding predictions.

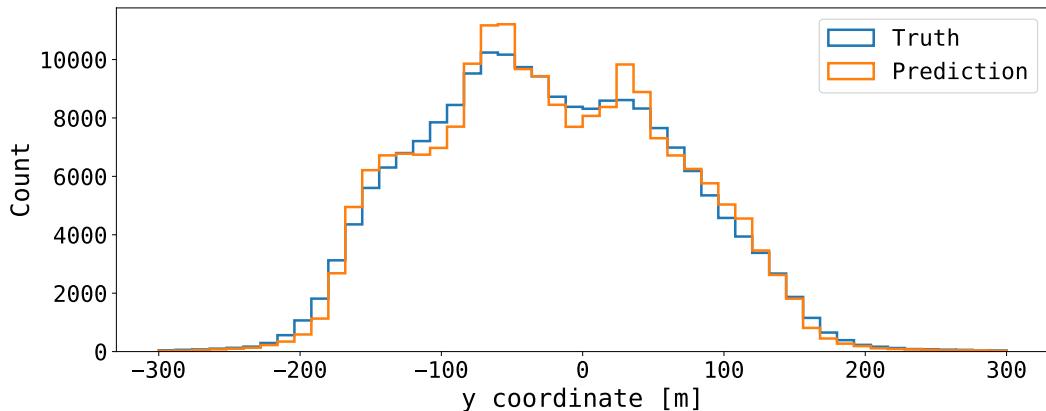


FIGURE 6.12: The distributions of the y-coordinates of the test set, and their corresponding predictions.

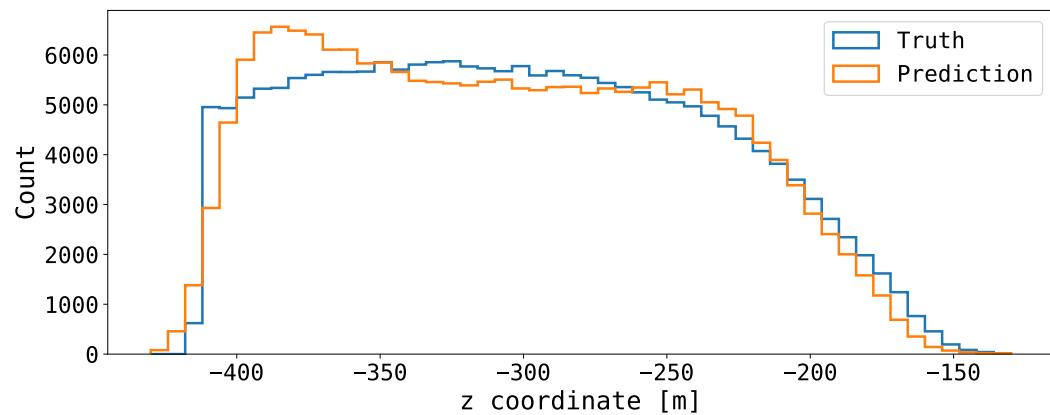


FIGURE 6.13: The distributions of the z-coordinates of the test set, and their corresponding predictions.

## 6.4 Pulse merging

In section 4.4, the different ways pulses may incorrectly be split in two were presented. These phenomena inspired the criteria of merging in the algorithm I created, called **PulseMerger**. The hypothesis is that these pulses can be removed or combined with signal pulses, potentially increasing training performance. Even if performance remains unchanged, it decreases the size of each event, allowing for faster training.

It is not trivial to identify the pulses that should be merged, and there are many different valid rules, but in this analysis, a simple charge criterion was used. If the splitting is the fault of WaveDeform, one might expect that it was a hit with a high charge that was split into a slightly lesser charge and a small charge, and one would require a small charge to be close in time to a significantly larger charge, to be merged. That is the logic behind the implementation of the charge cut. A criterion demanding that the smaller charge didn't have to be below 0.3PE, but just a certain fraction (e.g. 1/5) of the pulse it is merged onto was also investigated, (inspired by [60]), but abandoned by the simpler and more effective charge cut. A criterion to restrict a pair of pulses to be within a short time interval to be merged also just decreased the effect of the merging and was also abandoned.

The algorithm starts by identifying all the pulses with a low charge in the pulsemapper, iterates through them all, and merges them with the pulse that is closest in time and above the charge threshold. If two pulses that have low charge are closest to each other in time, they are paired with the pulses that are second closest in time.

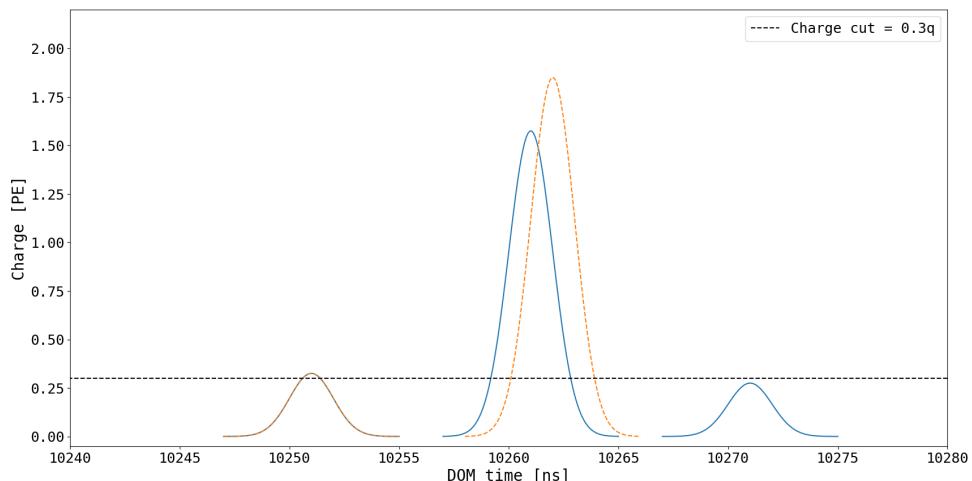


FIGURE 6.14: Visualization of three pulses recorded during an event before and after merging.

Figure 6.14 shows three pulses recorded by a DOM in an event. Since the charge cut is set to 0.3PE, any pulse with a charge below that threshold would be merged, given one of its neighbors is also a suitable merging candidate. In this plot, the blue peaks represent the original pulses, and the orange peaks are the same pulses after the pulse merging. Notably, the third pulse with a charge less than 0.3PE has been merged with the preceding larger pulse. The new charge is the sum of the two initial charges, and the time is the average of the two times, weighted by the charge of each pulse.

In figure 6.15, there is a clear MC/data discrepancy for HLC hits with charges less than 0.3PE, which then became the initial merging parameter.

### 6.4.1 The Effect of Merging on Charge Distribution

The merging of the MC and data set did not change the overall distribution a lot initially until the distributions were split into HLC and SLC<sup>1</sup> hits. It turned out that the merging had almost no effect on SLC hits, but had a significant effect on HLC hits. Immediately it is apparent that HLC hits have a larger MC/data discrepancy, and that the pulse merging alters the HLC distributions more than the SLC distributions, indicating that the pulse merging is mending a problem only appearing in HLC hits.

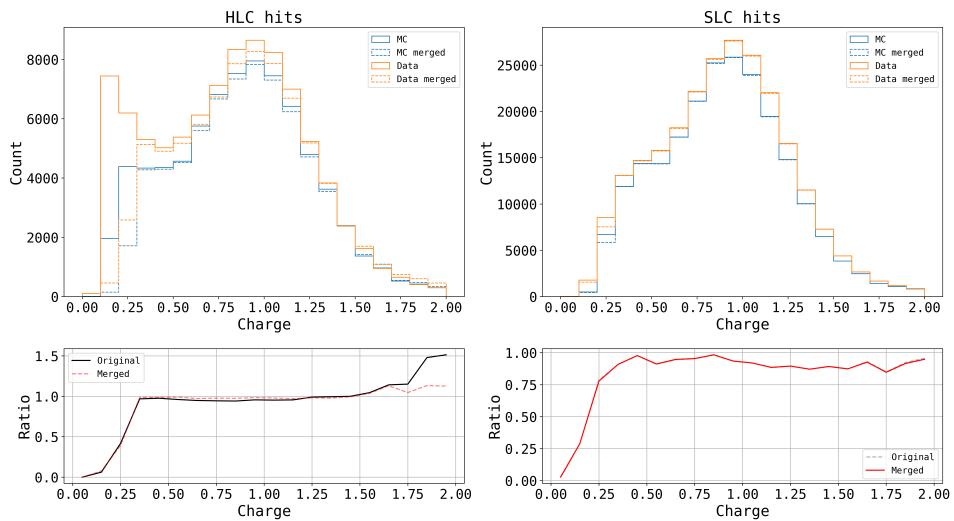


FIGURE 6.15: Charge distribution of HLC and SLC separately in MC and data, before and after merging. The plots in the bottom row are the ratios between MC and data before and after merging.

While the improvement is evident through visual inspection, it can also be seen from the plots in the bottom row. The lines are the quotient of the two corresponding histograms, and thus the closer that line is to 1, the more alike the MC and data histograms are. In figure 6.15, while merging has virtually no effect on the ratio plot of the SLC hits, the ratio of the merged pulses is slightly closer to 1 for most of the charge spectrum, and the ratio is less dependent on the charge.

### 6.4.2 Pulses Per DOM

MC differs noticeably from data in the number of pulses that are recorded per DOM during an event. This is something a pulse merging algorithm could potentially improve, as it lowers the number of pulses in a set. We see the results in figure 6.17. Upon inspection of the ratio plots in the bottom row, we see that the red line (the ratio of the merged sets) is closer to 1 up to 7 pulses per DOM, which also accounts for a large majority of all DOMs.

This plot was done without weighting MC because the very few DOMs had many pulses were weighted a lot higher, so the information of the plot was obscured by a bias in weighting.

<sup>1</sup>HLC and SLC are explained in section 3.2.2.

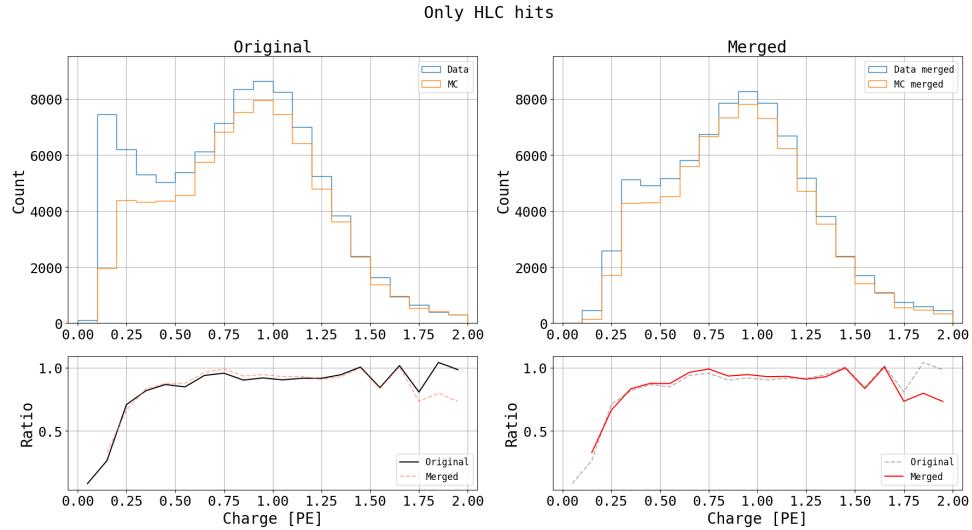


FIGURE 6.16: Charge distribution of HLC hits in MC and data before and after merging. The plots in the bottom row are the ratios between the distributions.

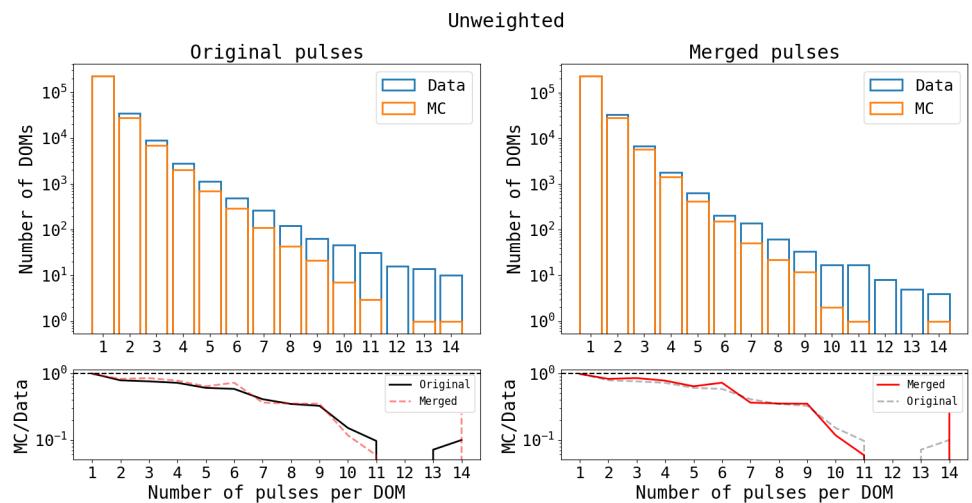


FIGURE 6.17: The number of pulses per DOM in MC and data before and after merging pulses.

## 6.5 Photon Reconstruction

Each of the many muons selected produces tens or hundreds of recorded photons. That gives us a very high statistics sample of photons of different angles and distances, that traverse the ice. Looking at how these photons are scattered and how often the DOMs are triggered tells us about the efficiency of the detector as well as the properties of the ice. Furthermore, they can be used to determine some timing properties of the detector.

Muons are well suited for this analysis because the low-energy neutrinos produce few pulses and are only measured in DeepCore, which is a relatively small volume, and high-energy neutrinos have low statistics.

The rest of the analysis chapter is based on the reconstruction of the Cherenkov photons emitted by the muons, and we will therefore start with a review of how the reconstructions are made.

In data and MC<sup>2</sup>, we have the predicted values for azimuth, zenith, and the coordinates of the endpoint of the particle. From these values a vector can be constructed, replicating the track of the muon, having the same angle and endpoint. The start point is defined to be 2km<sup>3</sup> away from the endpoint along the direction of the track to ensure that it falls outside the detector.

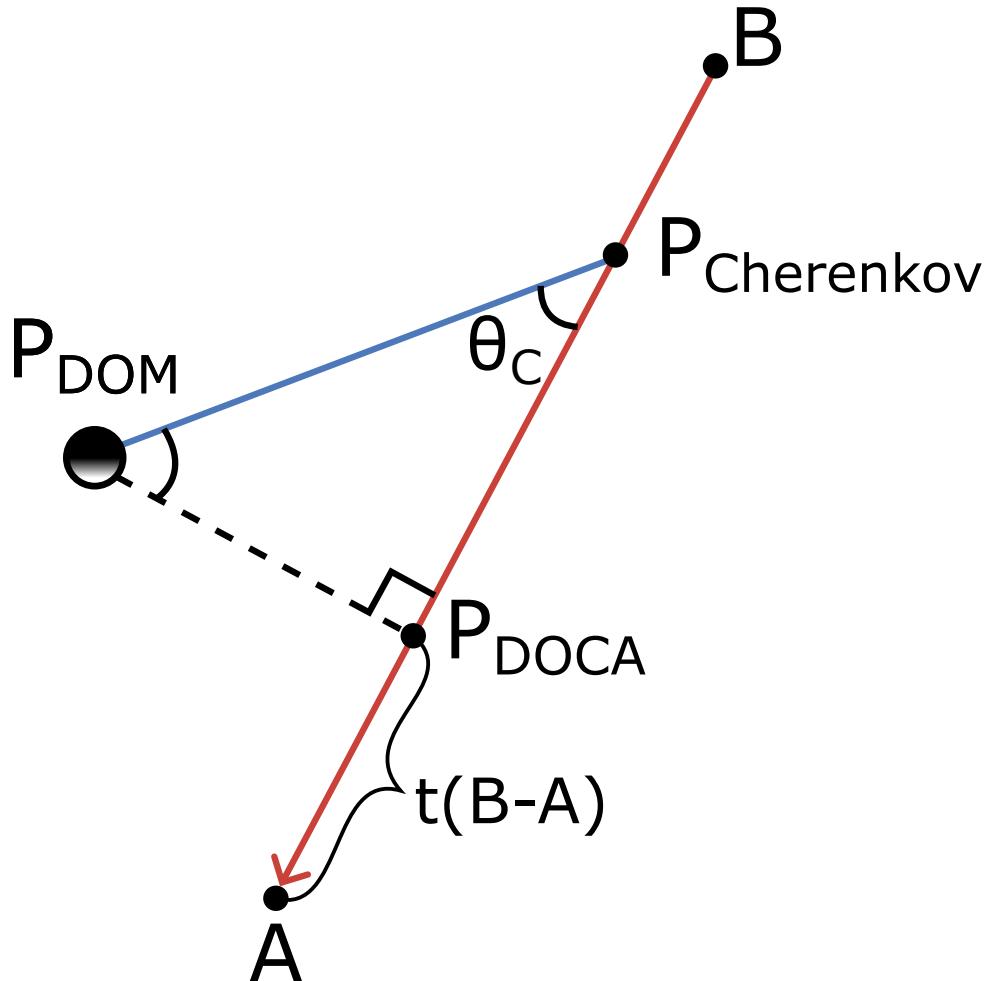


FIGURE 6.18: Illustration of the points and vectors needed for photon reconstruction.

<sup>2</sup>MC also contains the true values, but predictions were used for both data and MC for consistency.

<sup>3</sup>2km was chosen as the distance between the two furthest DOMs is 1557m.

Given a point  $P_{DOM}$  and a vector  $\overline{BA}$  going from point  $B$  to  $A$  in three dimensions, the point  $P_{DOCA}$ <sup>4</sup> on  $\overline{BA}$  that has the shortest distance to  $P_{DOM}$  is calculated. We start by projecting the vector  $\overline{AP_{DOM}}$  onto the vector  $\overline{BA}$ , by multiplying the length of the vector by the cosine of the angle  $\phi$  between the two vectors, and then dividing by the length of  $\overline{BA}$ , to get the result in units of the length of  $\overline{BA}$ ,

$$t = \frac{|\overline{AP_{DOM}}| \cdot \cos \phi}{|\overline{BA}|}. \quad (6.1)$$

where  $t$  is then the length of the projection of  $\overline{AP_{DOM}}$  onto  $\overline{BA}$ , or the distance from  $A$  to  $P_{DOCA}$  in units of the length of the vector  $\overline{BA}$ . Then  $P_{DOCA}$  is found by scaling  $\overline{BA}$  by the factor  $t$ , starting from  $A$ ,

$$P_{DOCA} = A - t\overline{BA}. \quad (6.2)$$

In the code, eq. 6.1 is rewritten, so  $t$  does not depend on  $\phi$ , but only the coordinates  $A$ ,  $B$ , and  $P_{DOCA}$ . First both the numerator and denominator are multiplied by the length of  $\overline{BA}$

$$t = \frac{|\overline{BA}| \cdot |\overline{AP_{DOM}}| \cdot \cos v}{|\overline{BA}|^2}. \quad (6.3)$$

Now the numerator and denominator can both be expressed as dot products

$$t = \frac{\overline{BA} \cdot \overline{AP_{DOCA}}}{\overline{BA} \cdot \overline{BA}} \quad (6.4)$$

and we only need the coordinates of the three points  $A$ ,  $B$ , and  $P_{DOCA}$  for the calculation.

The distance between two points  $P$  and  $Q$  is calculated using the Euclidean distance in three dimensions

$$\overline{PQ} = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}. \quad (6.5)$$

Because we know the point of absorption ( $P_{DOM}$ ), the fact that the point of emission ( $P_{Cherenkov}$ ) lies on the vector  $\overline{BA}$ , and the angle of emission  $\theta_{Cherenkov} = 40.75^\circ$ <sup>5</sup>, it is a matter of simple trigonometry to find  $P_{Cherenkov}$ . We will consider the triangle created by the three points  $P_{DOM}$ ,  $P_{DOCA}$ ,  $P_{Cherenkov}$ . Since the distance  $\overline{P_{DOCA}P_{DOM}}$  is minimized,  $\overline{P_{DOCA}P_{Cherenkov}}$  must make a right angle on  $\overline{BA}$  and we know that the photons are emitted with  $\theta_{Cherenkov}$ , and that the sum of the three angles is  $\pi$  rad or  $180^\circ$  so the last angle is

$$\angle P_{DOM}P_{DOCA}P_{Cherenkov} = \pi - \frac{\pi}{2} - \theta_{Cherenkov} = \frac{\pi}{2} - \theta_{Cherenkov} \quad (6.6)$$

So the distance from  $P_{DOCA}$  to  $P_{Cherenkov}$  is

$$|\overline{P_{DOCA}P_{Cherenkov}}| = |\overline{P_{DOCA}P_{DOM}}| \cdot \tan\left(\frac{\pi}{2} - \theta_{Cherenkov}\right). \quad (6.7)$$

Then it is possible to trace back a distance of  $\overline{P_{DOCA}P_{Cherenkov}}$  along  $\overline{BA}$ , to find  $P_{Cherenkov}$ , in the exact same way  $P_{DOCA}$  was found

$$P_{Cherenkov} = P_{DOCA} - k \cdot \overline{BA} \quad (6.8)$$

---

<sup>4</sup>Distance Of Closest Approach.

<sup>5</sup>This value is calculated in section 2.4.

where analogous to  $t$  we now have

$$k = \frac{|\overline{P}_{DOC} \overline{P}_{Cherenkov}|}{|\overline{BA}|}. \quad (6.9)$$

Now that  $P_{Cherenkov}$  is calculated, the photon vector can be defined as

$$\bar{v}_\gamma = \overline{P_{DOM} P_{Cherenkov}}. \quad (6.10)$$

The following properties of  $\bar{v}_\gamma$  are also calculated:

$$\text{photon\_zenith} = \arccos(\bar{v}_{\gamma,z}) \quad (6.11)$$

$$\text{photon\_azimuth} = \arctan\left(\frac{\bar{v}_{\gamma,y}}{\bar{v}_{\gamma,x}}\right) \quad (6.12)$$

$$\text{photon\_z} = \frac{1}{2}(P_{DOM,z} + P_{Cherenkov,z}) \quad (6.13)$$

$$\text{photon\_distance} = |\bar{v}_\gamma|. \quad (6.14)$$

The photon reconstruction of 100,000 events took a bit less than two hours or around 15 events per second. The pulsemmap after the photon reconstruction is shown in table 6.2

## 6.6 Efficiency Analysis

In this analysis, it is investigated how much the detector is triggered compared to how much it could have been i.e. its efficiency. To remove noise, only photons that travel less than 100m are selected for the efficiency analysis. Of all the unactivated DOMs in an event, some are not activated because they might be far away from the track, and some may not activate even if the track goes right by. Therefore it is sensible to consider the efficiency of the DOMs as a function of the distance the photons have to travel to be detected. The efficiency of the detector is defined as the ratio of DOMs within some distance that are activated. The photon vectors are reconstructed for all 5083<sup>6</sup> DOMs whether they are activated during an event or not, giving a total of 5083 distances per event, such that we also know the distances to the unactivated DOMs.

The histograms of all DOMs and activated DOMs are shown for data, MC, and truth in figure 6.19, with a bin width of 1m. The points in the bottom plot are the ratios of the heights of the bins in the top plot and show the efficiency of the detector as a function of the photon distance. The points have been fitted with an exponential function to characterize the efficiency. The denominator in the fraction in the exponent represents the effective absorption length of the measured photons. The absorption length is how long on average a beam of photons will travel before only  $\frac{1}{e}$  of the original photons are left. Because there is so much data, the statistical uncertainty of the fits is very low, meaning that the algorithm is very certain that these parameters are the best for fitting these points. In section 3.4 it was explained that the ice is not a homogeneous medium and that the properties varied greatly, and figure 3.8 illustrated that the absorption and scattering of the ice fluctuated a fair amount, and those are properties that affect the efficiency immensely. Figure 6.19 therefore describes the efficiency of the detector, averaged over the entire detector, including the clear ice near the bottom and the dust layer. The  $\chi^2$ -values are also very high for MC and truth, resulting in probability values of 0.000, whereas data does not have as low a  $\chi^2$  value and probability, albeit still pretty low.

---

<sup>6</sup>Of the 5160 DOMs in the ice, only 5083 appeared in the used event selection.

Feature name	Description
dom_x	x-coordinate of the triggered DOM
dom_y	y-coordinate of the triggered DOM
dom_z	z-coordinate of the triggered DOM
dom_charge	Charge of the pulse
dom_time	Time from readout start to trigger
QE	Quantum efficiency of the DOM
width	Resolution of the pulse (1.0 if HLC, 8.0 if SLC)
pmt_area	The area of the PMT
photon_distance	The distance travelled by the photon
photon_zenith	The zenith angle of the photon
photon_azimuth	The azimuth angle of the photon
cherenkov_x	x-coordinate of P_cherenkov
cherenkov_y	y-coordinate of P_cherenkov
cherenkov_z	z-coordinate of P_cherenkov
photon_z	$\frac{1}{2}(\text{dom}_z + \text{cherenkov}_z)$

TABLE 6.2: The features of the pulsemmap after photon reconstruction.

The longer a photon has to travel, the larger the probability of it being absorbed is, decreasing the number of hits. This effect is balanced out by the fact that by increasing the volume of the sphere of interest, the number of DOMs available to detect the photon also increases. Therefore the number of activated DOMs is approximately constant for distances between 20m and 100m, while the efficiency decreases.

The efficiency of the data is only around half of the efficiency MC, showing that events are simulated to activate more DOMs per event than what occurs in data. The structure of the curves in the top plot in figure 6.19 are very similar. This is also evident by the similar fits, where the absorption length parameter in data is very close to the MC and truth values, but the amplitude and offset are different. This shows that while the efficiency in data is about half of that of MC, the respective absorption lengths of photons are very similar.

### 6.6.1 Photon Zenith Angle

Figure 3.10 showed the angular efficiency of different simulations, and how uncertain the efficiency was in upward-going photons. As muons cannot pass very far through Earth, we only

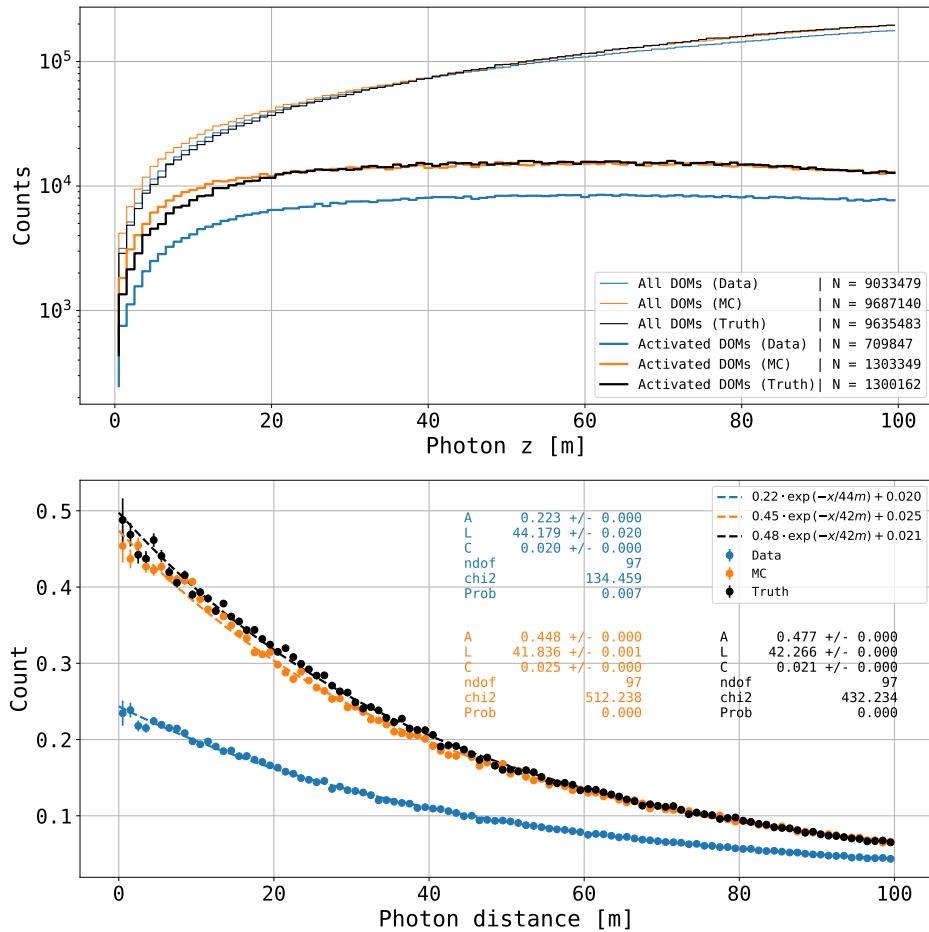


FIGURE 6.19: The top plot shows histograms of the photon distance to all DOMs and activated DOMs respectively, for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot.

see muons with an incident zenith angle well below 90°. Coupled with the constant Cherenkov angle of about 40°, the maximum photon zenith angle we will observe is below 130°. This analysis will therefore not be able to determine the efficiency of those angles. Were this experiment to be done with neutrinos, that absolutely can travel through Earth, we would be able to cover that angular region. The zenith efficiency of MC and truth in figure 6.20 seem to follow the same structure as those in figure 3.10, giving some credibility to the results. Data, on the other hand, has a significantly different structure. Where MC is steadily rising, data rises initially like MC but then plateaus for a large interval before rapidly increasing towards the end.

### 6.6.2 Photon z Coordinate

Many photons will span a long vertical distance, traversing many layers of ice. To consider the efficiency as a function of depth, the vertical movement will only smear the signal. Horizontal photons that only propagate through a slim layer of ice will therefore be considered in this analysis. By selecting horizontal photons where  $|p_{\text{Cherenkov}} - p_{\text{dom}}| < 10\text{m}$ , only photons that span a maximum of 10m vertically are used. Using  $\text{photon\_z}$  instead of  $\text{cherenkov\_z}$  or  $\text{dom\_z}$  the halfway point of the photon trajectory is used, and the efficiency as a function of the z-coordinate can be determined with a resolution of 10m. The number of meters can be

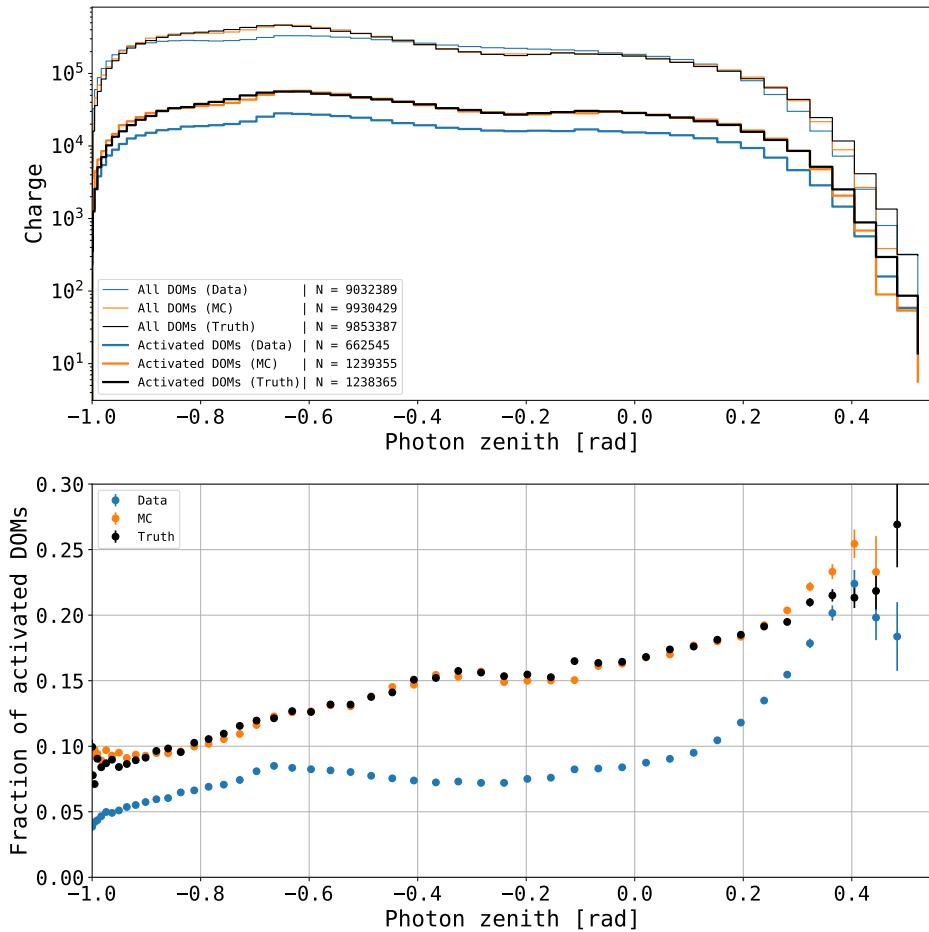


FIGURE 6.20: The top plot shows histograms of the photon distance to all DOMs and activated DOMs respectively, for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot.

adjusted to improve resolution, but that will also reduce the number of photons available for the analysis.

Figure 6.21 shows the efficiency as a function of the z-coordinate. The top plot shows how there is significantly more statistics in the DeepCore part of the detector, as all MC events are simulated to pass through DeepCore, where the detector density is also higher. The amount of Cherenkov radiation is proportional to the energy of the muon, and the muons have the least amount of energy right before they stop. We will therefore only see very low-energy muons in the bottom part of the detector, which will also produce the weakest Cherenkov radiation, resulting in a very low efficiency. There is also low statistics in the top part of the detector because only horizontal photons have been selected. The photons are emitted in a cone, of around 40 degrees around the track. If we then require photons to have a zenith angle of 90°, the muon would have to have a zenith angle of at least  $90^\circ - 40^\circ = 50^\circ$ . Since the muons need to have an angle of at least 50° and still hit the center bottom part, we get more of the "low" muons entering the detector through the sides and therefore miss the top. To put it more succinctly: A muon that passes through DeepCore with the right angle to produce horizontal Cherenkov radiation will at most pass through the upper part of the detector very briefly. In reality, the selected photons can have angles a bit lower than 90°, since we only require them to span a maximum of 10m in the z-direction, but the reasoning still stands.

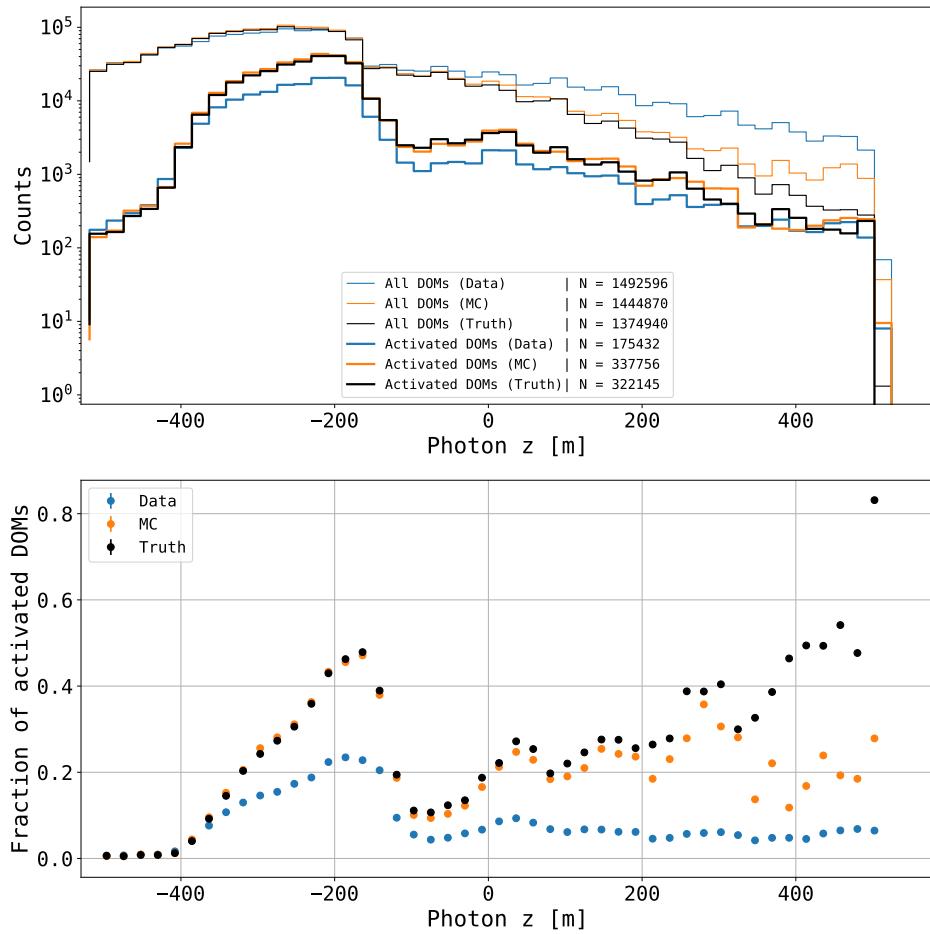


FIGURE 6.21: The top plot shows histograms of the photon  $z$  coordinate for all DOMs and activated DOMs respectively, for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot.

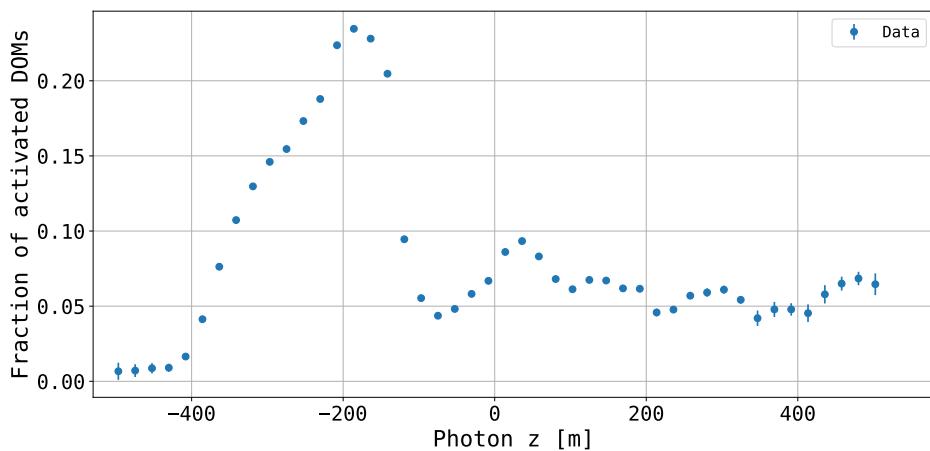


FIGURE 6.22: The efficiency of the detector in data as a function of the photon  $z$  coordinate.

Figure 6.22 shows a high efficiency in the DeepCore region and a dip in efficiency in the

dust layer as expected. The efficiency above 0m varies in a fashion that could be expected from the plotted absorption length in figure 3.8.

### 6.6.3 Photon Azimuth Angle

In section 3.4, the concept of birefringence was introduced along with the notion that the ice had a preferred direction along the axis of flow and that photons would be attenuated less along this axis than perpendicular to it. If this assumption is wrong, we would expect the efficiency to be uniform in azimuth, and if the assumption is right, we would expect the efficiency to be higher in this direction. This effect would reveal itself as a sinusoidal function with a period of  $\pi$  like in figure 3.11, with two peaks; one parallel with the flow direction and one anti-parallel, and two troughs for the two perpendicular directions. Furthermore, because this effect is only newly discovered, and has not been incorporated into the simulations yet, it should only appear in data.

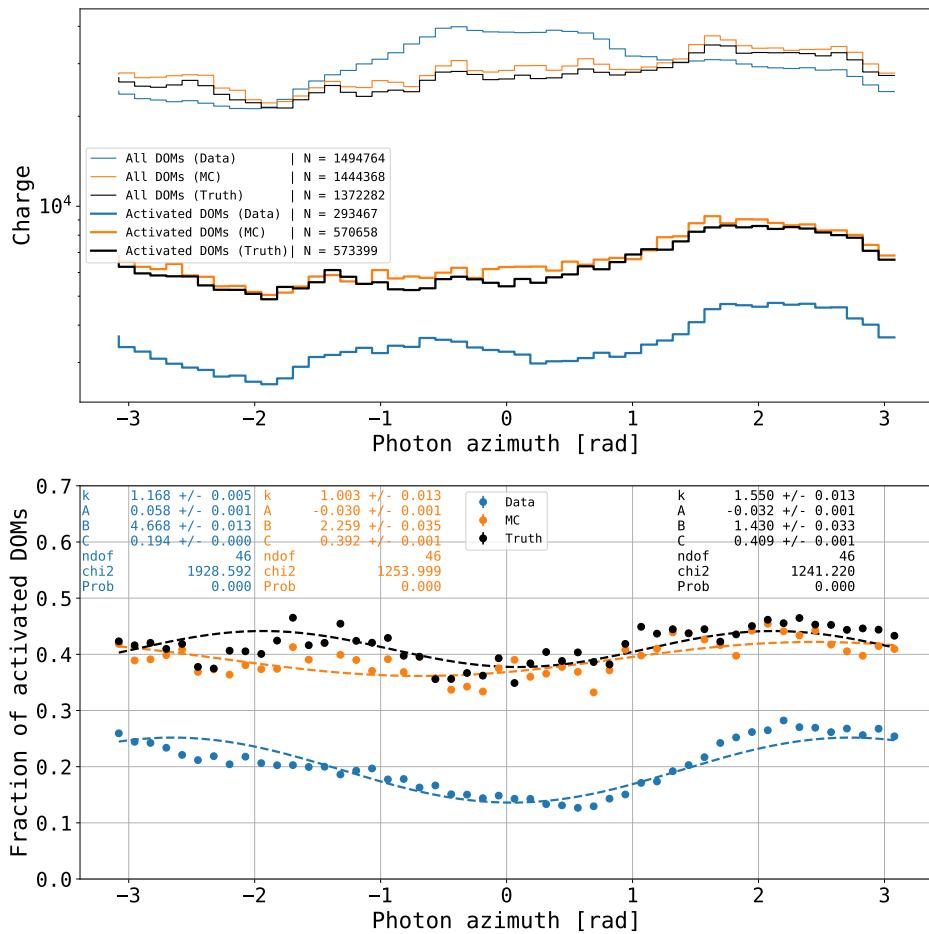


FIGURE 6.23: The top plot shows histograms of the number of DOMs and activated DOMs for data, MC, and truth. The bottom plot shows the ratio of the number of activated DOMs to the number of DOMs for each bin in the top plot.

Figure 6.23 shows the efficiency as a function of the azimuth angle. Because we expected a sinusoidal relationship, all three efficiencies have been fitted with sine functions. The fits have very high  $\chi^2$  values, thus very low probabilities, due to the very low statistical uncertainties on the points. The efficiency might be a product of multiple effects that cannot be modeled by a single sine function. Furthermore, our assumption was that the efficiency would be symmetric

with a phase of  $\pi$ , i.e. the efficiency parallel to any axis would be equal to the efficiency antiparallel to that axis, but that is not reflected in figure 6.23. Just like the fitting of the efficiency as a function of distance, the poor fit could be a result of averaging over too many effects, as opposed to e.g. considering one 10m layer at a time to have a more homogeneous sample and reduce the number of sources of errors.

Figure 3.11 has the charge ratio MC/data on the y-axis, not the ratio of activated DOMs, which we have plotted so far. We are trying to detect an effect that attenuates light based on the azimuth angle, and this attenuation will decrease the recorded charge of the pulse. If we only look at whether the DOM was activated or not, we miss this nuance of how much charge is detected. We will therefore weight each DOM with the sum of the charges of the pulses in the given DOM.

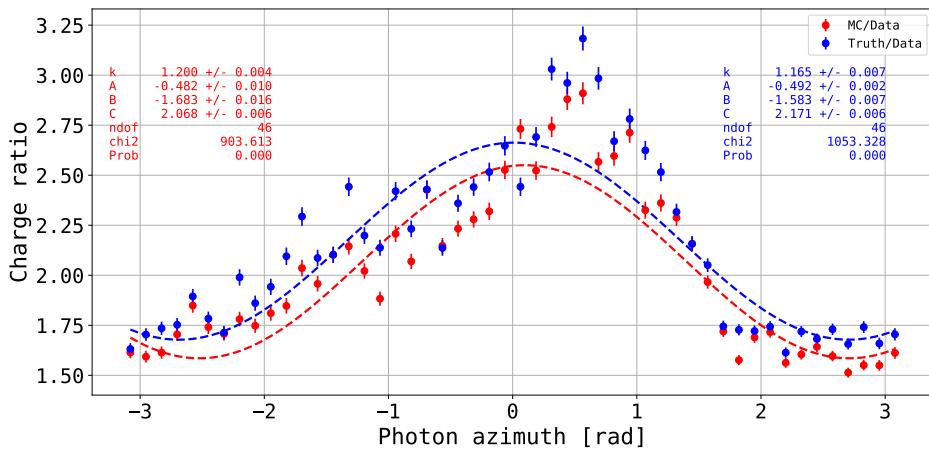


FIGURE 6.24: The ratio of recorded charge for data vs. MC and data vs. truth, as a function of the azimuth angle.

Figure 6.24 shows the ratios of the efficiencies of MC to data and truth to data respectively. From figure 3.11 we expected a sine curve with a period of  $\frac{2\pi}{k} = \pi \implies k = 2$ , but the fits have a  $k$ -parameter just above 1, corresponding to a period of a bit less than  $2\pi$ . The sine fits are very poor, but there could still be a sinusoidal component in the ratios that is simply obscured by some other effects.

Figure 6.24 also has much higher ratios than in figure 3.11 due to the efficiency in data being much lower than in MC, as mentioned earlier. The two figures are not directly comparable, since they measured the light from DOM to DOM, meaning the light would travel approximately 125m or more, while I selected photons to travel shorter than 100m, to get rid of noise. The birefringent effect was shown in figure 3.12, which bends light towards the flow axis. The figure also shows that if the detector is too close to the light source, the effect is smaller, because it has not had enough time to bend the light all the way towards the flow axis. Since figure 3.11 considered photons that travel a greater distance, there would be more time for the birefringent effect to accumulate and bend the light, towards the flow axis.

## 6.7 Hit Timing

In order to compare the timing of multiple pulses in different DOMs in different events, we need a common reference point for them all. One parameter each event has in common is that all photons are emitted from the same track, and we can therefore tie each pulse in an event to the physical track of the muon. The stopping point of the muon is used for this purpose. When we have reconstructed the photons, we know where they travel and when due to the timestamp on each pulse. We also know that the time it takes light to travel a distance  $d$  in the ice is  $dc/\eta$ , and the time it takes a muon to travel a distance  $d$  is  $dc$ . We can therefore calculate backwards to determine when the muon stopped from each pulse. In figure 6.25 the stopping point is labeled  $A$ , and we will refer to the time of stopping as  $t_A$ , which is equal to the time at which the muon arrives at point  $A$ . The calculation of  $t_A$  is shown in the box below.

### $t_A$ Calculation

```

distance_between_emission_and_stopping = |A P_Cherenkov|
time_between_emission_and_stopping = |A P_Cherenkov| ·  $\frac{1}{c}$ 
photon_travel_distance = |P_Cherenkov P_DOM|
photon_travel_time = |P_Cherenkov P_DOM| ·  $\frac{\eta_{ice}}{c}$ 
t_A = DOM_time - photon_travel_time + time_between_emission_and_stopping

```

Each event yields  $n_{pulses}$  values for  $t_A$ , which all should be the same value if not for scattering and uncertainties. The width of the distribution of  $t_A$  reveals the resolution of the timing.

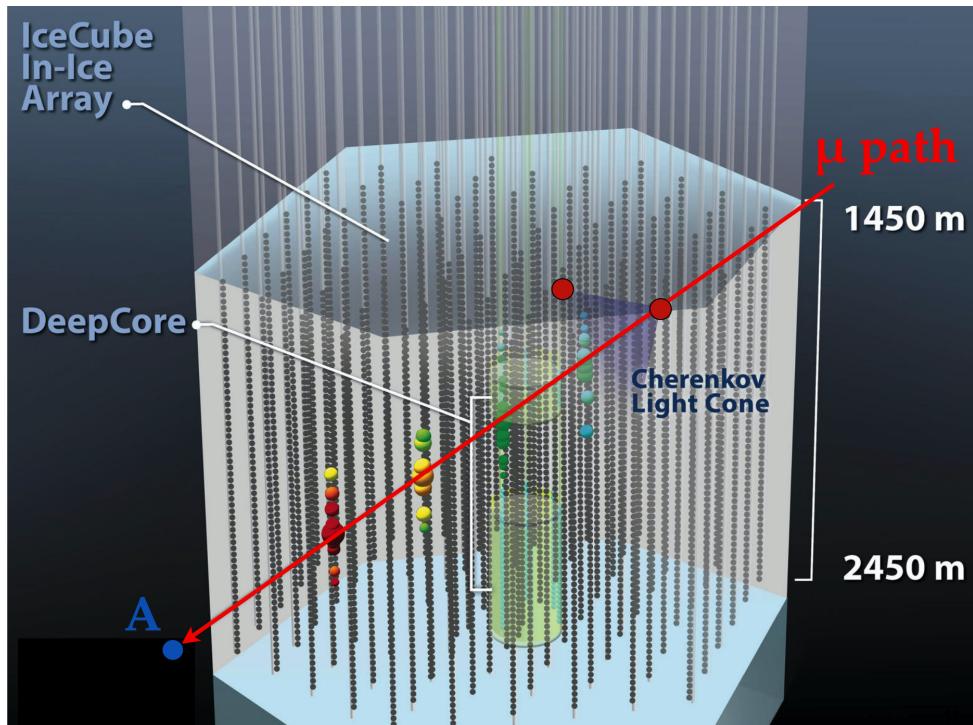


FIGURE 6.25: Figure showing the muon trajectory in the detector and the emitted Cherenkov radiation.

### 6.7.1 Pulse selection

To improve the signal, it was attempted to only select pulse series with a minimum of 15 pulses that were assumed to be direct hits, i.e. little to no scattering. This process is shown in figure 6.26, and was done by taking the median (red dotted line) of the timing of the pulses in the series, and creating a time frame starting 400ns before the median and ending 400ns after. 400ns was chosen as it seemed fitting upon visual inspection. This allows for more relative scattering at shorter distances, and the value could be changed to depend on the distance in further analysis. If a minimum of 15 pulses were in-time within this window, the pulse series was used.

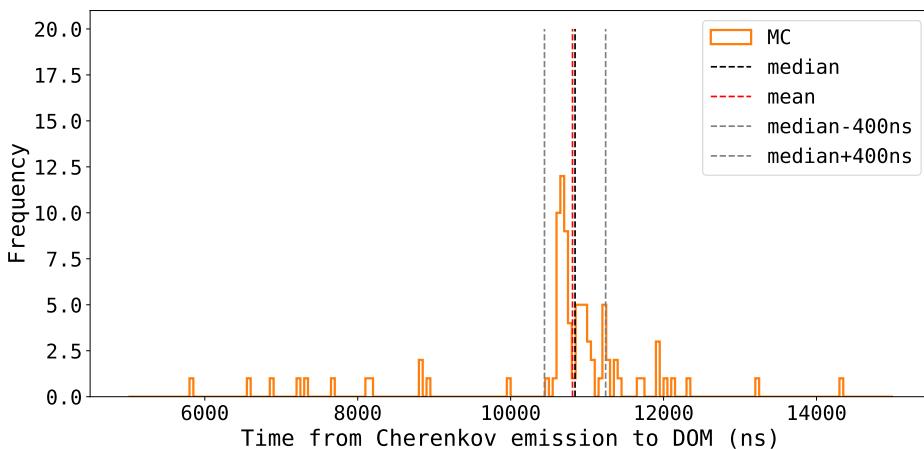


FIGURE 6.26: Histogram of the pulse series, where the median is marked in red, and the 800ns acceptance interval is shown by the grey dotted lines. The mean of the pulses within the interval is marked by the black dotted line.

The mean shown in figure 6.26 is then the time used in that DOM for the calculation of  $t_A$ .  $t_A$  was calculated from all the selected pulses and concatenated. The histogram is shown in figure 6.30, where all the pulse series have had their means subtracted, such that they can be compared.

This method of selecting suitable pulses is not perfect. While MuonGun simulates the muons separately, data might contain overlapping simultaneous events, and the median and mean of the pulse series might lie in between two separate pulse series that have been combined to one pulse series if they are recorded sufficiently close in time, as shown in figure 6.27.

The mean or median might therefore lie in between the two separate pulse series, and even though there are plenty of suitable pulses, they will not be registered. Figure 6.28 shows that the distribution of mean time of the signal is much wider for data than for MC, which could be caused by the overlapping events. Overlapping events could also extend the duration of an event such that more noise is recorded before or after the signal, causing the median to not coincide with the signal.

Figure 6.29 shows the distributions of all pulses and those that are in-time for data and MC. MC has more pulses initially and even more after the selection. The bump in the data distribution around 15-20 indicates that the distribution is a combination of two distributions, where the left part is the pulses where the signal was not within the interval, and the right part is where the signal was within the interval. If there are only a few pulses in the interval, it is very possible that it is just noise. If the data events did not have this effect, one could expect the blue distribution to undergo the same transformation as the orange one.

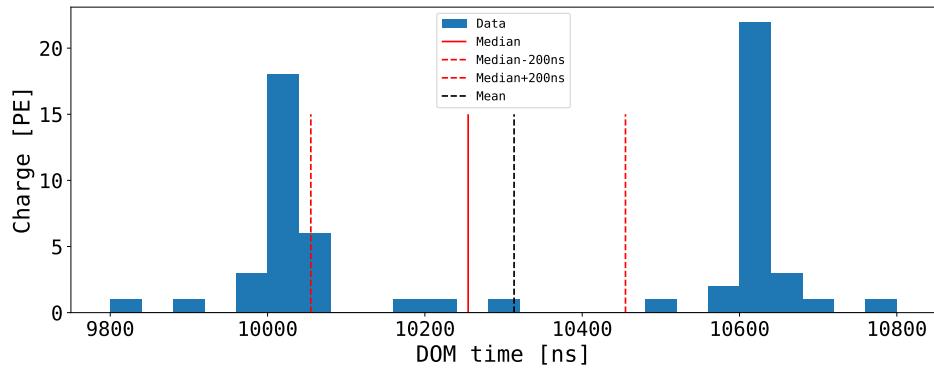


FIGURE 6.27: Fabricated example a pulse series containing two peaks, where the median of the series is located in between the peak, such that no signal is within the acceptance window.

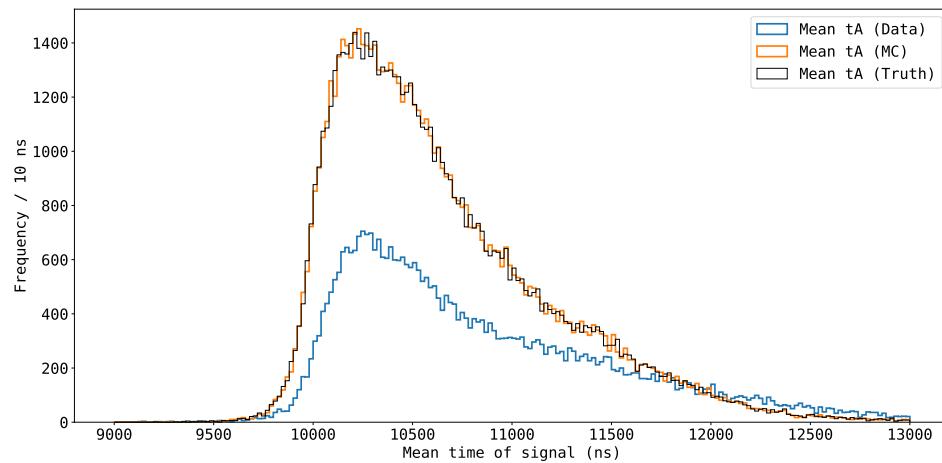


FIGURE 6.28: Histograms of the mean  $t_A$  for data, MC, and truth.

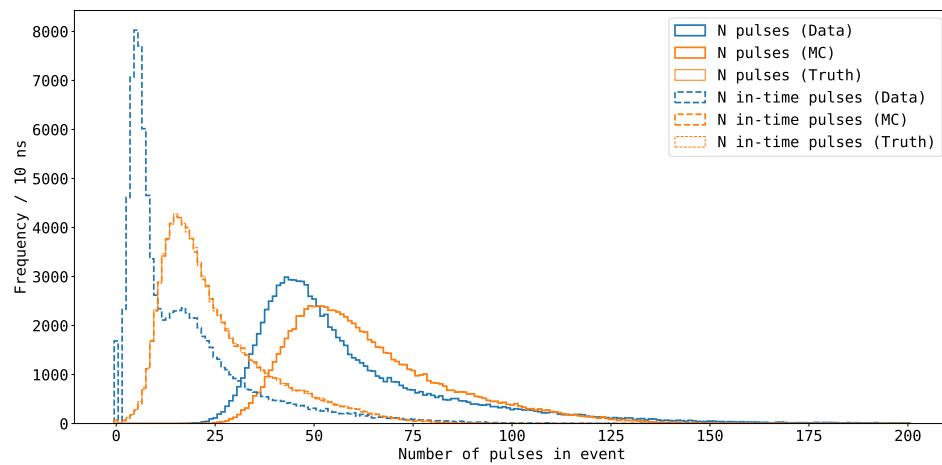


FIGURE 6.29: Histograms showing the distributions of all pulses and those that are in the allowed time interval for data

We showed in the previous section that the efficiency of a DOM decreased exponentially,

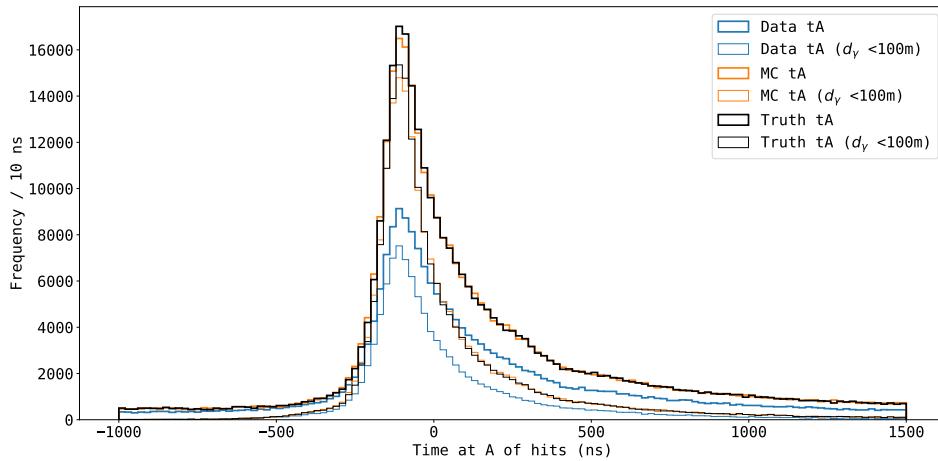


FIGURE 6.30: Histograms showing  $t_A$  for MC, data, and truth when applying no restriction, and selecting photons traveling less than 100m respectively.

and we therefore expect most data with photon distance greater than 100m to be noise. Figure 6.30 shows how selecting only photons traveling less than 100m reduces the background significantly. The right side of the distribution is reduced, because those are pulses that are delayed compared to 0, and the amount of scattering a photon undergoes is correlated with its propagation length. The left side looks relatively unchanged because the 100m+ photons are very unlikely to be early due to the increased scattering. The constant amount of pulses before  $-500\text{ ns}$  before cleaning shows how much background there is. This background is almost completely removed by the 100m criterion, and the ratio of signal to background has been increased significantly.

The stopping times for data and MC are calculated from the predictions on the angles and coordinates, which of course carries some uncertainty. Therefore they are plotted along with the black lines, which are calculated from the true values and therefore have 0 prediction uncertainty. The overlapping of the distributions of the times from predicted values and those from true values shows that the uncertainty in the reconstruction does not change the shape of the distribution very much.

## 6.7.2 Resolution Of The Timing

Ideally, the histogram of  $t_A$  should just be one sharp peak around 0 (because the mean has been subtracted from all pulse series, but the distribution has some width due to the timing error and scattering. There is also some asymmetry due to scattering, causing pulses to be delayed, stretching the distribution to the right. The histograms in figures 6.31, 6.32, and 6.33 have therefore been fitted with the **CrystalBall (CB) function** [61]

$$f_{CB}(x; \alpha, n, \bar{x}, \sigma) = N \begin{cases} \exp -\frac{(x-\bar{x})^2}{2\sigma^2} & \text{for } \frac{(x-\bar{x})^2}{\sigma^2} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n} & \text{for } \frac{(x-\bar{x})^2}{\sigma^2} \leq -\alpha \end{cases} \quad (6.15)$$

where the  $A, B$  variables are functions of  $\alpha$  and  $n$ . The CB function is a Gaussian in the interval  $]-\infty; -\alpha[$  and a potential function in  $[-\alpha; \infty[$ , such that the function and its first derivative is continuous. This makes it a suited function if we assume the timing has a Gaussian error, and the scattering component is potentially decreasing. Assuming not all background was removed by the 100m cut, a constant term is added to the CB function.

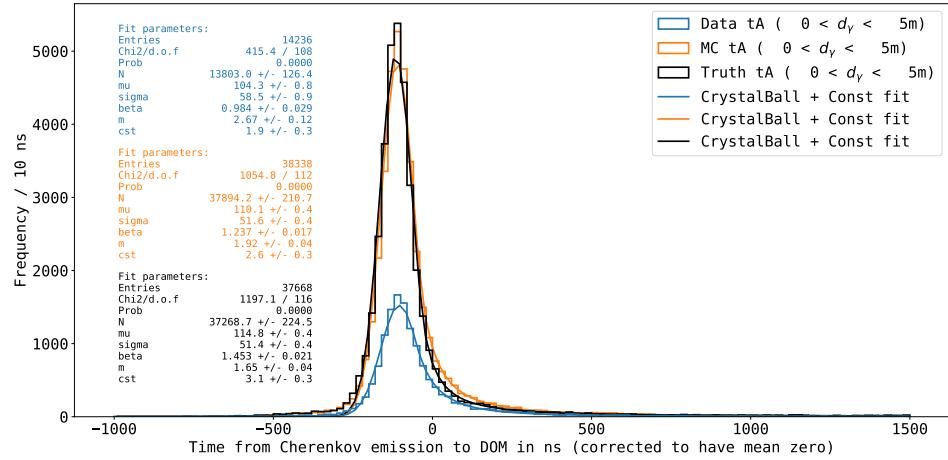


FIGURE 6.31: The distribution of  $t_A$  for photons traveling between 0m and 5m, fitted with a CrystalBall function.

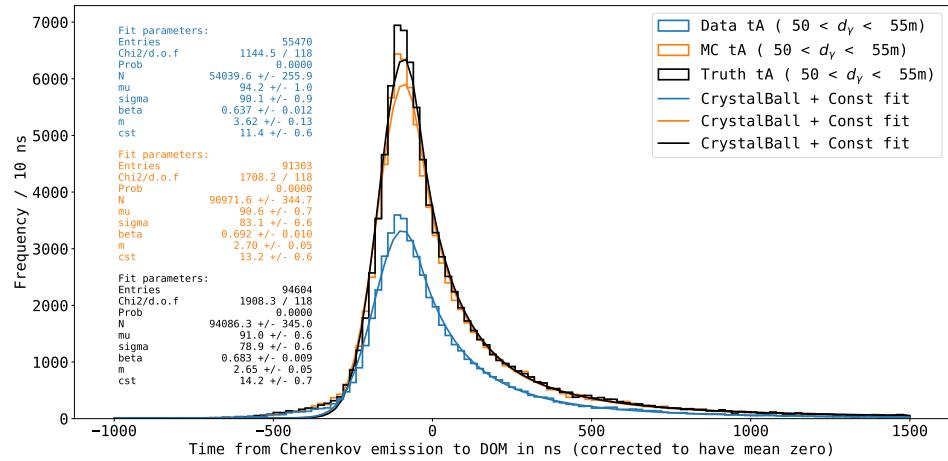


FIGURE 6.32: The distribution of  $t_A$  for photons traveling between 50m and 55m, fitted with a CrystalBall function.

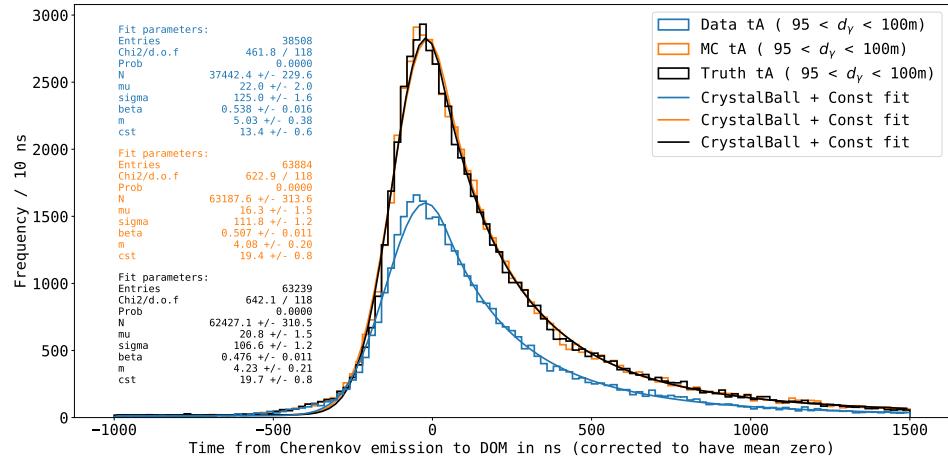


FIGURE 6.33: The distribution of  $t_A$  for photons traveling between 95m and 100m, fitted with a CrystalBall function.

The fits are seemingly decent, although the p-value is 0.0000, partly because the fit does not account for the prepulses that make up the small tail on the left side.

The width of the fitted distributions is the resolution of the timing, which increases with the distance as expected since the light has to travel farther, increasing the amount of scattering. If we plot the width of the distributions for each interval as a function of distance, we can describe how the resolution varies with distance.

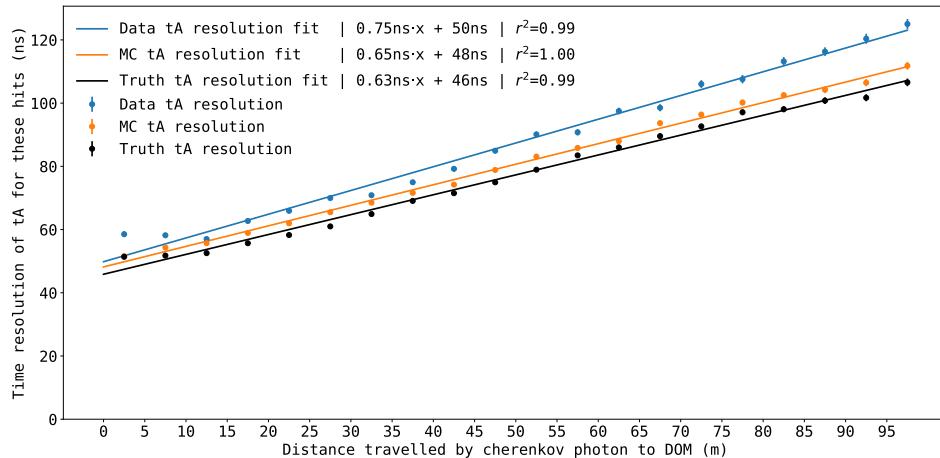


FIGURE 6.34: Plot of the resolution of the timing for data, MC, and MC truth, as a function of photon travel distance.

Figure 6.34 shows the resolution for every 5m interval. The resolutions seem to scale linearly, and all three resolutions have been fitted with linear functions where  $r^2 > 0.98$ , showing that the functions fit the points well, except for the resolution in data decreasing in the first three points. The intercept represents the resolution of the DOMs when the distance is equal to zero, which then necessarily can not be due to effects from the ice. The intercepts of the three lines match with our expectation that truth is lowest because there is no error from reconstruction, MC has some error from reconstruction, and data has a slightly higher error from reconstruction because the reconstruction is made from a model trained on MC events, not data.

The only difference in resolution between MC and truth must be due to the reconstruction being on predictions and truth respectively, so by subtracting the two, we get the uncertainty due to reconstruction. Assuming this effect is similar in data, we can subtract this effect from both MC and data.

The effects of the reconstruction are shown separately in figure 6.35. It is trending slightly upwards, because the absolute error on the length of the reconstructed photon vector increases with the length of the vector. The increase in resolution as the photons travel further slightly increases due to reconstruction, but mainly due to the ice.

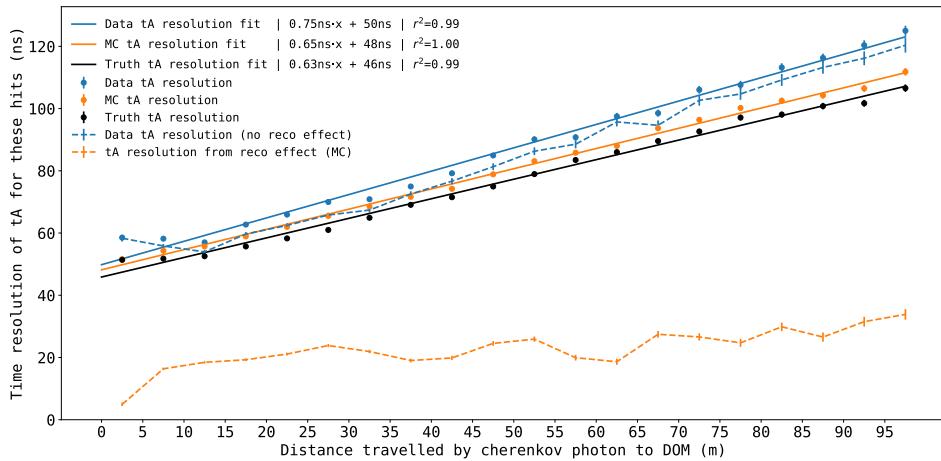


FIGURE 6.35: Plot of the resolution of the timing for data, MC, and MC truth, as a function of photon travel distance. The effects of the reconstruction are also plotted.

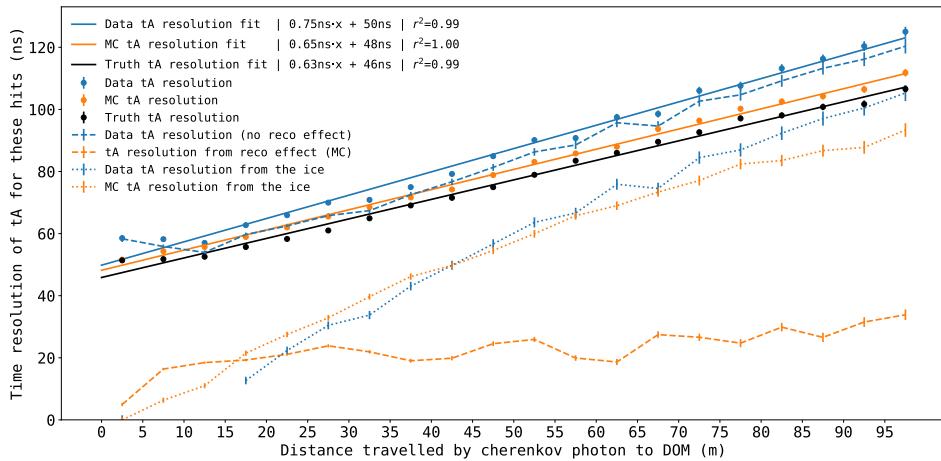


FIGURE 6.36: Plot of the resolution of the timing for data, MC, and MC truth, as a function of photon travel distance. The effects of the ice and the reconstruction are also plotted.

After subtracting the effects of the reconstruction and the ice, the dominating contributors to the resolution should be how the algorithm illustrated in figure 6.26 decides which dom\_time to use, and the timing uncertainty on the DOMs which was explained earlier to be 3.3ns and 8ns for HLC and SLC hits respectively. The resolution shown in figure 6.36 is still higher than expected from these two contributors, which could be explained by an offset in the timing between the DOMs, or something else entirely.

## Chapter 7

# Conclusion And Outlook

### 7.1 Conclusion

Starting with 62 million raw data events containing particles and noise events, the muons were first classified by Peter Andresen's [52] DynEdge implementation of the novel machine learning algorithm of graph neural networks. With a similar DynEdge model, I classified the stopped muons from that muon sample. The number of events per selection level is shown in table 7.1.

Selection	Number of events	AUC
None	61.7m	N/A
Muons	34.0m	0.994
Stopped muons	14.6m	0.991

TABLE 7.1: The number of events before and after both selections.

Three additional models were trained to predict the stopping position and the zenith and azimuth angle of the stopped muons.

Reconstructed variable	Resolution
Azimuth angle	0.211 rad
Zenith angle	0.054 rad
Stopping position x	19.8 m
Stopping position y	20.4 m
Stopping position z	19.0 m
Stopping position	21.5 m

TABLE 7.2: The resolutions of all the reconstructed variables.

These reconstructed muons were the basis of the analysis of this thesis. I developed an algorithm to merge pulses if there was reason to suspect they might be mistakenly split, in an attempt to restore the original signal. This algorithm improved the MC/data agreement in the charge distribution of HLC hits, while SLC hits were largely unchanged. The algorithm also improved the MC/data agreement in the number of pulses per DOM.

Another algorithm was developed to reconstruct the photons emitted through Cherenkov radiation from the track of the stopped muon. For each DOM, the point of photon emission, the photon travel distance, and the zenith and azimuth angles of the photon were calculated. This made it possible to calculate the efficiency of the detector. The efficiency's dependence on photon distance was fitted to an exponential function for both MC and data from which the absorption lengths of photons in the ice were determined to be 44m for data and 42m for MC, with great statistical precision. It must be noted that the detector is inhomogeneous, and the

absorption length describes an average of the entire detector. The fitted functions are shown in table 7.3.

Set	Fit
Data	$E(x) = 0.22 \exp -\frac{x}{44m} + 0.020$
MC	$E(x) = 0.45 \exp -\frac{x}{42m} + 0.025$
Truth	$E(x) = 0.48 \exp -\frac{x}{42m} + 0.021$

TABLE 7.3: The efficiency as a function of photon travel distance for data, MC, and truth respectively.

Furthermore, the azimuthal efficiency was found to be anisotropic, but the ratio of the efficiencies for MC and data did not correspond with our expectations, that it would be sinusoidal with a period of  $\pi$ , as in 3.11.

The efficiency as a function of zenith was calculated and was found to follow a different curve in data compared to MC. Because muons only come from above, and Cherenkov radiation has an obtuse angle, the poorly understood efficiency of upward-going photons could not be determined.

Lastly, the timing of the detector was investigated by calculating the expected stopping time from all pulses by taking their detection time, subtracting the photon's propagation time, and adding the muon propagation time from the point of photon emission to the stopping point. Only pulse series with at least 15 pulses within  $\pm 400$ ns of the median timing were used, to remove noise. All the accepted pulse series were divided by photon propagation distance in bins of 5m and fitted with a CrystalBall function. The width of these fits revealed the timing resolution of the detector. The widths were then fitted with a linear function describing the resolution as a function of photon propagation distance. The results of the fitting procedure are seen in the table below.

Set	Resolution fit
Data	$\sigma(x) = 0.75 \frac{\text{ns}}{\text{m}} \cdot x + 50\text{ns}$
MC	$\sigma(x) = 0.65 \frac{\text{ns}}{\text{m}} \cdot x + 48\text{ns}$
Truth	$\sigma(x) = 0.63 \frac{\text{ns}}{\text{m}} \cdot x + 46\text{ns}$

TABLE 7.4: The resolution as a function of photon travel distance for data, MC, and truth respectively.

The intercept of the linear fits shows the resolution when photons travel 0m through the ice, where the contribution from the ice would be 0. Thus the effect on resolution from the ice could be calculated and subtracted from the total resolution. By comparing MC to MC truth, the contribution to the resolution by the imperfect reconstruction could also be calculated and subtracted from the resolution.

Finally, it was found that even without these two effects, that were expected to be dominant, the resolution was still quite high. Then the largest expected contributors were the offsets in the DOM clocks and the algorithm with which the time of the direct hit of each DOM was chosen. Nevertheless, the uncertainty of the two dominant remaining contributors is expected to be well below the observed resolution.

## 7.2 Outlook

The analysis of this thesis has a broad scope, focusing more on doing multiple things well, than doing one thing perfectly. There is therefore plenty of potential for improvements.

### 7.2.1 Pulsemerging

The pulse merging algorithm is very simple in that it only considers one variable and the optimal merging parameters might demand the involvement of other parameters. Some of the possible causes of the additional pulses are fairly well known by IceCube, to the point where they know the expected time the effect might happen compared to the actual pulses. These expected times are not currently included in PulseMerger but could be a possible improvement.

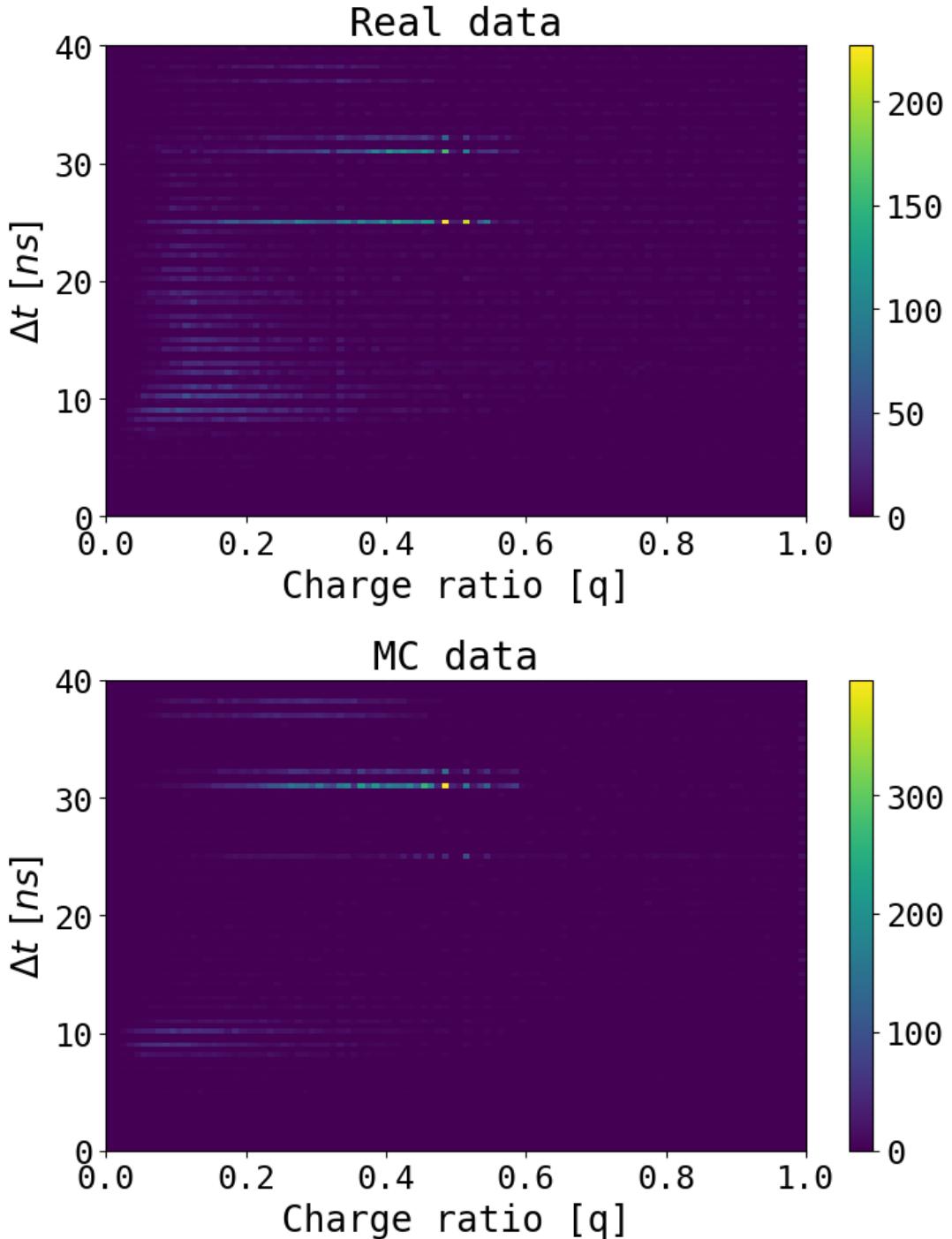


FIGURE 7.1: 2D histogram of the time difference and charge ratio between pulses in a DOM during an event for data and MC.

Figure 7.1 shows a 2-dimensional histogram for the time difference and charge ratio between any consecutive pair of pulses in a pulse series in one DOM during an event. The figure shows that there is a difference in data and simulation, e.g. the more pronounced peak around 25ns for data. While PulseMerger did not manage to fix this discrepancy, it can serve as inspiration for a potential improvement of the algorithm. More plots highlighting disagreements between MC and data can be found in Appendix A.

The ratio plots of the charge distributions are just one way of gauging MC/data agreement across the two sets. A more comprehensive test is seeing how well DynEdge can distinguish between the two. This approach also is the most logical, as it is DynEdge that is going to process the data to make predictions. In this case, one should train a model to predict whether a given event is simulation or data and note the AUC. Then an identical model should be trained on the same training data after the Pulsemerger algorithm has been applied to see if the AUC has improved. Another similar test would be to train classification and regression models on the original and merged pulsemaps, to see which produces the best results.

Lastly, the algorithm can be altered to a coarsening version, where hits are split into bins of increasing width e.g. [4,8,16,32,64,128,256,512]ns, summing the charge in each bin. This coarsening would reduce the amount of data to make the algorithm faster, and potentially improve MC/data agreement at the expense of removing some timing information.

## 7.2.2 Photon Vector Reconstruction

Because the points  $P_{DOCA}$  and  $P_{Cherenkov}$  are found by scaling  $\overline{BA}$ , they do not have the restriction of necessarily being on  $\overline{BA}$ . That is, in eq. 6.2, if  $t$  is negative, it corresponds to going along from  $A$  away from  $B$  in the direction dictated by  $\overline{BA}$ . That is not a problem, but if  $P_{DOCA}$  is so far past the  $A$  that  $P_{Cherenkov}$  also lies beyond  $A$ , it means that the light has been scattered, as it could not have hit the DOM if it maintained its initial trajectory. The algorithm to calculate these vectors should automatically exclude the vectors where this is the case.

## 7.2.3 Efficiency

The efficiency of the entire detector as a function of distance was fitted with an exponential function, that yielded a value for the absorption length, and it was explained that the absorption actually varies greatly in different parts of the detector. Therefore it would make sense to select only the horizontal photons, as done when calculating the azimuth and depth efficiency, and calculate the efficiency of each e.g. 10m layer. Then it would be possible to recreate figure 3.8, by plotting the absorption length for each layer.

The pulse selection algorithm used in the hit timing analysis was developed late in the writing of this thesis, and it was therefore not possible to implement it in the efficiency analysis due to time. This algorithm would remove noise and improve the signal, as it has in the timing analysis.

It was discussed in section 6.6 that the birefringent effect might not have appeared in figure 6.24 because the photons used in the analysis had the constraint that they must travel less than 100m, where figure 3.11 had the constraint that the minimum propagation distance was around 125. It was suggested that because the light traveled further, the birefringent effect would have more time to accumulate. In further analysis, the ratio of azimuthal efficiency in data and MC would be calculated for longer propagation distances, to see if the effect appears. If it does, one could even describe the magnitude of the birefringent effect as a function of propagation distance, which would be very useful to know when designing the simulation.

## 7.2.4 Scattering

In figure 6.30 we attributed the asymmetry of the distribution to the scattering of the photons, assuming scattering to be an effect that delays photons. Then we can also assume that without scattering, the distribution would be symmetric and that scattering only affects the right-hand side ( $t > t_{exp}$ ). If we replace the right-hand side with a mirror image of the left-hand side and subtract it from the original distribution, we are left with only the scattered photons.

## 7.2.5 Hit Timing

The pulse series selection algorithm was shown to work poorly, as the ratio of accepted pulses was significantly lower for data than MC due to overlapping events in data. This effect can easily be circumvented by using kernel density estimation (KDE) to determine where along the time-axis the density of charge is highest. Figure 7.2 shows how KDE accurately determines the times with the highest charge density, which is the signal we want, and how the other method misses both peaks.

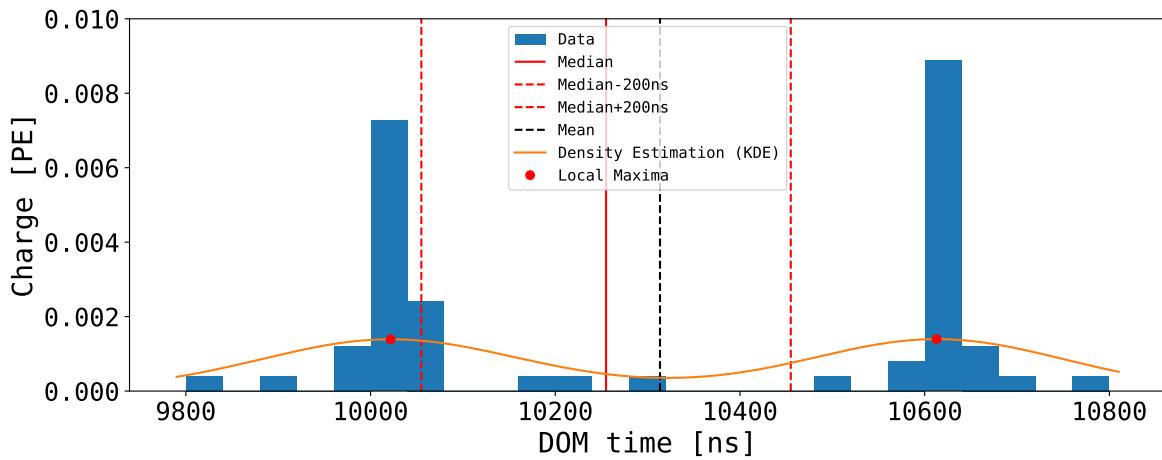


FIGURE 7.2: Plot showing the charges and timings of pulses in a DOM. The mean, median, and time interval are shown by the vertical line

Stopped muons, as opposed to through muons, were used for their predictability in this analysis. For the calculations of the timing of the pulses, a criterion that improved the signal was that at least 15 direct pulses were in the pulse series, and more pulses in general improved the signal. Through muons generally have more energy than stopped muons, causing them to produce more pulses, and therefore have more pulses per event, and they would then have more pulse series that have at least 15 direct hits. Through muons do not have a stopping point such that  $t_A$  can be calculated, but a predicted exit point could be used instead. Leon Bozianu [39] also defined the edges of the detector in terms of the IceCube coordinate system, such that one could determine when a muon was contained by the detector volume. The exit point could then be defined as the point where the muon leaves this volume. The exit point would furthermore always have one coordinate fixed, depending on which surface of the detector it leaves through, e.g. the bottom of the detector would have a constant z-coordinate, and the exit point would just vary in the x-y plane. This would also minimize uncertainty on the stopping position.

## Appendix A

# Cut-Plots

In the early stages of making this thesis, exploratory research was done to identify areas where data and MC would be different. Some of these findings did not need to be included in the thesis, but are nevertheless presented here.

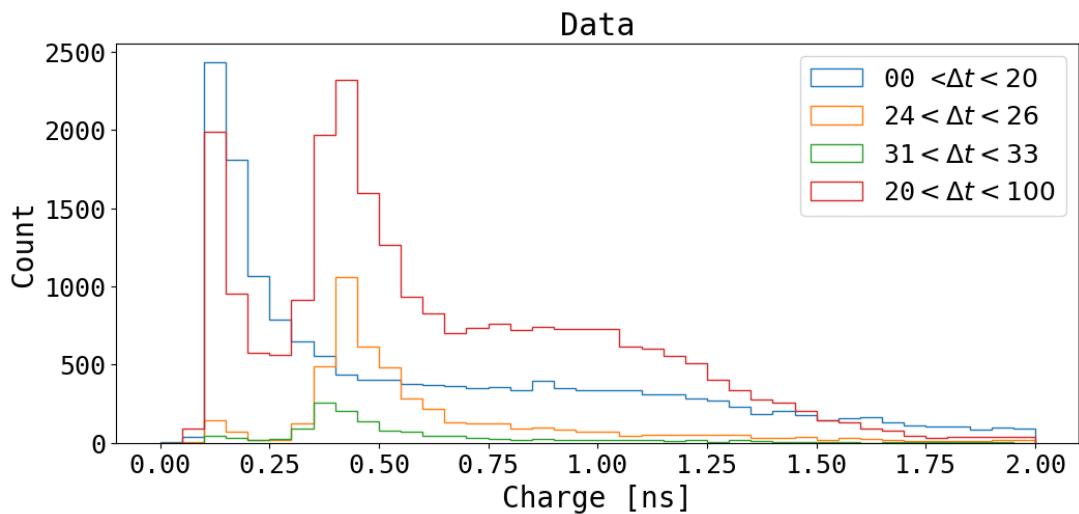


FIGURE A.1: The charge distribution of the first two pulses in a DOM, divided by the time difference between the two pulses for data.

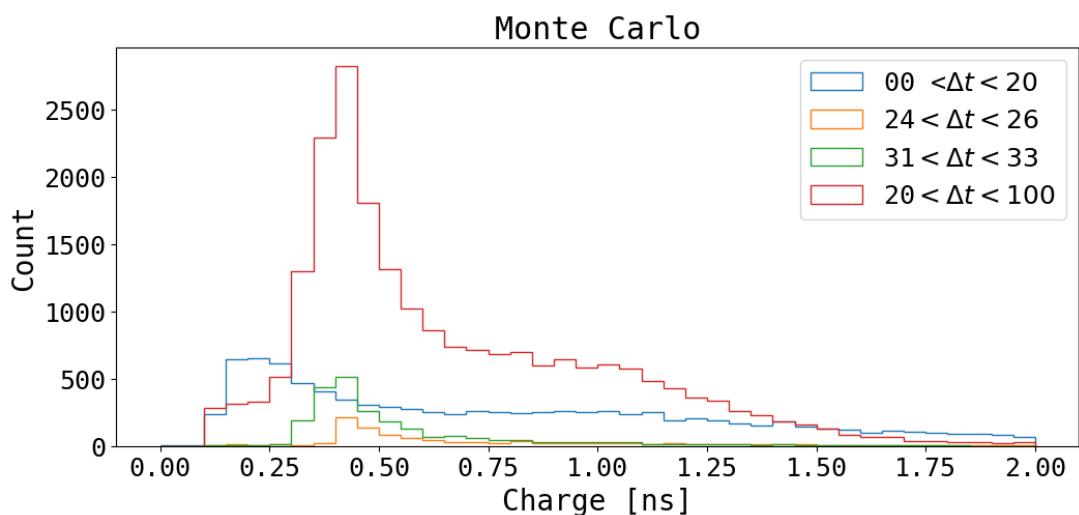


FIGURE A.2: The charge distribution of the first two pulses in a DOM, divided by the time difference between the two pulses for MC

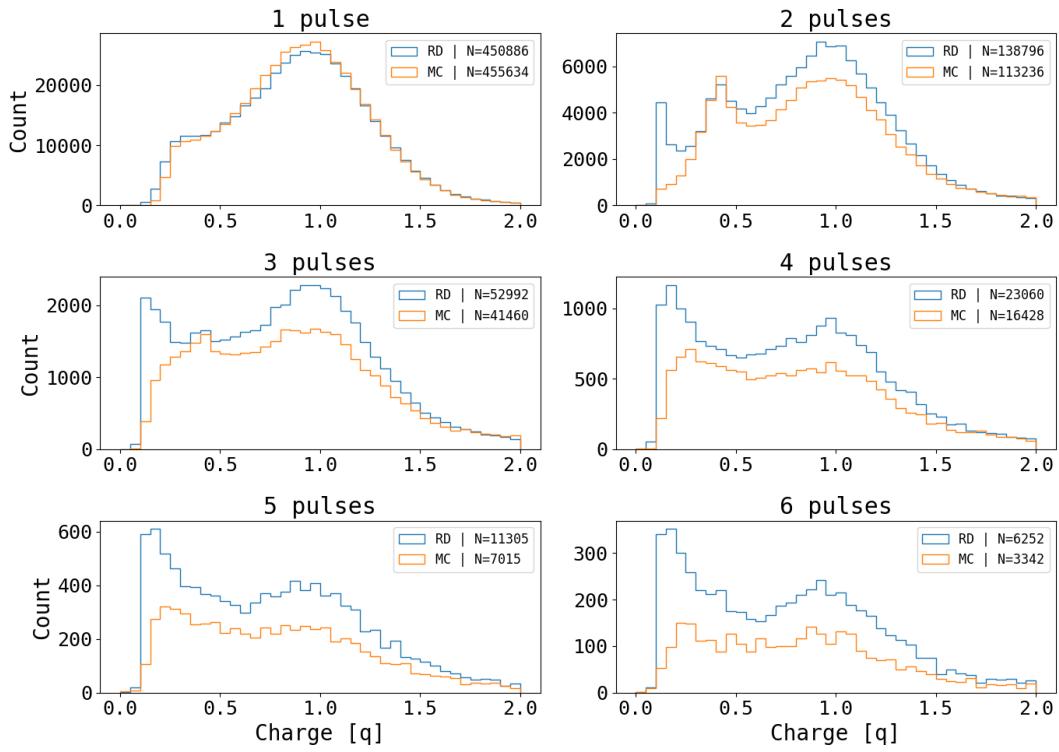


FIGURE A.3: The charge distribution for pulse series, divided into how many pulses are in the pulse series.

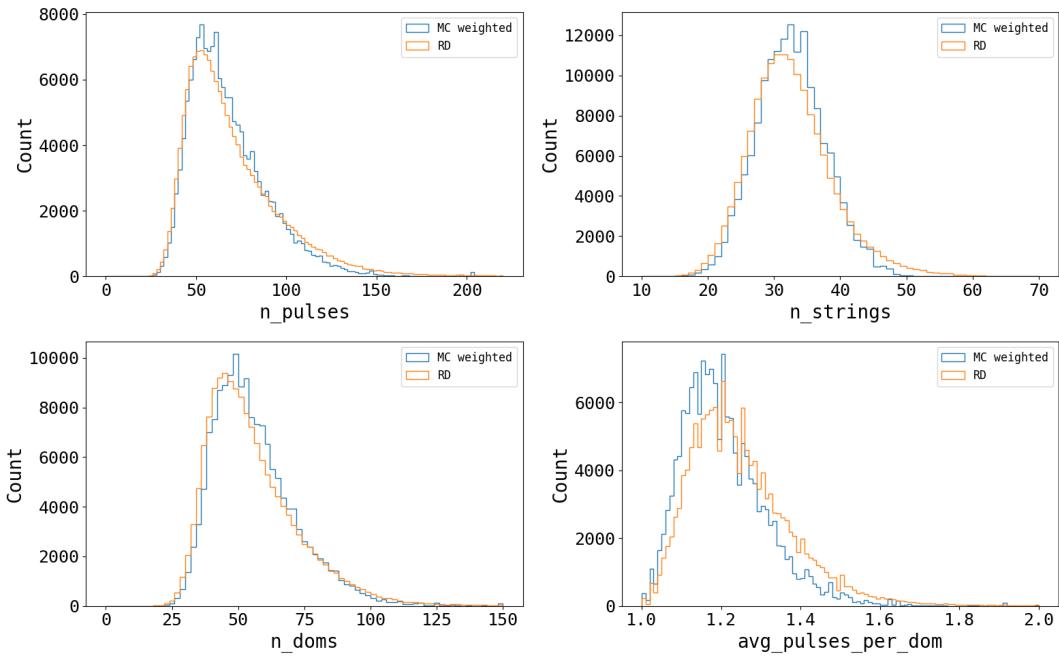


FIGURE A.4: The distribution of the number of pulses, number of strings, number of DOMs activated, and the average number of pulses per DOM in MC and data.

## A.1 Systematic Monte Carlo Sets

As the simulations are based on many parameters that are not well known, many simulations are run with slightly different parameters, such that slight variations can be tested, to see which fits data better. The so called systematic simulations, are a collection of 19 data sets that are simulated identically, except for a tweak of certain parameters, such as how much the light scatters in the ice. Any and all analysis could be done on these sets to compare.

Set Number	Scat	Abs	Ice Model Used
1x0500	+5%	+5%	spice 3.2.1/err s+.05 a+.05
1x0501	+5%	-5%	spice 3.2.1/err s+.05 a-.05
1x0502	-5%	+5%	spice 3.2.1/err s-.05 a+.05
1x0503	-5%	-5%	spice 3.2.1/err s-.05 a-.05
1x0504	-	+10%	spice 3.2.1/err s+.00 a+.10
1x0505	-	-10%	spice 3.2.1/err s+.00 a-.10
1x0506	+10%	-	spice 3.2.1/err s+.10 a+.00
1x0507	-10%	-	spice 3.2.1/err s-.10 a+.00
1x0508	-	+30%	spice 3.2.1/err s+.00 a+.30
1x0509	-	-30%	spice 3.2.1/err s+.00 a-.30
1x0510	+30%	-	spice 3.2.1/err s+.30 a+.00
1x0511	-30%	-	spice 3.2.1/err s-.30 a+.00
1x0512	-	+20%	spice 3.2.1/err s+.00 a+.20
1x0513	-	-20%	spice 3.2.1/err s+.00 a-.20
1x0514	+20%	-	spice 3.2.1/err s+.20 a+.00
1x0515	-20%	-	spice 3.2.1/err s-.20 a+.00
1x0516	-	+5%	spice 3.2.1/err s+.00 a+.05
1x0517	-	-5%	spice 3.2.1/err s+.00 a-.05
1x0518	+5%	-	spice 3.2.1/err s+.05 a+.00
1x0519	-5%	-	spice 3.2.1/err s-.05 a+.00

TABLE A.1: Table showing the different systematic sets, and which parameters have been changed. Table from [48]

# Bibliography

- [1] IceCube Collaboration et al. "The IceCube Neutrino Observatory - Contributions to ICRC 2017 Part V: Solar flares, Supernovae, Event reconstruction, Education Outreach". In: (Oct. 2017) (cit. on p. 1).
- [2] R. Abbasi et al. "Graph Neural Networks for Low-Energy Event Classification & Reconstruction in IceCube". In: 17.11 (Nov. 2022), P11003. DOI: [10.1088/1748-0221/17/11/p11003](https://doi.org/10.1088/1748-0221/17/11/p11003). URL: <https://doi.org/10.1088%2F1748-0221%2F17%2F11%2Fp11003> (cit. on pp. 2, 30, 38, 39).
- [3] The IceCube Webpage. *Machine learning method improves reconstruction and classification of low-energy IceCube events.* <https://icecube.wisc.edu/news/research/2022/11/machine-learning-method-improves-reconstruction-and-classification-of-low-energy-icecube-events/>. Accessed: 13-02-2023 (cit. on pp. 2, 30).
- [4] B. R. Martin and G. Shaw. *Nuclear and particle physics: an introduction*. Third edition. Hoboken, NJ: Wiley, 2019. ISBN: 978-1-119-34462-9 978-1-119-34461-2 (cit. on pp. 3–6).
- [5] Wikipedia. *Standard Model*. [https://en.wikipedia.org/wiki/Standard\\_Model](https://en.wikipedia.org/wiki/Standard_Model). Accessed: 09-08-2023 (cit. on p. 3).
- [6] Troels C. Petersen. *Higgs-partiklen*. <https://fysikleksikon.nbi.ku.dk/h/higgs/>. Accessed: 30-07-2023 (cit. on p. 4).
- [7] Wolfgang Pauli. *Pauli's letter to Meitner*. <https://cds.cern.ch/record/83282?ln=en>. Accessed: 26-07-2023. Dec. 1930 (cit. on p. 4).
- [8] C. L. Cowan et al. "Detection of the Free Neutrino: a Confirmation". In: *Science* 124.3212 (1956), pp. 103–104. DOI: [10.1126/science.124.3212.103](https://doi.org/10.1126/science.124.3212.103). eprint: <https://www.science.org/doi/pdf/10.1126/science.124.3212.103>. URL: <https://www.science.org/doi/abs/10.1126/science.124.3212.103> (cit. on p. 4).
- [9] John N. Bahcall and R. Davis. "Solar Neutrinos: A Scientific Puzzle". In: "Science" 191 (1976), pp. 264–267. DOI: [10.1126/science.191.4224.264](https://doi.org/10.1126/science.191.4224.264). URL: <https://www.science.org/doi/10.1126/science.191.4224.264> (cit. on p. 5).
- [10] The IceCube collaboration. *IceCube quick facts*. <https://icecube.wisc.edu/about-us/facts/>. Accessed: 04-07-2023 (cit. on pp. 5, 28).
- [11] The Niels Bohr Institute. *Neutrino Oscillation*. <https://nbi.ku.dk/english/research/experimental-particle-physics/icecube/neutrino-oscillation/>. Accessed: 06-04-2023 (cit. on pp. 5, 7).
- [12] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. "Remarks on the unified model of elementary particles". In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880. DOI: [10.1143/PTP.28.870](https://doi.org/10.1143/PTP.28.870) (cit. on p. 5).
- [13] MICHAEL F. L'ANNUNZIATA. "9 - CHERENKOV COUNTING". In: *Handbook of Radioactivity Analysis (Second Edition)*. Ed. by Michael F. L'Annunziata. Second Edition. San Diego: Academic Press, 2003, pp. 719–797. ISBN: 978-0-12-436603-9. DOI: <https://doi.org/10.1016/B978-012436603-9/50014-4>. URL: <https://www.sciencedirect.com/science/article/pii/B9780124366039500144> (cit. on p. 7).

- [14] P.B. Price and K. Woschnagg. "Role of group and phase velocity in high-energy neutrino observatories". In: *Astroparticle Physics* 15.1 (2001), pp. 97–100. ISSN: 0927-6505. DOI: [https://doi.org/10.1016/S0927-6505\(00\)00142-0](https://doi.org/10.1016/S0927-6505(00)00142-0). URL: <https://www.sciencedirect.com/science/article/pii/S0927650500001420> (cit. on p. 7).
- [15] IceCube Collaboration. "The response of an autonomous underwater telescope to high energy neutrinos for the observation of Gamma-ray bursts (GRBs)". In: *HNPS Advances in Nuclear Physics* (Mar. 2019). Accessed: 06-04-2023. DOI: <10.12681/hnps.1936> (cit. on p. 9).
- [16] IceCube collaboration. *IceCube und Neutrinos*. <https://masterclass.icecube.wisc.edu/de/learn-ger/neutrinonachweis-mit-icecube>. Accessed: 14-08-2023 (cit. on p. 9).
- [17] P. Kyberd et al. *nuSTORM: Neutrinos from STORed Muons*. 2012. arXiv: [1206.0294 \[hep-ex\]](1206.0294) (cit. on p. 10).
- [18] Donal E. Groom & Nikolai V. Mokhov & Sergei I. Striganov. "Muon Stopping Power And Range Tables 10 MeV—100 TeV". In: 62.1 (Jan. 2001), pp. 1–20. URL: <http://link.aip.org/link/?RSI/62/1/1> (cit. on p. 10).
- [19] CERN. *Cosmic rays: particles from outer space*. <https://home.cern/science/physics/cosmic-rays-particles-outer-space>. Accessed: 30-07-2023 (cit. on p. 11).
- [20] The AMANDA collaboration. *Antarctica: A Unique Laboratory for Science*. <https://www.nsf.gov/pubs/1996/nstc96rp/sb5.htm>. Accessed: 29-04-2023. 1996 (cit. on p. 13).
- [21] R. Abbasi et al. "The design and performance of IceCube DeepCore". In: *Astroparticle Physics* 35.10 (2012), pp. 615–624. ISSN: 0927-6505. DOI: <https://doi.org/10.1016/j.astropartphys.2012.01.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0927650512000254> (cit. on pp. 13, 14, 28).
- [22] "Low energy event reconstruction in IceCube DeepCore". In: 82 (Sept. 2022). DOI: <10.1140/epjc/s10052-022-10721-2>. URL: <https://doi.org/10.1140/epjc/s10052-022-10721-2> (cit. on pp. 14, 15, 30).
- [23] IceCube Collaboration. *IceCube Glaciology*. <https://icecube.wisc.edu/science/research/#glaciology>. Accessed: 27-07-2023 (cit. on pp. 14, 19, 20, 22).
- [24] R. Abbasi et. al. "The IceCube data acquisition system: Signal capture, digitization, and timestamping". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 601.3 (2009). DOI: <10.1016/j.nima.2009.01.001>. URL: <https://doi.org/10.1016/%2Fj.nima.2009.01.001> (cit. on pp. 16, 28).
- [25] M.G. Aartsen et al. "The IceCube Neutrino Observatory: instrumentation and online systems". In: *Journal of Instrumentation* 12.03 (2017), P03012. DOI: <10.1088/1748-0221/12/03/P03012>. URL: <https://dx.doi.org/10.1088/1748-0221/12/03/P03012> (cit. on pp. 16, 20, 26).
- [26] M.G. Aartsen et al. "In-situ calibration of the single-photoelectron charge response of the IceCube photomultiplier tubes". In: *Journal of Instrumentation* 15.06 (June 2020), P06032. DOI: <10.1088/1748-0221/15/06/P06032>. URL: <https://dx.doi.org/10.1088/1748-0221/15/06/P06032> (cit. on pp. 17–19, 26–28).
- [27] Hamamatsu Photonics. *PMT Handbook*. [https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99\\_SALES\\_LIBRARY/etd/PMT\\_handbook\\_v4E.pdf](https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/etd/PMT_handbook_v4E.pdf). Accessed: 28-07-2023. 2007 (cit. on p. 17).
- [28] R. Stokstad. "Design and Performance of the IceCube Electronics". In: (Sept. 2005). URL: <https://inspirehep.net/literature/704810> (cit. on p. 17).

- [29] The IceCube collaboration. *Beyond IceCube*. Accessed: 21-08-2023. URL: [\url{https://icecube.wisc.edu/science/beyond/}](https://icecube.wisc.edu/science/beyond/) (cit. on p. 17).
- [30] Étienne Bourbeau. "Measurement of Tau Neutrino Appearance in 8 Years of IceCube Data". PhD thesis. University of Copenhagen, 2020. URL: [https://www.nbi.dk/~petersen/IceCubeML/Theses/PhDThesis\\_EtienneBourbeau\\_5jan2021.pdf](https://www.nbi.dk/~petersen/IceCubeML/Theses/PhDThesis_EtienneBourbeau_5jan2021.pdf) (cit. on pp. 18, 29).
- [31] IceCube collaboration. *WaveDeform webpage*. <https://docs.icecube.aq/icetray/main/projects/wavedeform/index.html>. Accessed: 26-07-2023 (cit. on p. 18).
- [32] National Science Foundation. *About Ice Cores*. <https://icecores.org/about-ice-cores>. Accessed: 27-07-2023 (cit. on p. 19).
- [33] P. Buford Price et al. "Temperature profile for glacial ice at the South Pole: Implications for life in a nearby subglacial lake". In: *Proceedings of the National Academy of Sciences* 99.12 (2002), pp. 7844–7847. DOI: [10.1073/pnas.082238999](https://doi.org/10.1073/pnas.082238999). eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.082238999>. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.082238999> (cit. on pp. 19, 21).
- [34] IceCube Collaboration. *Drilling IceCube: a story of innovation, expertise and strong will*. <https://icecube.wisc.edu/news/detector/2014/12/drilling-icecube-story-of-innovation-expertise-and-strong-will/>. Accessed: 27-07-2023 (cit. on p. 20).
- [35] M. Ackermann et. al. "Optical properties of deep glacial ice at the South Pole". In: *Journal of Geophysical Research: Atmospheres* 111.D13 (2006). DOI: <https://doi.org/10.1029/2005JD006687>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006687>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006687> (cit. on pp. 21, 22).
- [36] M. Aartsen et al. *Searching for eV-scale sterile neutrinos with eight years of atmospheric neutrinos at the IceCube neutrino telescope*. May 2020 (cit. on p. 21).
- [37] Dmitry Chirkin and Martin Rongen. *Light diffusion in birefringent polycrystals and the IceCube ice anisotropy*. 2019. arXiv: [1908.07608 \[astro-ph.HE\]](https://arxiv.org/abs/1908.07608) (cit. on pp. 21, 22).
- [38] The IceCube collaboration. *Newly discovered optical effect allows IceCube to deduce ice crystal properties*. <https://icecube.wisc.edu/news/research/2022/10/newly-discovered-optical-effect-allows-icecube-to-deduce-ice-crystal-properties/>. Accessed: 29-07-2023. Oct. 2022 (cit. on pp. 22, 23).
- [39] Leon Bozianu. *Stopped Muon Reconstruction in IceCube DeepCore*. 2022. URL: [https://www.nbi.dk/~petersen/IceCubeML/Theses/MasterThesis\\_LeonBozianu\\_22May2022.pdf](https://www.nbi.dk/~petersen/IceCubeML/Theses/MasterThesis_LeonBozianu_22May2022.pdf) (cit. on pp. 24, 74).
- [40] D. Heck et al. "CORSIKA: A Monte Carlo code to simulate extensive air showers". In: (Feb. 1998) (cit. on p. 25).
- [41] Jakob van Santen. "Neutrino Interactions in IceCube above 1 TeV". PhD thesis. University of Wisconsin-Madison, 2014. URL: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjFu5vIrsiAAxW6RvEDHfKcCTwQFnoECBMQAQ&url=https%3A%2F%2Fdocushare.icecube.wisc.edu%2Fdsweb%2FGet%2FDocument-72536%2Fthesis.pdf&usg=A0vVaw0f-OG0eff03-xnMhhElW4r&opi=89978449> (cit. on p. 25).
- [42] G. Carminati et al. "Atmospheric MUons from PArametric formulas: a fast GEnerator for neutrino telescopes (MUPAGE)". In: *Computer Physics Communications* 179.12 (Dec. 2008), pp. 915–923. DOI: [10.1016/j.cpc.2008.07.014](https://doi.org/10.1016/j.cpc.2008.07.014). URL: <https://doi.org/10.1016/j.cpc.2008.07.014> (cit. on p. 25).

- [43] C. Andreopoulos et al. "The GENIE neutrino Monte Carlo generator". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 614.1 (2010), pp. 87–104. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2009.12.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900209023043> (cit. on p. 25).
- [44] The GENIE Collaboration. *Event Generator Global Analysis of Neutrino Scattering Data*. <http://www.genie-mc.org/>. Accessed: 14-08-2023 (cit. on p. 25).
- [45] The IceCube Collaboration. *The Vuvuzela noise module*. Accessed: 06-08-2023. URL: \url{https://docs.icecube.aq/icetray/main/python/icecube.vuvuzela.pregenerated\_noise.html} (cit. on p. 25).
- [46] P. Eller & J. Lanfranchi. *Retro Reco*. <https://github.com/icecube/retro>. Accessed: 28-04-2023 (cit. on pp. 26, 30).
- [47] David Freiherr Heereman von zuytdwyck. "HitSpooling: An Improvement for the Supernova Neutrino Detection System in IceCube". PhD thesis. Université Libre de Bruxelles, June 2015 (cit. on p. 27).
- [48] oscNext team. "oscNext - Simulations and sample (v00.07)". In: (2023) (cit. on pp. 28, 77).
- [49] Rasmus F. Ørsøe. "A Graph Neural Network Approach to Low Energy Event Reconstruction in IceCube Neutrino Observatory". MA thesis. Niels Bohr Institute, May 2021 (cit. on pp. 32, 38).
- [50] HandWiki.org. *Bessel function*. Accessed: 10-08-2023. June 2023. URL: \url{https://handwiki.org/wiki/Bessel\_function} (cit. on p. 32).
- [51] Alexander Amini et al. "Spatial Uncertainty Sampling for End-to-End Control". In: (May 2018) (cit. on p. 33).
- [52] Peter Andresen. *Observing The Universe Using The Antarctic Ice Sheet - Classifying And Reconstructing Neutrinos In Real Data From The IceCube Neutrino Observatory Using Graph Neural Networks*. 2023. URL: <https://www.nbi.dk/~petersen/IceCubeML/IceCubeML.html> (cit. on pp. 34, 41, 70).
- [53] Wikipedia.org. *Overfitting*. Accessed: 10-08-2023. June 2023. URL: \url{https://en.wikipedia.org/wiki/Overfitting} (cit. on p. 35).
- [54] Ramazan Gençay and Min Qi. "Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging". In: *Neural Networks, IEEE Transactions on* 12 (Aug. 2001), pp. 726 –734. DOI: <10.1109/72.935086> (cit. on p. 35).
- [55] Katrin Melcher. *A friendly Introduction to [Deep] Neural Networks*. Accessed: 03-08-2023. Aug. 2021. URL: \url{https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks} (cit. on pp. 36, 37).
- [56] Pytorch-geometric. *PyG Documentation*. Accessed: 10-08-2023. URL: \url{https://pytorch-geometric.readthedocs.io/en/latest/} (cit. on p. 38).
- [57] Yue Wang et al. "Dynamic Graph CNN for Learning on Point Clouds". In: *ACM Trans. Graph.* 38.5 (2019). ISSN: 0730-0301. DOI: <10.1145/3326362>. URL: <https://doi.org/10.1145/3326362> (cit. on p. 38).
- [58] Hongbin Pei et al. "Geom-GCN: Geometric Graph Convolutional Networks". In: (2020). arXiv: [2002.05287 \[cs.LG\]](2002.05287) (cit. on p. 38).
- [59] hep\_ml. *Reweighting algorithms*. Accessed: 08-08-2023. URL: \url{https://arogozhnikov.github.io/hep\_ml/reweight.html} (cit. on p. 43).

- [60] Alexander Trettin. "Verbesserung der Sensitivität auf Neutrino Mischungsparameter bei Messungen mit dem IceCube DeepCore Detektor". MA thesis. Institut für Physik der Humboldt-Universität zu Berlin, Oct. 2018 (cit. on p. 51).
- [61] John Erthal Gaiser. "Charmonium Spectroscopy From Radiative Decays of the  $J/\psi$  and  $\psi''$ ". PhD thesis. Aug. 1982 (cit. on p. 66).