UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Event Selection for the IceCube Upgrade

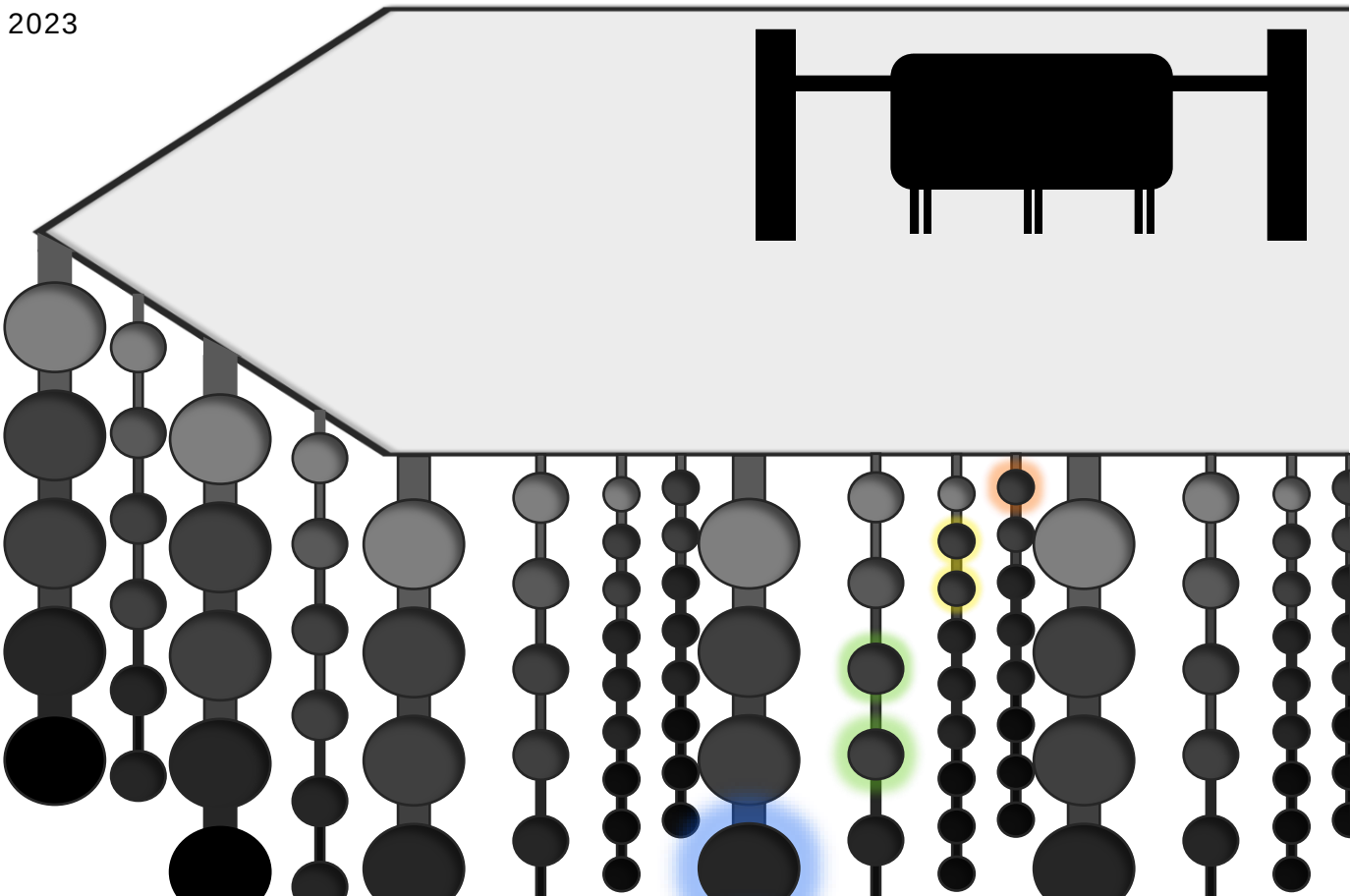
Machine Learning-Driven event selection for neutrino physics analysis using with the IceCube Upgrade

Master Thesis

Jorge Prado González

Supervisor: Jason Koskinen

May 22, 2023



A mis abuelos, a mis padres y a Daniel.

ABSTRACT

The IceCube Upgrade (ICU) is an extension of the IceCube neutrino detector, which was built with the goal of lowering the energy detection threshold of the current experiment to a few GeVs. Provided this is achieved, ICU will be capable of achieving world-leading constraints to oscillation parameters, unique measurements of ν_τ appearance, as well as providing some sensitivity to resolve neutrino mass ordering (NMO). NMO remains today one of the fundamental open questions in particle physics and relates to which one of the neutrino mass eigenstantes (ν_1, ν_2 or ν_3) is the lightest and which of them is the heaviest, being the scenario $\nu_3 > \nu_2, \nu_1$ the so-called Normal Ordering (NO), and $\nu_3 < \nu_2, \nu_1$ the Inverted Ordering (IO). ICU includes new detection modules with enhanced sensibilities that not only increase the capacity to detect neutrinos but also enhance instrumental noise by orders of magnitude. This, combined with the fact that ICU will detect low-energy neutrinos that will barely leave a trace in the detector, makes the cleaning and event selection procedure for IceCube a new paradigm. This thesis presents a complete event selection that, with the implementation of machine learning tools, achieves neutrino samples with purities over 99.7% and neutrino detection rates significantly higher than the ones currently achieved in IceCube. Finally, a study of the sensitivity of ICU to NMO after 3 years of data collection is depicted, presenting significance to resolve NMO over 3σ independently of whether NO or IO results to be true.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my supervisor, Jason Koskinen for providing me with the incredible opportunity to work on his project. I am immensely thankful for his constant involvement and interest in my work, pushing me to present results in collaborative calls and conferences, showing true excitement when things went well, and always being supportive when they did not. Furthermore, I am grateful for his understanding and support when physics was not the priority. Thank you very much for being what every supervisor should be.

I would also like to extend my appreciation to the entire IceCube group(+Kevin) at NBI: Markus, Tom, Kathrine, Moust, Linea, Amalie, Tania and Kevin, for their unwavering support and valuable feedback provided during our regular Friday group meetings, as well as for the pleasure of their company both inside and outside the office.

Individually, I would like to express my sincere gratitude to Kayla Leonard DeHolton, Jan Weldert, Rasmus Ørsøe, and Amalie Beate Albrechtsen for their invaluable assistance and contributions to the work presented in this project.

In addition, I would like to extend my heartfelt thanks to Laia, Lucia, Rasmus, Ben, Sotera, Story, Boglar, Alicia, Yannick, and Anna for becoming my second family. Your friendship, support, and companionship have made this journey memorable and meaningful. I especially want to express my gratitude to Bia, who has been the best listener, supporter, source of encouragement, and friend during the past two years.

STATEMENT OF CONTRIBUTIONS

The work described in this thesis is not solely the author's individual effort but rather a collaborative contribution involving daily collaboration and idea sharing with members of the Niels Bohr IceCube group, as well as a larger contribution from various universities affiliated with the IceCube Collaboration.

The majority of the time dedicated to this thesis was focused on developing an event selection from scratch. Although not explicitly presented in the thesis, a significant amount of work was devoted to step-by-step problem-solving, proposing and discussing solutions that may not have turned out to be optimal and therefore were not included in the final presentation.

The results achieved in collaboration or by other IceCube members that are presented in this thesis are the following: Chapter 3 introduces a pulse cleaning algorithm, and Chapter 4 discusses a classifier built with Dynedge, both developed by Rasmus Ørsøe. The author's contribution primarily involves studying the algorithm's performance. Additionally, machine learning techniques were initially applied to build QUESO Level 3, but they did not yield better results compared to the cuts on variables eventually used. The cuts on variables were presented by Jan Weldert and Kayla Leonard DeHolton, and due to their simplicity and excellent performance, they were adopted as QUESO Level 3. Finally, the NMO sensitivities reported in Chapter 5 were calculated by the author's office colleague, Amalie Beate Albrechtsen.

CONTENTS

INTRODUCTION

Neutrinos are elusive particles produced through various mechanisms, some of which are astrophysical, like the Sun or supernova explosions, while others are closer to us, such as nuclear reactors or particle showers induced by cosmic rays in Earth's atmosphere. Despite being the second most abundant particle in the universe after photons, detecting neutrinos is extremely challenging because they primarily interact only through the weak force (excluding gravity). When a neutrino interacts, it acquires a flavor (ν_e, ν_μ, or ν_τ) based on the lepton involved in the interaction. Neutrinos produced in one eigenstate can travel and interact in a different flavor state, known as "neutrino oscillations". Neutrino oscillations are an exciting field of research in particle physics, and in 2015, Takaaki Kajita and Arthur McDonald received the Nobel Prize for their work proving neutrino oscillation. The discovery of neutrino oscillations implies that neutrinos have a non-zero mass, challenging the Standard Model's prediction of massless neutrinos. Consequently, the study and characterization of parameters describing neutrino oscillations have become active research fields in particle physics [1][2].

Neutrinos can also be described by their mass eigenstates (ν_1, ν_2, and ν_3), which are the eigenstates of the Hamiltonian of a free particle. Various neutrino experiments have been conducted to determine the relative order of masses of these eigenstates, referred to as the neutrino mass ordering (NMO). Solar experiments have established that ν_2 is heavier than ν_1 [3]. However, it remains unresolved whether ν_3 is the heaviest state, known as normal ordering (NO), or the lightest, known as inverted ordering (IO). Determining the neutrino mass ordering is crucial for understanding the mechanism behind their masses.

The IceCube neutrino observatory is an atmospheric ice-Cherenkov neutrino detector situated at the South Pole. IceCube was initially designed to capture traces of high-energy neutrino interactions of astrophysical origin. However, IceCube's sub-array called DeepCore allows the study of low-energy neutrinos as well. The Earth is constantly bombarded by cosmic particles, which interact in the atmosphere and generate byproduct neutrinos, resulting in a high atmospheric neutrino flux. To further investigate atmospheric neutrinos, the IceCube detector will undergo an improvement during the Antarctic summer of 2024/2025, known as the IceCube Upgrade (ICU). ICU involves the installation of 7 new strings with sensitivity-enhanced Optical Detection Modules (DOMs) positioned closely together, enabling the detection of neutrinos with energies in the few GeV range. The ICU aims to achieve world-leading sensitivities in oscillation parameters and enhance the precision of tau neutrino detection, providing a

stringent test for the Standard Model [4][5]. Additionally, the ICU has the potential to shed light on neutrino mass ordering (NMO).

Implementing an upgrade of this magnitude requires a comprehensive understanding of the detector's operation and its new capabilities. This thesis presents an event selection for the ICU. Since neutrinos rarely interact with matter, the majority of signals captured by the detector arise from instrumental noise or atmospheric muons. In IceCube, the noise and muon rates were thousands of times higher than the neutrino rate [6]. For ICU, the number of noise events triggering the detector vastly outnumbers neutrino events by millions. Moreover, low-energy neutrinos leave minimal traces in the detector, making it exceedingly difficult to distinguish their signals from noise. Therefore, developing an event selection process has necessitated imaginative solutions and the application of state-of-the-art machine learning techniques.

The initial chapter of this work focuses on providing a background in particle physics. It begins with an overview of the standard model of particle physics, followed by an explanation of neutrino physics. The chapter concludes by discussing the atmospheric neutrino flux and the generation of muons in the atmosphere. These muons are relevant to this thesis as they contribute significantly to the non-instrumental noise in the IceCube detector.

Moving on to the second chapter, it offers an overview of the IceCube detector. It covers the current status of the project and provides details about the design and goals of the IceCube Upgrade. Additionally, the chapter explains how different events will be seen in the detector.

The third chapter is dedicated to explaining the initial processing steps that data from the IceCube Upgrade detector undergoes. It also delves into how simulated data is treated to facilitate further studies of the detector's capabilities before its deployment. The chapter includes a discussion on event cleaning using machine learning techniques and a preliminary event selection process to exclude events with a clear non-neutrino origin. Furthermore, it introduces machine learning techniques used for higher-level event selection.

Continuing the pipeline from the third chapter, the fourth chapter presents the final steps of the event selection process. Various approaches involving different machine learning techniques have been proposed and developed for these final steps. A comparison among these approaches is presented, ultimately selecting a model that combines predictions from Dynedge, a Graphical Neural Network (GNN), and a combination of Light Gradient Boosted Machines (LGBMs).

Finally, the fifth and last chapter, aside the conclusions, depicts a study of the NMO sensitivities with the IceCube Upgrade using the event selection presented in this work as a sample.

1

STANDARD MODEL

Contents

1.1 THE STANDARD MODEL OF PARTICLE PHYSICS

One of the main goals of science has historically been to explain the fundamental building blocks of our Universe. The **Standard Model** is a theory that represents our current understanding of the elementary particles, the fundamental constituents of the Universe, and their interactions, known as forces. Within the framework of the Standard Model, there are 12 elementary **fermions** with a spin of 1/2. Each fermion has a corresponding antiparticle. Fermions are categorized into six quarks and six leptons based on the type of charge they carry or, equivalently, how they interact. **Quarks** possess electric charge, weak isospin, and color charge, leading to their interactions through electromagnetism, weak force, and strong force. Quarks are tightly bound together due to a phenomenon called color confinement[1]. These color-neutral composite particles are classified as *mesons* (composed of a quark and an antiquark), such as kaons or pions, and *hadrons* (composed of three quarks), like protons or neutrons. Unlike quarks, **leptons** do not carry color charge, which means they do not interact via the strong force. Among leptons, there are

1 Color confinement refers to the phenomenon that color-charged particles cannot be observed in isolation but rather cluster together to form color-neutral particles (mesons and baryons) [7].

electrically charged particles: the **electron**, the **muon**, and the **tau**, which also carry weak isospin. These leptons differ from each other in mass. Additionally, there is an electrically neutral particle corresponding to each lepton, known as **neutrinos**. Neutrinos are very light particles influenced only by the weak nuclear force and gravity, making them difficult to detect.

The interactions between particles are mediated by gauge bosons with a spin of 1. **Photons** are massless spin-1 bosons that mediate the electromagnetic force between electrically charged particles. The $\mathbf{W}^{\pm}$ and **Z** gauge bosons mediate the weak interactions, and all three of them are massive. The $W^{\pm}$ boson also carries an electric charge and interacts electromagnetically. The strong interaction is mediated by eight massless particles called **gluons**. Finally, there is a boson with no spin, known as the **Higgs boson**, which plays a unique role in the Standard Model, as it explains why other elementary particles have mass[2] [2][8].

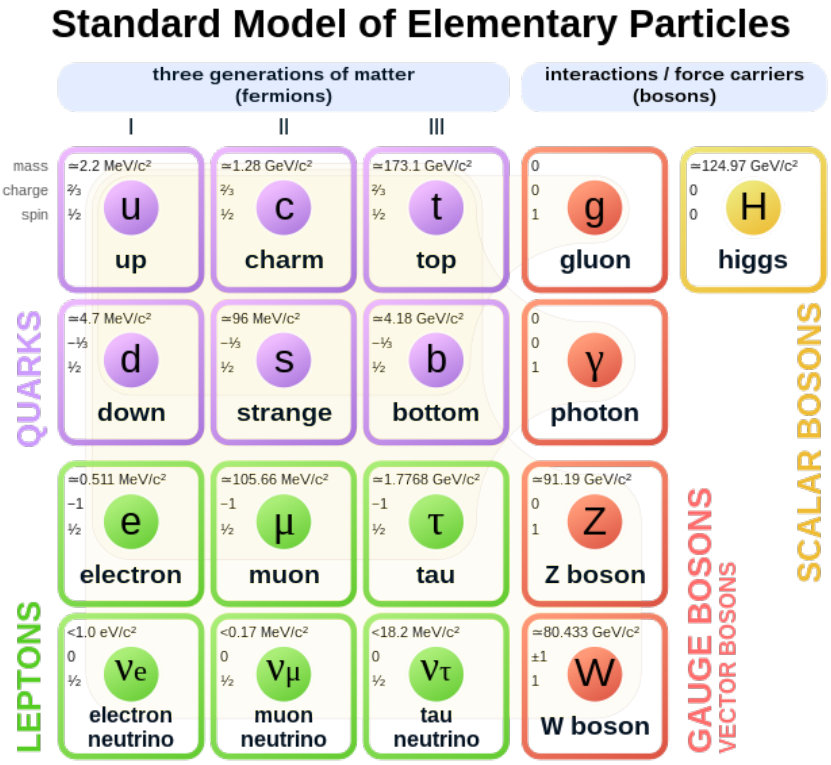


Figure 1.1: The elementary particles of the Standard model. *Image taken from* [8]

2 Neutrino interaction with the Higgs boson cannot explain by itself neutrino-tiny masses, and other mechanisms must be considered.

1.2 NEUTRINOS

Neutrinos are elusive particles. Unlike charged leptons that can be detected from the continuous track defined by the ionisation of atoms as they transverse matter, neutrinos are never directly observed, they are only detected through their weak interactions [2].
The existence of the neutrino was first postulated in 1930 by Wolfang Pauli with the purpose of explaining the energy and momentum conservation in beta decay. In beta decay, an electron or positron, the so-called *beta particles*, are emitted from an atomic nucleus, transforming the original nuclide into an isobar of that nuclide.

$$n \rightarrow p^+ + e^- + \bar{\nu}_e \tag{1.1}$$

$$p^+ \rightarrow n + e^+ + \nu_e \tag{1.2}$$

Where n represents a neutron, p^+ a proton, e^- and e^+ are beta particles, and ν_e and $\bar{\nu}_e$ are neutrino and antineutrino, respectively. If the black parts of equations (1.1) and (1.2) were the beta decay equations and no third body was involved, the spectra of beta particles emitted in the process would be expected to be discrete, as the beta particle will have acquired the energy difference of the process in the form of kinetic energy[3]. However, the spectra were continuous, and neutrinos were postulated to appear in the process and carry some of the energy, resulting in the observed distribution of energies of the beta particles. This frame led to the postulation of the existence of an electrically neutral particle (*neutr-*) that might have a very small mass (*-ino*). However, it was not until 1956 that Clyde Cowan, Frederick Reines, Francis B. Harrison, Herald W. Kruse, and Austin D. McGuire published confirmation that they had detected the neutrino [9]. The *Cowan-Reines neutrino experiment* was based on the beta capture process (see equation (1.3)). In the experiment, antineutrinos were generated in beta decay processes in a nuclear reactor, and they interacted with protons to give neutrons and positrons.

$$\bar{\nu}_e + p^+ \rightarrow n + e^+ \tag{1.3}$$

The positron quickly annihilates with an electron, emitting two detectable gamma rays. The neutron is captured by a nucleus that also emits a gamma ray. The coincidence of both events was a strong signature of an antineutrino interaction [10].

Today, it is admitted that neutrinos do indeed exist. There are three neutrino flavor states: the electron, muon and tau neutrinos (ν_e, ν_μ and ν_τ). These different neutrino flavors can only be identified by the

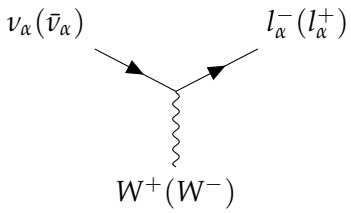
3 Equation (1.2) is by itself not energetically favourable and thus it cannot happen for a free proton, only inside nuclei.

flavors of the charged lepton produced in charged-current weak interactions. This means that, for instance, the electron neutrino ν_e is defined as the neutrino flavor produced in a charged-current weak interaction along with an electron. The muon neutrinos were first detected by Lederman, Schartz and Steinberg in 1962 [11]. In 1975, the tau lepton was discovered at the Stanford Linear Accelerator Center, and the existence of an associated neutrino was immediately theorized. However, it wasn't until 2000 that the DONUT collaboration at Fermilab detected tau neutrinos for the first time [2][10][12].

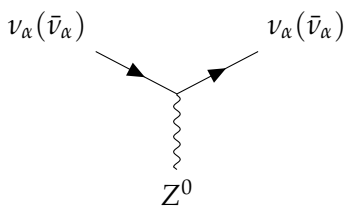
1.2.1 NEUTRINO INTERACTIONS WITH MATTER

Neutrinos can interact with both electrons and nucleons. The processes involving electron neutrinos are particularly relevant as they give rise to enhancements or suppressions in oscillation probabilities, providing IceCube with the capability to detect NMO. However, the cross sections of electron neutrino processes are significantly smaller compared to those involving nucleons.

All neutrino interaction processes occur through the weak interaction, which is mediated by either a $W^{\pm}$ boson in charged current (CC) processes or a Z boson in neutral current (NC) events (see Figure 1.2). In CC processes, a lepton of the same flavor as the initial neutrino is produced as a byproduct of the interaction. The identification of the lepton's flavor allows us to determine the flavor of the neutrino involved in the process. In NC interactions, an incoming neutrino transfers part of its energy to the Z boson and is transformed into another neutrino in the same flavor state. The outgoing neutrino leaves no trace, making it impossible to distinguish the flavor of the interacting neutrino [13].



(a) CC interaction vertex.



(b) NC interaction vertex.

Figure 1.2: Neutrino interaction vertex in cc (*left*) and nc (*right*) processes. The subindex α symbolizes a lepton flavor, ν a neutrino and l a lepton

Neutrino-nucleon interactions in the IceCube energy range are complex. The energy range $E_\nu \sim 0.1 - 20$ GeV is known as the "transition region". Below this range, the dominant process is quasi-elastic scattering, where the neutrino interacts with the nucleon as a coherent object. Deep inelastic scattering becomes dominant at higher energies, where the neutrino interacts with one of the constituent partons of the nucleon. In the transition region, several processes play important roles. Although there are adequate theories to describe each of these processes individually, there is no satisfactory unified theory that

describes the transition between them and how they should be combined. There are three main categories [14][15]:

- **Elastic and quasi-elastic scattering (QE):** As shown in Figure 1.3, the QE process is dominant below 1 GeV. In this process, neutrinos elastically scatter an entire nucleon. If the scattering occurs via charged current, where the neutrino interacts with a nucleon, exchanging a charged boson and emitting a charged lepton, it is referred to as "quasi-elastic scattering". On the other hand, if the neutrino scatters with the nucleon in a neutral current process, it is termed "elastic scattering".

- **Resonance production (RES):** This process is dominant for neutrino energies between 1 and 4 GeV. In this case, the neutrino excites the nucleon to a resonance state (Δ, N^*), which then decays into various mesonic final states.

- **Deep inelastic scattering:** With enough energy (the dominating process for neutrino energies above 10 GeV), the neutrino can scatter with an individual quark constituent of the nucleon. In this process, the nucleon disintegrates, producing a hadronic shower.

The contributions of each of these three scattering processes are shown in Figure 1.3. Because of the helicities of the states involved, which must be conserved for neutrinos[4], antineutrino-quark scattering is suppressed in comparison to neutrino-quark scattering. As a result, the neutrino-nucleon cross section is approximately twice as large as the antineutrino-nucleon cross section [2][13]. This is a crucial aspect of IceCube's ability to resolve NMO, as discussed in Section 2.2.1.

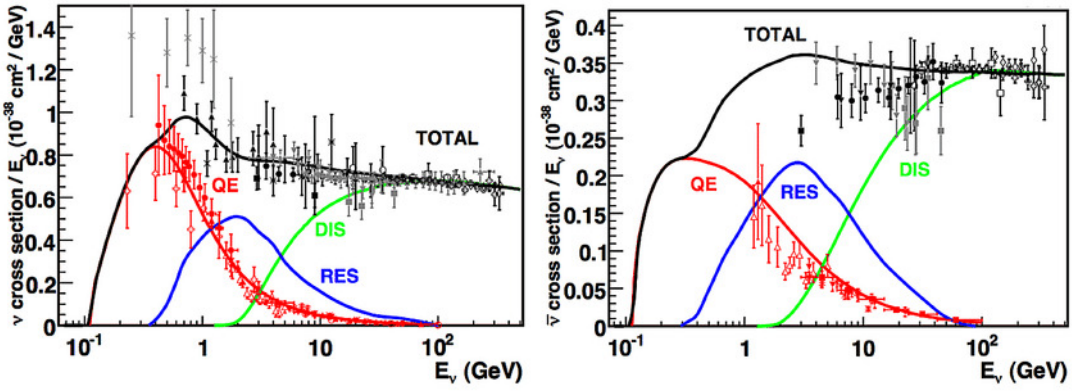


Figure 1.3: Neutrino *(left)* and and antineutrino *(right)* cross-sections as a function of the neutrino energy. *Image taken from* [16]

Even if the cross section of the interaction between electrons and neutrinos is significantly smaller than the one with nucleons, this interaction plays a crucial role in allowing IceCube to be sensitive to some of

4 Chirality and not helicity is what has to be strictly conserved, but, due to their small masses, helicity and chirality are almost equivalent in neutrino interactions [2].

the unsolved paradigms concerning neutrinos. This effect will be discussed as well in Section 2.2.1, after providing an explanation of neutrino oscillations and the neutrino mass ordering problem.

1.2.2 NEUTRINO OSCILLATIONS

Neutrinos were initially predicted to be massless by the Standard Model of Particle Physics. However, the observation of neutrino oscillations indicates that neutrinos have a non-zero mass. Neutrino oscillations have been observed in various neutrino experiments, and further research to constrain and better understand the physics behind neutrino oscillations will serve as a robust test for the Standard Model [3].

Neutrino oscillations are fascinating quantum mechanical phenomena that have attracted significant theoretical and experimental interest. Neutrinos exhibit oscillations when they are created in a specific flavor state or lepton number (electron, muon, or tau) but are later measured in a different one after traveling a certain distance. The first evidence of neutrino oscillations was discovered through the study of neutrinos produced in the Sun [17], followed by observations of atmospheric neutrinos, neutrinos produced in reactors, and neutrino beams generated in particle accelerators. These findings have provided a wealth of evidence for neutrino oscillations. This phenomenon can only occur if neutrinos interact on a different basis than the one on which they propagate, causing the eigenfunctions of the interaction Hamiltonian to differ from those of the propagation Hamiltonian.

Neutrinos have three mass eigenstates labeled ν_1, ν_2 and ν_3. The so-called mass eigenstates are the stationary states of the free-particle Hamiltonian that satisfy:

$$\hat{H}\Phi = E\Phi \tag{1.4}$$

However, there is no reason to think that these mass eigenstates are the same as the flavor eigenstates ν_e, ν_μ and ν_τ. In fact, the neutrino oscillation phenomenon is explained by stating that neutrinos propagate in one of their mass eigenstates, which are a linear combination of flavor eigenstates. Then, when a neutrino in a given mass eigenstate interacts, the mass eigenstate collapses to one of the flavor eigenstates producing the corresponding lepton. Thus, the mass eigenstates $|\nu_\alpha\rangle$ can be written in terms of the flavour eigenstates $|\nu_i\rangle$

$$|\nu_\alpha\rangle = \sum_{i\in 1,2,3} U^*_{\alpha i}\,|\nu_i\rangle \tag{1.5}$$

where $U_{\alpha i}$ are the elements of the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. The PMNS matrix is a 3×3 unitary matrix, meaning that $U^\dagger U = \mathbb{1}$. An $n \times n$ unitary matrix has n^2 free parameters, thus, under the consideration that the three neutrinos are Dirac particles, the PMNS matrix can be written in terms of three angles and six complex phases. However, five of these phases can be absorbed by the lepton

fields, so that the PMNS matrix can be written in terms of the three mixing angles θ_{12}, θ_{13} and θ_{23} and the complex phase δ_{CP}[5]:

$$\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.6)$$

With $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$.

The evolution of a flavor neutrino state $|\nu_\alpha\rangle$ in vacuum is described by the Schrödinger equation. However, the vacuum hamiltonian is not diagonal in the flavor basis. In order to solve the temporal evolution of the state, it is worth changing to the mass basis where the vacuum hamiltonian is diagonal ($H_{mass,vac}$ =Diag(E_1, E_2, E_3)). On this basis, the Schrödinger equation is:

$$i\frac{d}{dt}|\nu_i(t)\rangle = E_i\,|\nu_i(t)\rangle \quad (1.7)$$

Since the Hamiltomian is constant the differential equation can easily be solved by integrating at both sides of the equation finding

$$|\nu_i(t)\rangle = |\nu_i(t=0)\rangle\, e^{-iE_i t} \quad (1.8)$$

As it was of interest the evolution of the flavour state one can write it as a superposition of mass states using the relation with the PMNS matrix

$$|\nu_\alpha(t)\rangle = \sum_{i\in 1,2,3} U^\dagger_{\alpha i}\,|\nu_i(t)\rangle = \sum_{i\in 1,2,3} U^\dagger_{\alpha i} e^{-iE_i t}\,|\nu_i(t=0)\rangle \quad (1.9)$$

So the amplitude of the flavour transition $|\nu_\alpha\rangle \to |\nu_\beta\rangle$ after a time t is

$$\langle\nu_\beta|\nu_\alpha(t)\rangle = \sum_{i\in 1,2,3} U^\dagger_{\alpha i} U_{\beta i} e^{-iE_i t} \quad (1.10)$$

And the oscillation probability after a time t,

$$P_{\alpha\beta}(t) = |\,\langle\nu_\beta|\nu_\alpha(t)\rangle\,|^2 = \sum_{i,j\in 1,2,3} U_{\beta i} U^\dagger_{\beta j} U^\dagger_{\alpha i} U_{\alpha i} e^{-i(E_i - E_j)t} \quad (1.11)$$

For relativistic neutrinos, the energy of every mass state E_i

$$E_i = \sqrt{P_i^2 + m_i^2} = |P_i|(1 + m_i^2/2P_i^2 + \dots) \approx E + m_i^2/2E \quad (1.12)$$

5 If neutrinos were considered to be Majorana particles, they would need to include two additional complex phases.

Where the square root has been Taylor expanded to second order and the momentum of the mass state (P_i) with the energy of the flavour state (E). This is a good approximation since the masses of the neutrinos are very small in comparison with their momentum. Under this approximation

$$P_{\alpha\beta}(t) = \sum_{i,j\in 1,2,3} U_{\beta i} U^{\dagger}_{\beta j} U^{\dagger}_{\alpha i} U_{\alpha i} \exp\left(-i\frac{\Delta m^2_{ij} t}{2E}\right) \tag{1.13}$$

Where $\Delta m^2_{ij} = m^2_i - m^2_j$. In natural units, where the speed of light $c = 1$, and the velocity of the neutrinos is approximately the speed of light, the substitution t=L allows us to express the probability as a function of the distance traveled by the neutrino and not of the time. By doing so and also using the trigonometric identity $\cos(2A) = 1 - 2\sin^2(A)$ and the Euler identity[6] the exponential can be changed by:

$$\exp\left(-i\frac{\Delta m^2_{ij} t}{2E}\right) \approx 1 - 2\sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right) + i\sin\left(\frac{\Delta m^2_{ij} L}{2E}\right) \tag{1.14}$$

Giving a transition probability,

$$\begin{aligned} P_{\alpha\beta} =& \delta_{\alpha\beta} - 4\sum_{i>j} \mathfrak{Re}[U_{\alpha i} U^{\dagger}_{\alpha j} U^{\dagger}_{\beta i} U_{\beta j}] \sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right) \\ &+ 2\sum_{i>j} \mathfrak{Im}[U_{\alpha i} U^{\dagger}_{\alpha j} U^{\dagger}_{\beta i} U_{\beta j}] \sin\left(\frac{\Delta m^2_{ij} L}{2E}\right) \end{aligned} \tag{1.15}$$

From equation (1.15) it can be seen that the probability of the neutrino changing flavor depends on the energy of the flavor state, the distance (L) traveled by the neutrino, the square of the mass differences between flavor states, and the elements of the PMNS matrix [18] [19]. The derivation of antineutrinos oscillation probability is very similar to the one for neutrinos. The only difference is that the relation of the flavor states to the mass states (equation (1.9)) is not given by the hermitian but by the PMNS matrix instead. So, the oscillation probability for antineutrinos only differs from the one for neutrinos by a minus sign on the imaginary term in the equation (1.15).

1.2.3 THE NEUTRINO MASS ORDERING

The Neutrino Mass Ordering (NMO) or Neutrino Mass Hierarchy is an open question that pertains to determining the lightest and heaviest among the neutrino mass eigentstates (ν_1, ν_2, ν_3). Experimental data from various studies have confirmed very small differences in mass, $\Delta m^2_{21} \sim 10^{-5}\ eV^2$ and $|\Delta m^2_{3i}| \sim 10^{-3}\ eV^2$, where $i = 1, 2$. The sign of Δm^2_{21} has been resolved through solar neutrino experiments. The interaction of neutrinos with matter, which will be discussed later, affects the survival

6 The Euler identity states that $e^{i\phi} = \cos\phi + i\sin\phi$

probability of electron neutrinos produced in the Sun's nucleus ($P(\nu_e \rightarrow \nu_e)$) as they travel to the surface. This effect varies depending on the sign of Δm^2_{21}. Consequently, it has been established that $m_2 > m_1$ [20]. The NMO is determined by the sign of Δm^2_{3i}, as illustrated in Figure (1.4). A positive sign corresponds to Normal Ordering (NO), while a negative sign corresponds to Inverted Ordering (IO). Due to the small values of the mass splittings compared to the energies and baselines of modern experiments, and considering experimental resolution, the imaginary term in equation (1.15) averages out over multiple periods of the sine function, resulting in a vanishing effect. Therefore, in vacuum oscillations, a degeneracy arises between (Δm^2_{ij}, $-\Delta m^2_{ij}$). However, as explained in more detail in Section 2.2.1, this degeneracy is resolved when neutrinos propagate through matter.

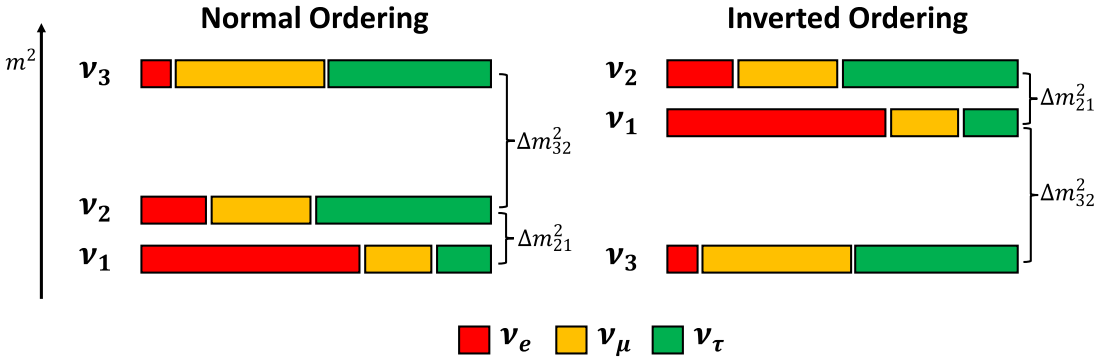


Figure 1.4: Neutrino mass eigenstates as a combination of flavor eigenstantes ordered in increasing mass value for normal (*left*) and inverted orderings. *Image produced by the author.*

Solving the question of NMO is crucial as it can shed light over important unsolved physics question, such as the mechanism by which neutrinos acquire their tiny masses or the relation among the lepton masses [18].

1.3 COSMIC RAYS

Different neutrino experiments have been designed to study neutrinos originating from various sources, such as the Sun, the Earth's atmosphere, or nuclear reactors. The term "atmospheric neutrino detectors" refers to those detectors focused on neutrinos produced as a result of cosmic ray interactions in the atmosphere. **Cosmic rays** are high-energy particles that reach the Earth from space. Approximately 99% of these particles consist of bare atomic nuclei, with around 90% being protons, 9% alpha particles[7], and 1% consisting of heavier element nuclei. Cosmic rays exhibit a wide energy range, spanning from approximately 10^8 eV to more than 10^{20} eV.

After being fully ionized by an astrophysical source, the cosmic ray nuclei travel vast distances, encoun-

7 Alpha particles consist of two protons and two neutrons bound together, forming a particle identical to a helium-4 nucleus [21]

tering galactic magnetic fields that cause their trajectories to deviate. This deflection results in a random distribution of cosmic ray arrival directions on Earth, leading to an isotropic cosmic ray flux [14][22] [23]. When cosmic rays enter the Earth's atmosphere, they collide with atoms and molecules, primarily oxygen and nitrogen, producing a cascade of particles (see Figure 1.5). This cascade mainly consists of mesons, which subsequently decay into secondary products, including atmospheric neutrinos and muons.

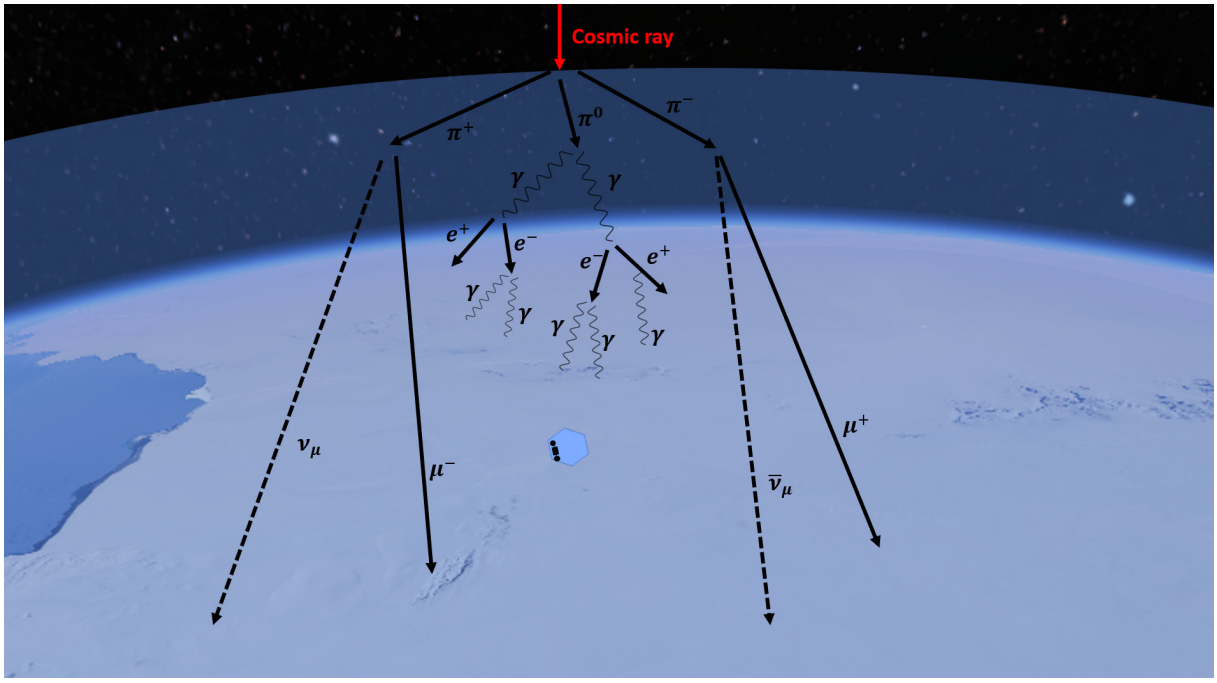


Figure 1.5: Cosmic ray entering Earth's atmosphere and the subsequent cascade of particles generated. *Self produced imaged with Earth picture taken from Google Earth [24] based on [25].*

1.3.1 ATMOSPHERIC NEUTRINOS

When the primary cosmic rays collide in the atmosphere, the mesons generated are mainly **pions** and **kaons**. Equation 1.16 represents the collision of a proton (p) with an atom N.

$$p + N \rightarrow \pi^+ + \pi^- + K^+ + K^- + X \tag{1.16}$$

Where X represents other hadrons.

The pions and kaons have short lifetimes and will decay rapidly into muons and muon neutrinos following Equation 1.17 and Equation 1.18.

$$\pi^+(K^+) \rightarrow \mu^+ + \nu_\mu \tag{1.17}$$

$$\pi^-(K^-) \rightarrow \mu^- + \bar{\nu}_\mu \tag{1.18}$$

These muons can decay into electrons and muon neutrinos before they reach Earth's ground, as described in Equations 1.19 and 1.20.

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \tag{1.19}$$

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \tag{1.20}$$

The above process is the main source of production of ν_e and $\bar{\nu}_e$ below 1 GeV. However, despite their short lifetime and due to their relativistic nature, muons over 1 GeV are Lorentz-boosted and can reach the ground before decaying. Thus, above 1 GeV, the ν_e and $\bar{\nu}_e$ dominant source is the decay of the long-neutral-kaon shown in Equation 1.21.

$$K_L^0 \rightarrow \pi^+(\pi^-) + e^-(e^+) + \bar{\nu}_e(\nu_e) \tag{1.21}$$

The neutrinos that are produced in the decay of pions and kaons are called conventional atmospheric neutrinos, and they do not include ν_τ.

In cosmic ray air showers, other heavier mesons can appear. The so-called prompt neutrinos are those produced in the decay of heavy mesons containing the charm quark. For instance, the D^+ meson can decay producing an electron or a muon neutrino as seen in Equations 1.22 1.23 (both processes with branching ratios[8] of 8.7%)

$$D^+ \rightarrow \bar{K}_0 + \mu^+ + \nu_\mu \tag{1.22}$$

$$D^+ \rightarrow \bar{K}_0 + e^+ + \nu_e \tag{1.23}$$

Prompt neutrinos also include ν_τ that come from the decay of strange $D_s^\pm$ mesons:

$$D_s^+ \rightarrow \tau^+ + \nu_\tau \quad \text{with the consecutive decay} \quad \tau^+ \rightarrow \mu^+ + \nu_\mu + \nu_\tau \tag{1.24}$$

$$D_s^- \rightarrow \tau^- + \bar{\nu}_\tau \quad \text{with the consecutive decay} \quad \tau^- \rightarrow \mu^- + \bar{\nu}_\mu + \bar{\nu}_\tau \tag{1.25}$$

8 Branching ratio is the percentage chance that a particle will decay in a certain way [26]

This implies that the total neutrino flux, which includes both the conventional and prompt fluxes, is expected to contain a contribution from tau neutrinos. However, the tau neutrino contribution to the overall flux is many orders of magnitude smaller than that of electron and muon neutrinos and has never been directly measured. The prompt contribution to the neutrino flux only becomes dominant at energies on the order of 10^5-10^6 GeV, and even at these energies, the tau neutrino contribution is still more than an order of magnitude lower than the contributions from other flavors. Therefore, for energies below 100 GeV, where the conventional flux prevails, the primary tau neutrino contribution can be neglected. The energy spectrum for neutrinos is depicted in Figure 1.6. The branching ratios of the mentioned processes along this energy range result in an approximate ratio of ν_e:ν_μ around 1:2 for energies in the range of a few GeV, with this ratio increasing for energies in the range of a few hundred GeV. The difference in the ratio between ν and $\bar{\nu}$ is due to the different amounts of positively and negatively charged mesons produced by cosmic rays, which are predominantly protons [13][27][28].

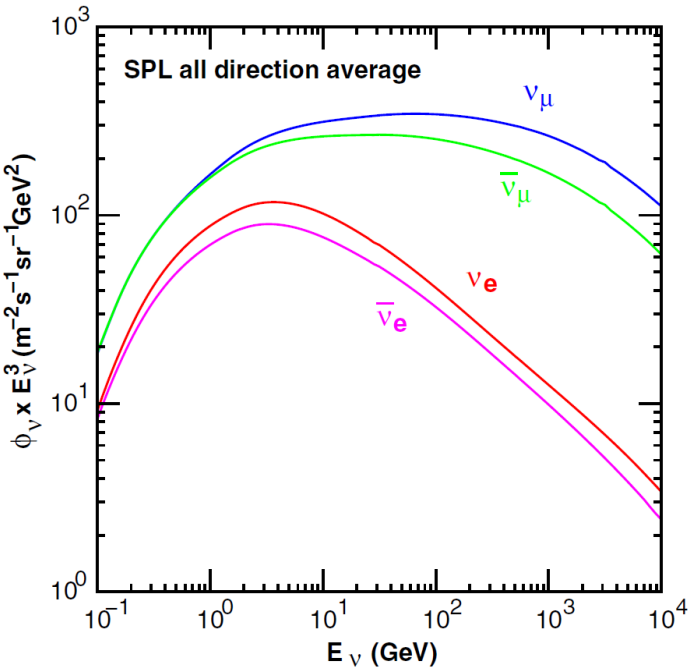


Figure 1.6: All-direction averaged atmospheric neutrino flux at the South Pole (SPL) averaged over a year. *Figure taken from [28]*

1.3.2 ATMOSPHERIC MUONS

It has been seen that in multiple meson decays (Equations 1.17-1.25) one of the byproducts is a muon. Muons are charged leptons with a lifetime of 2.2 μs. They are typically produced in air showers at altitudes of 10–15 km but can reach the Earth's surface before decaying due to relativistic effects. In fact, more energetic muons cannot only reach the ground but also penetrate, being the only products of cosmic

radiation, aside from neutrinos, that can be detected by "in-ice" IceCube[9]. Muons transversing matter deposit energy through continuous ionization and radiative processes. The energy loss rate depends on the energy of the muon and the medium it is traveling through. The energy loss of muon travelling in matter can be described by

$$-\frac{dE_\mu}{dX} = a + bE_\mu \tag{1.26}$$

where X is the distance that the muon travels in matter, a accounts for the ionization loss, and b for the fractional energy loss by three radiation processes: brehmsstrahlung, direct pair production, and photonuclear interactions. For IceCube ice, the measured values are a=0.163 GeVm^{-1} and b=0.192$\times$10^{-3} m^{-1}. Equation 1.26 can be integrated with these values, giving the needed energy of a muon to transverse certain distances through ice. For distances of $\sim$2 km, which is around the depths where the IceCube detectors are placed, the energy threshold for muons to reach is of the order of 400 GeV. Even though this energy threshold prevents most atmospheric muons from reaching IceCube, energetic enough muons are the primary non-instrumental source of noise, outnumbering neutrino rates by a factor of about 10^6 [13][29].

9 The IceCube neutrino observatory has a surface array of detectors called IceTop. "in-ice" IceCube refers to the detection modules deployed at ice depths greater than $\sim$1450 m.

2

ICECUBE

Contents

2.1 THE ICECUBE DETECTOR

The IceCube Neutrino Observatory, located at the South Pole, is an ice Cherenkov detector that spans approximately 1 cubic kilometer. It operates based on the principle of Cherenkov detection, making use of the phenomenon where charged particles generated in neutrinos' weak interactions with ice can travel faster through the medium than the phase velocity of light, resulting in the emission of photons known as Cherenkov radiation.

Construction of the detector was completed in December 2010, and it commenced operations in 2011. The large size of IceCube is attributed to its primary objective of detecting astrophysical neutrinos and subsequently investigating their sources. Additionally, IceCube has been extensively employed for other scientific purposes, including the indirect detection of dark matter, searches for exotic particles, studies on neutrino oscillation physics, and the detection of neutrino bursts originating from galactic core-collapse supernovae [30].

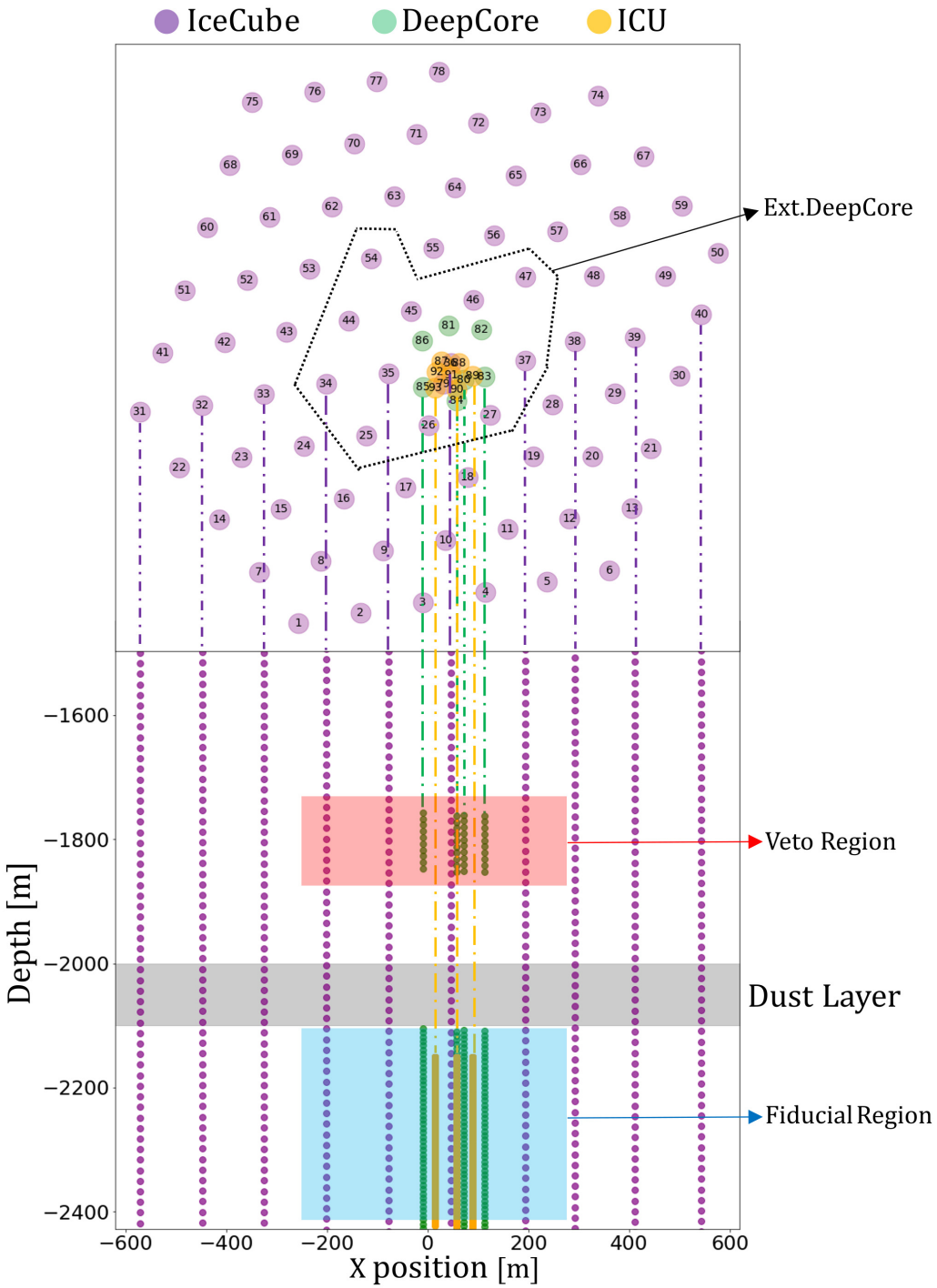


Figure 2.1: Top and side views of the IceCube, Deepcore and Icecube Upgrade arrays. The numbers in each circle in the top part of the figure indicate the string number. The position of the dust layer is shaded in gray. The veto and fiducial volumes are shaded in red and blue, respectively. For the sake of clarity, in the side view, only some of the strings are represented. *Image made by the author based on [31].*

2.1.1 ICECUBE ARRAY

The IceCube Neutrino Observatory consists of 5160 digital optical modules (DOMs) deployed in 86 vertical strings, positioned between depths of 1450m and 2450m beneath the ice surface. These DOMs

serve as the fundamental units for detecting neutrinos within IceCube. Equipped with photomultiplier tubes (PMTs), the DOMs have the capability to detect Cherenkov photons. All the PMTs in the IceCube DOMs are oriented downward, towards where the neutrinos more relevant for oscillation analyses arrive. When a photon strikes the photocathode in the PMT, electrons are ejected via the photoelectric effect. These electrons are generated within a region where a high voltage is applied, guiding them towards 10 dynodes (electron multipliers) that amplify the signal by a factor of 10^7.

The DOM Mainboard, which contains different electronics, serves as the central component of the IceCube data acquisition system (DAQ). It is responsible for reading, digitizing, processing, and buffering the PMT signals. To withstand the high pressure encountered, reaching up to 400 atmospheres during ice refreezing, the DOM components are enclosed within glass spheres.

Additionally, each DOM is equipped with 12 individually selectable LEDs that can emit light. These LEDs play a crucial role in studying and better understanding the properties of the surrounding ice [32].

Each string consists of 60 DOMs deployed along a single cable. Initially, the IceCube array consisted of 78 strings arranged in a hexagonal formation, with a spacing of 125m between strings and a vertical separation of 17m between DOMs within the same string. The configuration of DOMs was specifically designed to fulfill the primary scientific objective of IceCube: detecting astrophysical neutrinos. This necessitated an extensive detection volume, enabling the detection of astrophysical neutrinos despite their small interaction cross-section with ice and the extremely low expected fluxes at Earth [30].

2.1.2 DEEPCORE

DeepCore is a specialized component of the IceCube Neutrino Observatory designed to enhance its sensitivity to lower-energy neutrinos (below 100 GeV). This subset consists of 480 additional DOMs, densely distributed across eight new strings. The bottom 50 DOMs of these strings are positioned with a vertical spacing of 7 m, spanning depths from 2100 m to 2450 m. They are strategically placed to facilitate the detection and study of atmospheric neutrino oscillations.

To improve the rejection of downgoing atmospheric muons, the top 10 DOMs of each string are located at smaller depths than 2000 m. DeepCore was intentionally arranged to deviate from the hexagonal grid of sensors used in the main IceCube array. This departure was necessary to address the issue of certain regions lacking sensors within the previous symmetric layout, which could allow particles to pass through undetected. All DeepCore DOMs are deployed below the region known as the “dust layer”, which extends from depths of 2000 m to 2100 m. This region exhibits significantly higher levels of optical scattering and absorption, rendering it more opaque and thus more challenging for sensors to capture Cherenkov photons emitted by charged particles traversing through that area [30].

2.1.3 ICECUBE UPGRADE

The IceCube Upgrade (ICU) is a proposed project aimed at enhancing the IceCube/DeepCore detectors by introducing a new subset of sensor arrays. This new instrumentation will involve the deployment of 700 either new or upgraded sensors in a denser configuration. The ICU strings (strings 87-93) will be positioned approximately 20 m apart from each other, with a vertical spacing of around 3 m between modules. This arrangement will enable a reduction in the energy detection threshold to approximately 1 GeV, with improved capabilities for reconstructing events with energies in the few GeV range.

Unlike IceCube or DeepCore, not all DOMs in the ICU will be identical. The first type, known as PDOM, features a single PMT facing downwards, similar to the IceCube DOM, but with enhanced electronics. The second type is called D-Egg, consisting of 2×8" PMTs, one facing upwards and the other facing downwards. The third type, referred to as mDOM, comprises 24×3" PMTs. The inclusion of these new DOMs is expected to significantly enhance the directional sensitivity of the ICU, leading to improved rejection of atmospheric muons and background noise as well as more accurate reconstruction of neutrino directions. One PDOM, one D-Egg, and one mDOM are schematically represented in Figure 2.2.

The Upgrade will also introduce new capabilities for calibrating the existing detector to reanalyze archival data. Each optical module will be equipped with LEDs, CCD cameras, on-board sensors for pressure, temperature, and magnetic field, as well as accelerometers. Additionally, standalone light-emitting devices will be deployed with the Upgrade array for calibrating the glacial ice and detector responses. The calibration and study of ice properties conducted through the Upgrade will also contribute to addressing the challenges posed by IceCube-Gen2, the next-generation neutrino detector on a scale of approximately $\sim$8 km^3 [4][33][34][35].

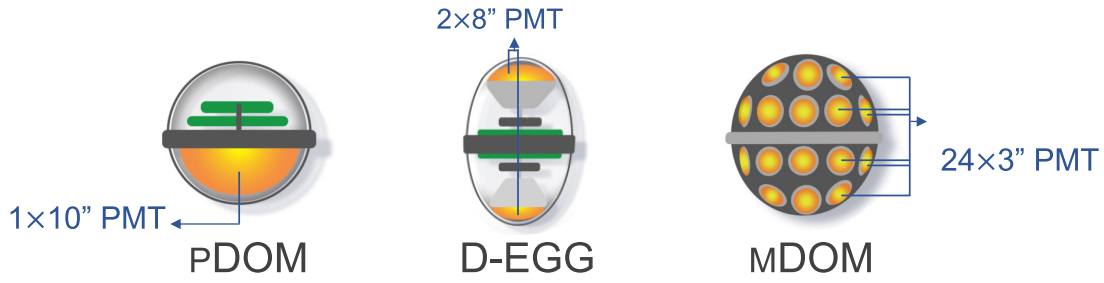


Figure 2.2: Different DOMs used in the IceCube Upgrade. *Image made by the author with DOM sketches taken from [36].*

2.2 SCIENCE CAPABILITY OF THE ICECUBE UPGRADE

As previously mentioned, the Upgrade project aims to enhance neutrino detection capabilities at the few GeV level by adding seven additional strings of densely spaced optical modules. This expansion not only lowers the energy threshold, enabling the detection of a greater number of neutrinos, but also enhances the capture of events that leave minimal information in the detector. This provides improved opportunities for characterizing and studying these events. Figure 2.3 illustrates the impact of the IceCube Upgrade on the detector's sensitivity, showing an increase in the number of pulses from 12 to 34 for the same event.

The newly constructed DOMs in the Upgrade will enhance the efficiency of Cherenkov photon detection and improve directional resolution. Additionally, the introduction of new calibration devices will contribute to a better understanding of ice properties and, consequently, reduce systematic uncertainties in the detector. This calibration improvement not only enables the ambitious goal of achieving world-leading sensitivities to neutrino oscillations but also facilitates unique measurements of tau neutrino appearance with high precision. Furthermore, the calibration advancements can be applied to the entire archival data collected by IceCube over the past decade, with anticipated improvements in angular and spatial resolution [4].

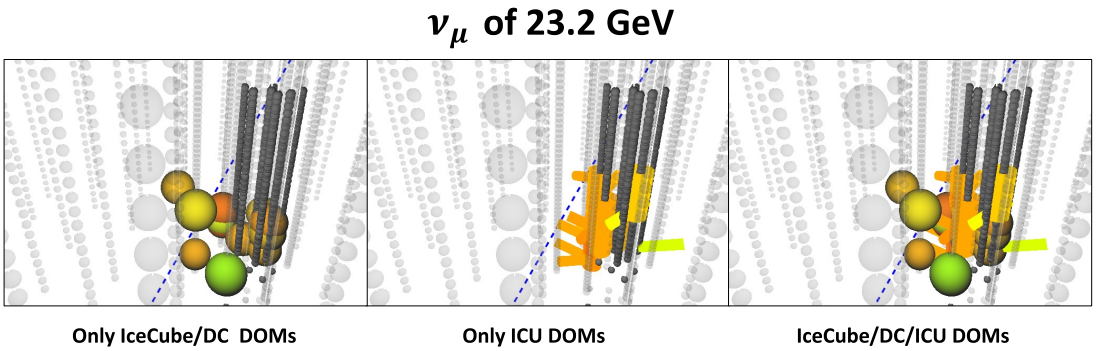


Figure 2.3: Detection of a ν_μ with an energy of 23.2 GeV seen by Icecube and Deepcore (*left*), only by ICU (*middle*) and by IceCube, DeepCore and ICU (*right*). For IceCube and DeepCore, the colored circles represent the pulses collected by the modules in the detector, with the circle size proportional to the charge collected in the pulse and the color of the circle a measure of the time in which the pulse is recorded (tending to red for early pulses and green-blue for late ones). For ICU, the pulses are represented in an analogous way, but the spheres are substituted by cylinders (for mDOMs) or cones (for DEggs) pointing towards the PMT direction. This event has 12 pulses without including ICU strings and 34 pulses with ICU strings. *Image produced by the author using [37].*

2.2.1 NMO SENSITIVITY WITH THE ICECUBE UPGRADE

The MSW (Mikheyev-Smirnov-Wolfenstein) effect provides an explanation for the modulation of neutrino oscillation probabilities in the presence of electrons in matter. This effect arises from the interaction of neutrinos with electrons via the charged current's elastic forward scattering [38]. In the presence of matter,

the vacuum Hamiltonian (as introduced in Equation 1.4) is modified to $H_M = H + V$, where H represents the vacuum Hamiltonian. Consequently, the mass eigenstates and eigenvalues differ between H_M and H, resulting in distinct dynamics for neutrino oscillations compared to those in vacuum. Moreover, the matter Hamiltonian can include a parametric enhancement of the oscillation probability when neutrinos traverse regions with density transitions within Earth's inner layers [18]. These physical processes also apply to antineutrinos; however, the sign of the potential is the opposite. As a result, any enhancement of oscillation probabilities for neutrinos is accompanied by a corresponding suppression of oscillation probabilities for antineutrinos, and vice versa [18].

IceCube (ICU) is capable of detecting both neutrinos (ν) and antineutrinos ($\bar{\nu}$), but it lacks the ability to distinguish between them. Consequently, for a 1:1 flux of neutrinos and antineutrinos, any enhancement in the oscillation probabilities for neutrinos would be counterbalanced by the corresponding suppression in the oscillation probabilities for antineutrinos. Fortunately, due to a charge imbalance in the incoming cosmic rays, the resulting particle showers generate a neutrino flux containing a higher proportion of $\nu_{e,\mu}$ compared to $\bar{\nu}_{e,\mu}$. Additionally, the charged current cross section is approximately 2 times higher for neutrinos than for antineutrinos. Therefore, the combined effect of these two factors ensures that the enhancement of oscillation probabilities in neutrinos or antineutrinos will not be fully canceled out by the suppression in their respective antiparticle [18].

Figure 2.4 shows the oscillation probabilities for a given neutrino energy and zenith angle in the NO and IO scenarios. The oscillation modulation effects induced by matter are stronger in the range of 2-10 GeV (with some contributions up to 15 GeV). Since IceCube's ability to detect NMO is based on a statistical sample of neutrinos in this range and not on being able to tell neutrinos from antineutrinos individually, the improved ability of the ICU to detect and reconstruct events with energies of a few GeV will boost the detector's ability to resolve NMO.

2.2.2 ν_τ APPEARANCE

Experiments focusing on neutrino appearance aim to detect neutrino flavors that were not initially present in the neutrino flux [13]. In the case of IceCube, the intrinsic tau neutrino component in the atmospheric neutrino flux is negligible, allowing for the study of the poorly constrained PMNS matrix element $U_{\tau 3}$ through the measurement of the signal strength of ν_τ neutrinos [31][18].

The τ lepton, being a heavy particle, requires a minimum energy of approximately 3.5 GeV for a ν_τ neutrino to produce a τ lepton. This characteristic places neutrino telescopes at the forefront of ν_τ appearance studies, as many neutrino accelerator experiments operate with neutrino beams below this threshold. The IceCube Upgrade, with its large oscillation baseline and dimensions, plays a crucial

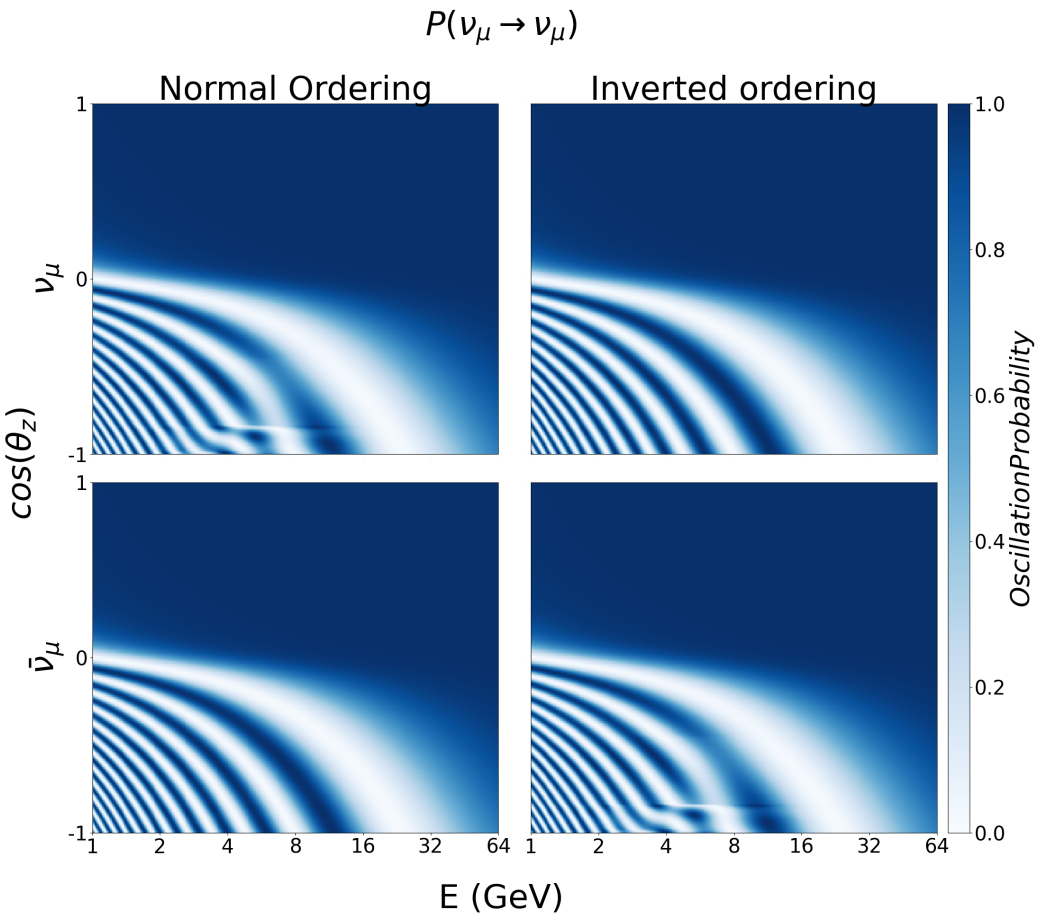


Figure 2.4: Oscillation probability from $\nu_\mu(\bar{\nu}_\mu) \to \nu_\mu(\bar{\nu}_\mu)$ as a function of the cosine of the zenith angle of arrival to the detector and the neutrino energy for Normal Ordering and Inverted Ordering. *Figure generated with* [39][40]

role as a world-leading experiment in this field. It is sensitive to the first oscillation maximum at an energy of around 24 GeV for baselines of up to 12,750 km. The detection efficiency peak of the IceCube Upgrade aligns well with the energy range of the ν_τ oscillation maximum, enabling the measurement of approximately 3000 ν_τ events per year [18][4].

2.3 SIGNATURES IN ICECUBE

Different particles and different particle interactions leave different traces in the detector. In IceCube, and especially in the IceCube Upgrade, several noise sources can cause a PMT to be triggered, leaving a signal as well. To build an event selection, it will be of great importance to distinguish and understand how different events look in the detector in order to choose the ones with a neutrino origin.

The light signatures resulting from particle passage through IceCube can be categorized into two groups: cascade-like and track-like events. **Cascade-like** events arise from hadronic and electromagnetic showers, which produce detection signals in the vicinity of the interaction vertex with a relative spheric shape, apart from some directionality. Cascade-like events are generated by all neutral current (NC) interaction processes, ν_e charged current (CC) interactions, most ν_τ CC interactions, and low-energy ν_μ interactions.

Track-like events occur when muons emitting Cherenkov photons traverse the detector. This results in a series of hits along the particle's trajectory, forming an elongated shape. The distinctive shape facilitates the reconstruction of directional information for track-like events, making them highly significant in neutrino source astronomy. Track-like events primarily originate from atmospheric muons, although some muons may enter the detector through unoccupied paths or dusty ice regions, leaving cascade-like traces. High-energy ν_μ CC events can produce muons that travel sufficiently long distances before decaying, thereby leaving a track signature. In contrast, tau leptons have a lifetime of only 3×10^{-13}s, traveling distances on the order of millimeters before decaying. When decaying, approximately 82.61% of the time, an electromagnetic or hadronic cascade, indistinguishable from the initial hadronic cascade, is produced. However, in 17.39% of cases, the decay follows the equations detailed in Equations 1.24 and 1.25, resulting in a muon that can be identified by its track signature in the detector [14][41]. Figure 2.5 schematically illustrates the various neutrino interactions, while Figure 2.6 provides an overview of the signatures left by track-like and cascade-like events within the detector

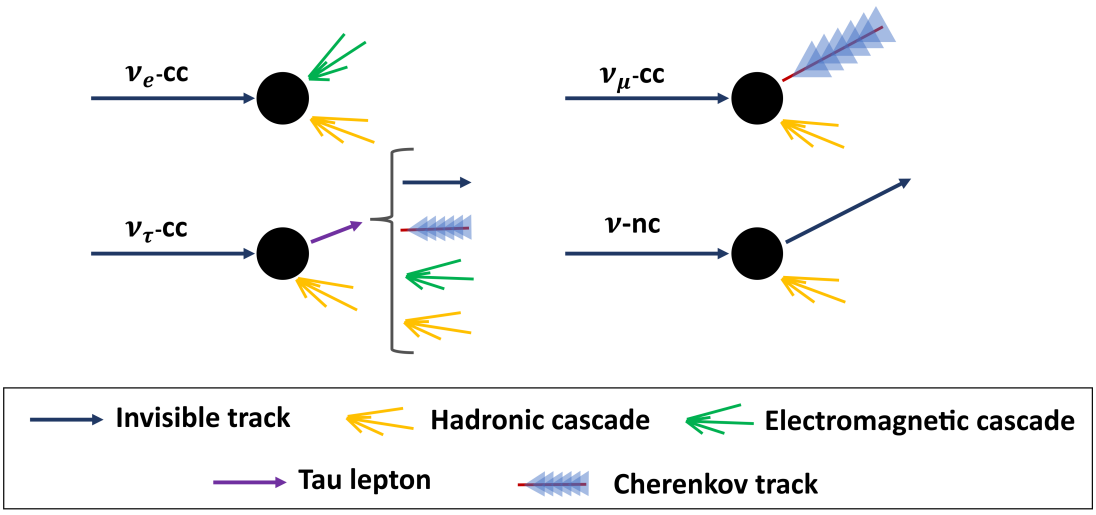


Figure 2.5: Neutrino interactions overview for different neutrino flavors and interaction types (cc or nc). *Self produced diagram based on [42]*

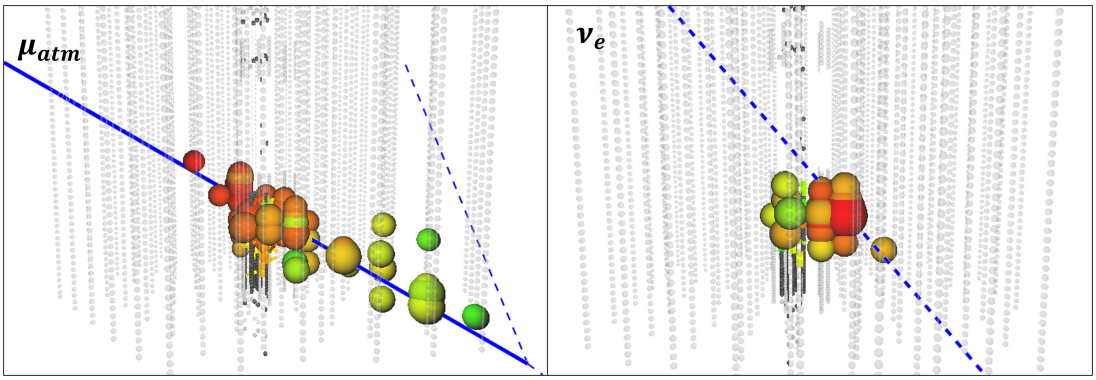


Figure 2.6: Typical track-like event signature in the IceCube detector caused by an incoming atmospheric muon (*left*) and typical cascade-like event caused by an electron neutrino (*right*). The colored circles represent the pulses collected by the modules in the detector, with the circle size proportional to the charge collected in the pulse and the color of the circle a measure of the time in which the pulse is recorded (tending to red for early pulses and green-blue for late ones). *Image generated by the author using [37]*

2.4 NOISE SOURCES IN ICECUBE

In the absence of an incident particle, certain IceCube photomultiplier tubes (PMTs) are still triggered, producing a signal known as dark noise (often referred to just as noise). Understanding the sources of noise is crucial, particularly for detectors that aim to detect rarely interacting particles like neutrinos. The overall noise profile of IceCube can be divided into three components: thermal noise, radioactive noise, and correlated scintillation noise.

Thermal noise arises from the thermionic emission of electrons from the detector hardware's surface, which is strongly influenced by the surface temperature. The emission of electrons follows a well-understood Poisson process and is simulated in the IceCube detector. Due to the operating temperatures of the DOMs ranging between -40°C and -20°C, the thermal noise is somewhat suppressed, with a thermal noise rate of approximately 20 Hz.

Radioactive decay noise results from the decay of isotopes present in the DOM pressure vessel. The decay process produces electrons, drawn from a Poissonian distribution, capable of triggering the detector's PMTs. The rate of radioactive noise remains relatively constant at around 250 Hz. Furthermore, the decay process leaves some energy in the glass, leading to the production of scintillating photons, which contribute to the correlated **scintillation noise**. The correlated noise does not follow a Poissonian distribution [30][43]. Figure 2.7 shows the hit index modulo 50 (counting up to 50 and then restarting the count) versus the hit time for a single IceCube DOM. The first hit registered in the DOM is given a hit index 1, the second is given a hit index 2, and so on. In the study of the DOM, a pure Poisson rate of 220 Hz was expected, as depicted in the figure by straight lines. The correlated noise hits appear in

the figure as clusters with a hit rate that dramatically increases over a very short period of time. This non-poissonian noise source has been studied but there is not fully agreement on its source [44].

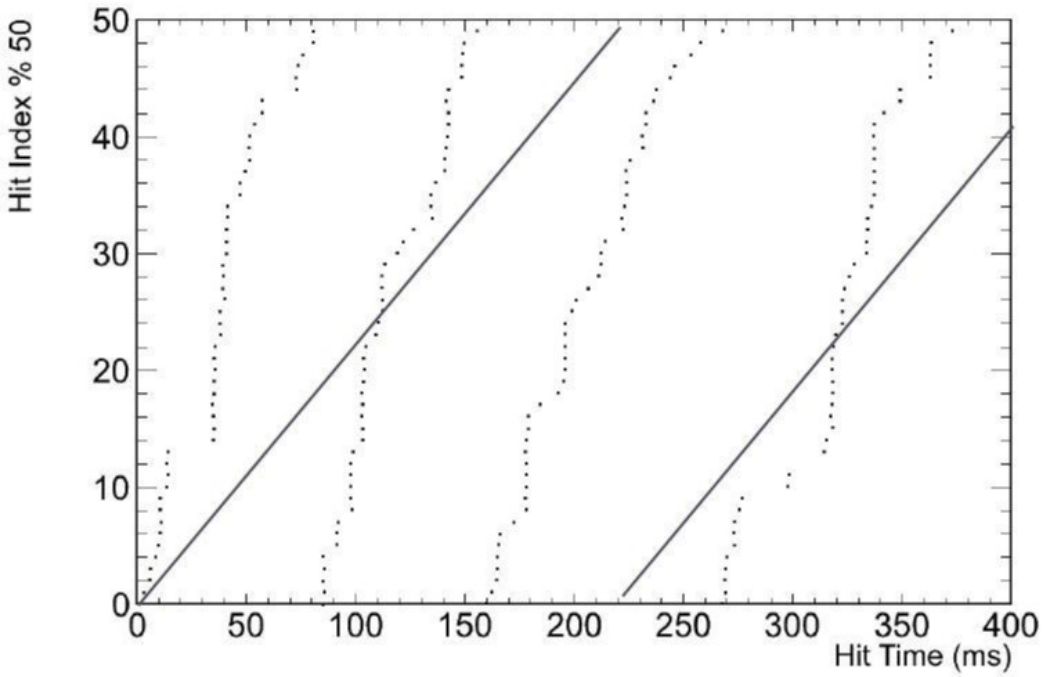


Figure 2.7: Hit index for a single DOM modulo 50 versus hit time. Diagonal lines represent a pure Poisson process with a rate equal to 220 Hz. *Figure taken from [44].*

Figure 2.8 illustrates the same event, a downgoing ν_e, presented on both the left and right sides. The left side displays the hits corresponding to the radiation emitted in the electromagnetic shower following the ν_e charged current interaction. On the right side, the same event is depicted, including all the hits caused by both the physical particle and the noise, which slightly obscures the shape of the actual event [14].

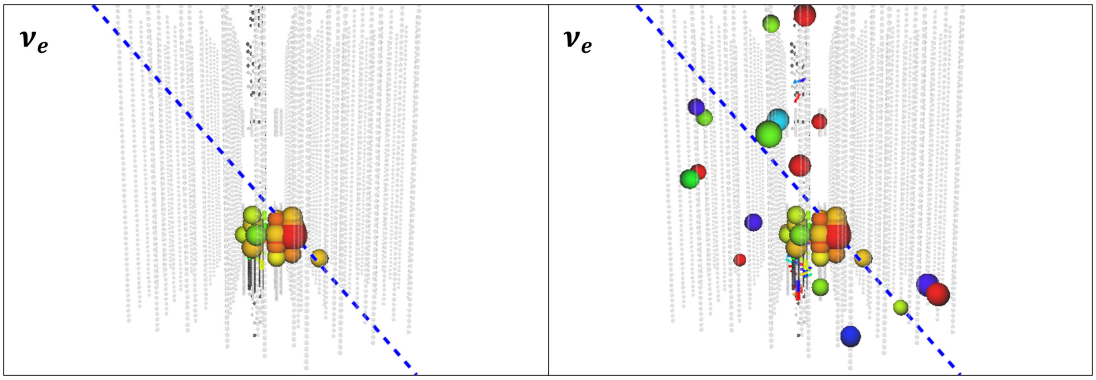


Figure 2.8: Signature of a noise-cleaned versus an uncleaned pulse series. Both subfigures represent the same event, with one only showing the pulses recorded by the detector with their origin in the radiation emitted after the particle interaction (*left*) and the other showing the pulses caused not only by the neutrino interaction but also noise hits (*right*). *Image generated by the author using [37].*

3

FROM DATA TO INFORMATION IN THE ICECUBE UPGRADE

Contents

The IceCube detector captures the signal of over one million atmospheric muons for each detected neutrino. Moreover, with ICU, the rate of noise events is millions of times higher than the rate of neutrino detection. Thus, it is essential to employ a cleaning algorithm that endeavors to retain all neutrino events while discarding those prompted by muons and noise. Furthermore, in instances where the detector is activated by a particle, a majority of the modules will record a signal resulting from noise rather than the particle, thereby blurring the authentic physical data obtained from the event.

The subsequent two chapters detail the methodology employed to convert the radiation detected by the DOMs into a sample dominated by neutrinos. The present chapter starts by addressing the simulation of data for the IceCube Upgrade. As will be demonstrated throughout this chapter, simulation plays a crucial role in the training of classification techniques that can then be applied to real-world data when the ICU starts to be operational. The simulation has been partitioned into four distinct phases, denoted as step1, step2, step3, and step4. The chapter ends by giving an introduction to the machine learning techniques used in high-level event selection processing happening after step 4.

3.1 SIMULATION. STEP1, STEP2 & STEP3

In several fields of physics, including particle physics, where many outcomes of many particle interactions have a probabilistic component, working with simulated data is a very helpful method for problem-solving. To create huge samples of probabilistic events and learn more about their eventual overall distributions, IceCube employs MonteCarlo (MC) simulations. As will be seen in Chapter 4, several machine learning algorithms are trained on simulated data in order to develop an event selection process for ICU, so that when ICU data-taking life begins, they are able to perform on real data. Consequently, it is crucial that simulated data faithfully reproduce both the physical process taking place inside the detector and a precise detector response.

The section gives an explanation of each simulation phase, which is summarized in Table 3.1

Step1	Neutrinos injected within a generation cylinder
event generation	Muons injected within a generation cylinder
	Neutrinos forced to interact within the generation volume
	Optionally targeted muons generation
	Optional KDE used for muon generation
	No detector included at this stage
Step2	Photon from interaction propagated to DOMs
photon propagation	Event cut if no photons reach a DOM
Step3	Noise is generated
detector simulation	Photon to PE conversion
	PMT readout
	Trigger simulation
Step4	Run pulse cleaning algorithms to remove noise pulses
Filtering & Pulse Cleaning	Run DeepCore-Upgrade filter

Table 3.1: Event simulation steps summary.

3.1.1 STEP1 - EVENT GENERATION

3.1.1.1 NEUTRINO SIMULATION

The neutrino simulation for IceCube is done with the software module GENIE [45]. The neutrino Monte Carlo event generator **GENIE** (Generates Events for Neutrino Interaction Experiments) simulates the initial interaction of a neutrino with a nuclear target. Inside a generating cylinder, the software produces

neutrinos (and antineutrinos). A spectral power law generates neutrinos, which are subsequently pushed to interact with matter inside the detector's volume via one of the methods described in Section 1.2.1. A *weight* that compensates for the physical probability of the event to match reality is then calculated using information such as the generation volume, energy range of the generation, and probability of interacting in a particular manner with a nucleon. [13][46].

3.1.1.2 ATMOSPHERIC MUON GENERATION

CORSIKA [47] (COsmic Ray SImulations for KAscade) is a software implemented to simulate particle showers produced by incident cosmic rays. The most accurate software to generate muons is CORSIKA since it accurately simulates the initial cosmic ray, its interactions with an atom or molecule in the atmosphere, the resulting particles generated in the interaction, and the particles generated in the decay of those. However, CORSIKA involves expensive computational resources to produce muons. As a result, in this thesis, an intrinsic software from IceCube, called MuonGun, was employed to mimic the atomospheric muons rather than CORSIKA. **MuonGun** is a quicker muon simulator that propagates the particles to a specific depth using a parametrization of the muon bundle flux created by simulating a large amount of cosmic rays with CORSIKA. Muons from the MuonGun are produced within the boundaries of a generating volume and directed towards the detector [48]. Large samples of muons that trigger the detector can be produced in this way. Furthermore, this program gives users the choice to produce muons using a kernel density estimator (KDE). Muons created using a KDE are made to have specific features, thus not being rejected in the first cleaning phase that seeks to eliminate them. At further steps in the event selection, classifiers must be trained on broad collections of data. This cannot be done with only a few muons to train, so efficiently producing MonteCarlo at some level of processing is necessary. The muons produced by a KDE are also assigned a probability of that event occurring, which is then used to compensate for the large number of muons at a given stage. [49][50][51].

3.1.2 STEP2- PHOTON AND SECONDARIES PROPAGATION

Secondary particles include the leptons generated in CC events, as well as the hadronic and electromagnetic showers resulting from electron neutrinos.The propagation within the medium of atmospheric muons, muons generated in ν_μ CC interactions, and those produced in τ decays is simulated using **PROPOSAL** [52]. The software accounts for various stochastic energy loss mechanisms, such as ionization, electron-pair production, bremsstrahlung, photo-nuclear interactions, and decays. The simulation of hadronic and electromagnetic showers, as well as the propagation of low-energy leptons and muons, is performed using

the software toolkit known as **GEANT4** [53].

Following the propagation of secondary particles within the medium, a simulation of Cherenkov photon propagation is carried out. Photons are tracked using **CLsim**, which uses pre-generated lookup splined tables that account for ice properties as a function of depth [54]. Afterwards, the photons that impact a photomultiplier tube (PMT) are collected and recorded in subsequent stages. [13][51].

3.1.3 STEP3-DETECTOR AND NOISE SIMULATION

3.1.3.1 NOISE SIMULATION

Prior to simulating the response of the DOM, it is necessary to consider the photons that originate from the noise sources, as discussed in Section 2.4. The noise simulation software used is referred to as **Vuvuzela** [43]. The software uses two distinct Poisson distributions to incorporate the thermal and radioactive sources of noise, with each distribution exhibiting a particular rate. The generation of glass scintillation noise follows a log-normal distribution, which has been inferred through data fitting and lacks a clear physical explanation.

3.1.3.2 DOM RESPONSE SIMULATION

The simulation of the response of the DOM to incoming photons is done through the utilization of an IceCube module named **DOMLauncher** [55]. The module emulates the response of a photomultiplier tube (PMT) to incident photons, converting them into photoelectrons (PEs)[1]. The PEs are then processed by the module, wherein they are assigned a weight corresponding to the pulse charge that photon yielded [13][51]. In the context of IceCube, the term **pulse** refers to the integrated electrical charge that is registered by a photomultiplier tube (PMT) over a specified time period. The present definition was introduced in light of the finite time and charge resolution of PMTs, which limits their ability to detect individual photons in certain instances. The charge of a pulse is then normalized and expressed in terms of number of PEs. An event consists of numerous pulses or hits that are gathered throughout the entirety of the detector. The terms **pulse map** or **pulse series** are commonly used to denote this collection of pulses [14].

1 Photoelectron (PE) is a unit of charge defined as the charge recorded by a DOM that corresponds to a single photon hitting the DOM.

3.1.3.3 LOCAL COINCIDENCES AND TRIGGERS

As seen in the preceding section, the response of the PMT to the incidence of photons is the generation of PEs. A DOM is considered to have experienced a **hit** when the charge accumulated by the PMT of the DOM exceeds the threshold of 0.25 PEs. The observed hit rate is dominated by noise hits, with rates per PMT of approximately $\sim$1000 Hz for IceCube DOMs, $\sim$500 Hz for PDOMs, $\sim$ 650 Hz for mDOMs and $\sim$ 490 Hz for DEggs [56].

The starting point for eliminating hits that are not attributed to particles involves keeping only hits that exhibit correlation with hits detected in adjacent DOMs. A Local Coincidence (LC) is defined as a combination of multiple adjacent DOMS registering a detection event within a temporal window of 1 μs. LC are further classified as LC1 when one module up and one module down also register a hit, LC2 when two modules up and two down register hits, and so on. If either LC1, LC2 or higher LC order conditions are met, the LC is said to be a **hard local coincidence** (HLC), with each hit in the HLC an "HLC hit". On the other hand, if a DOM is hit in isolation, it is called soft local coincidence (SLC) [13][57].

An event recorded every time an HLC is detected would result in enormous rates. As a result, some trigger conditions must be met before an event is created. The **simple multiplicity trigger** (SMT), albeit there are other trigger algorithms, is the principal trigger for IceCube. The SMTN searches for N or more HLC hits within a time window without imposing any further locality restrictions. An SMT8 (8 HLC hits required) within a 1.75 μs time window is needed if hits are recorded in ICU modules. The trigger for Deepcore is an SMT3 with a window of 2.5 μs, and for the remaining IceCube strings, it is an SMT8 inside 5 μs [58]. There are other triggers, such as the string trigger for very downgoing muons or the slow particle (SLOP) trigger for hypothetical magnetic monopoles, which are made to search for events with specific and different topologies, but they will not be relevant for this work. There is a specific *readout window* surrounding each trigger, meaning that a time frame before and a time frame after the trigger time are kept. The hits falling into the read-out window, including SLC hits, are all collected into one event that is stored. Multiple readout windows may overlap, and occasionally, multiple "physical" events are combined into a single event [57].

3.2 STEP4-FILTERING AND PULSE CLEANING

After step 3, atmospheric muons or noise still produce the vast majority of the hits, rather than neutrinos. In step 4, pulse cleaning, which tries to get rid of noise hits on events, and the DeepCore Upgrade filter, which tries to get rid of events that leave clear muon prints in the detector, are used.

3.2.1 PULSE CLEANING ALGORITHMS

Cleaning a pulse series means trying to eliminate any noise-induced hits from it. The IceCube and DeepCore experiments employed two distinct algorithms, namely the **static** or **dynamic Time Window** (TW) and the **Seeded Radius Time** (SRT) cleaning.

The TW algorithm selects a specific temporal point, denoted as t, and retains all relevant hits within a designated time frame surrounding t, while discarding any others outside of this temporal window. In dynamic time windowing, the series of pulses is scanned and the time window with the highest charge content is selected. Conversely, the static time window establishes the temporal boundaries surrounding the most significant triggers of the pulse series.

The SRT cleaning algorithm is illustrated in figure 3.1. In SRT cleaning, the HLC hits are initially stored and used as seeds. The algorithm searches for SLC hits that are located within a radial radius R and a time interval T from where the HLC hits seeds are recorded. The SLC hits found are then relabeled as seeds, and the algorithm is launched again. Upon completion of the process, if no additional SLCs are incorporated, the designated hits that were used as seeds will be retained, while the remaining hits will be discarded. Typical values of the radius R and the time interval T used are, for instance in DeepCore, R=100 m and t=400 ns [14][59].

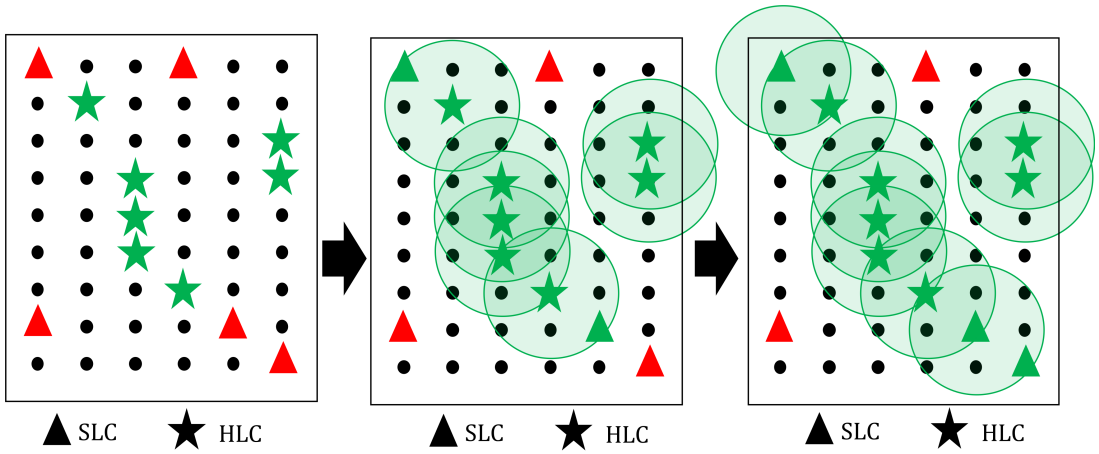


Figure 3.1: The SRT cleaning process is illustrated in three steps. In the first step (*left*), the HCL are identified and established as seeds. In the second step (*middle*), SLC hits within R and T are searched. The last step (*right*) depicts the situation when the process will be over, as there are no more SLC hits within R and T of any of the seeds. *Image generated by the author based on [59].*

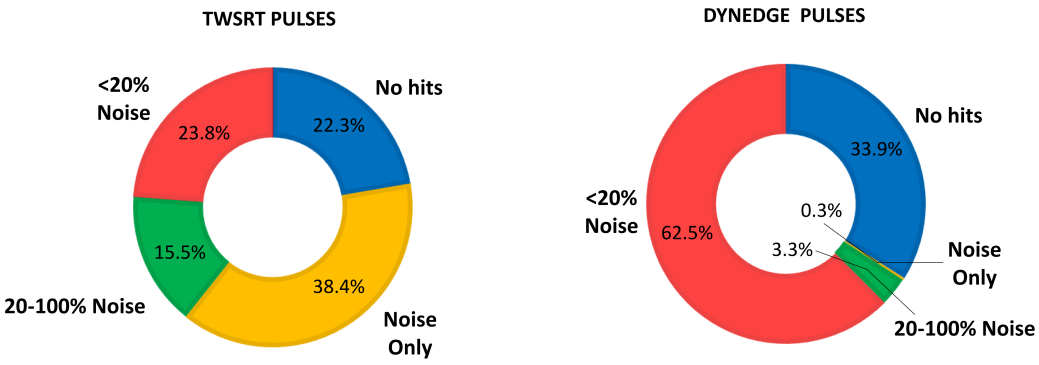
As aforementioned, the ICU will comprise several detection modules, several of which will be equipped with multiple PMTs. The average noise rates per PMT for mDOMs and DEggs are approximately 649 Hz and 488 Hz, respectively. Consequently, the noise rates per optical module for mDOMs and DEggs are approximately 15.5 kHz and 0.9 kHz, respectively [56]. The elevated rates, coupled with the proximity of numerous PMTs, significantly augment the chances of HLC noise hits persisting in the pulse series

despite undergoing TW-SRT pulse cleaning. Consequently, numerous efforts have been made to construct a pulse cleaning mechanism based on machine learning. These efforts have sought to incorporate various techniques, including random forests, convolutional neural networks (CNNs), and graph neural networks (GNNs). Among these techniques, GNNs have shown the best performance [60]. GNNs are machine learning algorithms that can be trained and then used to make predictions or classifications. Dynedge GNN, which will be further introduced in 3.3, is the tool used for ICU pulse cleaning. Dynedge is provided with pulses that are tagged as either noise pulses or particle pulses. Dynedge can then be trained with tagged pulse maps containing relevant information about the each pulse and PMT, such as pulse position and time or PMT orientation, to then being able to classify untagged pulses into noise or particle induced pulses [61].

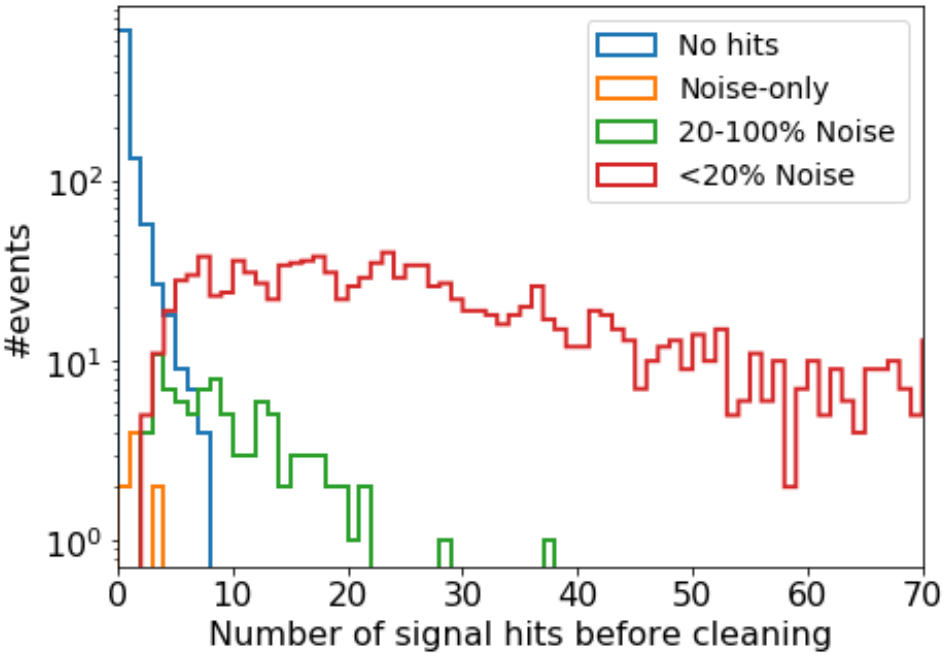
The diagram depicted in Figure 3.2a illustrates the fraction of noise present in the pulse series of various events following the application of TW-SRT cleaning (on the left-hand side) and Dynedge cleaning (on the right-hand side). The figure depicts that merely one-fifth of the TW-SRT pulse series have a dominance of physical hits over noise, with the rest either noise-dominated (more than 20%) or having been completely cleaned and thus removed as events. Later steps of the event selection were challenging due to the dominance of noise-like patterns in the majority of the data. Figure 3.2a illustrates that Dynedge cleaned pulses have two-thirds of the events primarily constituted by particle-induced hits. The remaining one-third of the events are predominantly composed of pulse series, where the algorithm has eliminated all the hits. The distribution of the true number of hits generated by particles is illustrated in Figure 3.2b. The figure depicts four distinct distributions, each of which corresponds to one of the four categories represented in the Dynedge pulse pie chart (right diagram of Figure 3.2a), and it indicates that the Dynedge pulse series categorized as “no hits” did not exceed 9 physical hits prior to undergoing the Dynedge pulse cleaning process. This suggests that the lost events contained minimal physical information within the detector. Thus, by eliminating no-hit events, minimal high-quality neutrino events are lost. [62].

3.2.2 UPGRADE-DEEPCORE FILTER

The goal of the DeepCore-Upgrade Filter is to eliminate events that are evidently caused by an atmospheric muon entering the detector. Initially, the filter scans for events that have been triggered by SMT3 or higher. A SRT cleaned pulse series is given and then separated into distinct series of Fiducial and non-fiducial (Veto) pulses. The Fiducial volume is defined as the set of DOMs located within the Depths of z=-2432 m and z=-2160m (with z=0 representing the ice surface) in the extended DeepCore (EDC), as illustrated in Figure 2.1. The DOMs located outside this particular area are referred to as constituents of the Veto region. The application of the filter involves the computation of the mean and standard deviation of the



(a) Classification of the pulse series according to the fraction of the hits corresponding to noise.



(b) Distribution of the true number of particle hits.

Figure 3.2: Figure 3.2a compares how clean the pulse series are after two different pulse cleaning algorithms are run: the TWSRT pulse cleaning algorithm (*left*) and the Dynedge pulse cleaning (*right*). *Image produced by the author based on [62].* Figure 3.2b shows the distribution of the true number of particle hits for the four different categories of pulses displayed in the Dynedge pulses pie chart. *Plot generated by Jan Weldert taken from [62].*

fiducial hits times, denoted as $\bar{t}$ and σ_t, respectively. Subsequently, the hits that fall within one standard deviation of the mean time are chosen $t \in [\bar{t} - \sigma_t, \bar{t} + \sigma_t]$. The computation of the center of gravity (CoG) of these hits involves the calculation of the average position of said hits ($\vec{r}_{CoG}$). A time for the CoG is then calculated by assuming a simple cascade hypothesis with light propagating from the CoG without scattering, which is the CoG time:

$$t_{CoG} = \frac{\sum_i^N t_i - \frac{|\vec{r}_{CoG} - \vec{r}_i|}{c_{ice}}}{N} \tag{3.1}$$

Where the summation from i to N stands for all the hits within one standard deviation, r_i and t_i for the times and positions of each hit, and c_{ice} is the light speed in ice. The next step is to compute the relative velocity from each of the veto hits to the CoG, defined as:

$$v_{veto,i} = \frac{|\vec{r}_{CoG} - \vec{r}_{veto,i}|}{t_{CoG} - t_{veto,i}} \tag{3.2}$$

Where the subindex i goes over all the veto hits. Finally, if the velocity from the i-veto hit is within the range of 0.25 m/ns to 0.40 m/ns, the the i-veto hit is discarded as it could be causally connected to a muon crossing the detector [6][63].

Table 3.2 sums up the rates of each event type at different points of step 4. After Dynedge pulse cleaning, noise rates have been reduced by a factor $\sim 10^5$. A remarkable drop in muon rate would be expected before and after the DeepCore-Upgrade Filter is run, but the number of muons rejected is almost negligible. The algorithm presented uses the same SRT-clean pulse series and parameters as the ones for the old DeepCore and IceCube-only implementations, but just includes the hits in the ICU DOMs. The noise contamination in the SRT pulse series combined with the high ICU noise rates makes this filter highly inefficient, requiring optimization of the filter and of the ICU tiggers. This poor performance was not of great concern given that later event selection succeeded in removing all these muons. Indeed, the study in this thesis can help to build future triggers and filtering tools that are more strict and can be tuned to cut and lower these extremely high rates at the trigger level while also losing neutrinos that were going to be lost anyway in later steps of the event selection.

	Rate Trigger level (mHz)	**Rate Dynedge Pulse Cleaning (mHz)**	**Rate DCU Filter (mHz)**
$\nu_e + \bar{\nu}_e$ cc	8.84	3.423	3.3048
$\nu_\mu + \bar{\nu}_\mu$ cc	18.4	8.879	8.3070
$\nu_\tau + \bar{\nu}_\tau$ cc	0.539	0.296	0.2828
ν nc	4.1538	1.421	1.36
μ_{atm}	7.62×10^5	2.33×10^5	2.30×10^5
Noise	4.57×10^8	1218	1216

Table 3.2: Rate of events for different neutrino flavors and interaction type, muons and noise at different stages of the step 4 cleaning.

3.3 Machine learning Techniques for IceCube event selection

Following the implementation of the DeepCore-Upgrade Filter and Dynedge pulse cleaning algorithms, the identification of events resulting from pure noise and muons has become more challenging, necessitating a more advanced methodology for event classification. Through the construction of variables associated with the event, such as, for instance, the charge accumulated in distinct regions of the detector, the interaction vertex of the particle, or other relevant factors as elaborated in Section 4.1, discernible variations in the distribution of said variables can be observed among neutrino, noise, and muon events. In certain cases, the variable distributions of neutrinos and noise or muons are separated from each other. Then, by applying a simple cut to the variable value and keeping only events with a higher (or lower) value, the rejection of a significant proportion of noise and muon events is possible. In cases where the distributions lack clear separation, it may be necessary to employ machine learning techniques for classification, such as **Boosted Decision Trees** (BDT). BDT classifiers are computational models that integrate the forecasts of multiple decision trees to construct a definitive classification model.

It is a common practice to construct models by combining multiple decision trees to enhance separation efficacy and mitigate the risk of overfitting. If consecutive decision trees are going to be trained, in order to avoid constructing one identical decision tree after the other, it is necessary to perform a re-weighting of the events rather than retraining on the same data. The process of re-weighting is commonly referred to as boosting. BDTs are constructed through the training of multiple decision trees on various boosted samples. The resulting model is a weighted combination of the individual decision trees. While following the same underlying principle, there are different boosting techniques. The Gradient Boosting (GB) method was employed for the selection of the ICU event. The idea of GB is to train each tree to correct the errors made by the previous one by training a decision tree on the residuals (i.e., the difference between the actual target values and the predicted target values) made by the previous trees. Through iterative processes, each subsequent tree acquires and incorporates information that was previously unaccounted for by its predecessors. The name “gradient” arises from the fact that each tree provides the following one with the gradient of the loss function in the different steps taken, so that the direction and magnitude of the corrections needed are quantified. A loss function is a mathematical mechanism for calculating the discrepancy between a predictor’s output and the expected value. By minimizing the gradient of the loss function, the algorithm seeks to minimize the difference between the actual target values and the predicted target values. Gradient-boosted decision trees have proven to perform well in event classification, as will be shown in Chapter 4. Furthermore, it is of significant importance, considering the vast quantity of data that Icecube accumulates, that once a model is trained, the computations for predictions are quite fast

[64][65][66]. In this work, the gradient-boosted decision trees were constructed using python's library LightGBM [67], in which are named *Light Gradient Boosted Machines* (LGBM), with *Light* present in the name only to remark how fast to train and perform these machines are.

Unlike BDTs, which require reconstructed variables, **Graph Neural Networks** (GNNs) only rely on pulse maps for event classification. In neural networks, a neuron is represented as a numerical value that is linked to other neurons through a mathematical function. Neurons are arranged in different layers, with each neuron connected to all the neurons in the following layer. The value of a given neuron is dependent upon the values of the neurons in the preceding layer, as well as the connections between them. The initial layer of the neural network is referred to as the input layer, where the network receives an input signal. Conversely, the final layer is designated as the output layer, which generates the network's response. Through the process of training a neural network, the parameters of the mathematical function that governs the connections between layers are adjusted and optimized in order to improve the accuracy of the output. The loss function is responsible for assigning a numerical value, referred to as a cost, which quantifies how wrong the prediction is. The process of improving predictions involves modifying connections through the minimization of the loss function.

GNNs are a type of neural network that is founded on the principles of mathematical graph theory. Graphs are structures formed by a collection of *nodes* related by *edges*. Graphs possess the inherent ability to naturally represent complicated data structures. For instance, in IceCube, the nodes are the DOM positions, with attributes such as the charge the DOM collected or the orientation of the PMTs assigned to the nodes. The edges would then define how a DOM relates to its neighbors. The utilization of graphs in neural networks enables easy incorporation of irregular geometries into the edge specification of the graph, freeing the algorithm from artificial constructions. GNNs are thus a suitable machine learning approach to dealing with IceCube problems due to their built-in nature [68].

The Graph Neural Networks for Neutrino telescope Event Reconstruction (**GraphNet**) is a software application that has been developed for the purpose of training and constructing GNNs that can be used in various atmospheric detectors [69]. The model presented in this thesis used for pulse cleaning, event reconstruction, and classification within GraphNet is referred to as **DynEdge**. In this model, graphs are used as well as a data structure. During the training process, a node's acquired value is the result of convolving the values of its neighboring nodes in the graph. Dynedge stands out by its ability to generate a new graph after each convolution by changing the edges linking the nodes and enabling the algorithm to adaptively determine the optimal connectivity of each event in the latent state. [70][61]. In this event selection, Dynedge is used multiple times. Dynedge pulse cleaning has proven to be the most

suitable for the ICU. Additionally, Dynedge has the capacity to be trained for particle ID classification as well as to perform event reconstructions. Trained models are very fast to implement, allowing for event reconstructions at very early stages of event selection when the event rates are still very high. This capacity is an important asset in this thesis, as these reconstructed quantities can be included as variables inside LGBMs, enhancing their performance.

4

ICECUBE UPGRADE EVENT SELECTION

Contents

All the pre-processing chain for IceCube Upgrade events have been covered in Sections 3.1 and 3.2. The rates listed in Table 3.2 indicate that the rate of muons and noise remains significantly greater than that of neutrinos. Once step 4 is run, all the algortihms employed to continue with the event selection are englobed in a project under the name QUESO, which stands for Quick Upgrade Event Selection for Oscillation (QUESO). QUESO is divided into two levels. The first one, called QUESO Level 3, eliminates more than 99% of the noise and muons through the implementation of straight cuts on variables, outlined in Section 4.1. Following QUESO Level 3, the subsequent step, QUESO Level 4, aims to further eliminate any remaining atmospheric muons present in the sample. This task is more challenging, as the remaining muons are the most difficult to separate from neutrinos. QUESO level 4 implements BDTs in conjunction with GraphNet to build a classification model to finally isolate the neutrino sample.

4.1 QUESO LEVEL 3

QUESO Level 3 aims to eliminate residual noise and muons that retain their muon-like shape following step 4. Thankfully, it is possible to achieve satisfactory outcomes by simply making cuts to certain variables. The used variables are listed here:

- **Vertex Guess Z:** It is the z-component of the position of the first hit in the Dynedge clean pulse series. Low-energy interacting neutrinos are expected to leave hits mostly in the fiducial volume. Muons, on the other hand, are expected to leave long tracks, and thus, downgoing ones will leave hits in the top part of the detector. This can be seen in Figure 4.1, where a significant fraction of the muons interact with a reconstructed interaction vertex at the top part of the detector. By keeping the sample of events with a Vertex Guess Z variable smaller than z=-120 m in the Icecube coordinate system[1] one can reject $\sim$ 90% of the muon rate while keeping over 99% of the neutrino one [71].

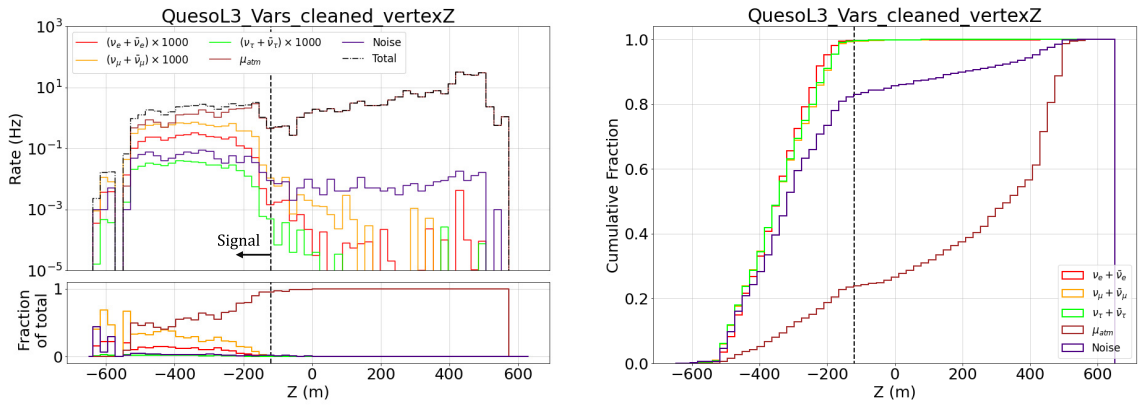


Figure 4.1: Distribution of the vertex guess Z variable for different particles (*left*). Cumulative and normalized distribution of the variable (*right*).

- **Uncleaned length:** This variable is defined as the difference in time between the last and the first pulse in the uncleaned pulse series (i.e., the pulse map at trigger level). Muon events will tend to last longer than neutrino events. The reason is that a muon event's starting time will be close to the moment when the muon enters the detector and will last until it decays or leaves the detector. On the other hand, because neutrinos enter the detector undetected, the modules won't begin to capture light until the neutrino interacts and creates a particle or radiation shower. For this same reason, coincident events in which a neutrino and muon event or two muons interact simultaneously last longer as well. Within the oscillation working group, there is no interest in these events, and thus, a cut that gets rid of them is also beneficial. It can be seen in Figure 4.2 that neutrino events are distributed towards shorter times compared to muons. By only keeping events with an uncleaned length smaller than 15000 ns, almost 50% of the muon rate is eliminated while keeping over 90% of the neutrino rate[71].

- **Cleaned Length:** Analogous to the uncleaned length, the cleaned length is the difference between the time of the last and the first pulse in the cleaned pulse series. Even if a cut in this variable attempts to get rid of events of long lengths, just like the uncleaned length cut, the difference

1 The depth d from the ice surface relates to the z component of the IceCube coordinates via d=1948.07 m-z

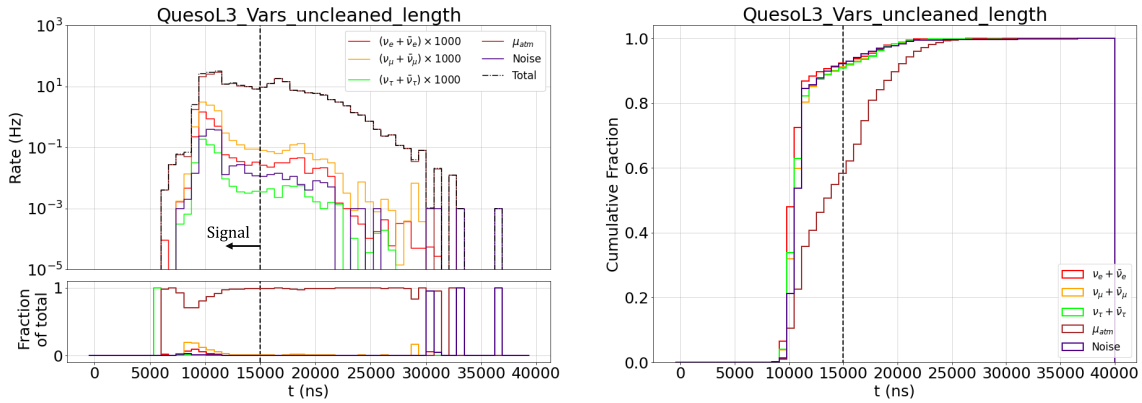


Figure 4.2: Distribution of the uncleaned length variable for different particles (*left*). Cumulative and normalized distribution of the variable (*right*).

between the clean and the uncleaned pulse series is notable, and Dynedge is not attempting to clean hits caused by muons, just pure noise ones. Thus, even targeting the same events, by only keeping events with an uncleaned length shorter than 3000 ns, 25 % of the muons are cut while keeping over 98% of the neutrinos [71].

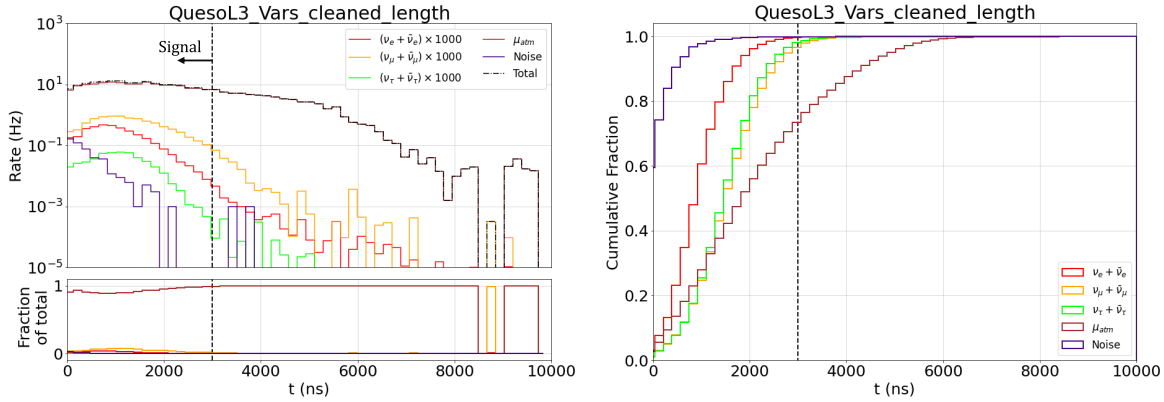


Figure 4.3: Distribution of the cleaned length variable for different particles (*left*). Cumulative and normalized distribution of the variable (*right*).

- **Number of DOMs with a hit:** While the previous variables attempted to get rid of muon events, the target now is the events caused by pure noise. Most of the events caused by pure noise come from a few noise hits within a short time window with nearby DOMs that have triggered the detector. Thus, most of the noise events are likely to have only a few hits. By keeping events that have hit more than 4 different DOMs, 97% of the noise is rejected while keeping more than 93% of the neutrinos [71].

- **Number of hits in the fiducial volume:** In an analogous way and based on the same principle, a cut in the number of PMTs' hits can target noise events. In this variable, the number of DOMs hit

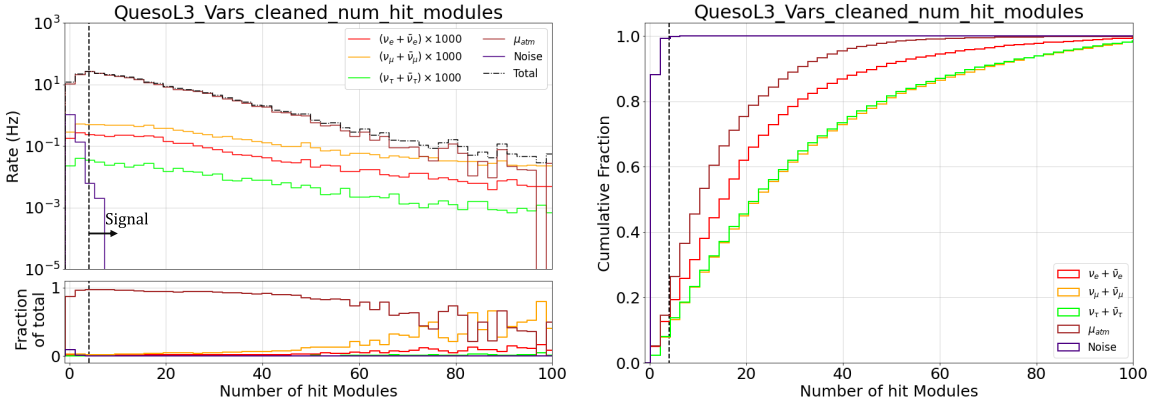


Figure 4.4: Distribution of the number of hit DOMs variable for different particles (*left*). Cumulative and normalized distribution of the variable (*right*).

is replaced by the number of PMTs with a hit (which for the ICU differs as there are DOMs with multiple PMTs) and only the fiducial volume ones are taken into account. Figure 4.5 shows how by rejecting the events with less than 6 hits in the fiducial volume, over a 99% of the noise is vanished while keeping over 85% of the neutrinos [71].

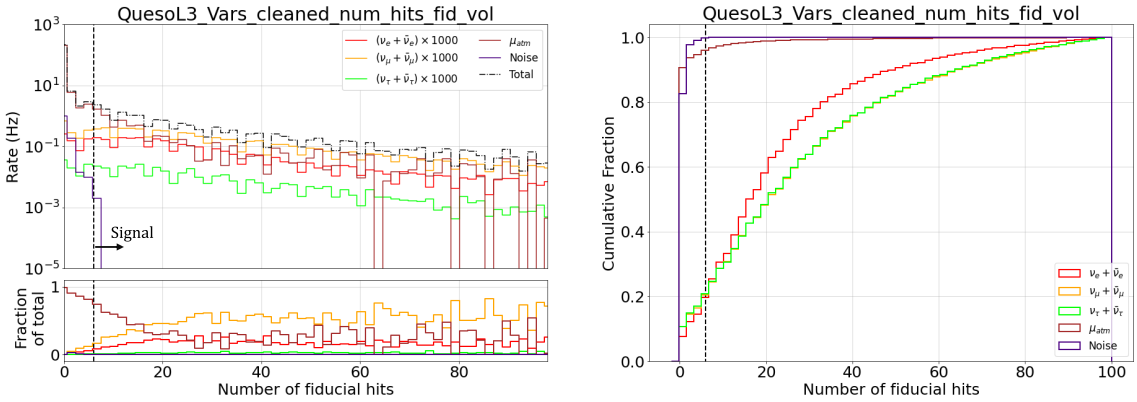


Figure 4.5: Distribution of the number of fiducial hits variable for different particles (*left*). Cumulative and normalized distribution of the variable (*right*).

	Rates at step 4	Rates at QUESO Level 3
ν_e+ CC	3.30 mHz	2.54 mHz
ν_μ CC	8.31 mHz	6.34 mHz
ν_τ CC	0.28 mHz	0.22 mHz
ν NC	1.36 mHz	1.05 mHz
μ_{atm}	230573	783 mHz
Noise	1216 mHz	3 mHz

Table 4.1: Rate of events caused by different neutrino flavours and interaction type, muons and noise after step 4 and after QUESO Level 3.

Variable	Target	Cut
Vertex Guess Z	Muons	$<$ -120 m
Uncleaned Length	Muons and Coincidents mHz	$<$15000 ns
Cleaned Length	Muons and Coincidents mHz	$<$3000 ns
Number of DOMs hit	Noise	$\geq$ 4
Fiducial hits	Noise	$\geq$ 6

Table 4.2: Summary of the variables used in of QUESO level 3 cuts, with the targeted event classes and the value used for the cut.

4.2 QUESO LEVEL 4

As indicated by the data presented in Table 4.1, the rate of atmospheric muons remains nearly two orders of magnitude greater than the rate of neutrinos. The subsequent section presents a detailed description of the methodology employed to eliminate muons while minimizing the exclusion of neutrinos.

The event selection process for the ICU involves a noise simulation consisting of 1000 files, each containing 100 time windows of 10 ms duration. It can be inferred that each file is representative of a duration of one second in the simulation of the detector. Thus, one event in one of the thousand files corresponds to roughly 1 mHz of noise rate. The decrease in the frequency of noise from millions of mHz at the trigger level to approximately 3 mHz at QUESO Level 3 indicates that just three noise events remain from the million of simulated events. The QUESO Level 4 model has been designed to effectively reject muons and keep neutrinos, forgetting about noise presence. QUESO Level 4 will eliminate the aforementioned three events even if it was not trained for the purpose, but it is possible that with an increased number of Monte Carlo noise simulations, the final noise rates may not reach zero even if reported as such in the section.

With the aim of serving as a muon-neutrino classifier for QUESO Level 4, two different models were trained: one using Dynedge and another using two consecutive BDTs.

4.2.1 DYNEDGE CLASSIFIER

GraphNet has demonstrated significant usefulness and suitability as a tool for addressing IceCube-related issues. Dynedge has the capability to undergo training in order to construct a model that effectively distinguishes the residual muons from the neutrinos. To train the model, a large amount of muons were generated using a KDE, so that they could pass Level 3. The Dynedge model, now referred to as Dynedge Level 4, was developed to take a pulse series as input and generate a score between zero and one. A score closer to one indicates a higher likelihood of the input being a neutrino, while a score closer to zero suggests a higher probability of it being a muon. Then, a specific score value is established to act as a threshold for selection. The distributions of scores assigned by the classifier to muons and neutrinos are illustrated in Figure 4.6. The phenomenon of muon accumulation in proximity to zero is observed, while neutrinos tend to accumulate in the vicinity of one. The graphical representation depicted in Figure 4.7 illustrates the proportion of the rate retained for muons and neutrinos across various score values. By adjusting the classifier cut to approximately 0.87, the muon rate can be reduced to match that of the neutrino rate, resulting in the rejection of approximately 99% of muons while retaining approximately 85% of neutrinos. By implementing a cut setting with greater aggression, specifically by setting the score threshold to 0.998, it is feasible to obtain a final sample containing ten times more neutrinos than muons while retaining nearly 60% of the neutrinos.

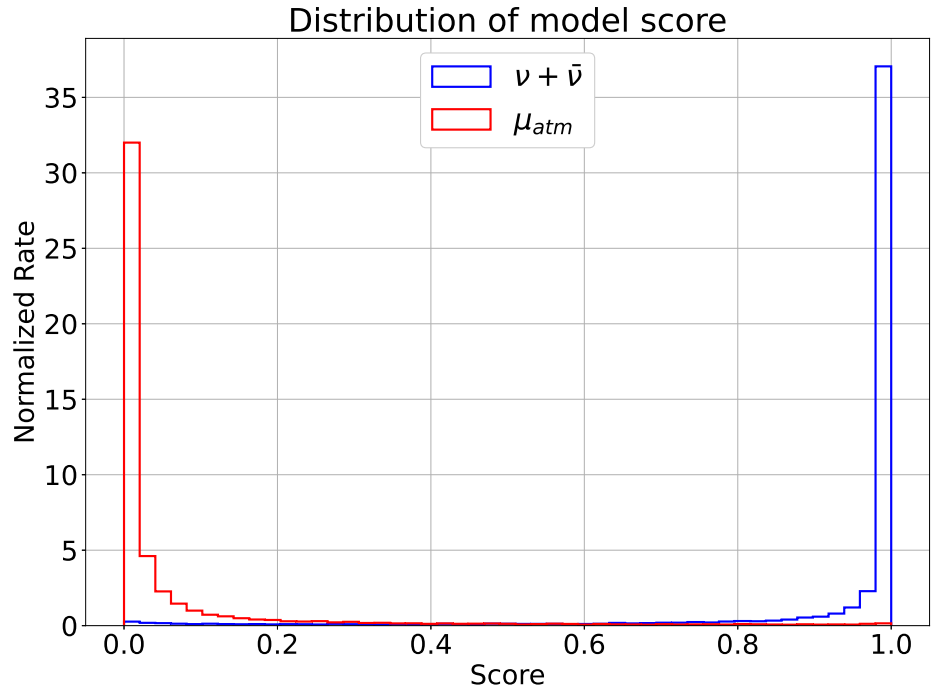


Figure 4.6: Distribution of the Score given by the Dynedge Level 4 classifier.

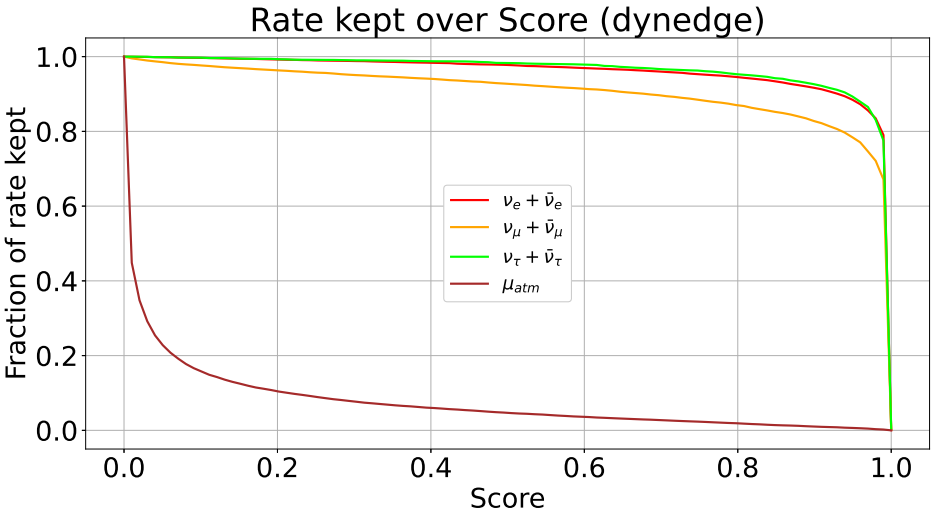


Figure 4.7: Fraction of the rate kept for each event class as a function of the score given by Dynedge level 4 classifier.

4.2.2 TWO LGBMS CLASSIFIER

In parallel with the development of the dynedge classifier model, an algorithm using LGBMs for predictive purposes was also developed. Despite the differences between this machine learning technique and that of

Dynedge, both of them ultimately produce an event score that distinguishes between neutrinos and muons in an analogous way. In contrast to dynedge, a LGBM is unable to generate predictions only based on pulse map data. Instead, it requires quantities that are computed over the pulse map in a way similar to those computed for the cuts in QUESO Level 3. One of the primary benefits of performing computations directly on the pulse series is that it reduces the loss of information that occurs during the conversion from raw data to variables. On the other hand, the utilization of distinct variables that represent the data at this particular stage of processing facilitates checking for agreement between the real data and the simulated one used for training the algorithms. The present study does not permit a comparison as the ICU has not yet been deployed. However, it is crucial for future investigations to verify the agreement at various stages, as failure to do so may result in inadequate performance of the trained models when exposed to authentic data.

As the section title suggests, two consecutive LGBM models were trained for sequential utilization, with the selection process based on two thresholds, one for each individual model's score. The idea behind this approach was to develop numerous variables that could potentially determine whether the source of the event was or was not a neutrino. These variables differ in their resolution power, with some possessing a high degree of resolution while others have a more subtle one. Furthermore, the variables account for distinct characteristics of the event (charge-related, position-related, shape-related, etc.). When constructing a singular LGBM model, the algorithm employed a selection process that prioritized variables with higher classification power, resulting in a limited consideration of certain variables that also possessed classification power. Consequently, using only one LGBM led to a loss of information. The strategy followed involved training a model, referred to as **1stLGBM**, followed by training another model using the muons that the 1stLGBM struggled to identify, referred to as **2ndLGBM**.

4.2.2.1 1stLGBM

The 1stLGBM comprises a set of nine different variables, shown in Figure 4.8. Five of them correspond to the differing amounts of charge that are accumulated in distinct regions of the detector, such as the veto and fiducial volumes, or at the upper part of the detector. The Dynedge pulse cleaning algorithm, as discussed in Section 3.2, provided a quick first reconstruction of zenith angle and energy, along with a preliminary classification system that distinguishes between neutrino and non-neutrino events (very broad and thus not discussed), as well as track and cascade event types. The following is a description of the variables being used:

- **L4 FiducialCharge:** This variable represents the amount of charge collected in the cleaned pulse series for an event within the fiducial volume extended DeepCore. Among the presented variables, it has been found to have the least classification power. However, it still takes into account the

fact that the majority of low-energy neutrinos will be primarily detected inside the fiducial volume, unlike muons [6].

- **QAbove200:** This variable quantifies the collected charge in the detector at heights above Z > -200m in the IceCube coordinate system, prior to the DeepCore trigger in the cleaned pulse series. Since muons, being downgoing particles that traverse the ice and emit radiation, are expected to deposit some energy in the upper parts of the detector when they reach and trigger DeepCore or ICU, the distribution of QAbove200 for muons tends to peak at higher values of charge, as shown in Figure 4.8b [6][72].

- **L4 RTVeto250Charge:** This variable is computed similarly to the SeededRT pulse cleaning method, which was explained in Section 3.2.1, resulting in different cleaned pulse series. For the computation of this variable, the pulse series includes hits that have a neighboring hit within a radius of R = 100 m and a time window of T = 250 ns. The total charge collected from the hits in the Veto region is computed, and this total charge is referred to as "L4 RTVeto250Charge". Since muons produce hits along their path, this variable tends to have higher values for muons compared to neutrinos, as depicted in Figure 4.8c [6][72].

- **L4 VICH nch:** VICH stands for Veto-Identified causal Hits, and nch represents the number of cleaned hits, i.e., the number of hits in the cleaned pulse series. The algorithm is given a value for the radius and time window and determines whether the hits in the veto region are causally connected, indicating that a muon is more likely to have produced them than noise. Similar to the previous variables, the number of hits for muons tends to peak at higher values compared to neutrinos. Refer to Figure 4.8d for the distribution of this quantity [6][72].

- **L4 vetoFiducial RatioCharge:** This quantity represents the ratio of the total charge collected in the veto region to the total charge collected in the fiducial region, both using the Dynedge cleaned pulse series. Low-energy neutrinos primarily interact and produce hits around the interaction position, typically inside the fiducial volume, due to the lower energy detection threshold resulting from the higher density of DOMs. However, muons leave hits not only in the fiducial volume but also along their path. Consequently, the fraction of veto charge collected relative to the fiducial charge tends to be higher for muons than for neutrinos. Figure 4.8e illustrates the distribution of this quantity for muons and neutrinos. The y-axis of the plot is logarithmic, as most events appear as neutrinos, while identifying muons is more challenging. This implies that the pulse maps of most events show minimal charge collected outside the fiducial volume [6][72].

- **Dynedge energy:** This variable represents the predicted energy value by Dynedge. The Dynedge reconstruction is a fast process that can be performed early in the event selection chain, even when

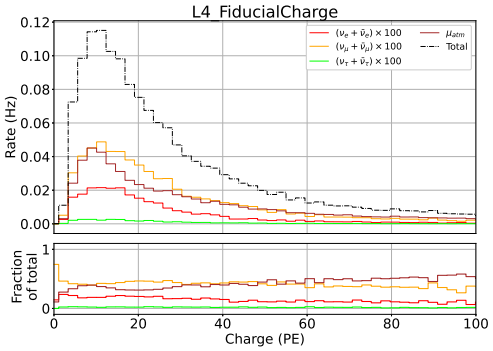
the event rate is high. Using a reconstructed variable in event selection is highly advantageous and extremely useful (see Figure 4.8f).

- **Dynedge nu pred:** This variable is a score ranging from 0 to 1 and operates similarly to the one described in Section 4.2.1. The model used to generate this score was trained on a data sample at step 4, resulting in lower muon rejection capability at Level 3 compared to the models presented in this section. However, this variable can still be employed within a LGBM model to enhance its ability to distinguish between muons and neutrinos.

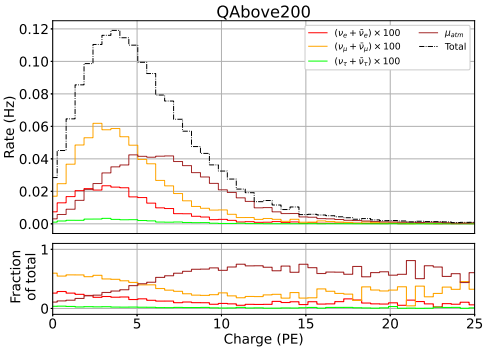
- **Dynedge track pred:** This quantity is also a score between zero and one. A value close to zero suggests that the event is likely to be a cascade event, while a value close to one indicates a higher likelihood of being a track event. Muons leave track-like signatures in the detector, making them more likely to be classified as track-like events. This provides the variable with the capability to distinguish muons.

- **Dynedge zenith pred:** Neutrinos can arrive at the detector from all directions, whereas muons produced in the atmosphere can only reach the detector from the top. Due to their absorption by the Earth, muons cannot arrive from the bottom portion of the detector. Figure 4.8i demonstrates that muons rarely have a reconstructed zenith angle of $zen > \pi/2$, which would indicate arrival from the bottom part of the detector. Therefore, the reconstructed variable exhibits excellent classification power in distinguishing muons from other event types.

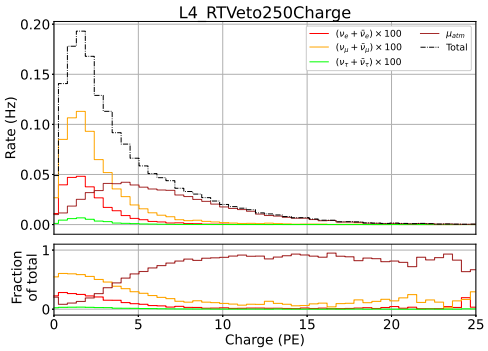
The model is tuned through the modification of its hyperparameters (see Table A.1) to prevent overfitting while maximizing classification accuracy. The training of the model was conducted on 65% of the entire dataset, while the remaining 45% was reserved for testing purposes. The model's predictive performance was evaluated on two distinct sets of data: the training sample itself and the testing sample. The accuracy of the model on the training sample was found to be 94.08%, while the accuracy on the testing sample was 93.85%. This reported accuracies are the ones of the final model provided, however, with the aim of checking for overfitting, the prediction accuracy was computed on 20 different models trained using the same hyperparameters but over 20 different randomly chosen training and testing samples. If the accuracy of prediction on the different train and test samples varied, differing notably between train and test, it could be a sign of overfitting. However, as seen in Figure A.1a, there was not such a variation.
Each of the events has a weight accounting for the feasibility of that event being detected in IceCube. These weights are used when training the algorithm so that the neutrinos that will have a greater impact on the total rate are prioritized by the algorithm to be kept. Muon also have a weight accounting for their feasibility, but the algorithm is trained to have a preference to reject and not accept those having a bigger contribution to the total muon rate.



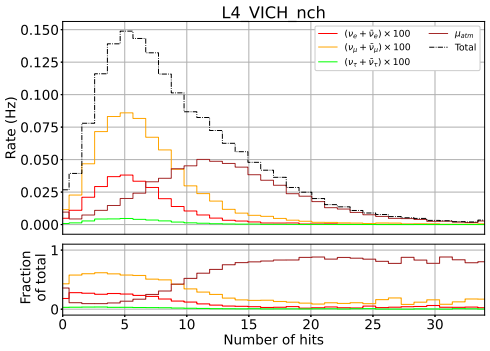
(a) QUESO level 4 Fiducial Charge



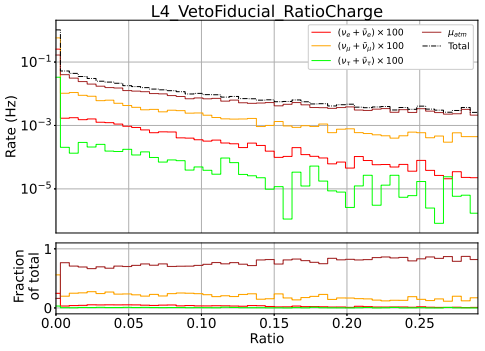
(b) QUESO Level 4 QAbove200



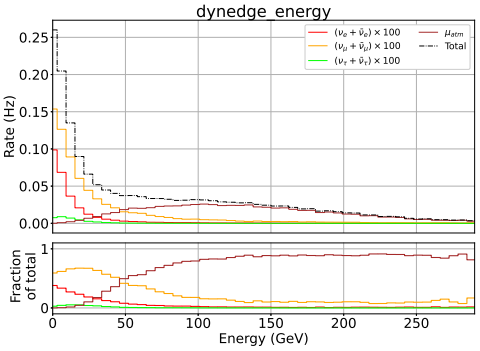
(c) QUESO Level 4 RTVeto250Charge



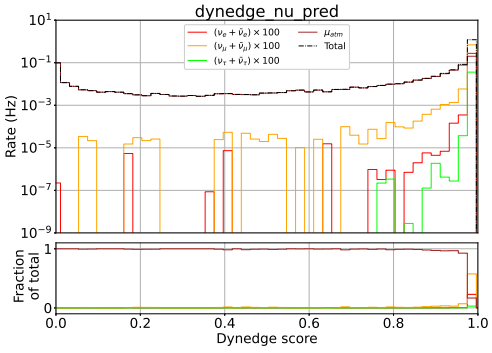
(d) QUESO Level 4 VICH nch



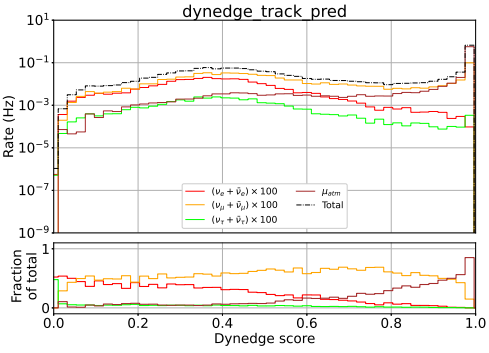
(e) QUESO Level 4 VetoFiducial Ratio Charge



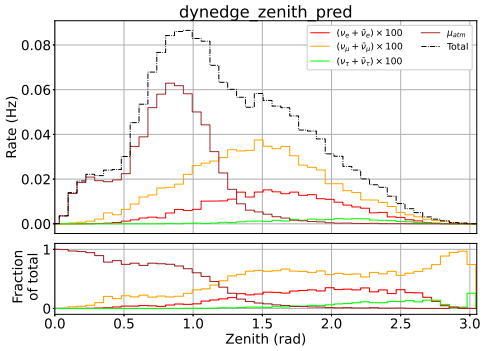
(f) QUESO Level 4 dynedge energy



(g) QUESO Level 4 Dynedge nu prediction



(h) QUESO Level 4 Dynedge track prediction



(i) QUESO Level 4 Dynedge Zenith prediction

Figure 4.8: 1^{st}LGBM variable distributions.

Figure 4.9 shows the values of the feature importances each of the variables acquires in the model. Feature importances is a way to quantify the impact of each feature on the model's output. The method used to compute their values is the so-called "split gain" method. The split value gain represents the reduction in the loss function achieved by spitting the data on a particular feature at a particular value. The algorithm iteratively evaluates different split points and selects the one that minimizes the loss function. The split gain is accumulated through all the iterations and then used to compute the importance score, which is a useful metric to quantify which variables are more influential in the model. From the aforementioned variables, as expected, the reconstructed variables and the Dynedge predictions are the ones acquiring greater importance. The ones related to charge distribution are also of significant importance and are taken into account by the classifier.

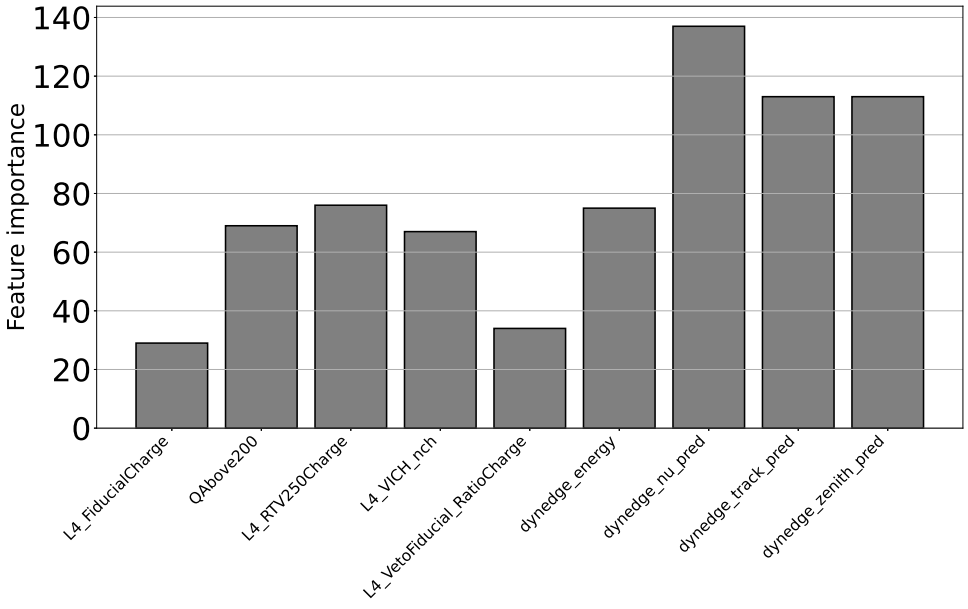


Figure 4.9: 1^{st}LGBM variable feature importance

The performance of the 1^{st}LGBM can directly compete with the Dynedge Level 4 classifier, even if not all the muons have been targeted yet. The 1^{st}LGBM looks for muons that were clearly identified as not neutrinos by Dynedge at a pulse cleaning level, plus the ones that really left many hits outside the fiducial volume but were not properly cleaned in previous steps due to a poor performance of the DeepCore-Upgrade Filter. Not all the muons in the remaining sample after 1^{st}LGBM are necessarily neutrino-like events, and by introducing the 2^{nd}LGBM, downgoing muons, long tracks, and long duration events are then targeted. So before comparing to Dynedge, let's introduce the 2^{nd}LGBM.

4.2.2.2 2^{nd}LGBM

The construction of the model follows the same approach as the 1^{st}LGBM model, albeit with a modification in the used variables and the data over which it was trained, which are the events with a score given by the $1st$LGBM higher than 0.85. This second classifier is comprised of a total of twelve distinct features. Eight variables have been designed to differentiate events based on their shape and location within the detector. Additionally, there's a variable that accounts for the quantity of DOMs hit. The tenth variable pertains to the temporal shape of the event, while the eleventh and twelfth variables are reconstructed dynedge quantities that have been reused due to their underutilization in the previous classifier. It is crucial to bear in mind that this second classifier was trained on a sample set with less muon presence, as the training sample is the one passing the 1^{st}LGBM. Furthermore, it was not designed to function independently as a classifier, but rather to eliminate the muons that the initial 1^{st}LGBM model struggled to identify. The twelve variables are:

- **L4 ToIParams evalratio:** The acronym "ToI" stands for Tensor of Inertia, which is a mathematical concept used to describe the distribution of mass in a rigid body [73]. In the context of an IceCube event, the charge distribution along the detector can be characterized using this mathematical framework. By computing the eigenvalues of the tensor, we can obtain a measure of how elongated or spherical the charge is distributed along three orthogonal axes. The smallest eigenvalues correspond to the longest axes, and vice versa. This variable is defined as the ratio of the smallest eigenvalue to the largest eigenvalue [63].

- **L4 first hlc rho:** This variable represents the radial distance, measured on the x-y plane, from the first HLC to string 36, which is assumed to be the center of IceCube. For muons entering from the lateral edges, the first HLC is typically located further away from the center of the detector compared to a neutrino event that first interacts within the fiducial volume. Figure 4.10b demonstrates this characteristic. The irregular shape of the distribution is due to the absence of DOMs at certain radial distances, resulting in the impossibility of having an HLC at those distances [6].

z sigma: The variable represents the root mean square (RMS) of the vertical position of all the hits in the cleaned pulse series, computed using Equation 4.1.

$$z_{RMS} = \sqrt{\frac{1}{n}(z_1^2 + z_2^2 + ... + z_n^2)} \tag{4.1}$$

with z_{RMS} the RMS value of the z-component of the position and z_i the z-component of the position of each of the individual hits. The neutrino distributions shown in Figure 4.10c are narrower and centered towards lower z-distances from the z=0 point of the IceCube coordinates [74].

- **z travel:** It accounts for the total distance from the first hit in the clean pulse series to the last in chronological order. Negative values of the value of the variable mean that the last hit is below the first one. While neutrinos can arrive at the detector isotropically, leaving more symmetrical signatures in the detector, muons are always downgoing and leave hits along their trajectory. Thus, negative values are expected for muons, while for neutrinos, the distributions are expected to be centered somewhere close to zero. This behavior is indeed observed in Figure 4.10d [6].

- **cog z:** The variable stands for the z-component of the center of gravity (CoG) of the clean pulse series. Neutrino distributions tend to peak either at the top, mid, or bottom part of the fiducial volume (see Figure 4.11a). The higher rates of neutrinos with CoGs at the bottom of the fiducial volume were expected as there are no upgoing muons and the ones reaching such depths would have most likely left hits on the upper parts of the detector, causing the CoGs to be at upper locations. The peak at depths corresponding to the middle of the fiducial volume was expected for all particles. The higher concentration of events with a CoG at the upper edge of the fiducial volume corresponds to events with hits at the top part of the detector but with the majority of their hits in the ICU and DeepCore DOMs.

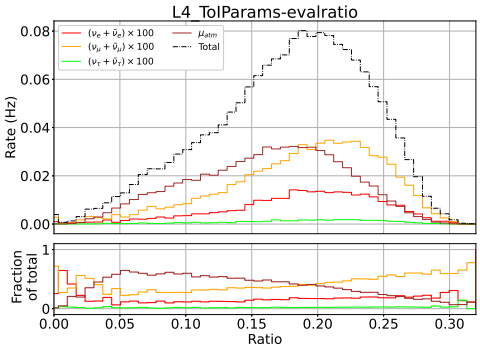
- **L4 event duration:** It is defined as the final time t_f minus the initial time t_i and then divided by t_f. Here, t_f is defined as the time taken by the detector to collect 95% of the total change of the event, and t_i the time taken to collect 5% of the total charge. In muon events, more time is expected to pass from the first to the last hit, as seen in Figure 4.11c.

- **L4 3Zones median slope:** The variable divides the time elapsed from the first to the last hit in the clean pulse series into three equal time intervals. Then, for all the hits contained in one of the time intervals, the median position is computed, obtaining one point for each of the three time intervals. The variable value is nothing but the slope of the linear fit of these three points. For the same reason that for the z travel variable, the distribution for muons is expected to get negative values while the

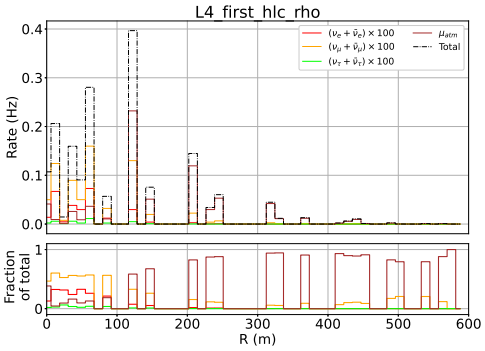
ones for neutrinos are expected to pile up near zero. The spike in Figure 4.11d is due to the fact that events without hits in one of the time intervals are considered to have a slope equal to zero.

- **L4 separation in cogs:** In an analogous way to the previous variable, for the separation in cogs, the time elapsed from the first to the last hit in the cleaned pulse series is divided into four different time intervals of equal length. The distance between the CoG of the hits in the first time interval and the CoG of the hits in the fourth time interval is what is called separation in cogs. The distance will be bigger for track-like events and thus for muons, as can be seen in Figure 4.11e [72].

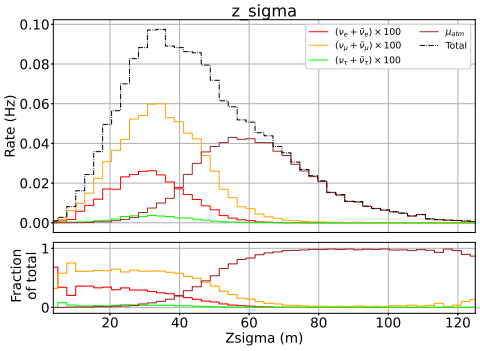
As mentioned, the model predicts using 12 variables. The remaining four variables are **QuesoL3 Vars cleaned num hit modules**, **QuesoL3 Vars cleaned vertexZ**, **dynedge track pred** and **dynedge zenith pred**, which were used in previous steps of the event selection and described in previous sections.



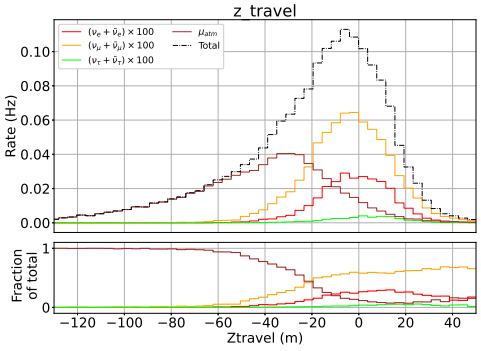
(a) QUESO Level 4 ToI evalRatio



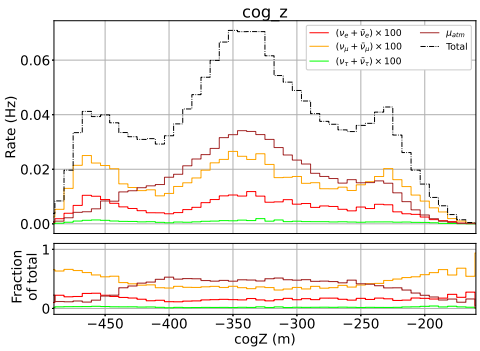
(b) QUESO Level 4 first hcl rho



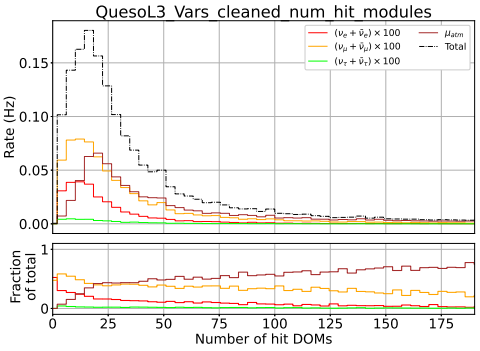
(c) QUESO Level 4 z sigma



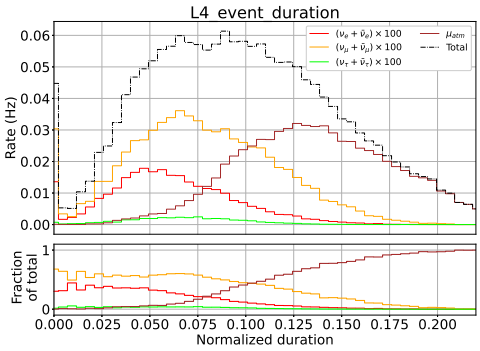
(d) QUESO Level 4 z travel



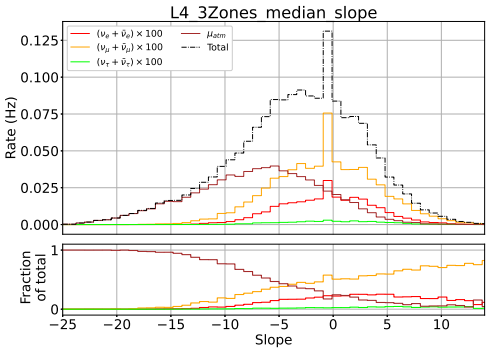
(a) QUESO Level 4 cog z



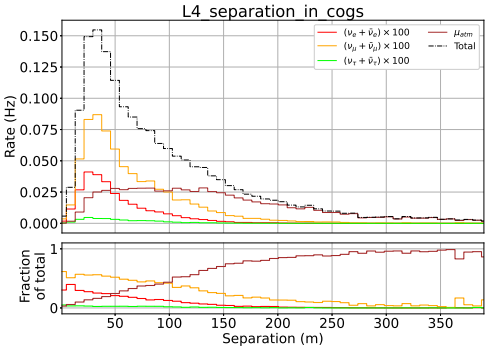
(b) QUESO Level 3 Cleaned number of hit modules



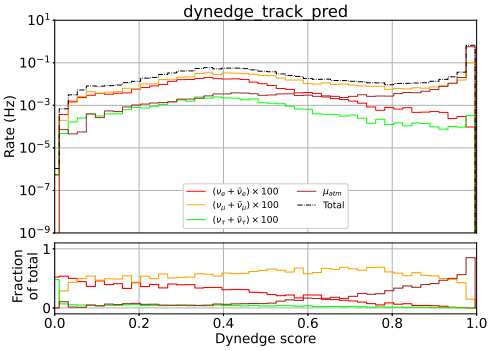
(c) QUESO Level 4 Event Duration



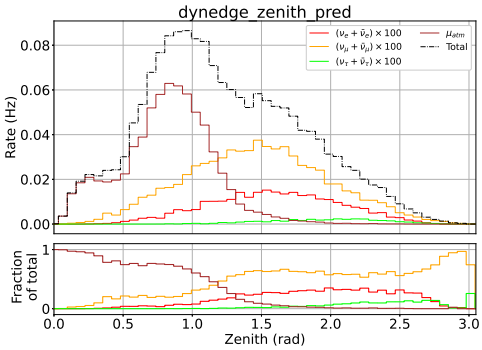
(d) QUESO Level 4 3Zones Median Slope



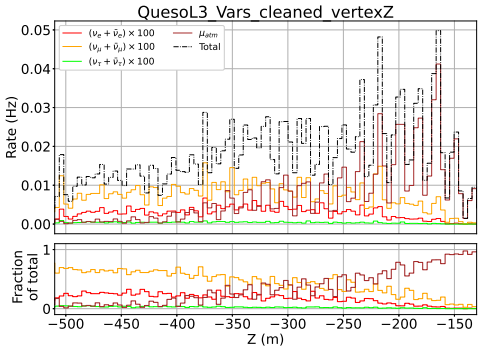
(e) QUESO Level 4 Separation in CoGs



(f) Dynedge track prediction



(g) Dynedge zenith prediction



(h) QUESO Level 3 Vertex Guess Z

Figure 4.11: 2^{nd}LGBM variable distributions.

The hyperparameters chosen to train the model are shown in Table A.1. Again, the train and test samples were divided so that 65% of the total sample was used to train the model, while the remaining 45% was used to test the model's performance. The same test to check for overfitting that was performed for the 1^{st}LGBM is done for this model without finding any clear sign of it (see Figure A.1b). The accuracy of the model's predicting over the training sample was 78.99%, with a very similar one for the testing sample, which was 78.92%. The feature importances of the variables of the 2^{nd}LGBM model are illustrated in Figure 4.12. The event duration is identified as the most important variable. It does not come as a surprise; it is seen in Figure 4.11c that the distribution of this quantity for muons differs notable from the one for all the other neutrino flavors. Also, even if the features importance is not a measure of correlation, this is the only temporal variable in the two LGBMs, with none of the other variables attempting to distinguish muons and neutrinos using times.

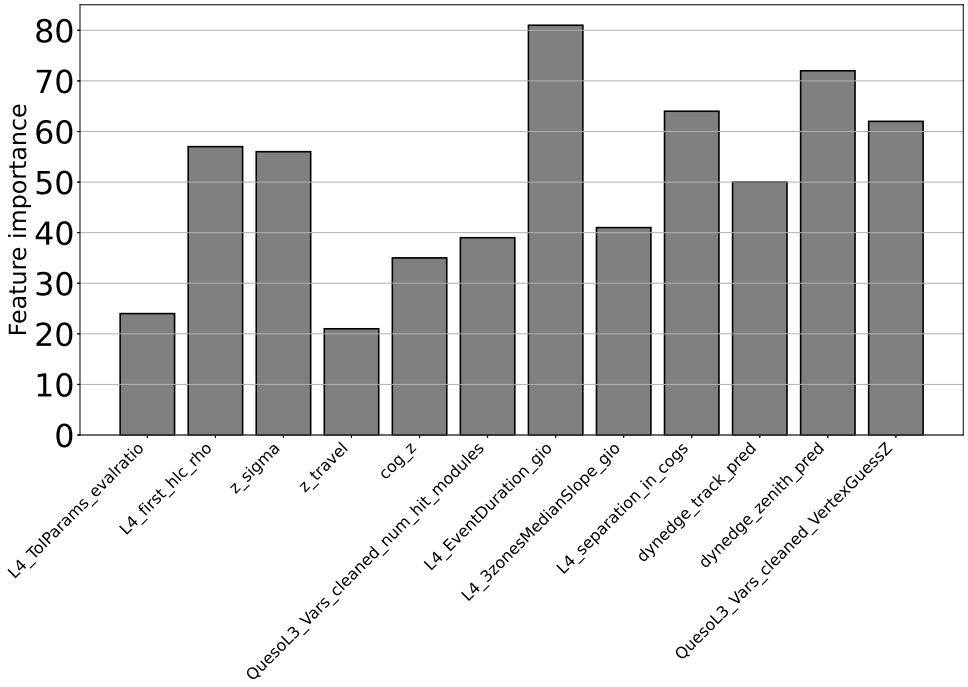


Figure 4.12: Feature Importances of the 2^{nd}LGBM.

4.2.3 COMPARING AND COMBINING THE MODELS

It is not easy to compare the two classification paths, the first using the Dynedge Level 4 classifier score and the second combining two scores, one for each LGBM. Because the second classification path depends on the combination of two scores, now referred to as *score1* and *score2* for the 1^{st} and 2^{nd}LGBM, respectively, it is necessary to first identify the combination that would directly compare to a value in the Dynedge Level 4 Classifier score, now referred to as scoreDYN. QUESO Level 4's objective is to keep the

most neutrinos while getting rid of the most muons. Thus, once a specific muon rate threshold has been reached, the neutrino rate can be compared from a one-to-one perspective, regardless of the classification path taken. For example, if Model A achieves a muon rate of 1 mHz and a neutrino rate of 10 mHz by keeping events with a scoreA$\geq$0.9, it can be said that it is performing better than Model B, which, when it achieves a muon rate of 1 mHz by keeping events with scoreB1$\geq$0.6 and scoreB2$\geq$0.4, it is keeping only a neutrino rate of 9 mHz.

Figure 4.13 depicts the neutrino rates obtained for 20 different muon rates, spread evenly in a logarithmic scale ranging from 10 mHz to 0.01 mHz. Two kinds of cuts were performed to obtain the results: one with a cut set in the scoreDYN and the other with the optimal combination of score1 and score2. By manipulating either score1 or score2, it is possible to decrease the muon rate when using two scores. An instance of a combination score1$>$0.8 and score2$>$0.2 may result in a reduction of the muon rate to 1 mHz while maintaining a neutrino rate of 2 mHz. However, this particular combination may not be the most effective if another combination, such as score1$>$0.9 and score2$>$0.1, also reduces the rate to 1 mHz while maintaining a higher neutrino rate of 3 mHz. Various permutations of these two scores can yield the intended muon rate while producing varying rates of neutrinos. To look for the best combination of the scores, an iterative scan for every possible combination of score1 and score2 that could effectively reduce the muon rate to the desired level is carried out. The optimal combination is then chosen based on the highest neutrino rate. The lower graph shows the scoreDYN value for the neutrino rate in the top one, identified as Dynedge Level 4 Classifier. The score1 and score2 values in the lower graph are for the neutrino rate identified in the top graph as 2LGBMs Level 4 Classifier.

The results show that the Dynedge Level 4 classifier works a little bit better than the 2LGBMs when it comes to getting a final sample with muon rates of about 7 mHz or higher. The ultimate objective of the event selection process is likely to be to achieve muon rates that are below 7 mHz. In this regime, the performance of the 2LGBMs model clearly outperforms the Dynedge one.

The way in which Dynedge predicts is very different from the way a LGBM does. Consequently, it is possible that one of the algorithms encountered difficulty rejecting a specific subset of muons that the other algorithm was able to reject with ease. The Venn diagram depicted in Figure 4.14 illustrates a comparative analysis of the final samples that have successfully passed QUESO Level 4 for both the Dyendge and 2LGBMs pathways when applying cuts that reduced the muon rate to 1 mHz. The circular shapes in orange correspond to the subset of the sample that successfully passed the 2LGBMs-Level 4 cut on both the left and right sides of the depicted figure. The blue ones represent the subset of the population that successfully met the criteria for Dynedge-Level 4. The numbers represent the number of events in each group. The diagram depicted in the left portion of the figure illustrates the neutrinos that passed through Level 4, being the intersection of the two circles of neutrinos passing through both Dynedge and

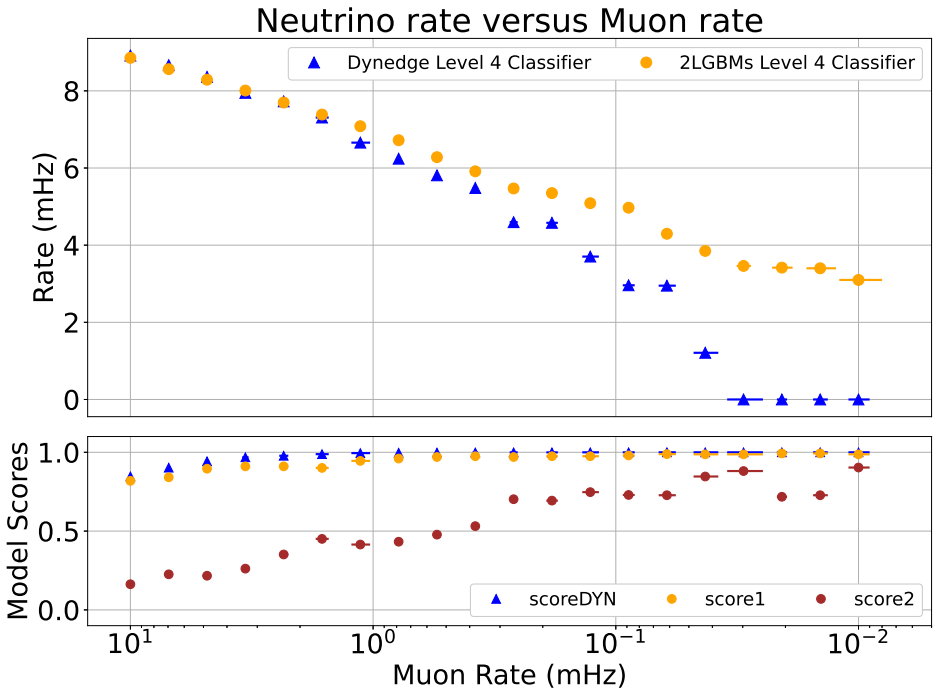


Figure 4.13: The rate of neutrinos kept with the Dynedge Level 4 model and the combination of the two LGBMs as a function of a muon rate. The bottom plot shows the values of scoreDYN, score1 and score2 used for the cuts to obtain the above neutrino rates.

the LGBMs cuts. The right-hand side of the depicted figure illustrates the analogous scenario, but for the subset of muons that would successfully reach Level 4. The bulk of the events in the neutrino sample were found to be situated at the intersection of both models, indicating that a significant proportion of them would remain unaffected even if both cuts were implemented. In contrast, the application of both cuts would result in only a minor proportion of the muons being able to pass through. The presented diagram gives confirmation that different algorithms exclude different muons. Consequently, the effort to combine the three models and create a QUESO level 4 was worthwhile.

Two different approaches were taken to combine the three models. The first one combines the three scores in the most efficient way to lower the muon rate as much as possible while keeping as many neutrinos as possible. In a way similar to what was done when comparing the Dynedge model to the 2LGBM model, certain muon rates are chosen, and the highest neutrino rate is found by searching iteratively for combinations of scoreDYN, score1, and score2 that lower the muon rate to the desired value. The second one was to first make a cut using the 1^{st}LGBM model and, on the remaining sample, retrain a second LGBM using the same variables as before but also including among them the score generated by the Dynedge Level 4 model. As Dynedge really was targeting different muons than the LGBMs, by including into the second LGBM variables the scoreDYN, these muons that were better found by Dynedge can start to also be rejected by the second and improved LGBM, which now on will be

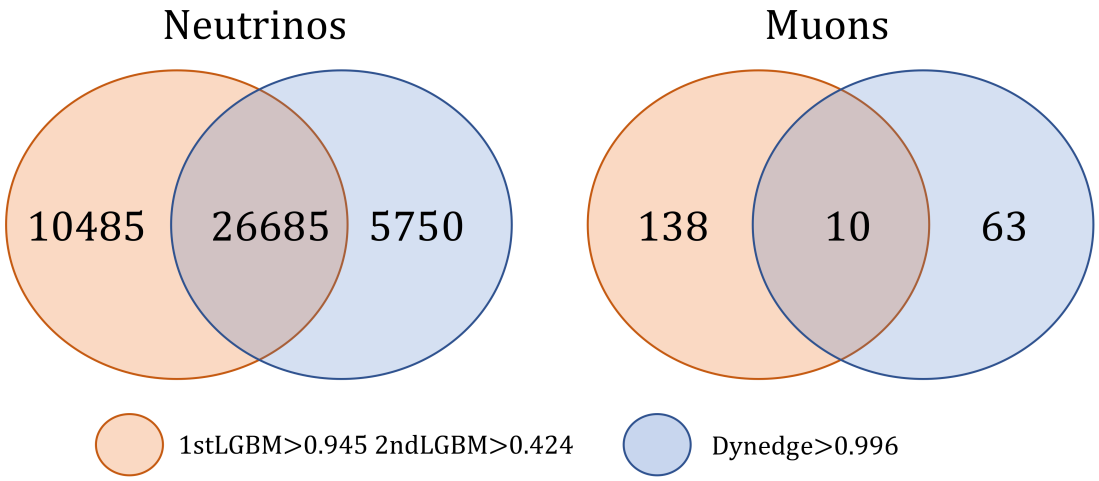


Figure 4.14: Population of neutrinos (*left*) and muons (*right*) that passed the cuts on the Dynedge Level 4 classifier (*blue*), the cuts on the two LGBM (*orange*), or both (*the intersection*) when tuning both classification paths to lower the muon rates to 1 mHz.

called 2^{nd}LGBM-v01. 2^{nd}LGBM was trained on data that passed score1$>$0.85. The 2^{nd}LGBM-v01, with the inclusion of scoreDYN as a variable, was competitive one-to-one with 1^{st}LGBM. Thus, the better performance of the combination of both was not achieved by setting a first cut that high, but by training this second algorithm on the sample that survived score1$>$ 0.4.

Table 4.3 shows for muon rates of $\sim$10 mHz, $\sim$1 mHz, $\sim$0.5 mHz, $\sim$0.1 mHz and $\sim$0.01 mHz the neutrino rates obtained by only using the predictions of L4-Dynedge, by using the combination of the two LGBMs, by combining the scores of the three models, and by using the best combination of the 1^{st}LGBM and 2^{nd}LGBM-v01. Highlighted are the values of the neutrino and muon rates for each classification path that would be more likely real settings for most of the physics analyses, as for these rates of muons, the samples resulting from the combination of the three scores have a background of $\sim$1.5% (for a muon rate of 0.1 mHz), or $\sim$0.3% (for a muon rate of 0.01 mHz), which would unlikely be a significant background.

	Rate of ν (mHz)	Rate of μ (mHz)	ScoreDYN	Score1	Score2(-v01)
L4-Dyendge[2]	8.91	10	0.8474	0	0
	6.48	1	0.9966	0	0
	5.80	0.5	0.9979	0	0
	3.03	**0.1**	0.9998	0	0
	0.32	**0.01**	0.9999	0	0
L4-2LGBMs	8.86	10	0	0.8192	0.1629
	7.07	1	0	0.9454	0.4243
	6.24	0.5	0	0.9742	0.4236
	4.92	**0.1**	0	0.9796	0.7293
	3.10	**0.01**	0	0.9871	0.9033
L4-Dyn+2LGBMs	9.17	10	0.565	0.5997	0.1000
	7.81	1	0.8919	0.8564	0.2907
	7.35	0.5	0.9664	0.8639	0.2907
	6.47	**0.1**	0.9861	0.8757	0.4814
	4.50	**0.01**	0.9314	0.9871	0.6957
L4-Dyn+2ndLGBM-v01	9.20	10	0	0.5302	0.1719
	7.17	1	0	0.7823	0.6554
	7.18	0.5	0	0.8175	0.7969
	5.75	**0.1**	0	0.9537	0.9356
	4.44	**0.01**	0	0.9769	0.9782

Table 4.3: Maximum neutrino rate achieved when lowering the muon rate to 10 mHz, 1 mHz, 0.5 mHz, 0.1 mHz or 0.01 mHz using different combinations of the models and the combination of scores used to achieve the result. The label Score2(-v01) is the score of the 2ndLGBM, or 2ndLGBM-v01 for the last classifier combination.

For almost all muon rate thresholds, the best performance is obtained when a cut is based on the combination of the three scores. The event selection for the NMO analyses presented in Chapter 5 will consequently be done by combining Dynedge Level 4, 1stLGBM adn 2ndLGBM models. L4-Dyn+2ndLGBM-v01, however, performs only marginally worse. This path has the advantage that instead of dealing with three scores, which can occasionally be messy and difficult for new users to tune, it is only necessary to deal with two.

It is true that neutrino rates are a good indicator of how good an event selection is, but neutrino rates do not perfectly translate into physics analysis sensitivities. The NMO sensitivities with the presented event selection are shown in Chapter 5, along with comparisons to an earlier event selection developed for the ICU as well as the sensitivity with IceCube and DeepCore (but no Upgrade).

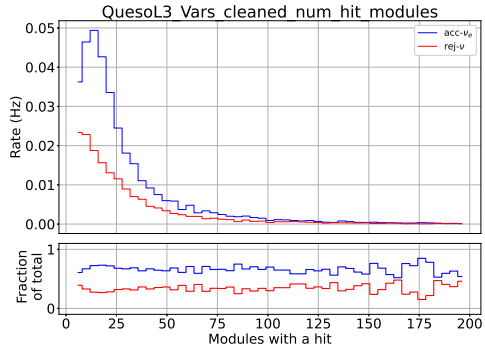
4.2.4 STUDY OF NEUTRINOS REJECTED

When using machine learning classifiers, it is important to keep track of the events that we want but fail to pass the classifier. Especially if the goal of the event selection is to get a clean data sample to check

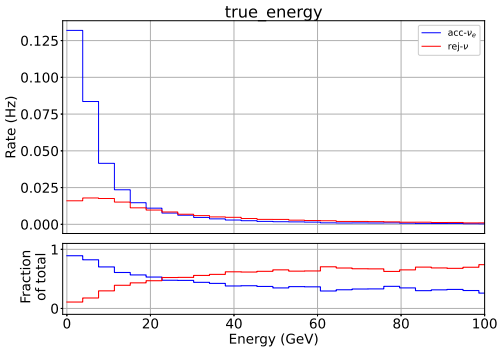
2 A muon rate within 0.005 mHz of the 0.01 mHz value cannot be found as the muon rate does not drop as a continuous so an interpolated value for the neutrino rate is given

the sensitivity of the IceCube Upgrade to NMO, it is important to make sure that the classifier hasn't learned any pattern that would hurt the upgoing low energy sample since it would be the one providing sensitivity in NMO analyses. Figure 4.15 shows the distribution for the true energy, zenith, azimuth, the radius on the x-y plane from string 36 of the interaction vertex as well as its z-component, and the number of hit DOMs of the event. In terms of the arrival direction, the model seems not to have any preference when rejecting or accepting events with a given azimuth. However, it does prefer to reject more neutrinos with zenith angles of zen$< \pi/2$ (downgoing). This result was expected as there are no upgoing muons and, with some directional information, it is easier to make a mistake by tagging a downgoing neutrino as a muon. This bias is not alarming, as the neutrinos of interest to resolve NMO are the ones passing through the Earth (zen$> \pi/2$). As expected, given the nature of the variables used in the model and the characteristics of muon signatures in the detector, the model tends to reject neutrinos interacting at larger radii and in the upper regions of the detector.There seems to be no preference to accept or reject events with a given number of hit DOMs, even if it tends to accept neutrinos of lower energies while rejecting more of the higher-energy ones. Normally, there would be a correlation between the energy of the neutrino and the ability to reconstruct and study the event in the future. However, there is an even tighter correlation between the number of DOMs hits and this capacity. Thus, this second bias is again not of great concern, as it will maybe even benefit the NMO analyses, as it keeps neutrinos within the energy range of interest for oscillations, and it does not seem as if events that could be better reconstructed have been missed.

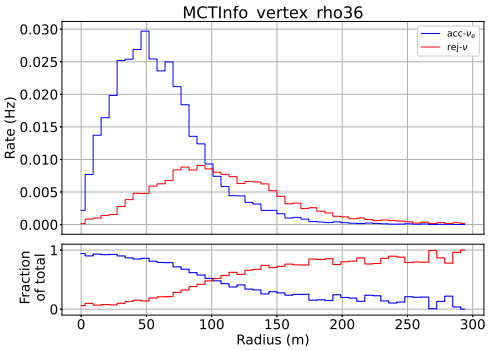
In the reconstruction of a variable, Dynedge predicts by generating a probability distribution, whose mean is taken as a prediction, and it also provides a value kappa defined as $\kappa = 1/\sigma^2$ which is a measure of how certain the model is of being accurate in that prediction. High values of κ mean that the variance of the probability function is small and therefore the probability function is narrow, so the model is sure about its prediction, and vice versa. To check in further detail if there was a bias towards losing highly reconstructable events in the event selection, the distribution of the zenith kappas was studied (see Figure 4.16). What is seen in the figure is that the event selection tends to reject events with lower values of zenith kappas. This means that it is more likely to reject events in which Dyndege has doubts about the accuracy of the reconstructed quantity. Additionally, there is a little bump in the distribution of the rejected neutrinos for κ values over 100. These events were checked individually, and the pulse map did not exhibit any unusual features. Indeed, events with this high value of κ were expected to be long tracks in the detector, whose direction is easy to reconstruct even with the naked eye. However, this was not the case, and cascade events with many hits in the fiducial volume were the dominant events with $\kappa \geq 100$. It is still unclear why these events have such high kappa values.



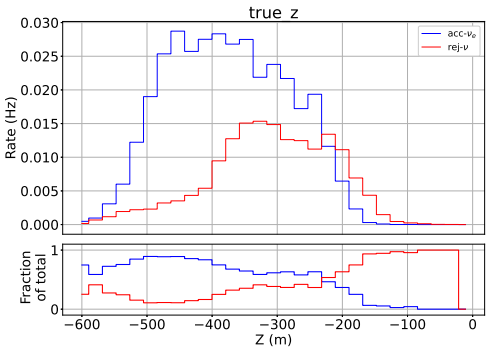
(a) Cleaned number of hit modules



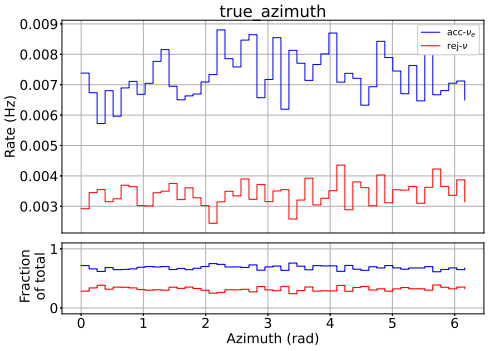
(b) True neutrino energy



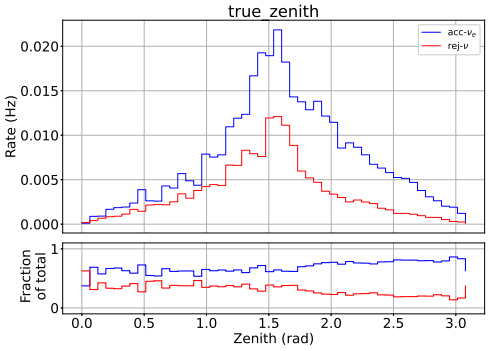
(c) Radial distance to string 36



(d) True z-position of interaction vertex



(e) True azimuth



(f) True zenith

Figure 4.15: True variable distributions for the neutrinos passing QUESO Level 4 when lowering the muon rate to 0.1 mHz.

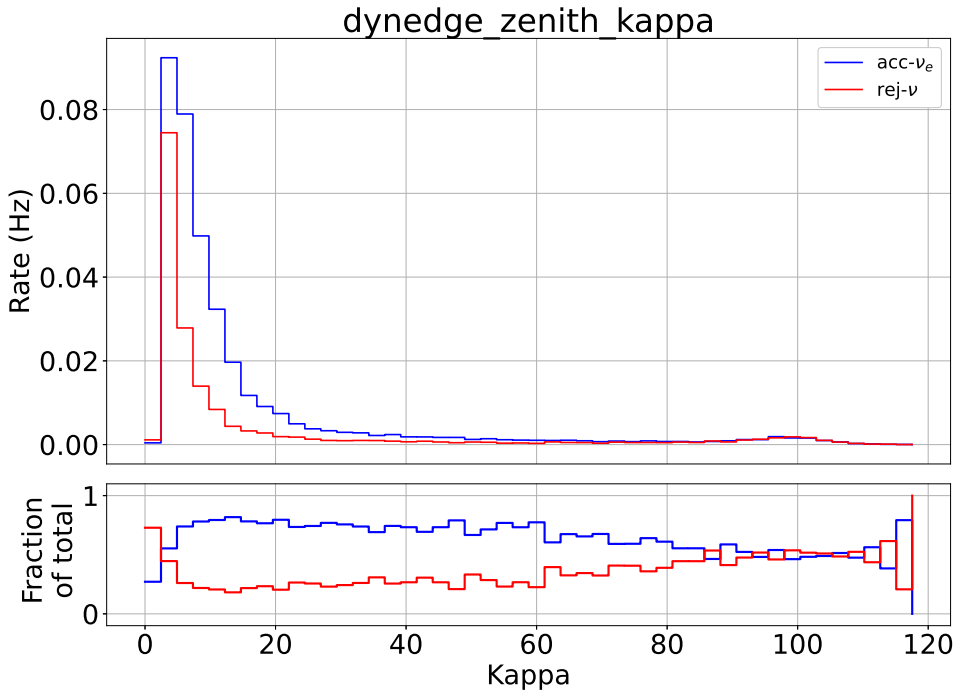


Figure 4.16: Distribution of the Dynedge κ for the neutrinos passing QUESO Level 4 when lowering the muon rate to 0.1 mHz

4.3 QUESO RESULTS RECAP

This section collects the results of the complete event selection presented for ICU for the sake of completeness. The QUESO Level 4 final selection can be tuned to produce either a sample with a higher rate of neutrinos, recognizing that the sample will be slightly more contaminated by muons, or a sample with a lower neutrino rate but higher neutrino purity[3]. The criteria used for the final sample presented in this section were to keep the muon rate below 0.1 mHz. This sample corresponds to the last column of Table 4.4. The neutrino rate purity of the sample is 98.5%, with a rate ν_τ-cc higher than the muon rate. As expected, the majority of the neutrinos lost are ν_μ-cc, because the interaction produces a muon byproduct, causing the signature of these neutrinos to be similar to the muon's. The low-energy oscillations group event selection for IceCube (not ICU) was reaching a neutrino rate slightly above 2 mHz with a muon contamination rate almost 1 mHz, with a ν_τ-cc rate roughly 13 times less than the muon contamination rate [6]. In this study for the IceCube Upgrade, at a muon rate of $\sim$1 mHz, the neutrino rate is close to 8 mHz, about a factor 4 higher than for IceCube alone, with the ability to drop the muon rate below the ν_τ-cc one while maintaining more than 6 mHz of total neutrino rate.

3 Defining the neutrino rate purity of a sample as the rate of neutrinos over the total rate of events.

	Rates at step 4	Rates at QUESO Level 3	Rates at QUESO Level 4
ν_e+ CC	3.30 mHz	2.54 mHz	1.88 mHz
ν_μ CC	8.31 mHz	6.34 mHz	3.75 mHz
ν_τ CC	0.28 mHz	0.22 mHz	0.15 mHz
ν NC	1.35 mHz	1.04 mHz	0.69 mHz
μ_{atm}	230572	782 mHz	0.099 mHz
Noise	1216 mHz	3 mHz	0

Table 4.4: QUESO performance summary. The first column from the left shows the rates after step 4, the mid column the rates after QUESO level 3, and the right column the rates after QUESO Level 4 lowering the muon rate below 0.1 mHz.

Figure 4.17 schematically illustrates the rates of the different event types over the different steps of the event selection starting at a trigger level (step 3). The final values of the rates represented in the figure at QUESO Level 4 are those presented in the last column of Table 4.4, with a noise line going to values very close to zero.

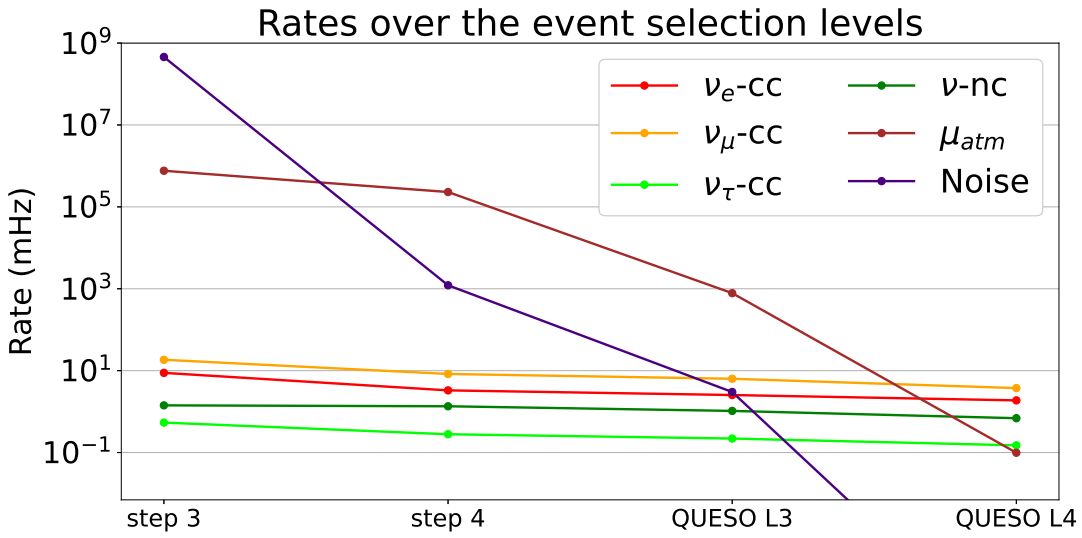


Figure 4.17: Rate of events at different processing levels along the IceCube Upgrade Event Selection.

5

ICECUBE UPGRADE NMO SENSITIVITIES WITH QUESO

This project has mainly focused on developing the QUESO event selection. The good results presented in Table 4.4 allowed us to generate different final samples of events and study the sensitivity that the IceCube Upgrade would have to resolve the NMO. My colleague Amalie Beate Albrechtsen master's thesis project was dedicated to how to compute sensitivities to NMO for ICU, with the statistics used to derive the presented sensitivities explained in detail in [75].

Neutrino telescopes have a dependence on their sensitivity to NMO to the value of θ_{23}. That is the reason the sensitivities for these atmospheric neutrino detectors are often reported as a function of θ_{23}. The value of $\theta_{23} = 45^\circ$ is what is called maximal mixing, corresponding to values of $\sin^2(\theta_{23}) = 0.5$. The hypothesis of inverted ordering with a value of $\sin^2(\theta_{23}) \sim 0.5 - 0.6$ is moderately degenerated with the hypothesis of normal ordering with a value of $\sin^2(\theta_{23})$ in the opposite octant[1] [18]. The degeneracy can be appreciated in the dips near maximal mixing for both, No and IO sensitivities in Figures 5.1a and 5.1b

To compute the sensitivities, it is assumed that the sample has only neutrinos. It is a fair assumption when the event selection can achieve samples with a neutrino rate purity of very close 100%, but it is also true that by doing so, the rate of neutrinos also drops. In a study of how much this reduction in rate would affect the sensitivities of ICU to NMO two event samples were generated (see Table 5.1). In the first one, *sample 1*, the assumption of no muons in the sample is softer, and a rate of 0.5 mHz of muons would have been part of the sample. In the second one, *sample 2*, the assumption is made quite firmly, and less than 0.01 mHz of muons is demanded. The difference in the total neutrino rates is almost 3 mHz, which is indeed a big difference. However, the only neutrinos that affect NMO sensitivities are those that passing through the Earth ($\cos(zen) < 0$) and in the correspondent energy range, considered in Table 5.1 as $E \leq 25$ GeV. The rate difference of neutrinos within the NMO-sensitive region is $\sim$0.7 mHz. If this difference induces a huge drop in the NMO sensitivity, a muon-contaminated sample would have to be

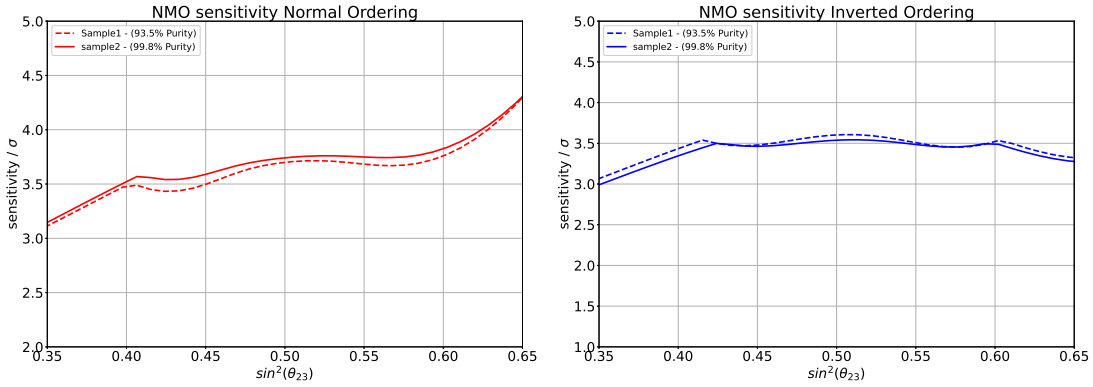
1 It is said to be in one or another octant if $\theta_{23} > 45^\circ$ or $\theta_{23} < 45^\circ$.

used to study how the presence of muons affects the sensitivity. However, if the difference in sensitivities is negligible, the high-purity sample can be used, and in the reported NMO sensitivities, the hypotesis of no muons holds [2].

	Sample 1 Rates	**Sample 2 Rates**
ν_e+ CC	2.09 mHz	1.48 mHz
ν_μ CC	4.30 mHz	2.39 mHz
ν_τ CC	0.18 mHz	0.12 mHz
ν NC	0.78 mHz	0.51 mHz
μ_{atm}	0.51 mHz	0.0099 mHz
Noise	0 mHz	0 mHz
NMO-ν	3.06 mHz	2.39 mHz
Total-ν	7.35 mHz	4.49 mHz
Purity	93.5%	99.8%

Table 5.1: Rates for different events of the two different samples where the NMO sensitivities were studied.

Figure 5.1 shows the sensitivities of the IceCube Upgrade to NMO with the first 3 years of ICU data if the true ordering was normal ordering (Figure 5.1a) and if the true ordering was inverted ordering (Figure 5.1b). The results for both event selections are remarkable, as within 3 years ICU has the potential to achieve more than 3σ significance for NMO, independently of the true mass order. The difference in sensitivities between the two samples is not significant, which is also important as it means that even if we choose to make more agressive cuts in the event selection, the amount of neutrinos contributing to NMO sensitivities is not that significant. In this scenario, the NMO sensitivity with sample 2, with a purity of 99.8% really illustrates the NMO sensitivity under the confident assumption of no muons in the sample.



(a) Normal ordering sensitivities

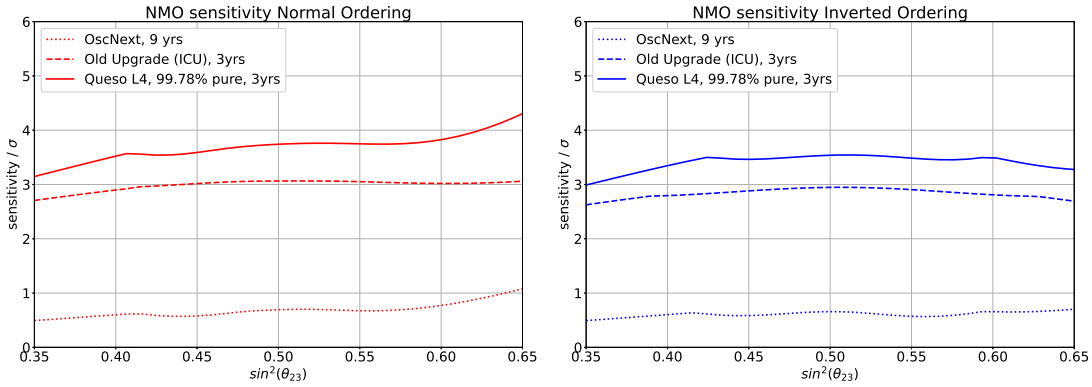
(b) Inverted Ordering sensitivities

Figure 5.1: ICU sensitivities to resolve NMO with 3 years of data collection using data samples 1 and 2.

2 The reported sensitivities do not include detector systematics.

To understand the impact the event selection has had on the sensitivities, Figure 5.2 illustrates the comparison of NMO-sensitivities for three different data samples. oscNext 9 years stands for IceCube and DeepCore (no ICU) after 9 years of operation and an event selection detailed in [6]. Old Upgrade ICU (3 years) relates to ICU data with the preliminary event selection [76]. This event selection reported a total neutrino rate of 6.8 mHz for a sample with a purity lower than 87%, meaning a remaining muon rate of 0.892 mHz. For that muon rate, the presented event selection achieves a neutrino rate of 7.74 mHz, almost 15% higher. In the figure, detector systematics have been accounted for when computing the oscNext 9 years and the old Upgrade 3 years, but not for the new ICU sample. This will lead to a slight decrease in the sensitivities presented for the new ICU pure sample.

Figure 5.2 really demonstrates the potential the ICU has to resolve NMO compared to IceCube-alone, independently of the event selection chosen. Between the lines for the ICU there might not be a huge difference, but the old Upgrade selection has a muon contamination in the sample of over 13%, while this number drops to 0.22% for the ICU event selection.



(a) Normal ordering sensitivities

(b) Inverted Ordering sensitivities

Figure 5.2: Comparison of IceCube-alone DeepCore NMO sensitivity with 9 years of data (*OcNext, 9yrs*), the one for ICU and old event selection with 3 years of data (*Old Upgrade (ICU), 3 yrs*) and with the one for ICU with the new event selection presented in this work (*Queso L4, 99.78% pure, 3yrs*).

6

CONCLUSION

This work presents a comprehensive event selection for the IceCube Upgrade. The high noise rates at trigger levels with noise over 10^8 mHz, along with pulse maps for muons and neutrinos that are heavily dominated by noise hits, necessitated the exploration of new tools and procedures for event selection. The introduction of machine learning for pulse cleaning has proven to be essential for a detector with such high noise rates. This approach enables the analysis of pulse maps dominated by signals and avoids the rejection of neutrino events with more than nine hits in the detector.

The efficiency of the DeepCore-Upgrade Filter has been found to be insufficient, and additional efforts to fine-tune this tool may be necessary in the near future, especially before the ICU begins its operation, as the extremely high muon rates in step 4 pose significant challenges. Although this was not a major concern in this thesis, the QUESO performance was remarkable and reduced the muon rate by over 10^7 orders of magnitude while preserving the majority of relevant neutrinos. Indeed, the work presented here should aid in the development of stringent triggers and filters at step 4 that reduce the muon rate, even at the expense of losing neutrinos that would have been lost anyway when QUESO was run.

The new event selection for ICU clearly outperforms the old one. By using the exact same initial data, the new event selection achieves a final neutrino rate approximately 1 mHz higher while maintaining the same muon rate. To put this improvement in perspective, the final samples for IceCube-alone and DeepCore had total neutrino rates of approximately 2 mHz [6]. Furthermore, the new event selection allows for a final sample with muon and noise contributions of less than 0.22%, resulting in improved NMO significance compared to the old event selection, which had muon contamination of over 13%. Additionally, the new event selection improves the sensitivities of 9 years of IceCube-alone and DeepCore by approximately 2σ within three years of analysis.

The presented event selection has provided the opportunity to study the NMO sensitivities with data samples that have negligible muon or noise contamination. The reported significances, while not accounting for detector systematics, shed light on the true potential of ICU for studying NMO. The reported significances exceeding 3σ within the first three years of data collection were far beyond what

was initially expected for the IceCube Upgrade. Even though the primary goal of ICU is the study of tau neutrino appearance, these results also position the experiment as one of the most competitive worldwide in the near future for resolving NMO.

A

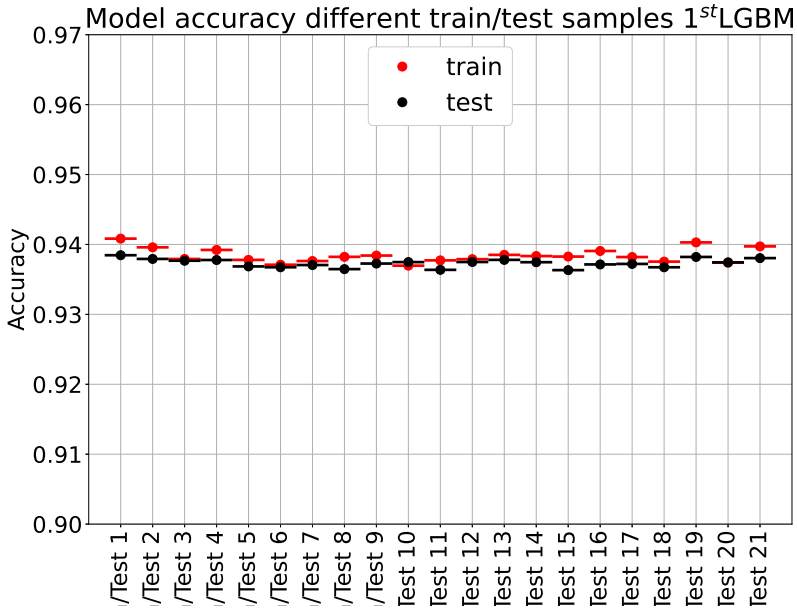
LGBMS

A.1 LGBMs HYPERPARAMETERS

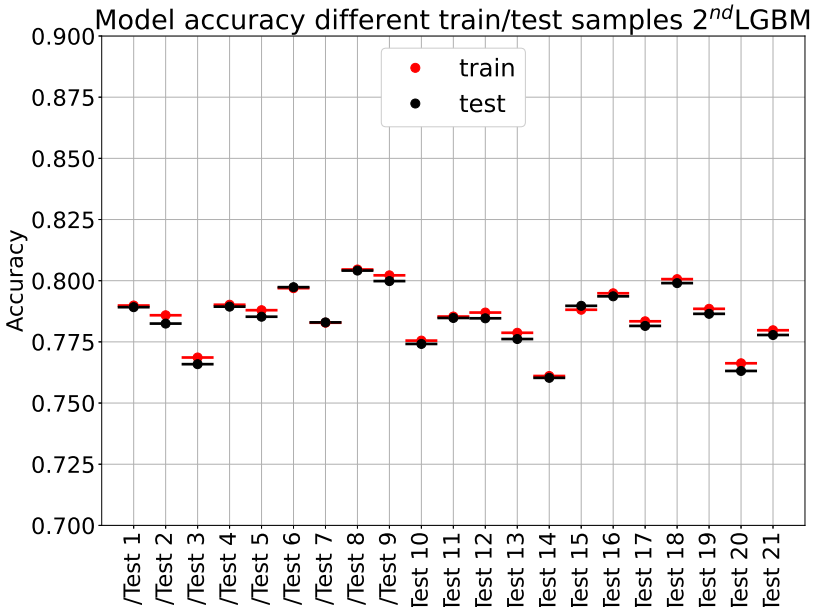
Parameter	1^{st}LGBM	$2^{n}d$LGBM	2^{nd}LGBM-v01
boosting_type	gbdt	gbdt	gbdt
objective	binary	binary	binary
max_depth	6	4	8
num_leaves	25	25	25
is_unbalanced	False	False	False
max_bin	64	64	64
min_data_leaf	15	50	15
lambda_l1	2	2	4
lambda_l2	2	1	2
min_gain_to_split	2	8	2

Table A.1: Hyperparameters used to tune the 1^{st}LGBM, 2^{nd}LGBM and 2^{nd}LGBM-v01.

A.2 TRAIN/TEST ACCURACIES TO CHECK OVERFITTING ON THE LGBMS



(a) Train/test accuracies 1stLGBM



(b) Train/test accuracies 2ndLGBM

Figure A.1: Train/test accuracies of the 1st and 2ndLGBMs for 20 different random choices of training and testing samples

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

[1] S. Bilenky, 'Neutrino oscillations: from a historical perspective to the present status', Nuclear Physics B **908**, Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015, 2 (2016).

[2] M. Thomson, *Modern particle physics* (Cambridge University Press, 2012).

[3] C. A. Argüelles et al., *Measuring oscillations with a million atmospheric neutrinos*, 2023.

[4] A. Ishihara, *The icecube upgrade – design and science goals*, 2019.

[5] W. Y. Ma et al., 'Physics potential of the icecube upgrade', Journal of Physics: Conference Series **1468**, 012169 (2020).

[6] S. B. et. al (oscNext team), 'Oscnext - simulations and sample (v00.06)', (2022).

[7] W. contributors, *Standard model-wikipedia, the free encyclopedia*, `https://en.wikipedia.org/wiki/Color_confinement`.

[8] W. contributors, *Standard model-wikipedia, the free encyclopedia*, `https://en.wikipedia.org/wiki/Standard_Model`.

[9] C. L. Cowan et al., 'Detection of the free neutrino: a confirmation', Science **124**, 103 (1956).

[10] W. contributors, *Neutrino, the free encyclopedia*, `https://en.wikipedia.org/wiki/Neutrino`.

[11] G. Danby et al., 'Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos', Phys. Rev. Lett. **9**, 36 (1962).

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

[13] F. Huang, 'Measurement of atmospheric tau neutrino appearance with IceCube/DeepCore', PhD thesis (The Pennsylvania State University The Graduate School, 2018).

[14] E. B. Hansen, 'Early Atmospheric Muon Rejection with IceCube-PINGU', Master's thesis (University of Copenhagen, Faculty of Science, 2016).

[15] J. A. Formaggio et al., 'From ev to eev: neutrino cross sections across energy scales', Rev. Mod. Phys. **84**, 1307 (2012).

[16] P. Kyberd et al., 'Nustorm: neutrinos from stored muons', (2012).

[17] S. Fukuda et al. (Super-Kamiokande Collaboration), 'Solar ^8B and hep neutrino measurements from 1258 days of super-kamiokande data', Phys. Rev. Lett. **86**, 5651 (2001).

[18] J. Koskinen, 'Standard neutrino oscillations', in *Probing particle physics with neutrino telescopes* () Chap. Chapter 5, pp. 109–139.

[19] I. J. Martínez Soler, 'Neutrino Oscillations in Particle Physics and Astrophysics', PhD thesis (U. Autonoma, Madrid (main), U. Autonoma, Madrid (main), 2018).

[20] M. Nakahata, 'History of solar neutrino observations', Progress of Theoretical and Experimental Physics **2022**, 12B103, 10.1093/ptep/ptac039 (2022).

[21] W. contributors, *Standard model-wikipedia, the free encyclopedia*, https://en.wikipedia.org/wiki/Alpha_particle.

[22] HEASARC, *Cosmic rays*, https://en.wikipedia.org/wiki/Alpha_particle.

[23] J. R. Hörandel, 'Cosmic-ray composition and its relation to shock acceleration by supernova remnants', Advances in Space Research **41**, 442 (2008).

[24] Google, *Google earth*, https://earth.google.com/web/@-77.41575703,-106.86404255,2662.95688668a,3251542.47665942d,35y,163.87184054h,53.51887281t,0r.

[25] M. Usner, 'Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector', PhD thesis (Mathematisch-Naturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin, 2018).

[26] W. contributors, *Branching fraction - wikipedia, the free encyclopedia*, https://en.wikipedia.org/wiki/Branching_fraction#:~:text=In%20%5C%20particle%20%5C%20physics%5C%20and%5C%20nuclear,the%5C%20decay%5C%20of%5C%20elementary%5C%20particles..

[27] M. Tanabashi et al. (Particle Data Group), 'Review of particle physics', Phys. Rev. D **98**, 030001 (2018).

[28] M. Honda et al., 'Atmospheric neutrino flux calculation using the nrlmsise-00 atmospheric model', Phys. Rev. D **92**, 023004 (2015).

[29] J. Yanez, 'Status of standard oscillation physics with icecube deepcore', Journal of Physics: Conference Series **1468**, 012122 (2020).

[30] M. Aartsen et al., 'The IceCube neutrino observatory: instrumentation and online systems', Journal of Instrumentation **12**, P03012 (2017).

[31] M. Aartsen et al., 'Measurement of atmospheric tau neutrino appearance with icecube deepcore', Physical Review D **99**, 10.1103/PhysRevD.99.032007 (2019).

[32] K. Hanson et al., 'Design and production of the icecube digital optical module', Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **567**, Proceedings of the 4th International Conference on New Developments in Photodetection, 214 (2006).

[33] M. DuVernois, 'Generation-2 IceCube Digital Optical Module and DAQ', PoS **ICRC2015**, 1148 (2016).

[34] A. Ishihara et al., 'Overview and performance of the D-Egg optical sensor for IceCube-Gen2', PoS **ICRC2017**, 1051 (2017).

[35] L. Classen et al., 'The mDOM - A multi-PMT digital optical module for the IceCube-Gen2 neutrino telescope', PoS **ICRC2017**, 1047 (2017).

[36] I. Collaboration, *Artistic schematic for three new sensor module designs for the icecube upgrade.* https://icecube.wisc.edu/gallery/nsf-approves-funding-for-icecube-upgrade/.

[37] I. Collaboration, *Icetray. a swiss army knife for icecube analysis*, https://github.com/icecube/icetray.

[38] W. contributors, *Mikheyev–smirnov–wolfenstein effect-wikipedia, the free encyclopedia*, https://en.wikipedia.org/wiki/Mikheyev-Smirnov-Wolfenstein_effect.

[39] R. Wendell, *Prob3++ software for computing three flavor neutrino oscillation probabilities.* https://github.com/rogerwendell/Prob3plusplus.

[40] IceCube Collaboration, *NuFlux: a library for calculating atmospheric neutrino fluxes*, CFF Version: 1.2.0. Available at: https://doi.org/10.5281/zenodo.5874708, 2022.

[41] J. A. Formaggio et al., 'From eV to EeV: neutrino cross sections across energy scales', Reviews of Modern Physics **84**, 1307 (2012).

[42] E. Bourbeau, 'Measurement of Tau Neutrino Appearance in 8 Years of IceCube Data', PhD thesis (University of Copenhagen, 2021).

[43] M. J. Larson, 'Simulation and identification of non-poissonian noise triggers in the IceCube neutrino detector', PhD thesis (Department of Physics in the Graduate School of The University of Alabama, 2013).

[44] N. STANISHA, 'Characterization of low-dt non-poisson noise in the IceCube neutrino detector', PhD thesis (The Pennsylvania State University Schreyer Honors College, 2014).

[45] C. Andreopoulos et al., 'The GENIE Neutrino Monte Carlo Generator', Nucl. Instrum. Meth. **A614**, 87 (2010).

[46] M. Alam et al., 'Genie production release 2.10.0', (2015).

[47] D. Heck et al., *CORSIKA: a Monte Carlo code to simulate extensive air showers.* (1998).

[48] M. J. Larson, 'A Search for Tau Neutrino Appearance with IceCube-DeepCore', PhD thesis (University of Copenhagen, 2018).

[49] J. van Santen, *Muongun*, https://events.icecube.wisc.edu/event/126/contributions/7262/attachments/5693/6668/MuonGun.pdf.

[50] J. Weldert et al., *Upgrade simulation - production status*, https://events.icecube.wisc.edu/event/164/contributions/9122/attachments/6941/8706/Upgrade$%5C%$20Pre-meeting$%5C%$20-$%5C%$20Simulation.pdf.

[51] I. Collaboration, *Upgrade simulation*, https://wiki.icecube.wisc.edu/index.php/Upgrade_Sim#File_contents.

[52] J. H. Koehne et al., 'PROPOSAL: A tool for propagation of charged leptons', Comput. Phys. Commun. **184**, 2070 (2013).

[53] S. Agostinelli et al., 'Geant4—a simulation toolkit', Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).

[54] D. C. for the IceCube Collaboration, 'Photon propagation with gpus in icecube', http://dx.doi.org/10.3204/DESY-PROC-2014-05/40 (2014).

[55] T. I. Collaboration, *Domlauncher*, https://docs.icecube.aq/icetray/main/projects/DOMLauncher/index.html.

[56] J. Weldert et al., *Upgrade noise-only simulation tests*, https://docs.google.com/presentation/d/1t1CZ3G-mXLN_W4tK-LK3X2iLx5BCh_oGHS_KyI3_zzI/edit#slide=id.p.

[57] J. L. Kelley et al., 'Event triggering in the IceCube data acquisition system', AIP Conference Proceedings **1630**, 154 (2014).

[58] J. Weldert et al., *Upgrade trigger studies*, https://drive.google.com/file/d/1VhJAx48nHWDs2qd1Yjz98DSRnhoqyzH6/view.

[59] T. Sttutard, *Introduction to event selections. low-energy workshop (atlanta)*, 2018.

[60] T. Kozynets et al., *Gnn-assisted pulse cleaning for the icecube-upgrade*, https://drive.google.com/file/d/1bIAC83Hq7aGLMgo7R-aygYMXUMAcVjUR/view.

[61] R. Abbasi et al., 'Graph neural networks for low-energy event classification &amp reconstruction in IceCube', Journal of Instrumentation **17**, P11003 (2022).

[62] R. Ørsøe, *Graphnet for icecube upgrade*, https://events.icecube.wisc.edu/event/164/contributions/9201/attachments/6934/8695/Aachen2023$%5C%$20Upgrade$%5C%$20PreMeeting.pdf.

[63] I. Collaboration, *Icecube docs*, https://software.icecube.wisc.edu/.

[64] J. Koskinen, *Lecture 10: multivariate method - boosted decision tree*, https://www.nbi.dk/~koskinen/Teaching/AdvancedMethodsInAppliedStatistics2021/Lecture10_MVA.pdf.

[65] G. Ke et al., 'Lightgbm: a highly efficient gradient boosting decision tree', in Advances in neural information processing systems, Vol. 30, edited by I. Guyon et al. (2017).

[66] D. Nelson, *Gradient boosting classifiers in python with scikit-learn*, https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/.

[67] Microsoft, *Lightgbm*, https://lightgbm.readthedocs.io/en/latest/index.html.

[68] B. Sanchez-Lengeling et al., 'A gentle introduction to graph neural networks', Distill, https://distill.pub/2021/gnn-intro, 10.23915/distill.00033 (2021).

[69] A. Søgaard et al., *Graphnet*, version v0.2.0, June 2022.

[70] K. E. Iversen, *The Good, the Bad and the Noisy. Characterization of IceCube Neutrino Events Using Graph Neural Networks*, Master's thesis, 2022.

[71] K. L. DeHolton, *Quick upgrade event selection for oscillations (queso)*, https://events.icecube.wisc.edu/event/164/contributions/9124/attachments/6930/8687/Upgrade$%5C%$20Pre-meeting$%5C%$20-$%5C%$20Queso.pdf.

[72] I. Collaboration, *Low energy event selection*, https://wiki.icecube.wisc.edu/index.php/Low_Energy_Event_Selection.

[73] S. J. Peraire, *Lecture notes in dynamics. lecture 26. 3d rigid body dynamics: the inetrtia tensor, (mitopencourseware)*, https://ocw.mit.edu/courses/16-07-dynamics-fall-2009/dd277ec654440f4c2b5b07d6c286c3fd_MIT16_07F09_Lec26.pdf.

[74] W. contributors, *Root mean square-wikipedia, the free encyclopedia*, https://en.wikipedia.org/wiki/Root_mean_square.

[75] A. B. Albrechtsen, *Sensitivity to the Neutrino Mass Ordering in the IceCube Upgrade*, Master's thesis, 2023.

[76] T. Sttutard, *Oscnext for the upgrade*, 2019.