Draft version April 23, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

MODIFYING MACHINE LEARNING ALGORITHMS FOR IMPROVED TAU NEUTRINO SENSITIVITY IN THE DUNE DETECTOR

P. BHAURA,[1] Supervisor: PROF. N. ILIC,[1] and Supervisor: W. DALLAWAY[1]

[1] *Department of Physics, University of Toronto, McLennan Physical Laboratories, 60 St George St, Toronto, ON M5S 1A7*

(Dated: April 22, 2024)

ABSTRACT

This report encompasses a concise historical overview of advancements in neutrino physics and an introduction to neutrino oscillation theory. Next, the specifics of Liquid Argon Time Projection Chambers (LArTPCs) and their proposed application in the Deep Underground Neutrino Experiment (DUNE) are explored. Additionally, a gentle introduction to Graph Neural Networks (GNNs) and their relevance in High-Energy Physics (HEP) is offered. Expanding upon this, a specific GNN called NuGraph2 and the modifications made to enhance its performance specifically for tau neutrino event classification is discussed. These enhancements have resulted in improved tau neutrino event detection capabilities within the framework of DUNE.

1. INTRODUCTION

Neutrino physics is a branch of particle physics that focuses on the study of neutrinos, which are extremely elusive subatomic particles characterized by their weak interactions with matter. Neutrinos are part of the Standard Model (SM) of particle physics and come in three distinct flavors: electron neutrinos (ν_e), muon neutrinos (ν_μ), and tau neutrinos (ν_τ), each paired with a corresponding charged lepton (electron (e), muon (μ), and tau (τ) respectfully).

The Deep Underground Neutrino Experiment (DUNE) stands as a prominent international experiment in neutrino research, meticulously crafted to explore various aspects of neutrino physics including neutrino oscillations, interactions, and the search for proton decay. DUNE's primary objective is to conduct precise measurements of neutrino attributes using extensive Liquid Argon Time-Projection Chamber (LArTPC) detectors, strategically positioned deep underground to minimize cosmic ray disturbances. While DUNE's scope encompasses all neutrino flavors, a particular emphasis will be placed on the investigation of ν_τ, the least understood among the three flavours (Machado et al. [2020]). Understanding the oscillations of ν_τ is crucial for advancing our understanding of neutrino physics.

In this project, we focus on the modification of a Graph Neural Network (GNN) called Nugraph2. NuGraph2, developed by the ExaTrkX collaboration, is a GNN designed for reconstructing particle interactions in neutrino physics detector environments. This report will go over the specific modifications that were made to this GNN with the goal of better classifying tau neutrino events in the DUNE far detector.

2. NEUTRINOS

2.1. *A Brief History*

The discovery of neutrinos, a remarkable episode in particle physics history, emerged from decades of experiments and theoretical developments. The concept of neutrinos was introduced by Wolfgang Pauli in the early 1930s to address energy and momentum conservation discrepancies in certain radioactive beta decay processes. Although he famously wagered that neutrinos could never be detected, in 1956, Clyde Cowan and Frederick Reines confirmed their existence through an experiment in South Carolina, USA, where they observed antineutrinos interacting with protons to produce positrons and neutrons.

Subsequently, in the 1960s and 1970s, a shortfall in the expected number of solar neutrinos (ν_e) led to the investigation of what became known as the "solar neutrino problem." Raymond Davis Jr. and John N. Bahcall's experiment in South Dakota, USA, contributed to this research by using chlorine atoms to detect solar neutrinos deep underground. The observed deficit compared to predictions and raised questions about neutrino physics and solar processes.

The puzzle was resolved in the late 1900s and early 2000s with the discovery of neutrino oscillations (see section 2.2.1), wherein neutrinos change their flavor as they travel long distances (Bellini et al. [2014]). The Super-Kamiokande and Sudbury Neutrino Observatory exper-

iments provided critical evidence for these oscillations, fundamentally altering our understanding of neutrinos. It was revealed that neutrinos, previously assumed to be massless, indeed possessed mass (Bellini et al. [2014]). This realization challenged the established SM of particle physics, prompting the need for significant revisions and extensions to accommodate the newfound knowledge about neutrinos. It also opened new avenues for exploring neutrinos' fundamental properties and their role in the universe.

2.2. *Underlying Neutrino Physics*

Neutrinos, existing in three flavors — ν_e, ν_μ, and ν_τ — correspond to charged leptons e, μ, and τ respectively. These elusive particles possess distinctive properties. Neutrinos are fermions, characterized by a half-integer spin (specifically 1/2), shared with electrons, quarks, and protons. Unlike other fermions, neutrinos bear no electric charge, leading to minimal electromagnetic interactions and making them challenging to detect (Machado [2023]).

Their primary interaction occurs through the weak nuclear force, one of nature's fundamental forces, and it's considerably weaker than electromagnetic or strong nuclear forces. Consequently, neutrinos can traverse substantial matter with minimal interaction and can even penetrate the Earth largely undisturbed.

Unveiling precise neutrino mass values and their hierarchy remains a focal point of particle physics research. Although we confirm neutrinos possess mass, it's notably small compared to SM particles like electrons and quarks. Furthermore, the three neutrino flavors exhibit slight differences in estimated mass values.

2.2.1. *Neutrino Oscillations*

Neutrino oscillations are a phenomenon in particle physics that occur when neutrinos change from one flavor to another as they travel through space (Bellini et al. [2014]). The theory of neutrino oscillations was first proposed by Bruno Pontecorvo, Ziro Maki, Masami Nakagawa, and Shoichi Sakata in the 1960s and was later experimentally confirmed, leading to significant advancements in our understanding of neutrinos and their properties.

Neutrino oscillations arise from the fact that neutrinos are quantum mechanical particles described by wavefunctions, and they have mass. In quantum mechanics, a particle's wavefunction is a superposition of its different mass eigenstates. In the case of neutrinos, these mass eigenstates are different from their flavor eigenstates (Machado [2023]). As neutrinos travel through space, they evolve in time, and this evolution is governed by the differences in the masses of the mass eigenstates (Machado [2023]). The key to neutrino oscillations is that the flavor eigenstates are not constant during this evolution. Instead, they change as a linear combination of the mass eigenstates, resulting in a periodic oscillation between different neutrino flavors.

The PMNS matrix, also known as the Pontecorvo-Maki-Nakagawa-Sakata matrix, is a mathematical construct used in the field of particle physics to describe the mixing of different neutrino flavors. The PMNS matrix is a 3×3 unitary matrix that describes the relationships between the flavor eigenstates ($|\nu_e\rangle$, $|\nu_\mu\rangle$, and $|\nu_\tau\rangle$) and the mass eigenstates ($|\nu_1\rangle$, $|\nu_2\rangle$, and $|\nu_3\rangle$) of neutrinos (Giganti et al. [2018]). In its simplest form, it is expressed as the following unitary transformation:

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} \, |\nu_i\rangle \, . \tag{1}$$

Here, the subscript α indicates flavor eigenstates as $\alpha = e, \mu, \tau$ and the subscript i indicates mass eigenstates, m_i, such that $i = 1, 2, 3$. $U_{\alpha i}$ represents the coefficients of the PMNS matrix (coefficients of the linear combination of mass eigenstates). When the standard three-neutrino theory is considered, the matrix is 3×3. If only two neutrinos are considered, a 2×2 matrix is used. If one or more sterile neutrinos (hypothetical neutrino particles) are added, it is 4×4 or larger (Giganti et al. [2018]). In the 3×3 form, it is given as follows:

$$\begin{aligned} U &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} . \end{aligned} \tag{2}$$

Here, c_{ij} and s_{ij} represent $\cos\theta_{ij}$ and $\sin\theta_{ij}$ respectfully, with θ_{ij} representing the mixing angle. The phase factor, δ, is non-zero only if neutrino oscillations violate charge-partity (CP) symmetry (Giganti et al. [2018]). CP symmetry combines both charge and parity symmetries. It suggests that the laws of physics should be the same if you simultaneously replace all particles with their antiparticles (charge reversal) and invert the spatial coordinates (parity reversal) (Machado [2023]). In other words, CP symmetry postulates that the fundamental laws of physics should be invariant under this combined transformation. For neutrinos, we are concerned with if neutrinos and antineutrinos oscillate with the same probabilities. Neutrino oscillation experiments have provided strong evidence for the existence of CP violation in the neutrino sector, implying a non-zero value for δ (Bellini et al. [2014]).

The matrix elements, $U_{\alpha i}$, in Eq. 1 and 2 determine the probability amplitudes for a neutrino of a specific flavor eigenstate to be in a specific mass eigenstate (Giganti et al. [2018]). Eigenstates with different masses propagate at varying frequencies (Giganti et al. [2018]). The more massive ones exhibit quicker oscillations when contrasted with their lighter counterparts. Given that mass eigenstates also result from combinations of flavor eigenstates, this difference in oscillation frequencies leads to interference between the corresponding flavor constituents of each mass eigenstate. When constructive interference occurs, it becomes feasible to witness a neutrino, initially of a particular flavor, altering its flavor during its journey. The likelihood of a neutrino, originally of flavor α, ultimately being detected as having flavor β (for the case of two neutrinos) is as follows:

$$P(\nu_\alpha \to \nu_\beta) = \sin^2(2\theta_{ij}) \sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right). \quad (3)$$

Here, Δm^2_{ij} represents the the mass difference (which is defined as $m^2_i - m^2_j$), L represents the oscillation distance, and E represents neutrino energies. This formula is commonly employed when analyzing the transition $\nu_\mu \leftrightarrow \nu_\tau$ in atmospheric mixing, as the electron neutrino has minimal involvement in this scenario (Wester et al. [2023]). It is also applicable to the solar case $\nu_e \leftrightarrow \nu_\chi$, where ν_χ is a mix/superposition of ν_μ and ν_τ (Maltoni & Smirnov [2016]). These simplifications are possible due to the exceedingly small mixing angle θ_{13} and the close proximity in mass of two of the mass states compared to the third one.

Experiments are able to measure $P(\nu_\alpha \to \nu_\beta)$ and E while L is usually set or known. The focus then becomes calculating θ_{ij} and Δm^2_{ij} (which I will refer to from now on as Δm^2). Determining Δm^2 is especially important because using it we are able to create a mass hierarchy for neutrinos. The neutrino mass hierarchy refers to the relative ordering of the masses of the three known flavors of neutrinos. Experiments involving neutrino oscillations have provided important insights into their masses and their mass hierarchy. There are two possible neutrino mass hierarchies: the normal hierarchy (NH) and the inverted hierarchy (IH). In the NH, the mass ordering of neutrinos is as follows: the lightest neutrino is ν_1, the next heaviest is ν_2, and the heaviest is ν_3 (Qian & Vogel [2015]). This hierarchy is also sometimes referred to as the "1-2-3 hierarchy." In the IH, the mass ordering is as follows: the lightest neutrino is ν_3, the next heaviest is ν_1, and the heaviest is ν_2 (Qian & Vogel [2015]). This hierarchy is also sometimes referred to as the "3-1-2 hierarchy."

Experimental efforts, such as those conducted at neutrino observatories and particle physics facilities, have been dedicated to measuring the neutrino mass hierarchy. By carefully measuring the oscillation patterns and the differences in masses, scientists have made significant progress in determining the neutrino mass hierarchy, although it remains an active area of research. The precise determination of the neutrino mass hierarchy can help us refine our understanding of the SM of particle physics and may have implications for theories beyond the SM, such as those related to dark matter and the nature of the universe's earliest moments.

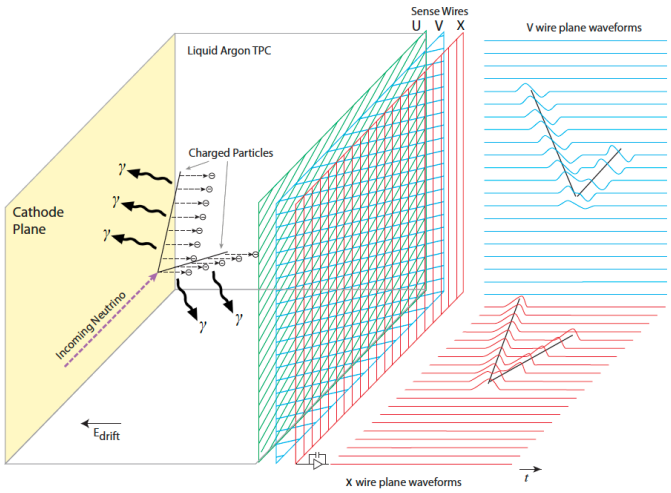

Figure 1. The operational concept (as shown in (Abi et al. [2020])) of the single-phase LArTPC involves the migration of negatively charged ionization electrons generated by a neutrino interaction. These electrons drift horizontally, opposite to the electric field, within the liquid argon and accumulate on the anode, composed of U, V, and X sense wires. On the right-hand side, a depiction illustrates the two-dimensional time projections as the event unfolds. The timing of the interaction's occurrence is determined by light detectors (not depicted), providing the event's t_0.

3. LIQUID ARGON TIME PROJECTION CHAMBERS

LArTPCs are advanced detectors used primarily in neutrino physics and particle physics experiments. They utilize liquid argon as a target material for detecting particles.

The structure of a standard LArTPC consists of three components (see Figure 1). Positioned on one side of the detector lies a cathode plane under high voltage, essential for creating a drift electric field throughout the Time-Projection Chamber (TPC). While the specific electric potential depends on the detector's geometry, this high-voltage cathode typically generates a drift

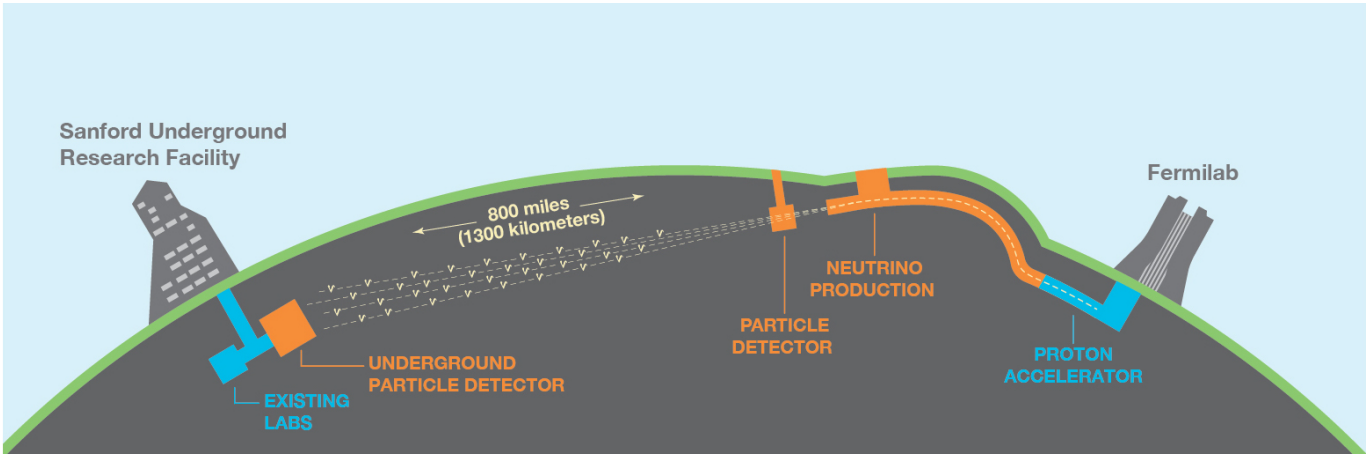

Figure 2. The configuration of the LBNF beamline at Fermilab, Illinois, and the DUNE detectors in Illinois and South Dakota, separated by 1300 km. (source: dunescience.org)

field of 500 V/cm across the entire detector (Abi et al. [2020]).

Opposite to the cathode plane, there are multiple anode wire planes positioned with potentials significantly higher (less negative) than that of the cathode. Each plane is separated from adjacent ones by a narrow gap, typically around 1 cm wide (Abi et al. [2020]). These planes are composed of numerous parallel conducting wires spaced just a few millimeters apart. Collectively, these planes detect signals from the drifting electrons. The inclusion of multiple planes with diverse wire orientations enables two-dimensional event reconstruction, while the third dimension is determined from the drift times of electrons.

The third component consists of a field cage positioned between the cathode and anode. This cage is responsible for upholding a consistent electric field between the two, ensuring that the paths of drifting electrons remain as close as possible to the shortest route between the point of ionization and the anode plane. The primary aim is to minimize any deviation in particle trajectories during event reconstruction.

A supplementary light-collection system is frequently integrated with the standard LArTPC to enhance event data extraction through scintillation light. Additionally, it holds significance in triggering events as it gathers scintillation light nanoseconds after particles traverse the detector (Abi et al. [2020]). This time frame is notably shorter (by approximately a factor of 1000) than the duration required for liberated electrons to drift to the wire planes (Abi et al. [2020]). Consequently, it is commonly utilized to designate the collection time of scintillation photons as the trigger time (t_0) for an event.

Liquid argon serves as an advantageous medium for various reasons. Its noble nature renders it electrically neutral, ensuring that electrons generated by particle interactions can freely drift towards the detector readout without being absorbed. Additionally, when an energetic charged particle traverses through, argon emits scintillation photons in proportion to the energy it absorbs from the passing particle (Majumdar & Mavrokoridis [2021]). Furthermore, liquid argon proves to be cost-effective, facilitating the feasibility of large-scale projects. However, the primary incentive for employing liquid argon as a sensitive medium lies in its density. With a density of approximately 1.4 g/cm^3, liquid argon heightens the likelihood of particle interactions within the detector. This characteristic holds particular significance in neutrino physics, where interaction cross sections between neutrinos and nucleons are typically small (Majumdar & Mavrokoridis [2021]).

LArTPCs offer several advantages over other detector technologies, including excellent spatial resolution, good energy resolution, and the ability to distinguish between different types of particles based on their interactions within the argon volume. They are key components of many current and future neutrino experiments, such as DUNE in the United States and the ProtoDUNE detectors at CERN.

3.1. *DUNE*

The dominance of matter over antimatter in the early universe, the intricate behaviors of supernova neutrino bursts responsible for creating the vital elements for life, and the possibility of proton decay are all mysteries central to the forefront of particle physics and astrophysics. These puzzles are essential for gaining insights into the initial development of our universe, its present condition, and its ultimate destiny. DUNE stands as a world-class international experiment committed to unraveling these mysteries.

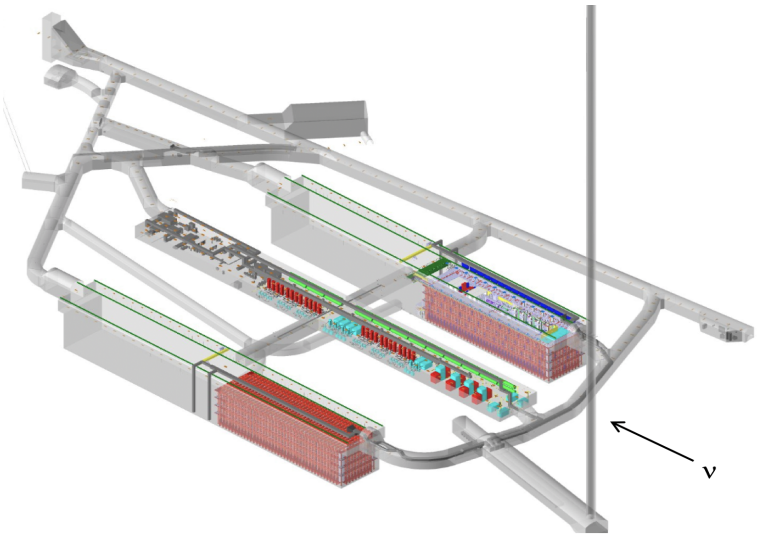

Figure 3. At SURF in South Dakota, underground chambers have been prepared to house the DUNE FD and its associated cryogenics systems. In the illustration (as shown in (Abi et al. [2020])), you can see the cryostats (red), which are designated for the initial two FD modules. On the right-hand side, the Ross Shaft, a vertical passage, will serve as the entry point to the subterranean area dedicated to DUNE. Each cryostat has dimensions of 65.8m in length, 18.9m in width, and 17.8m in height. The two detector caverns, accommodating these cryostats, measure 144.5m in length, 19.8m in width, and have a height of 28.0m, offering additional space around the cryostats.

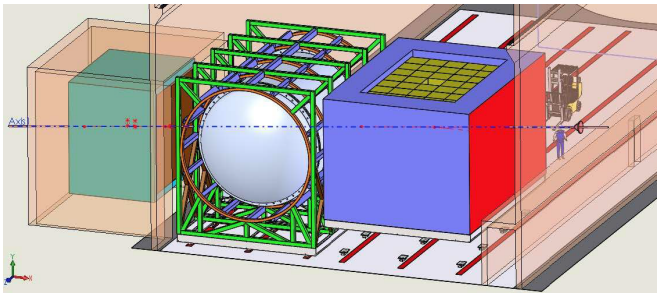

Figure 4. A setup of the DUNE ND as shown in (Abud et al. [2021]). The beam is shown as it enters from the right. Neutrinos first encounter the LArTPC (right), the MPD (center), and then the on-axis beam monitor (left).

The DUNE experiment is made up of three central elements: a state-of-the-art high-intensity neutrino source produced by a proton accelerator at Fermilab, delivering megawatt-class power, a far detector (FD) positioned 1.5km underground at the Sanford Underground Research Facility (SURF) in South Dakota, and a composite near detector (ND) strategically installed just ahead of the neutrino source (Abi et al. [2020]) (see Figure 2). The Long Baseline Neutrino Facility (LBNF) project, hosted by Fermilab, will provide the beamline and the civil construction for both detectors of the DUNE experiment.

3.1.1. *Near Detector*

Although project work focuses more on the FD, it is still essential to highlight the pivotal role of the ND in the experiment. The ND plays a fundamental part by serving as the experiment's control, effectively constraining systematic errors and measuring the initial, unaltered energy spectra of ν_e and ν_μ (and their antineutrinos) (Abud et al. [2021]). This comparison of energy spectra near the beam source, before any oscillation occurs, and at the far site is crucial for untangling the various energy-dependent effects that influence the beam spectrum. It also helps to diminish systematic uncertainties to the extent necessary for the detection of CP violation (Abud et al. [2021]). Furthermore, the ND plays a significant role in precisely measuring neutrino-argon interactions using both gaseous and liquid argon (Abud et al. [2021]). This measurement further mitigates the systematic uncertainties associated with modeling these interactions, enhancing the accuracy of the experiment.

Furthermore, the ND will have its own distinct physics program, separate from the FD. This program encompasses the measurement of neutrino interactions, delving into two fundamental aspects of the SM: electroweak physics and quantum chromodynamics (Abud et al. [2021]). In addition, it will embark on explorations beyond the SM, actively seeking non-standard interactions, sterile neutrinos, dark photons, and other exotic particles (Abud et al. [2021]). Positioned 574m downstream from the neutrino beam source, the ND will consist of three primary detector components: a LArTPC called ArgonCube, a High-Pressure Gaseous Argon Time Projection Chamber (HPgTPC), surrounded by

an Electromagnetic Calorimeter (ECAL) collectively referred to as the Multi-Purpose Detector (MPD), and a on-axis beam monitor called System for On-Axis Neutrino Detection (SAND) (Abud et al. [2021]) (see Figure 4).

3.1.2. *Far Detector*

The DUNE FD comprises four LArTPC detector modules, each housing a minimum of 10 *kt* of liquid argon within the cryostat's sensitive area, situated approximately 1.5 km below the surface (Abi et al. [2020]). The design of the four equally sized modules is adaptable enough to accommodate staged construction and the ongoing evolution of LArTPC technology (Abi et al. [2020]) (see Figure 3). DUNE is currently planning and developing two LArTPC technologies: single-phase (SP), where all detector components inside the cryostat are submerged in liquid, and dual-phase (DP), where certain elements function within a layer of gaseous argon above the liquid.

- Within the SP technology, ionization charges traverse the liquid argon horizontally, guided by an electric field towards a vertical anode for subsequent readout (Abi et al. [2020]). This configuration mandates the use of exceptionally low-noise electronics to achieve effective readout with a favorable signal-to-noise (S/N) ratio, as there is no in-cryostat signal amplification (see Figure 1).
- DP technology is not as well-established as the SP technology, and although it comes with some challenges, it offers several benefits. In DP, ionization charges ascend vertically within the liquid argon and then transition into a layer of gaseous argon above the liquid (Abi et al. [2020]). Special devices known as large electron multipliers (LEMs) amplify these signal charges in the gas phase before they reach a horizontal anode. This amplification in the gas phase reduces the requirements on electronics noise and allows for an extended drift length, which, in turn, necessitates a higher voltage (see Figure 5).

4. GRAPH NEURAL NETWORKS IN HIGH ENERGY PHYSICS

Machine learning (ML) has gained significant traction within the realm of neutrino physics and the broader domain of high-energy physics (HEP). Initially, ML applications in particle physics were focused on traditional classification and regression techniques, such as boosted decision trees, support vector machines, and shallow neural networks (NNs), utilizing physics-based

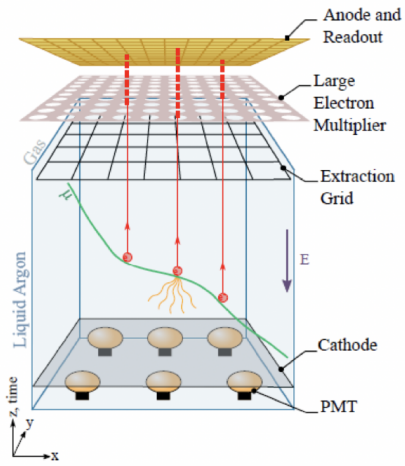

Figure 5. The fundamental concept (as shown in (Abi et al. [2020])) behind the DP LArTPC is as follows: ionization charges ascend vertically within the liquid argon and then transition into a layer of argon gas above the liquid, where they undergo amplification before being gathered on the anode. Positioned beneath the cathode are the light detectors, known as photomultiplier tubes.

high-level features (Thais et al. [2022]). However, recent advancements have ushered in a new era where more sophisticated deep learning architectures, including deep NNs, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), are being employed. These approaches leverage low-level information, such as energy deposits in detectors, rather than derived variables, leading to the exploration of diverse data representations like images and sequences. Moreover, the integration of cutting-edge ML methods has fostered closer collaboration between the particle physics and ML communities, providing physics researchers with opportunities to contribute directly to the advancement of state-of-the-art ML architectures.

In recent years, geometric deep learning (GDL) has risen as a significant subset of ML, concentrating on understanding non-Euclidean data structures such as sets, groups, graphs, and manifolds. These techniques have proven effective across various scientific and societal areas, including representing knowledge dependencies, modeling physical systems (Thais et al. [2022]), discovering chemicals and drugs (Gaudelet et al. [2021]), mapping community connections (Liu et al. [2020]), and beyond. Notably, GNNs stand out as a powerful category of GDL algorithms tailored for graph-based operations.

Graphs serve as a data framework illustrating objects (depicted as nodes) and their interconnections (depicted as edges). These structures capture intricate relation-

ships and dependencies between objects, crucial for accurately modeling physical data. In the context of particle physics data, graph-based representations offer numerous benefits compared to alternative formats: unlike vector or grid arrangements, graphs accommodate variable-sized data, preserving information without necessitating padding or loss (Thais et al. [2022]); moreover, graphs excel in handling sparse and diverse detector data, which is challenging to depict using image-based formats (Thais et al. [2022]). Graphs effectively encapsulate a wide spectrum of particle physics data, encompassing energy distributions within detectors, individual physical entities like tracks or missing energy, and individual particles or clusters of particles (Shlomi et al. [2020]).

Following the success of CNNs for grid-structured data (i.e pixels in an image), GNNs aim to extend many of the same powerful techniques to irregular, graph-like data structures. Convolutional techniques have demonstrated exceptional performance in the "spatial" realm (Thais et al. [2022]). Typically, a spatial convolution entails two main steps: aggregation and updating. Although the boundary between these steps may seem arbitrary, it proves beneficial for most graph neural network tasks focused on learning hidden representations of graph nodes.

Adopting (Battaglia et al. [2018])'s formalism, a graph is defined as follows:

$$G = (\mathbf{u}, V, E) \tag{4}$$

as a collection of node features $v_i = V$, edge features $e_{ij} = E$, and graph features $u_g = \mathbf{u}$. In general, for the $(l+1)^{th}$ convolution iteration, a node $i's$ hidden representation v_i^{l+1} can be computed by:

$$Aggregate : v_i^{l'} = \rho(e_{ij}^{l+1}), \text{ where } e_{ij}^{l+1} = \phi^e(v_i^l, v_j^l, e_{ij}^l)$$

$$Update : v^{l+1} = \phi^v(v_i^{l'}, v_i^l, u^l)$$

where $j \in \eta_i$. In other words: Features are aggregated around a node $i's$ neighbourhood η_i by first calculating each e_{ij}, called the "message" on the edge connecting node i and node j (Thais et al. [2022]). The message function ϕ^e is dependant on the choice of the architecture and ρ stands for any permutation-invariant aggregation (Thais et al. [2022]). The new node features can be combined with previous node features, or some higher features u belonging to the graph, and passed through. This node update is represented by ϕ^v. The whole process can be represented in Figure 6.

4.1. *NuGraph2*

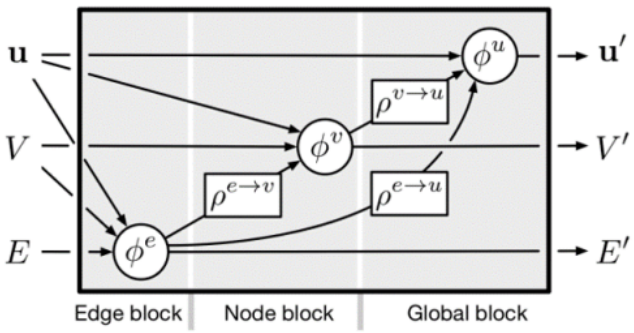

Figure 6. A GNN convolution as defined by (Battaglia et al. [2018]). ϕ^e defines a message function, ρ is an aggregation around nodes, ϕ^v is a node update, and ϕ^u is a graph update.

NuGraph2, developed by the ExaTrkX collaboration, serves as a GNN tailored for the task of low-level particle reconstruction from simulated neutrino interactions within a LArTPC detector. This innovative system functions by categorizing the detector plane hits based on particle characteristics, tapping into the interconnectedness between planes to incorporate crucial 3D contextual details. The framework detailed in (Hewes et al. [2024]) delivers dual outputs – a binary filter to sift out non-primary physics interaction hits and a semantic classifier to label each hit with its corresponding particle type. However, the fundamental message-passing mechanism, as elaborated in (Thais et al. [2022]), holds potential for diverse applications like event classification, clustering, and beyond.

Datasets provided by an experimental setup as in Figure 1 are used to construct heterogeneous graph objects representing each neutrino interaction. Signals detected on TPC wires are condensed into distinct Gaussian hits, pinpointing particular wires and times. Each wire plane generates its own subgraph, where hits on a particular plane serve as the foundation for its graph nodes. For every graph node, four input characteristics are established: the wire index, time stamp, and both the integral and root mean square width of the Gaussian pulse.

Edge connections between hits within the subgraph of each detector plane are established via a Delaunay triangulation algorithm (Pourya et al. [2022]), employing the wire and time coordinates of each node. This method produces a fully linked graph comprising both short-range and long-range connections, facilitating consistent performance even in scenarios featuring spatially isolated points caused by detector anomalies like dead wire areas or physical phenomena like disconnected electromagnetic shower segments.

Apart from the planar subgraphs created from detector hits, a "nexus" subgraph is also generated (Hewes

et al. [2024]). This is achieved through an initial algorithm that clusters 2D hits spanning multiple detector planes. These clusters of planar hits serve as the basis for establishing graph edges connecting planar nodes to nexus nodes. These nexus nodes essentially exist as "virtual" entities; they are solely defined by the graph edges linking them to planar hits. They lack dedicated input features and internal graph edges connecting one nexus node to another.

Graph objects are created from simulated events through a comprehensive workflow. The NuML package facilitates the generation of event-level HDF5 files, compiling tables that encapsulate essential low-level data such as simulated particles, energy deposits, and detected hits. Subsequently, these event HDF5 files are streamlined into graph objects using the PyNuML package. PyNuML is specifically engineered to provide a versatile and efficient solution for numerous ML tasks in HEP. It offers algorithms capable of generating ground truth labels from a simulated particle tree and subsequently transferring them onto the observed detector hits.

The NuGraph2 network architecture consists of an encoder, an iterative message-passing engine, and a set of decoders; a diagram of the network structure is provided in Figure 7. NuGraph2's **encoder** serves to take the input graph node features, x_i^{in}, and generate the initial categorical node embedding. After initial encoding to produce graph node features, the network utilizes an iterative two-step message-passing mechanism as its core engine. This engine consist of a set of **planar** blocks, that operate independently on each 2D detector plane's subgraph, and a **nexus** block that exchanges information between 2D planes to allow for degeneracy breaking (Hewes et al. [2024]). The **planar** block is responsible for propagating information internally inside each detector plane, while the **nexus** block is responsible for mixing information between detector planes. The output of the final message-passing iteration is passed into two **decoders**, which are responsible for producing the filter and semantic outputs.

5. MODIFICATIONS TO ARCHITECTURE & RESULT

It was discovered that several potential modifications could be made to the existing NuGraph2 architecture. The first of which included a modification to the initial encoding step and a modification to the frequency of classes within the dataset. As mentioned, the initial encoding step generates a learned embedding for each graph node. Currently, this is done by a two-layer sequential network layer with one linear layer

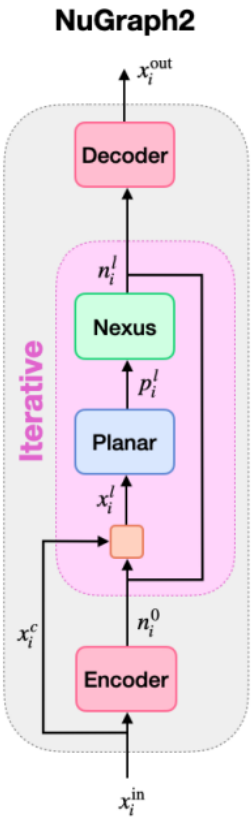

Figure 7. The overall structure of the NuGraph2 network architecture as presented in (Hewes et al. [2024]). The i and l subscripts denote planar nodes and the current iteration of message passing respectfully

and one non-linear activation function layer given by $z = \tanh(Wx + b)$ where W, b are learned parameters. In ML, activation functions are used within NNs to introduce non-linearity into the model, allowing it to learn complex patterns in data (Mehlig [2021]). A linear activation function simply outputs the input without any transformation. In other words, it preserves the linearity of the input. Using only a linear activation function results in a model that is no more powerful than a single-layer perceptron (one of the simplest forms of NNs) (Mehlig [2021]). Non-linear activation functions introduce non-linearity into the model, enabling it to learn and represent more complex relationships in the data (Mehlig [2021]).

We modified the initial encoding step by adding an additional layer to the existing two-layer sequential network in order to create a four-layer sequential network. This is given by $z = \tanh(W_2 \tanh(W_1 x + b_1) + b_2)$ where W_i, b_i are again learned parameters. This allowed us to introduce more learned parameters specifically for the purpose of generating initial learned embeddings for ν_τ.

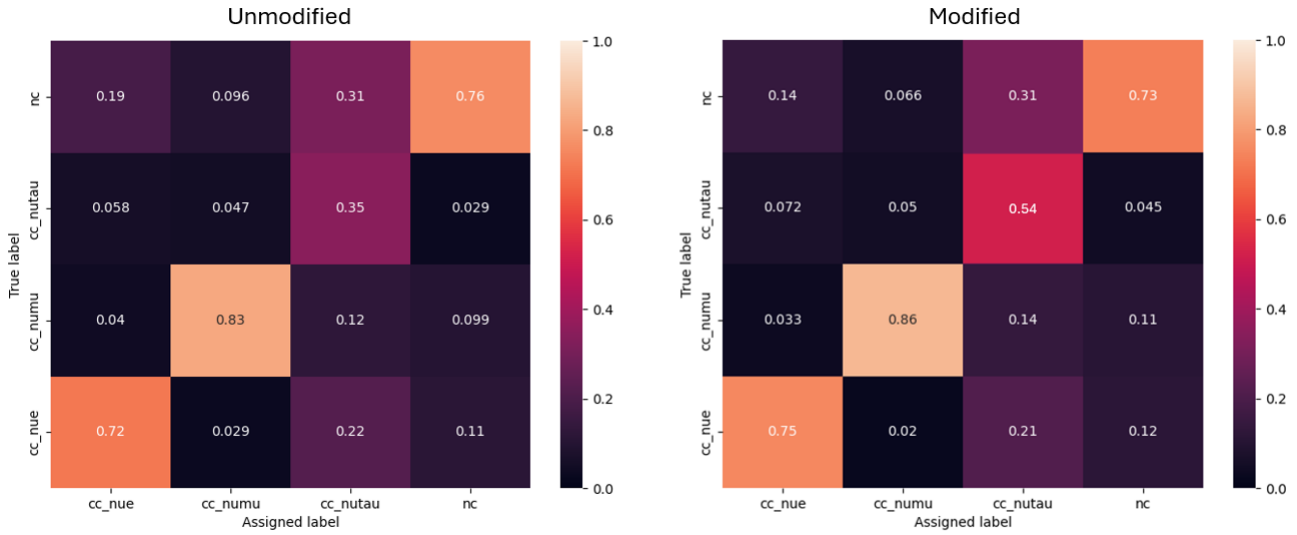

Figure 8. Precision matrices, both unaltered and adjusted, are normalized according to the type of neutrino event. An improvement in classifying ν_τ events can be seen, showing a modest gain of approximately 20%, whereas the classification of other event types show minimal variation. The latter is expected due to the inherent variability in training outcomes.

Figure 9. Values of the loss function, both in the modified and unmodified structures, were compared. The weighted loss function (orange) shows a reduction in value at a rate approximately 50% faster, indicating enhanced efficiency. This was attributed to the inclusion of specific weights, which also led to increased precision in ν_τ labeling.

Class imbalance is a fundamental problem in applications such as semantic segmentation (Tian et al. [2022]). Specifically, uneven class distributions in a training dataset often result in unsatisfactory performance on under-represented classes (Tian et al. [2022]). Nugraph2 breaks the entire dataset into the (total number of samples)/(batch size) number of batches. Each of these batches contains approximately the same relative frequency of each class. The typical batch (for a batch size of 32) contains approximately 7 or 8 ν_μ and ν_e samples, around 16 NC (neutral current) samples, and typically only 1 (or sometimes none at all) ν_τ sample. This implies the under-representation of ν_τ samples in the training. In order to overcome this issue, we artifi-

cially balanced the relative frequency of each batch by randomly oversampling the ν_τ samples each time. Now, the model sees each ν_e, ν_μ, and NC samples exactly once per epoch (one complete pass through the entire training dataset), but will see ν_τ samples much more frequently at approximately 7 times per epoch. This allows the model to have more exposure to the ν_τ samples. The result of the two previous modifications can be seen in Figure 8.

As mentioned, the output of the final message passing iteration is passed on into the two decoders. Alongside the decoders, NuGraph2 also outputs the performance metrics **Recall** (efficiency), **Precision** (purity), and **Loss** (accuracy between prediction and truth) (Hewes et al. [2024]). A loss function in ML is a measure of how well a model's predictions match the actual targets (Mehlig [2021]). It quantifies the difference between predicted values and the actual values. The goal of training a ML model is to minimize this loss function to zero, which essentially means making the model's predictions as close to the actual targets as possible.

Currently, NuGraph2 utilizes binary cross-entropy for the hit filtering decoder and Recall Loss for the semantic decoder (Hewes et al. [2024]). The choice of the loss function depends on the type of problem being solved. For regression problems (where the task is to predict a continuous value), common loss functions include mean squared error (MSE) and mean absolute error (MAE). For binary classification problems (where the task is to classify instances into one of two classes, as in our case of filtering), the cross-entropy loss function is often used. For multi-class classification problems (where the task is to classify instances into one of several classes, as in our case for semantic & neutrino event type classification), categorical cross-entropy (such as Recall Loss) or softmax loss functions are commonly used (Mehlig [2021]). Some of these loss functions are summarized in Appendix A.

We modified the Recall Loss function structure by introducing weights specifically for ν_τ labels. This entices the model to assign ν_τ labels to a given sample by returning a smaller cross entropy when it does so, and penalizes the loss when it doesn't. This resulted in more tau labeling overall, which resulted in slightly larger training precision and loss function efficiency (see Figure 9).

6. CONCLUSION

The development and application of NuGraph2 for particle reconstruction in neutrino physics has been a dynamic project embraced by DUNE and its scientific network. Nugraph2 offers 98.0% efficiency in filtering background hits and 94.9% efficiency in semantically labelling hits, while producing 2D representations that are 94.8% consistent (Hewes et al. [2024]).

Our attention was directed towards enhancing the classification of neutrino events, specifically ν_τ events, using the Nugraph2 framework. In its original state, the assigned labels for neutrino event classification exhibit an accuracy that is in the neighborhood of 20% lower (excluding ν_τ event classifications, which are approximately 60% lower) than that of filtering and semantic labelling (as detailed in Figure 8).

Adjustments were implemented to the established Nugraph2 framework with the specific aim of enhancing the accuracy of ν_τ event classification using simulated DUNE data. These modifications resulted in an approximate 20% increase in the accuracy of assigned ν_τ event labels during the most effective model training session.

NuGraph2 is engineered to be highly adaptable across different detector technologies. It can process any data structure comprising of detector hits and provides flexibility regarding the number of detector planes, truth labelling schemes, and input features for graph nodes. This versatility simplifies its application to various detector geometries, including LArTPCs and others. Furthermore, NuGraph2's core message-passing engine serves as a versatile solution for various tasks. Active development is ongoing to explore several additional modifications and enhancements.

ACKNOWLEDGEMENTS

I want to express gratitude to Murdock Aubry and Yijie Wang, my colleagues on this project, as well as Professor Nikolina Ilic and William Dallaway, my supervisors, for their invaluable support throughout this project. They have been instrumental in helping grasp concepts that were initially unfamiliar. A heartfelt appreciation goes out to the ExaTrkX collaboration, V Hewes, and all others who contributed to the development of the Nugraph2 framework, and for providing abundant resources to help deepen the understanding of these topics.

REFERENCES

Abi, B., Acciarri, R., Acero, M. A., et al. 2020, Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE, doi: https://doi.org/10.48550/arXiv.2002.02967

Abud, A. A., Abi, B., Acciarri, R., et al. 2021, Deep Underground Neutrino Experiment (DUNE) Near Detector Conceptual Design Report, doi: https://doi.org/10.48550/arXiv.2103.13910

Battaglia, P. W., Hamrick, J. B., Bapst, V., et al. 2018, Relational inductive biases, deep learning, and graph networks, doi: https://doi.org/10.48550/arXiv.1806.01261

Bellini, G., Ludhova, L., Ranucci, G., & Villante, F. 2014, Advances in High Energy Physics, 2014, doi: https://doi.org/10.1155/2014/191960

Gaudelet, T., Day, B., Jamasb, A. R., et al. 2021, Briefings in Bioinformatics, 22, bbab159, doi: 10.1093/bib/bbab159

Giganti, C., Lavignac, S., & Zito, M. 2018, Progress in Particle and Nuclear Physics, 98, 1, doi: https://doi.org/10.1016/j.ppnp.2017.10.001

Hewes, V., Aurisano, A., Cerati, G., et al. 2024, NuGraph2: A Graph Neural Network for Neutrino Physics Event Reconstruction, doi: https://doi.org/10.48550/arXiv.2403.11872

Liu, F., Xue, S., Wu, J., et al. 2020, in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, ed. C. Bessiere (International Joint Conferences on Artificial Intelligence Organization), 4981–4987, doi: 10.24963/ijcai.2020/693

Machado, P. 2023, Neutrino Properties and Interactions, 9–53, doi: 10.1142/9789811282645_0002

Machado, P., Schulz, H., & Turner, J. 2020, Phys. Rev. D, 102, 053010, doi: 10.1103/PhysRevD.102.053010

Majumdar, K., & Mavrokoridis, K. 2021, Applied Sciences, 11, doi: 10.3390/app11062455

Maltoni, M., & Smirnov, A. Y. 2016, The European Physical Journal A, doi: https://doi.org/10.1140/epja/i2016-16087-0

Mehlig, B. 2021, Machine Learning with Neural Networks: An Introduction for Scientists and Engineers (Cambridge University Press), doi: https://doi.org/10.1017/9781108860604

Pourya, M., Goujon, A., & Unser, M. 2022, Delaunay-Triangulation-Based Learning with Hessian Total-Variation Regularization, doi: https://doi.org/10.48550/arXiv.2208.07787

Qian, X., & Vogel, P. 2015, Progress in Particle and Nuclear Physics, 83, 1, doi: https://doi.org/10.1016/j.ppnp.2015.05.002

Shlomi, J., Battaglia, P., & Vlimant, J.-R. 2020, Machine Learning: Science and Technology, 2, 021001, doi: 10.1088/2632-2153/abbf9a

Thais, S., Calafiura, P., Chachamis, G., et al. 2022, Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges, doi: https://doi.org/10.48550/arXiv.2203.12852

Tian, J., Mithun, N., Seymour, Z., Chiu, H.-P., & Kira, Z. 2022, Striking the Right Balance: Recall Loss for Semantic Segmentation, doi: https://doi.org/10.48550/arXiv.2106.14917

Wester, T., Abe, K., Bronner, C., et al. 2023, Atmospheric neutrino oscillation analysis with neutron tagging and an expanded fiducial volume in Super-Kamiokande I-V, doi: https://doi.org/10.48550/arXiv.2311.05105

APPENDIX

A. LOSS FUNCTIONS

- **Mean Squared Error**:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

N is the number of data samples, y_i is the true value of the ith sample, and $\hat{y}_i$ is the predicted value of the ith sample.

- **Mean Absolute Error**:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

N is the number of data samples, y_i is the true value of the ith sample, and $\hat{y}_i$ is the predicted value of the ith sample.

- **Binary Cross Entropy**:

$$CE = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i))$$

N is the number of data samples, y_i is the true value of the ith sample, and $\hat{y}_i$ is the predicted value of the ith sample.

- **Multi-Class Cross Entropy**:

$$CE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

N is the number of data samples, C is the number of classes, $y_{i,c}$ is the true value that is 1 if sample i belongs to class c and 0 otherwise, $\hat{y}_{i,c}$ is the predicted value that is 1 if sample i belongs to class c and 0 otherwise.