

Università degli Studi di Genova

SCUOLA DI SCIENZE MFN

Department of Physics

Master of Science in Physics



Università di Genova

Exploring Machine Learning techniques for event reconstruction with Lenses in the LAr target of **SAND** at the DUNE neutrino experiment

Supervisors:

Alice Campani, PhD

Alessio Caminata, PhD

Co-supervisor:

Andrea Cocco, PhD

Candidate:

Leonardo Lena, BSc

Session of 14/10/2026

Academic Year 2025–2026

Only I cut on the timber – only I carved on the stone:
“After me cometh a Builder. Tell him, I too have known.”

— Rudyard Kipling, *The Palace*

Acknowledgements

I primi ringraziamenti vanno alla mia famiglia: a papà Andrea, a mamma Cristina, e a Ludovica. Mi avete cresciuto e sostenuto, insegnato valori, e priorità. Siamo diversi e determinati. Siamo spine nel fianco e aghi nel pagliaio. Siamo unicamente noi, e se non ci fossimo, dovrebbero inventarci.

In secondo luogo ringrazio l'Opera Pia Conservatorio Fieschi, in particolare nella persona del Dr. Crosa, il cui sostegno ha reso possibile la mia piena dedizione allo studio per l'intera durata del percorso.

Ringrazio gli amici della Divisione DIFI e quelli del Conservatorio: mi siete stati affianco, abbiamo condiviso pranzi e cene, consigli, preoccupazioni, tante risate e tante lacrime e quella parte di me di ciò è grata.

Ringrazio i relatori Alice e Alessio per la quantità di tempo e attenzioni dedicatemi in questi mesi, e tutto il gruppo DUNE-Genova per avermi calorosamente accolto.

Contrariamente ad una tendenza—oramai frequente—di ringraziare solamente se stessi, ringrazio infine chiunque abbia mai interagito con me: in bene e in male. Siamo il risultato di piccoli passi compiuti ogni giorno; passi allungati, accorciati, o deviati dai piccoli obbiettivi e ostacoli quotidiani. Un libro non è solo il titolo: è ogni capitolo, ogni pagina, ogni frase, parola, lettera... e tu, sì tu che stai leggendo questi ringraziamenti, dedicando del tempo a leggere i miei pensieri, sei una lettera del mio libro.



Abstract

Neutrinos are the most abundant of known matter particles in the Universe, and the study of their properties has brought to light a range of interesting phenomena. Among these, neutrino flavour oscillations provide key evidence of physics beyond the Standard Model, demonstrating that neutrinos are light, yet massive particles. Within the lively experimental program dedicated to advancing our understanding of neutrino nature and mass ordering, Deep Underground Neutrino Experiment (DUNE) will play a central role. DUNE is a next generation long-baseline experiment that employs a high-intensity neutrino beam produced at Fermi National Accelerator Laboratory and directed to a liquid argon-based underground detector at Sanford Underground Research Facility. The comparison of neutrino events originating from the same source and measured at the near and far site by detectors based on the same technology and target will allow precision measurements of the CP-violating phase in the lepton sector and the neutrino mass ordering, thanks to the presence of the matter effect that breaks the degeneracy between them. The presence of a near detector complex is essential to this effort, providing precise measurements of the un-oscillated neutrino flux and constraining the systematic uncertainties that drive the experiment's sensitivity. It comprises three detectors. System for on-Axis Neutrino Detection serves as the primary beam monitor and features a magnetised detector equipped with a high resolution tracker, a high performance electromagnetic calorimeter, and a one-tonne liquid-argon cryostat instrumented with photodetectors and innovative imaging devices for a spatial reconstruction of neutrino-argon interactions. GRanular Argon for Interactions of Neutrinos (GRAIN) will also allow the measurement of the total energy released in the medium and provide a time reference for each event. Two technologies are explored for the imaging element of GRAIN cameras: coded aperture masks and lenses. While a first prototype of a lens-based optical detector has been tested and characterised, and a preliminary version of the software for the event reconstruction with traditional methods exists, the aim of this thesis is to explore the application of Machine Learning techniques to the spatial reconstruction of particle interactions in GRAIN.

Contents

Acknowledgements	iii
Introduction	viii
1 Neutrino physics	1
1.1 Brief history of neutrino interactions	1
1.2 Neutrinos in the Standard Model	2
1.2.1 Right-handed components	3
1.3 Neutral particle oscillations	3
1.3.1 Flavour oscillations	4
1.3.2 Oscillations in vacuum	4
1.4 Neutrino masses	6
1.4.1 Dirac mass term	6
1.4.2 Majorana mass term	7
1.4.3 Neutrino mass hierarchy	8
1.5 Propagation through matter	9
1.5.1 Oscillations through a medium	10
1.6 Neutrino cross sections	12
1.7 Argon as a target	13
1.7.1 Ionisation	13
1.7.2 Scintillation	14
1.8 Other neutrino experiments	14
1.8.1 Oscillation experiments	14
1.8.2 Other neutrino studies	16
2 The DUNE experiment	19
2.1 DUNE physics	19
2.1.1 The PMNS complex phase δ_{CP}	20
2.1.2 Neutrino mass hierarchy	22
2.1.3 Secondary physics program	22
2.2 LBNF beam	23
2.3 Far Detector	24
2.3.1 DUNE-HD	25
2.3.2 DUNE-VD	26
2.4 Near Detector	27
2.4.1 DUNE-PRISM	28
2.5 SAND	29
2.5.1 Tracker	30
2.5.2 Calorimeter	32
2.6 GRAIN	32
2.6.1 Cryostat and cryogenics	33
2.6.2 Optical detectors	33
3 Machine Learning	37
3.1 Brief history of machine learning	37
3.2 Introduction to supervised statistical learning	38
3.3 Linear regression: least squares	40
3.3.1 Regularisation, the first hyper-parameter	41
3.4 Binary classification: soft Support Vector Machines (SVMs)	42

3.4.1 Stochastic Gradient Descent (SGD) also known as “ <i>training</i> ”	43
3.5 Non-linear learning	44
3.5.1 Kernel methods	44
3.5.2 Neural networks	45
3.6 Geometrical Deep Learning	47
3.6.1 The blueprint	47
3.6.2 A specific case: Convolutional Neural Networks	48
3.6.3 ResNet	49
4 Lens-based reconstruction	51
4.1 Introduction	51
4.1.1 Out-of-focus imaging	53
4.2 Hough transform reconstruction	53
4.2.1 Hough transform	55
4.2.2 Multi-Otsu thresholding	55
4.2.3 DBSCAN clustering	55
4.3 Truth definition and CNN	56
4.3.1 Truth propagation to the 2-dimensional level	57
4.3.2 Dataset assembly	59
4.3.3 Architecture	60
4.3.4 Results	63
4.4 Next steps	65
4.4.1 GRAIN layout optimisation	65
4.4.2 2-dimensional improvements	66
4.4.3 3-dimensional rework	67
4.4.4 Adoption of a Machine Learning (ML) suite	67
Conclusions and outlook	69
Glossary	72
Bibliography	74

Introduction

Neutrinos are massive neutral leptons, which means that they interact exclusively through the gravitational and weak interactions. The mass of neutrinos is at least two orders of magnitude smaller than the mass of all other leptons. This fact led physicists to initially postulate they didn't have any mass at all [Section 1.4]. However, the discovery of *flavour oscillations*, experimentally confirmed for the first time in 1998 by the Super Kamiokande experiment [1], proved the model wrong. The whole phenomenon can be explained in simple words in the following terms: neutrinos interact as *flavour* eigenstates but propagate as *mass* eigenstates. Therefore, a neutrino produced at the source with a specific flavour is necessarily in a superposition of mass eigenstates and thus has a non-zero probability of interacting at the observation point as a different flavour [Section 1.3.2]. The name of this phenomenon comes from the fact that this probability oscillates as a function of the ratio between the distance travelled and the neutrino energy L/E .

By observing, measuring and extensively studying neutrino oscillations it is possible to find answers to long-standing questions in particle physics; namely the mass ordering (MO) of neutrinos [Section 1.4] and the possible matter-antimatter asymmetry in this sector of the Standard Model (SM) [Section 1.3.2 and Section 1.5]. The Deep Underground Neutrino Experiment (DUNE) [Section 2] is a next-generation neutrino oscillation experiment designed to achieve sensitivities high enough to answer these [Section 2.1] and many more questions [Section 2.1.3]. DUNE has the structure of a standard neutrino oscillation experiment: a Near Detector (ND) [Section 2.4] is placed around 60 m underground at the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois where a high-intensity neutrino beam [Section 2.2] will be produced and directed towards the Far Detector (FD) [Section 2.3] at Sanford Underground Research Facility (SURF) in Leads, South Dakota, for a total distance of almost 1300 km.

The FD will host four Liquid Argon Time Projection Chambers (LArTPCs) for a total of 68 kt of liquid argon, the largest LAr detector ever built.

The ND has the task of accurately measuring the un-oscillated neutrino beam, in order to fully characterise its flux and alignment in near-real time. One part of the near detector, referred to as Precision Reaction Independent Spectrum Measurement (PRISM) [Section 2.4.1], will be placed on top of a track, allowing it to move to different positions away from the primary beam axis in order to boost the long baseline oscillation analysis. Thus, the other ND component, called System for on-Axis Neutrino Detection (SAND), will instead permanently stay on the beam axis, and act as the experiment's main beam monitor. It is composed of a magnetised high precision tracker [Section 2.5.1], a calorimeter [Section 2.5.2] and an instrumented cryostat filled with Liquid Argon (LAr) called GRanular Argon for Interactions of Neutrinos (GRAIN) [Section 2.6].

Because of GRAIN's unique geometry and the presence of a magnetic field, a standard LArTPC technology, mostly based on ionisation charge collection [Section 1.7], would not work. GRAIN will therefore be instrumented with various custom-built light detection devices [Section 2.6.2]. As for the imaging element, two technologies are currently being investigated: coded aperture masks [Section 2.6.2.1] and lenses [Section 2.6.2.2].

It is fair to assume that the reader will probably have already heard the terms *artificial intelligence* or Machine Learning (ML). Despite their recent abuse [Section 3.1], these concepts constitute a sizeable part of applied mathematics and, if correctly used, have shown a large potential in boosting scientific research. For the purpose of this present work, whenever the words Machine Learning are used, one

should interpret them as “*the numerical approximation of an unknown and complicated function*”. The main objective of this thesis is to prove a concept, that is, that machine learning-based algorithms can be exploited to increase the performance of event reconstruction in the lens-based GRAIN neutrino detector. To this extent, the original work presented in this thesis consists of the development and testing of a Convolutional Neural Network (CNN), a specific kind of ML algorithm, with the objective of boosting event reconstruction in GRAIN.

The thesis is structured as follows:

- Section 1 will introduce the main aspects of neutrino physics with a focus on neutrino oscillations specifically. This is required to understand the need for a project as daring and large as the DUNE experiment.
- Section 2 will provide an overview of the DUNE experiment and its role in the current landscape of neutrino research. The thesis will place particular emphasis on the Near Detector and in particular to GRAIN.
- Section 3 aims at providing a mathematical introduction to Machine Learning. A historical introduction and an explanation of its main drawbacks will hopefully help demystify this phenomenon.
- Section 4 will summarise the original work—the development of a Machine Learning-based reconstruction algorithm for GRAIN and its comparison to the standard reconstruction algorithm already in place.

Neutrino physics

Designed to provide the reader with a brief review of current open questions, this first section outlines the theoretical foundations of neutrino physics and related phenomena. This is a cardinal requirement to understand why neutrino physics is important and how current and upcoming neutrino experiments aim to discover the missing pieces. Section 1.1 and Section 1.2 will briefly yet formally introduce neutrinos, outlining their history, why, and how they were first postulated and formalised in the coming years. Section 1.3 will focus on the theory and phenomenology behind neutrino oscillations, followed by Section 1.4 which will provide an overview of the questions around neutrino masses. Section 1.5 will explain the *matter-effect*, a phenomenon that has a cardinal role in achieving DUNE’s physics goals and Section 1.6 attempts to provide a rough idea of the phenomenology expected at the LBNF energy range. Finally, Section 1.8 consists of an overview of the main neutrino experiments currently running or in construction outlining their purpose and physics objectives.

§1.1 Brief history of neutrino interactions

Neutrinos are the most abundant massive particles of the known universe. Indeed, the fact that they have mass at all is, without a doubt, one of the most important discoveries in particle physics of the last 30 years. Neutrinos were first postulated by Pauli in 1930 in a letter [2] sent to a group of colleagues he addressed as “Dear Radioactive Ladies and Gentlemen” to explain the observed continuous β -decay spectrum [3], something that is incompatible with the two-body final state hypothesis because of simple energy and momentum conservation arguments. Beta decay is one of the three possible kinds of atomic decay and, in the simplest case—e.g. the case of tritium—it is the process through which a neutron decays into a proton and an electron [4, p. 100] and it can be described by the reaction

$$n \rightarrow p^+ e^- \bar{\nu}_e . \quad (1.1)$$

Since the neutrino has—as the name suggests—no electric charge, is very light and has a very small cross-section, resulting in a low chance of interaction, physicists initially thought this process was a two-body decay. In the letter, Pauli claims he “do(es) not dare to publish anything about this idea” since “one probably would have seen those neutrons, if they exist, for a long time” and—quoting Debye—how “it’s better not to think about this at all, like new taxes.” Clyde Cowan and Frederick Reines originally planned to detonate a 20-kiloton nuclear explosive in order to detect these elusive particles using a detector able to withstand the explosion they referred to as “El Monstro” [5]. The actual discovery is dated 1956, when the two physicists decided to change the neutrino source to a nuclear reactor [6] not because of environmental or safety concerns, but because this modified setup allowed better noise rejection [5]. This discovery was awarded the Nobel prize in Physics in 1995.

The first theoretical framework of β -decay that included the neutrino was proposed by E. Fermi [7] in 1934. In the article, Fermi postulates a point-like interaction with no mediator in a Hamiltonian formalism. It was only in the ‘50s that the nowadays widely known Lagrangian formulation of this

hypothesis was first introduced by E. C. G. Sudarshan and R. E. Marshak [8] and shortly after independently by R. P. Feynman and M. Gell-Mann [9]:

$$\mathcal{L} = -\frac{G^{(\beta)}}{\sqrt{2}} \bar{p}\gamma^\alpha(1 - a\gamma_5)n \bar{e}\gamma_\alpha(1 - \gamma_5)\nu_e. \quad (1.2)$$

In this equation:

- $G^{(\beta)}$ is a constant known as the Fermi constant and it expresses the strength of the weak interaction.
- γ^α are the Dirac matrices defined by the relation $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\mathbb{1}_4$ with $\{, \}$ the anticommutation operator, $\eta^{\mu\nu} = \text{diag}(+, -, -, -)$ the mostly-negative Minkowski metric and $\mathbb{1}_4$ the four-dimensional identity matrix.
- The matrix $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is an explicit manifestation of the chiral nature of this interaction, mainly through the fact that $P_L := \frac{1-\gamma^5}{2}$ can be proven to be a projection operator that selects the left component of a spinor.
- Finally, the a coefficient takes into account the fact that protons and neutrons are not elementary particles.

§1.2 Neutrinos in the Standard Model

Formally,¹ the Standard Model (SM) is a gauge theory based on the local symmetry group $\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y$, where the subscripts C, L and Y represent colour, left-handed chirality and weak hypercharge respectively.

Since the $\text{SU}(3)_C$ symmetry is unbroken and since it does not mix with the $\text{SU}(2)_L \times \text{U}(1)_Y$ sector it is possible, for the purpose of this thesis, to focus exclusively on this latter part that represents the Electroweak Sector.

The symmetry group $\text{SU}(2)_L$ is referred to as *weak isospin*, its subscript L denotes that its elements act in a nontrivial way only on the left-handed components of fermion fields and that right-handed components are singlets under weak isospin transformations. The group has three generators T_a ($a = 1, 2, 3$), that satisfy the angular momentum commutation relations

$$[T_a, T_b] = i\varepsilon_{abc}T_c. \quad (1.3)$$

The symmetry group $\text{U}(1)_Y$ is called *weak hypercharge* and it is generated by the hypercharge operator Y . This operator is connected to T_3 and to the charge operator Q by the Gell-Mann-Nishijima relation

$$Q = T_3 + \frac{Y}{2}. \quad (1.4)$$

In order to lay the ground for a gauge theory one needs to add three vector gauge boson fields A_a^μ associated with the three generators T_a and one vector gauge boson field B^μ associated with the generator Y . In gauge theories it is common to introduce the *covariant derivative* as a way to absorb interactions in the kinetic term. For the electroweak theory the covariant derivative is

$$D_\mu := \partial_\mu - igT_a A_\mu^a - ig'\frac{Y}{2}B_\mu \quad (1.5)$$

¹The following theoretical sections have all been liberally adapted from the book by C. Giunti and C. W. Kim [10].

where g and g' are two independent coupling constants.

The left-handed chiral components of the fermion fields are grouped into *weak isospin doublets*. This choice has been guided from past experiences and has no formal reasoning. In fact from a purely theoretical point of view, one could develop this theory using the dimension three representation of $SU(2)_L$ and defining the lepton as an isospin triplet. Nevertheless, since the choice of a doublet leads to the observed phenomenology, there is currently no reason to assume the existence of a third isospin state.

The doublet of a generic lepton flavour α is defined as

$$L_L^\alpha := \begin{pmatrix} \nu_{\alpha L} \\ \alpha_L \end{pmatrix}. \quad (1.6)$$

The generators of $SU(2)_L$ are therefore fixed to be $T_a = \frac{\sigma_a}{2}$ where σ_a are the Pauli matrices and the doublets satisfy the relations

$$T_a L_L^\alpha = \frac{\sigma_a}{2} L_L^\alpha, \quad Y L_L^\alpha = -L_L^\alpha \quad \forall \alpha, a. \quad (1.7)$$

In the SM, neutrinos are therefore defined as the up-component of a $SU(2)_L$ weak isospin doublet used to provide a phenomenological description of leptons.

§1.2.1 Right-handed components

In the SM, the right handed components of charged leptons α_R are implemented as a singlet of isospin and right-handed neutrinos do not exist. This second fact was mainly guided by the two-component theory of neutrinos that was proven phenomenologically correct by C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson [11], with the first experimental proof of parity violation, and further confirmed by M. Goldhaber, L. Grodzins, and A. W. Sunyar [12], with the first experimental proof of neutrinos interacting exclusively as a left-handed chiral field. This has an indirect consequence. Indeed, the generic mass term of a fermion in the SM, generally referred to as Dirac mass term, is

$$\mathcal{L}_m = -m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \quad (1.8)$$

hence, the postulate that neutrinos do not have a right-handed component, also implies that they are massless. However, neutrino oscillations—described in the next section—are well-explained by introducing neutrino masses, thus a brief introduction on SM extensions that add them will be given in Section 1.4.

§1.3 Neutral particle oscillations

Neutral particle oscillations are the transmutation of a particle with no electric charge into a different neutral particle due to the change of some quantum number by means of an interaction that does not conserve said quantum number. This phenomenon was first investigated by M. Gell-Mann and A. Pais [13] in 1954.

There are two kinds of neutral particle oscillation:

- *Particle-antiparticle* oscillations, also often referred to as *mixing*, of neutral mesons $K^0 \leftrightarrow \bar{K}^0$, $D^0 \leftrightarrow \bar{D}^0$ and $B^0 \leftrightarrow \bar{B}^0$ [4, §13].

- *Flavour* oscillations of neutrinos $\nu_e \rightleftharpoons \nu_\mu \rightleftharpoons \nu_\tau$, experimentally confirmed in 1998 [1].

For the purpose of this thesis only an overview of this latter kind will be provided.

§1.3.1 Flavour oscillations

Light neutrinos in the SM are three. This was experimentally first proven in 1989 by the ALEPH experiment at CERN [14] by measuring the decay width of the Z^0 boson. Hence, to describe neutrino physics it is necessary to choose a triple of neutrinos. This is indeed a *choice*: since quantum mechanics relies fundamentally on linear algebra, describing neutrino physics is much like choosing a basis for a three-dimensional vector space: there are infinitely many valid triples of states that can be selected to form a complete basis. The standard procedure is to select special triples called *operator eigenstates*, that is, some specific triple of neutrinos that keeps some specific measurable quantity—formally called an observable—unchanged under some specific transformation—formally called an operator.

To describe neutrino oscillations just two triples are required: eigenstates of **interaction** (ν_e, ν_μ, ν_τ)—also referred to as *flavour eigenstates* since they keep their flavour constant under the interaction operator—and eigenstates of **propagation** (ν_1, ν_2, ν_3)—also referred to as *mass eigenstates* since they keep their mass constant under the propagation operator.

It is worth noting that the definition of neutrino flavour is *dynamic*. Indeed, since neutrinos cannot be directly observed, neutrino flavour is determined exclusively through interactions: a neutrino of flavour α is defined as the one produced with a α^+ anti-lepton or that produces a α^- lepton when interacting with matter [15].

The object to transform one triple in the other is a three-by-three complex-valued matrix called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [4, §14.3]

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (1.9)$$

This equation implies that a neutrino produced with flavour α is a linear superposition of all mass eigenstates whom all independently evolve under the propagation operator. The next subsection provides a formal description of neutrino oscillation probabilities in vacuum and shows how this superposition alone is enough to enable oscillations.

§1.3.2 Oscillations in vacuum

Let H_0 be the Hamiltonian of an interaction-less theory and $|\nu_i\rangle$ its eigenstates of eigenvalue E_i . To simplify the calculations two hypotheses will be assumed: that the eigenstates have the same momentum ($p_i = p$) and that the energy range is highly relativistic ($p \sim E \gg m_i$). Hence it is possible to define the propagation eigenstates as

$$H_0|\nu_i\rangle = E_i|\nu_i\rangle \quad \text{where} \quad E_i = \sqrt{m_i^2 + p^2} \simeq E + \frac{m_i^2}{2E}. \quad (1.10)$$

Furthermore, it is possible to state that the physical description of a neutrino with flavour α can be written as a coherent superposition of the previously-defined propagation eigenstates

$$|\nu(t=0)\rangle = |\nu_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle \quad (1.11)$$

where $U_{\alpha i}^*$ are the elements of the complex conjugate of the PMNS matrix. In the plane-wave approximation, considering these hypotheses, mass eigenstates evolve by changing their phase independently

$$|\nu(L)\rangle = \sum_{i=1}^3 U_{\alpha i}^* e^{-iE_i L} |\nu_i\rangle \quad (1.12)$$

where $L \sim t$ is the distance travelled by the neutrino.

The Born rule enables the calculation of the probability of a neutrino produced with a flavour α to interact as a neutrino with flavour β as a function of the travelled distance

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= |\langle \nu_\beta | \nu(L) \rangle|^2 = \\ &= \left| \sum_{i=1}^3 U_{\alpha i}^* U_{\beta i} e^{-iE_i L} |\nu_i\rangle \right|^2 = \sum_{i,j=1}^3 U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* e^{-i(E_i - E_j)L}. \end{aligned} \quad (1.13)$$

By applying the approximations of Equation 1.10 to this phase, it can be written in function of a new quantity usually referred to as the *squared-mass difference*

$$E_i - E_j = \left(E + \frac{m_i^2}{2E} \right) - \left(E + \frac{m_j^2}{2E} \right) = \frac{m_i^2 - m_j^2}{2E} =: \frac{\Delta m_{ij}^2}{2E}. \quad (1.14)$$

Notice how the oscillation probability depends on the quartic product $U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*$ that is independent from the PMNS matrix parametrisation. For reasons that will be clear shortly, it is interesting to separate the real and imaginary part of Equation 1.13

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \delta_{\alpha\beta} - 4 \sum_{i>j} \Re(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin^2 \Delta_{ij} \\ &\quad + 2 \sum_{i>j} \Im(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin \Delta_{ij} \end{aligned} \quad (1.15)$$

$$\text{where } \Delta_{ij} := \frac{\Delta m_{ij}^2 L}{4E}.$$

Transition probabilities therefore *oscillate* as a function of L/E and become relevant only when $\Delta_{ij} \approx \mathcal{O}(1)$. Accelerator experiments operating with $\mathcal{O}(1 \text{ GeV})$ neutrinos can therefore probe different ranges of Δm^2 depending on whether they have a short ($L \approx 10^2 \text{ m}$) or long ($L \approx (10^5 - 10^6) \text{ m}$) baseline [4, Table 14.1]. The Short Baseline Neutrino (SBN) program [Section 1.8.1.3], currently running at Fermilab, is an example of the former, whereas DUNE [Section 2] falls wholly within the latter category together with the NO ν A and T2K experiments [Section 1.8.1.4].

Antineutrino case If one were to repeat the calculations using antineutrinos and therefore begin from

$$|\bar{\nu}(t=0)\rangle = |\bar{\nu}_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i} |\bar{\nu}_i\rangle \quad (1.16)$$

the only difference encountered would be the exchange $U \leftrightarrow U^*$. The oscillation probability for antineutrinos is thus

$$\begin{aligned}
 P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) &= \delta_{\alpha\beta} - 4 \sum_{i>j} \Re(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin^2 \Delta_{ij} \\
 &\quad - 2 \sum_{i>j} \Im(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*) \sin \Delta_{ij}.
 \end{aligned}
 \tag{1.17}$$

Comparing Equation 1.15 and Equation 1.17 reveals that the sole difference between the oscillation probability of neutrinos and antineutrinos is the sign of the imaginary part. Differences in measurements of oscillation probabilities in the neutrino and antineutrino sector can therefore enable the study of the imaginary part of the PMNS matrix, a possible hint towards CP-violation (CPV) in the lepton sector. All neutrino oscillation experiments have attempted to measure CP-driven asymmetries, parametrised as

$$A_{\text{CP}} = \frac{P(\nu_\alpha \rightarrow \nu_\beta) - P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)}{P(\nu_\alpha \rightarrow \nu_\beta) + P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)}, \tag{1.18}$$

but current and past experimental efforts have not managed to determine with enough statistical confidence whether the asymmetry is present. In fact, the measurement of this quantity is one of the main physics drivers behind future generation neutrino experiments such as Hyper-Kamiokande (Hyper-K) [Section 1.8.1.1] and DUNE [Section 2.1].

§1.4 Neutrino masses

In Section 1.3 it was explained how the introduction of a neutrino mass is a valid way to explain the phenomenology of neutrino oscillations. Nevertheless, as explained in Section 1.2, it is important to take into account that in the SM right-handed neutrino fields do not exist, and that hence the standard fermionic mass term

$$\mathcal{L}_m = -m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \tag{1.19}$$

cannot be created. The immediate consequence of this hypothesis is that in the SM neutrinos are postulated as massless. Currently there are two main SM extensions that attempt to introduce neutrino masses: one generates the so-called *Dirac mass term* while the other the so-called *Majorana mass term*.

§1.4.1 Dirac mass term

The extension to the SM required to form a neutrino mass term in the Dirac formalism is the introduction of the right-handed components $\nu_{\alpha R}$. It is worth noting that the introduction of one field per flavour is not the minimal extension and that the presence of only one sterile neutrino field cannot be excluded, since not even right-handedness is required. Indeed, a right handed spinor can be obtained from a left-handed field by means of Equation 1.25.

The *minimally extended Standard Model* theory extends the standard Yukawa coupling to the Higgs field by introducing a term akin to the one responsible for the mass of up-type quarks [4, §11.2]

$$\mathcal{L}_Y = - \sum_{\alpha, \beta=e, \mu, \tau} Y_{\alpha\beta}'^\ell \bar{L}_{\alpha L} \Phi \ell_{\beta R}' - \sum_{\alpha, \beta=e, \mu, \tau} Y_{\alpha\beta}'^\nu \bar{L}_{\alpha L} \tilde{\Phi} \nu_{\beta R}' + h.c. \tag{1.20}$$

In this equation:

- Y'^ℓ is a three-by-three complex matrix. The prime notation distinguishes it from Y^ℓ , which conventionally represents its diagonalised form.

- $L_{\alpha L}$ is the left-handed doublet of $SU(2)_L$ defined in Equation 1.6.
- $\ell'_{\beta R}$ and $\nu'_{\beta R}$ are the fields of a right-handed lepton and neutrino of defined flavour β .
- Finally Φ is the Higgs field doublet acting as the mediator between left and right components, and $\tilde{\Phi} := i\tau_2 \Phi^*$ intrinsically allows to select the up component of the $L_{\alpha L}$ doublet.

This equation can be rewritten in the unitary gauge by setting:

$$\begin{aligned} \Phi &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \quad \tilde{\Phi} = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H \\ 0 \end{pmatrix} \\ \therefore \mathcal{L}_Y &= - \left(\frac{v + H}{\sqrt{2}} \right) [\bar{\ell}'_L Y'^\ell \ell'_R + \bar{\nu}'_L Y'^\nu \nu'_R] + h.c. \end{aligned} \quad (1.21)$$

where objects without explicit flavour index represent the collections containing all of them.

The standard Higgs mechanism continues with diagonalising the Yukawa matrices to find their eigenvectors $y_k^{\ell/\nu}$, rewriting the Lagrangian in this new basis, expressing it in terms of the Dirac fields $\psi = \psi_L + \psi_R$, and finally finding a term similar to Equation 1.19 where the mass is evident. This procedure shows that, in this framework, neutrino masses are given by

$$m_k = \frac{y_k^\nu v}{\sqrt{2}} \quad (k = 1, 2, 3). \quad (1.22)$$

It is worth noting that despite being proportional to the Higgs Void Expectation Value (VEV) $v \approx 250$ GeV just like charged leptons and quarks, neutrino masses are much smaller, which calls for very tiny Yukawa couplings. In fact, the SM Higgs mechanism does not constrain the values of any Yukawa coupling. This, coupled with the naturalness problem posed by the strikingly small values of neutrino masses compared with those of their electrically charged partners, leads to the belief that the SM could be an effective theory obtained from the low-energy limit of a more complete theory.

§1.4.2 Majorana mass term

The Dirac equation for a fermion field $\psi = \psi_L + \psi_R$

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \quad (1.23)$$

is equivalent to the system of two equations

$$\begin{cases} i\gamma^\mu \partial_\mu \psi_L = m\psi_R \\ i\gamma^\mu \partial_\mu \psi_R = m\psi_L \end{cases} \quad (1.24)$$

meaning that the spatial evolutions of chiral fields ψ_L and ψ_R are coupled by the mass. Since the equations are decoupled when $m = 0$, a massless fermion can be described by a single chiral field called a *Weyl spinor*. It is tempting to assume that therefore the description of a massive particle requires a four-component spinor, but this statement is actually false. Indeed, one can show that the spinor defined as

$$\psi_L^C := \mathcal{C} \overline{\psi}_L^T \quad (1.25)$$

where $\mathcal{C}$ is the charge conjugation operator, is a valid right-handed spinor and that the Majorana field defined as

$$\psi := \psi_L + \psi_L^C \quad (1.26)$$

can be used to introduce a new mass term called *Majorana mass*

$$\mathcal{L}_M^{\text{mass}} = -\frac{1}{2}m\overline{\nu_L^C}\nu_L + h.c. \quad (1.27)$$

It is worth noting that the Majorana spinor is intrinsically an eigenstate of charge conjugation

$$\mathcal{C}\psi = \psi \quad (1.28)$$

which implies the equality of particle and antiparticle. The natural consequence is that only neutral fermions can be described by a Majorana field.

The full Majorana Lagrangian density consisting of the kinetic terms for a neutrino can be written as².

$$\mathcal{L}_M = \overline{\nu_L} i \not{\partial} \nu_L - \frac{m}{2} (-\nu_L^T \mathcal{C}^\dagger \nu_L + \overline{\nu_L} \mathcal{C} \overline{\nu_L}^T) \quad (1.29)$$

where the kinetic term has the same form as that of a massless neutrino in the SM.

§1.4.3 Neutrino mass hierarchy

From an experimental point of view, to date, there have been no measurements of neutrino masses and attempts at direct measurements thereof have only been able to provide upper limits. Latest results from the KATRIN experiment [Section 1.8.2.3], set the limit for the β -decay neutrino at

$$m_\beta < 0.45 \text{ eV} \quad (90\% \text{ C.L.}) \quad [16]. \quad (1.30)$$

A set of three masses can be wholly described using two independent squared-mass differences separating the determination of the absolute scale from the neutrino oscillation phenomenology that depends solely on these former quantities as shown in Equation 1.15 and Equation 1.17. The *solar* mass splitting Δm_{21}^2 was measured to be positive by the SNO collaboration in 2002 [17] by exploiting the asymmetry between neutrinos and anti-neutrinos caused by the Mikheyev-Smirnov-Wolfenstein (MSW) effect [Section 1.5], whereas the same study on atmospheric neutrino oscillations has not yet been possible due to the much quicker oscillation frequency. This measurement is one of the main physics goals of DUNE. Nevertheless, the values of these oscillation parameters are known with fairly good accuracy: the NuFit collaboration [18] performs and provides combined studies using currently-available data from all neutrino experiments. The most recent results for the mass splittings are provided in Table 1.

Normal ordering	Inverted ordering
$\Delta m_{21}^2 = (7.537_{-0.10}^{+0.094}) \cdot 10^{-5} \text{ eV}^2$	$\Delta m_{21}^2 = (7.537_{-0.10}^{+0.094}) \cdot 10^{-5} \text{ eV}^2$
$\Delta m_{31}^2 = +(2.521_{-0.020}^{+0.021}) \cdot 10^{-3} \text{ eV}^2$	$\Delta m_{32}^2 = -(2.483 \pm 0.020) \cdot 10^{-3} \text{ eV}^2$

Table 1: Current best fit results on neutrino mass splittings, from NuFit 6.1 [18].

²The operator $\tilde{\partial}_\mu$ is defined as $\frac{\partial_\mu - \tilde{\partial}_\mu}{2}$ where $\psi \tilde{\partial}_\mu = \partial_\mu \psi$ is a derivative that acts on the left. Writing the Dirac equation this way has no practical purpose thanks to Gauss' theorem, but it makes the Lagrangian explicitly real "as a Lagrangian should be." [10, p. 7]

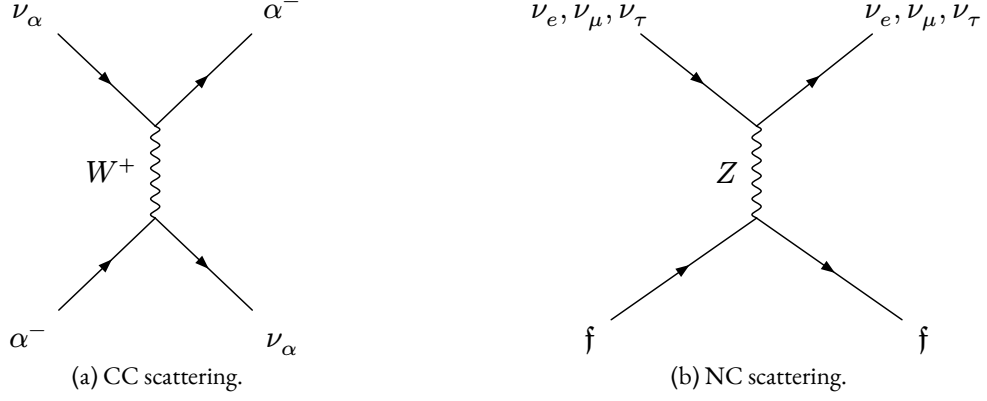


Figure 1: Feynman diagrams of neutrino Charged Current (CC) and Neutral Current (NC) interactions. NC interactions happen for every flavour of neutrino, whereas CC interactions require the same flavour lepton to be present and participate.

§1.5 Propagation through matter

Neutrinos, being neutral leptons, interact exclusively through the weak and gravitational interactions with this latter rendered almost irrelevant by their small masses. Additionally, effects from incoherent scatterings are relevant only if the nucleon density is of the order of 10^{35} cm^{-3} , that is in neutron stars and supernova cores [10, Intr. to Chap. 9]. Since DUNE's neutrino beam travels through the Earth, this effect should be considered only for neutrinos with energies exceeding 10^5 GeV , that is four orders of magnitude bigger than the DUNE beam's $\approx 2.5 \text{ GeV}$. Thus, in order to study neutrino propagation through matter, it is sufficient to consider exclusively coherent weak scattering.

The effective low-energy charged-current Hamiltonian density corresponding to Figure 1a is

$$\mathcal{H}_{\text{eff}}^{(\text{CC})}(x) = \frac{G_F}{\sqrt{2}} [\bar{\nu}_e(x) \gamma^\mu (1 - \gamma_5) e(x)] [\bar{e}(x) \gamma_\mu (1 - \gamma_5) \nu_e(x)] . \quad (1.31)$$

In order to separate neutrino and electron contributions one can perform what is called a Fierz transformation

$$\mathcal{H}_{\text{eff}}^{(\text{CC})}(x) = \frac{G_F}{\sqrt{2}} [\bar{\nu}_e(x) \gamma^\mu (1 - \gamma_5) \nu_e(x)] [\bar{e}(x) \gamma_\mu (1 - \gamma_5) e(x)] . \quad (1.32)$$

Modelling neutrino propagation through matter begins by finding the average of this Hamiltonian density over the electron background in the rest frame of the medium through which the propagation happens

$$\begin{aligned} \overline{\mathcal{H}_{\text{eff}}^{(\text{CC})}}(x) &= \frac{G_F}{\sqrt{2}} [\bar{\nu}_e(x) \gamma^\mu (1 - \gamma_5) \nu_e(x)] \int d^3p_e f(E_e, T) \\ &\quad \times \frac{1}{2} \sum_{h_e=\pm 1} [\langle e^-(p_e, h_e) | \bar{e}(x) \gamma_\mu (1 - \gamma_5) e(x) | e^-(p_e, h_e) \rangle] \end{aligned} \quad (1.33)$$

where $|e^-(p_e, h_e)\rangle$ is a one-electron state with definite momentum and helicity and $f(E_e, T)$ is the statistical distribution of electron energy E_e which also depends on the temperature T and is normalised by

$$\int d^3p_e f(E_e, T) = n_e V \quad (1.34)$$

with n_e the electron density and V the volume such that $n_e V$ is the total number of electrons. This average amounts to

$$\overline{\mathcal{H}_{\text{eff}}^{(\text{CC})}}(x) = V_{\text{CC}} \bar{\nu}_{e,L}(x) \gamma^0 \nu_{e,L} \quad (1.35)$$

where the charged-current potential V_{CC} is given by

$$V_{\text{CC}} = \sqrt{2} G_F n_e. \quad (1.36)$$

Analogously, the neutral-current potential of neutrinos propagating in a medium with density n_f of fermions f can be calculated: the effective NC Hamiltonian corresponding to the Feynman diagram in Figure 1b is

$$\mathcal{H}_{\text{eff}}^{(\text{NC})} = \frac{G_G}{\sqrt{2}} \sum_{\alpha=e,\mu,\tau} [\bar{\nu}_\alpha(x) \gamma^\mu (1 - \gamma^5) \nu_\alpha(x)] \sum_f [\bar{f}(x) \gamma_\mu (g_V^f - g_A^f \gamma^5) f(x)]. \quad (1.37)$$

where g_V^f and g_A^f are coupling constants that take into account effects from non-elementary particles like protons and neutrons.

One could perform the calculations, or, by comparing this with Equation 1.36, convince themselves that the neutral-current potential for a neutrino ν_α due to coherent interaction with fermions f is

$$V_{\text{NC}}^f = \sqrt{2} G_F n_f g_V^f. \quad (1.38)$$

Since $g_V^e = -g_V^p$ and supposing the medium to be electrically neutral, the contributions of electrons and protons cancel themselves out. The contribution from neutrons, for which $g_V^n = -\frac{1}{2}$, is therefore the only remaining one.

Summarising, the effective neutrino potential Hamiltonian density in low temperature and low density environments, as a result of charged and neutral current interactions, is

$$\overline{\mathcal{H}_{\text{eff}}}(x) = \sum_{\alpha=e,\mu,\tau} V_\alpha \bar{\nu}_{\alpha,L}(x) \gamma^0 \nu_{\alpha,L}(x) \quad (1.39)$$

where $V_\alpha = V_{\text{CC}} \delta_{\alpha e} + V_{\text{NC}} = \sqrt{2} G_F \left(N_e \delta_{\alpha e} - \frac{n_n}{2} \right).$

By repeating analogous calculations, it is possible to describe the behaviour of antineutrinos travelling through the earth's mantle. The only difference is that the effective potential switches its sign

$$\bar{V}_\alpha = -V_\alpha. \quad (1.40)$$

This sign change has one important phenomenological consequence [see Section 1.5.1].

§1.5.1 Oscillations through a medium

The total Hamiltonian that describes neutrino oscillations through matter is therefore the sum

$$H = H_0 + H_{\text{eff}} \quad (1.41)$$

where H_0 is the free Hamiltonian whereas H_{eff} is the newly-calculated interaction Hamiltonian that is defined as diagonal on flavour eigenstates $H_{\text{eff}} |\nu_\alpha\rangle = V_\alpha |\nu_\alpha\rangle$.

The transition amplitude

$$\psi_{\alpha\beta}(t) := \langle \nu_\beta | \nu_\alpha(t) \rangle \text{ with } \psi_{\alpha\beta}(0) = \delta_{\alpha\beta} \quad (1.42)$$

is such that its modulus squared quantifies the probability that a neutrino born at $t = 0$ with flavour α is found to have flavour β at a time t

$$P(\nu_\alpha \rightarrow \nu_\beta; t) := |\psi_{\alpha\beta}(t)|^2. \quad (1.43)$$

In the Schrödinger picture, time evolution is described by the Schrödinger equation

$$i \frac{d}{dt} |\nu_\alpha(t)\rangle = H |\nu_\alpha(t)\rangle, \quad (1.44)$$

Taking the time derivative of Equation 1.42 yields

$$i \frac{d}{dt} \psi_{\alpha\beta}(t) = \sum_\eta \left(\sum_i U_{\beta i} E_i U_{\eta i}^* + \delta_{\beta\eta} V_\beta \right) \psi_{\alpha\eta}(t). \quad (1.45)$$

In Section 1.3 we discussed how neutrino oscillations arise from *relative* phases between the different flavour components. As a result, any global phase shift is phenomenologically irrelevant. We can use this fact to discard V_{NC} contributions—since they are equal for all flavours—and one term of the sum, for example $i = 1$ to express this result in terms of squared mass differences. Equation 1.45 then becomes

$$i \frac{d}{dt} \psi_{\alpha\beta}(t) = \sum_\eta \left(\sum_i U_{\beta i} \frac{\Delta m_{i1}^2}{2E} U_{\eta i}^* + \delta_{\beta\eta} \delta_{\eta e} V_{\text{CC}} \right) \psi_{\alpha\eta}(t). \quad (1.46)$$

This equation shows how neutrino oscillations in matter, just like neutrino oscillations in vacuum, depend on the squared-mass differences, not on the absolute mass values. We can rewrite this in matrix form

$$i \frac{d}{dt} \Psi_\alpha = \mathbb{H}_F \Psi_\alpha \quad (1.47)$$

which has the shape of a Schrödinger equation with the effective Hamiltonian matrix $\mathbb{H}_F$ in the flavour basis being

$$\mathbb{H}_F := \frac{1}{2E} (U \mathbb{M}^2 U^\dagger + \mathbb{A}) \quad (1.48)$$

In the case of three neutrino mixing, the components of this equation are defined thusly

$$\Psi_\alpha := \begin{pmatrix} \psi_{\alpha e} \\ \psi_{\alpha \mu} \\ \psi_{\alpha \tau} \end{pmatrix}, \quad \mathbb{M}^2 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Delta m_{21}^2 & 0 \\ 0 & 0 & \Delta m_{31}^2 \end{pmatrix}, \quad \mathbb{A} = \begin{pmatrix} A_{\text{CC}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (1.49)$$

where

$$A_{\text{CC}} := 2E V_{\text{CC}} = 2\sqrt{2} E G_F n_e. \quad (1.50)$$

It is worth noting that this quantity, given its dependence on V_{CC} which has opposite sign for neutrinos and antineutrinos (see Equation 1.40), introduces a matter-antimatter asymmetry wholly independent of the vacuum oscillations one A_{CP} . In fact, one can prove that, in the simpler two-neu-

trino case, under the hypothesis of a vacuum mixing angle θ and mass difference Δm^2 , the measured mixing angle becomes equal to

$$\tan 2\theta_{\text{eff}} = \frac{\tan 2\theta}{1 - \frac{A_{\text{CC}}}{\Delta m^2 \cos 2\theta}}. \quad (1.51)$$

This phenomenon is called the Mikheyev-Smirnov-Wolfenstein (MSW) effect and is also known as the *matter effect*. This asymmetry has a resonance when $A_{\text{CC}}^R = \Delta m^2 \cos 2\theta$ meaning that this effect can lead to maximal mixing even in the case of extremely small or absent vacuum mixing. Furthermore, this effect explicitly depends on the sign of the squared mass difference, playing a crucial role in unveiling the mysterious neutrino mass ordering.

§1.6 Neutrino cross sections

The phenomenology of neutrino interactions strongly depends on their energy scale. A fairly dated, but in-depth review of neutrino cross sections written by J. A. Formaggio and G. P. Zeller [19] separates them into five different energy ranges: threshold-less processes with $E_\nu \approx (0 - 1)$ MeV, low energy nuclear processes with $E_\nu \approx (1 - 100)$ MeV, intermediate energy cross sections with $E_\nu \approx (0.1 - 20)$ GeV, high energy cross sections with $E_\nu \approx (20 - 500)$ GeV, and ultra high energy neutrinos with $E_\nu \approx (0.5 - 1000)$ TeV. The Long Baseline Neutrino Facility (LBNF) spectrum range is thus wholly contained within the *intermediate* energy range, where neutrino scattering mechanisms fall into three main categories:

- *Elastic (NC) and quasi-elastic (CC) scattering*: neutrinos scatter off an entire nucleon through the exchange of a neutral Z or charged W boson.
- *Resonance production*: neutrinos excite the nucleon creating a baryonic resonance that decays to a variety of possible mesonic final states.
- *Deep Inelastic Scattering (DIS)*: the neutrino can resolve the individual quark constituents of the nucleon with the result of a hadronic shower.

For energies below ≈ 2 GeV, neutrino-hadron interactions are predominantly quasi-elastic: the (anti)neutrino scatters off of a target neutron (proton) converting it to a proton (neutron)

$$\nu_\mu n \rightarrow \mu^- p, \quad \bar{\nu}_\mu p \rightarrow \mu^+ n. \quad (1.52)$$

Such interactions with simple topology were extensively studied both theoretically and experimentally on old deuterium-based setups, but modern-day experiments generally use complex nuclei as their neutrino target. A compelling and sufficient reason for this choice is that complex nuclei are denser, allowing for high-mass detectors with significantly smaller volumes. This kind of interaction is generally studied starting from a relativistic Fermi Gas model that is then corrected with a variety of different techniques. A much more in-depth and updated review of the phenomenological aspects of neutrino cross sections was written by S. Dolan, L. Pickering, P. Stowell, C. Wilkinson, and C. Wret [20].

For what concerns experimental studies, the Main Injector Experiment to study ν -A (MINERvA) experiment, between 2009 and 2019, has devoted time and resources to the study and measurement of neutrino interactions on various nuclei at the GeV energy range including C, CH, Pb and Fe, providing invaluable data that is used to the present day [21]. Because of DUNE's strict requirements, the use of literature-derived data and models on neutrino-nucleus interactions is not enough to reach the required measurement uncertainties [22, §1.3]. Two argon targets at the DUNE ND, GRAIN

[Section 2.6] and ND-LAr [Section 2.4.1.1], will directly measure $\nu - \text{Ar}$ cross sections, reducing systematics on DUNE long baseline oscillation studies.

§1.7 Argon as a target

Argon is the 18th element of the periodic table; it is a noble gas with simple chemistry and it constitutes around 0.9 % of the earth's atmosphere making it commercially available and relatively cheap. In recent years argon has gathered increasing attention from physicists for its properties in rare-event searches like neutrino oscillation and dark matter experiments. All the data and information provided in this section are extracted from an in-depth study of liquid argon detectors written in 2024 by W. M. Bonivento and F. Terranova [23].

§1.7.1 Ionisation

When a charged particle propagates through a medium, a phenomenon called *ionisation* takes place: if the ionising particle deposits an energy greater than a material-dependent threshold (≈ 25 eV for argon) an electron is extracted from a target nucleus forming an electron-ion pair. If the target material is then subject to an electric field, the ionisation electrons will drift towards the anode responsible for generating the field where they can be detected. A two dimensional detection at the anode plane, plus measurement of detection time, coupled with knowledge of the medium properties driving electron motion and attenuation, is enough to produce a three-dimensional spatial reconstruction of the physical interaction. A detector able to perform this measurement is called a Time Projection Chamber (TPC), a technology that has seen widespread use in particle physics from its inception to the modern day [4, §35.6.4].

A Liquid Argon Time Projection Chamber (LArTPC) is a TPC in which the selected target material is Liquid Argon (LAr) first introduced in 1977 by C. Rubbia [24]. The choice of LAr as the target material for a TPC offers several upsides. Namely, its widespread availability and contained costs allow the construction of large instrumented volumes at reasonable prices. Furthermore, its radiation length $X_0 = 14.0$ cm ensures that a volume of about $\approx 1 \text{ m}^3$ of LAr is enough to construct a homogeneous Electromagnetic CALorimeter (ECAL). Nevertheless, when performing precision studies—as is often the case when this kind of detector is used—an accurate knowledge of systematic downsides is required. For example, the drift velocity in LAr of ionisation electrons subject to a typical electric field with an intensity of $\mathcal{O}(10^2 \text{ V cm}^{-1})$ is $\approx 1.6 \text{ m ms}^{-1}$, making this kind of detector unsuitable if quick detector response is needed. Ways of mitigating this systematic effect do exist, like the ArgonCube technology on which the ND-LAr detector is based [Section 2.4.1.1]. In addition, the ionisation electrons drifting towards the anode will inevitably diffuse in the plane orthogonal to the motion. This transverse diffusion has been measured to be normally distributed with a standard deviation of ≈ 1.3 mm after 1 m of drift, making anode segmentation below 1 mm both impractical and unnecessary. Finally the presence of pollutants in the medium can strongly impact the signal quality and most of all intensity. Indeed, the concentration of drift electrons $[e]$ decreases exponentially in time as

$$[e](t) = [e](t=0)e^{-k_A[S]t} \quad (1.53)$$

where $[S]$ is the concentration of impurities and k_A is the *attachment rate constant*, which is specific to each kind of impurity [25]. Argon purity is typically expressed in terms of the electron lifetime, defined as $\tau = 1/(k_A[S])$. This lifetime usually spans a few ms. For example, data from ICARUS Run 2 yielded average values of 4(8) ms in the East(West) cryostat [26, §V.A]. Argon purity is thus an

extremely important parameter to keep track of during data acquisition as it can directly imply some signal loss.

§1.7.2 Scintillation

The passage of an ionising particle in argon does not just generate electron-ion (Ar^+) pairs, but also excited (Ar)_{ex} atoms in the ratio $(\text{Ar})_{\text{ex}}/(\text{Ar}^+) = 0.21$. In a process known as *self-trapping*, lasting approximately $(1 - 10)$ ps, these excited atoms collide with other Ar atoms forming the Ar_2^* excited dimer, often referred to as *excimer*. The excimers can also form from the (Ar^+) ions through various subsequent processes including *electron recombination*, albeit in a slightly longer time-scale of around 100 ps. Said excimer can exist in two almost degenerate (same energy) states: the *singlet* $^1\Sigma_u^+$ and the *triplet* $^3\Sigma_u^+$. The excimer de-excitation happens through the gamma radiation process $\text{Ar}_2^* \rightarrow \gamma + \text{Ar} + \text{Ar}$ with a photon spectrum well-described by a Gaussian bell peaked at 128 nm and a Full Width Half Maximum (FWHM) of 6 nm. This light emission has two characteristic times: a short one $\tau_s \approx (2 - 6)$ ns and a long one $\tau_l \approx (1.1 - 1.6)$ μs associated with the singlet and triplet lifetimes respectively. These time-distributed pulses have been shown to depend on the nature of the ionising particle, enabling Particle IDentification (PID) techniques based on pulse shape analyses [23, Figure 2]. The fairly good photon yield of $\mathcal{O}(10^5 \gamma/\text{MeV})$ coupled with its quick emission time enables the use of LAr scintillation light for trigger purposes, providing for example the t_0 required for 3-dimensional reconstruction in a LArTPC.

Nevertheless the signal is wholly contained in the Void Ultra Violet (VUV) part of the electromagnetic spectrum, posing challenges connected to its detection. Indeed, older light-detection technologies like photo-multiplier tubes (PMTs) have negligible quantum efficiency at the VUV range, requiring the use of expensive and, most of all fragile, wavelength-shifting coatings. Newer technologies like silicon photomultipliers (SiPMs) —preferred for their in general cheaper cost, increased sturdiness, and operation simplicity—do reach higher, but still unsatisfactory, detection efficiencies at these wavelength ranges. To this extent, xenon-doping of LAr, is a very simple and cheap solution to this conundrum. As better explained by E. Segreto [27] slight xenon-doping of LAr, even at the small tens of ppm range, is able to completely shift the scintillation light from 128 nm to 174 nm, where the detection efficiency of standard Ultra Violet (UV) light-detection devices is at the level required by most implementations. For example, in the ProtoDUNE-SP prototype [Section 2.3.1], a doping of just 5.8 of Xe doubled the light collection efficiency [23, §VII.A].

§1.8 Other neutrino experiments

This section provides a non-exhaustive overview on running and planned neutrino experiments other than DUNE, which is the main subject of Section 2. The first part [Section 1.8.1], introduces other experiments that are studying neutrino oscillations specifically, whereas the second [Section 1.8.2] will focus on a selection of experiments that aim to increase the available knowledge about neutrinos using different phenomenons.

§1.8.1 Oscillation experiments

Hyper-Kamiokande Hyper-Kamiokande (Hyper-K) is an upcoming long baseline neutrino oscillations experiment set to begin data-taking in 2028 [28]. The far detector consists of a new 258kton water Cherenkov detector instrumented with photo-multiplier tubes (PMTs), whereas the near detectors will be recycled from the Tokai to Kamioka (T2K) experiment. Hyper-K physics target includes both

solar and beam neutrinos produced at the Japan Proton Accelerator Research Complex (J-PARC) with the objective of investigating T2K’s evidence of maximal CP-violation. Its recent analysis [29] showed Hyper-K will reach an expected 5σ sensitivity on 60% δ_{CP} values after 10 years of data-taking and an expected mass-ordering sensitivity of $3.5 - 4.5\sigma$, depending on oscillation parameters, after 6 years of data-taking. Hyper-K is therefore on track to achieve the discovery of maximal CP-violation in the lepton sector and to provide evidence for either neutrino mass ordering. However, it lacks the statistical significance required to claim the discovery of the neutrino mass ordering, which is the primary physics goal of DUNE [30, §2.1.1].

JUNO The Jiangmen Underground Neutrino Observatory (JUNO) is a 20 kt liquid scintillator detector located 52.5 km from multiple reactor cores designed to resolve the interference pattern of reactor neutrinos with sub-percent precision. Neutral particle oscillation spectrums are the product of two oscillating components: a slow and a fast one. As shown in Figure 1 of A. Cabrera et al. [31], the quick oscillation depends on neutrino mass ordering, hence it is possible to determine it by measuring this spectrum with high-enough energy resolution. This effect is indeed one of the main drivers behind the construction of this medium baseline neutrino oscillation experiment that has recently shown its ability to reduce uncertainties on reactor neutrino oscillations with just two months of data-taking [32].

SBN program The Short Baseline Neutrino (SBN) program is a neutrino physics experiment currently running at Fermilab. Its structure is that of a standard beam oscillations experiment with the Short Baseline Near Detector (SBND) placed 110 m from the neutrino source and the repurposed Imaging Cosmic And Rare Underground Signals (ICARUS) detector 600 m from it. The primary physics objective of SBN is the search of a sterile neutrino with mass at the eV scale hinted at by anomalies in various oscillation (LSND, MiniBooNE) and gallium-based solar neutrino experiments (GALLEX, SAGE) [33]. The SBN program’s detectors are aligned with the Booster Neutrino Beam (BNB), but ICARUS’s location also falls about 5.6° off-axis the more powerful NuMI beam used to fuel the NO ν A experiment, allowing it to combine data from multiple beams while also providing valuable data on $\nu - \text{Ar}$ cross sections at an energy range also relevant for DUNE. ICARUS has recently released a ν_μ disappearance analysis of BNB-only data from its second physics run [26]. While no evidence of 3+1 oscillation is observed, this analysis is systematics-limited as cross section and flux effects are largely unconstrained due to the lack of near detector data—SBND started its physics run in 2024.

NO ν A and T2K The NuMI Off-Axis ν_e Appearance (NO ν A) and Tokai to Kamioka (T2K) experiments are current-generation long baseline neutrino oscillation studies located in the USA and Japan respectively. NO ν A studies the Neutrinos at the Main Injector (NuMI) neutrino beam produced at Fermilab using two functionally identical detectors: the smaller 0.3 kt near detector is located 1 km downstream from the neutrino production target, whereas the bigger 14 kt far detector is located 810 km away in Minnesota. Both are centred 14.6 mrad off the central beam axis yielding a narrow-band neutrino beam peaked at 1.8 GeV. The detectors consist of highly segmented tracking calorimeters made out of long plastic cells filled with a liquid scintillator. T2K studies the Japan Proton Accelerator Research Complex (J-PARC) neutrino beam with a primary magnetised tracker located 280 m from the target and 43.6 mrad off-axis dubbed ND280, and the Super-Kamiokande 50 kt water Cherenkov light detector located 295 km downstream. In September 2025 the two experiments published a joint analysis combining data from the two experiments [34]. This analysis revealed a result

often cited as the “NOvA-T2K tension”: the measured values of δ_{CP} from the two experiments do not agree under the hypothesis of normal mass ordering [34, Extended Data Fig. 9]. Nevertheless the tension is not present under the assumption of inverted ordering and neither experiments nor their combined data show any strong preference for either mass ordering hence the interpretation of these results will be strongly influenced by future mass ordering measurements.

§1.8.2 Other neutrino studies

CUORE/CUPID The Cryogenic Underground Observatory for Rare Events (CUORE) is an experiment searching for neutrinoless double-beta decay ($0\nu\beta\beta$) that recently concluded its data-taking phase. This hypothesised extremely rare interaction, more properly expressed as $(A, Z) \rightarrow (A, Z + 2) + 2e^-$ has never been observed and it can only be explained by mechanisms Beyond the Standard Model since it breaks its only residual symmetry, $B - L$ [35]. The observation of this interaction would prove the Majorana nature of neutrinos explained in Section 1.4.2. CUORE has operated at the Laboratori Nazionali del Gran Sasso (LNGS) underground facility for 10 years and is now being decommissioned in view of the upgrade to Cuore Upgrade with Particle IDentification (CUPID), where the same infrastructure is used to host cryogenic calorimeters with scintillation properties to provide a substantial reduction of CUORE’s dominant background source, degraded alpha particles [36], thus reaching a sensitivity of $\sim 10^{27}$ yr on the half-life of $0\nu\beta\beta$ [37]. Since the decay rate is proportional to $Q_{\beta\beta}^5$, switching from $^{130}\text{TeO}_2$ crystals ($Q_{\beta\beta} = 2.6$ MeV) to $\text{Li}_2^{100}\text{MoO}_4$ crystals ($Q_{\beta\beta} = 3034$ keV) allows for a higher chance of signal while also moving the Region Of Interest (ROI) past most of the β/γ background. The new Ge-based light readout to discriminate α events enabled by the scintillating properties of Li_2MoO_4 , the increase of the polypropylene neutron shield and the addition of a muon veto system will allow CUPID to reach a background of $\approx 10^{-4}$ counts/keV/kg/year (ckky), the main component of which comes from $2\nu\beta\beta$ pileup that can be rejected thanks to the light detection device’s fast response of around 0.5 ms [36].

LEGEND The Large Enriched Germanium Experiment for Neutrinoless $\beta\beta$ Decay (LEGEND) experiment is taking data for the search for $0\nu\beta\beta$ at LNGS. To pursue this search, LEGEND utilises High Purity Germanium (HPGe) detectors enriched with (86 – 92) % ^{76}Ge acting as both radioactive source and active read-out. The LEGEND-200 detector started data-taking in 2023. A combined analysis of its early results with those from two other past experiments (GERDA and the Majorana Demonstrator) increased the half-life of $0\nu\beta\beta$ decay to $T_{1/2}^{0\nu} > 1.9 \cdot 10^{26}$ yr (90% C.L.) corresponding to an upper limit on the effective Majorana mass in the range $m_{\beta\beta} < (75 - 200)$ meV [38]. To tag and reject background signals, the HPGe detectors are immersed in a cryostat filled with LAr that is itself immersed in a water tank instrumented with 63 PMTs for a total estimated background level of $(5_{-2}^{+3}) \cdot 10^{-4}$ ckky. After 5 years of data-taking the LEGEND-200 detector is planned to be decommissioned to make way for the LEGEND-1000 detector. LEGEND-1000 will work very similarly to its successful predecessor sporting an increase to 1 t of Germanium and an expected background reduced by an order of magnitude with the objective of pushing the sensitivity on $0\nu\beta\beta$ decay to 10^{28} yr corresponding to $m_{\beta\beta} < (20 - 30)$ meV [39].

KATRIN The KARlsruhe TRItium Neutrino (KATRIN) experiment aims to directly measure neutrino masses by performing precision spectroscopy of tritium β -decay close to its kinematic endpoint. The molecular tritium β -decay observed by KATRIN

$$T_2 \rightarrow {}^3\text{HeT}^+ + e^- + \bar{\nu}_e \quad (1.54)$$

has a kinematic endpoint $E_0 \approx 18.6$ keV and a three-body final state. Therefore, from conservation arguments, the electron energy can at most equal the difference between the reaction's endpoint and the neutrino mass m_β . By applying a well-known electrostatic potential U , the KATRIN experiment is able to select exclusively electrons with an energy $E_e \in [E_0 - 300 \text{ eV}, E_0 + 135 \text{ eV}]$ that are then collimated on a 148-pixel silicon-PIN-diode array. For the neutrino mass analysis only electrons with energies above $E_0 - 40 \text{ eV}$ are selected: the region above the endpoint is used to constrain the background rate and the points below the ROI serve calibration and monitoring purposes. The latest results [40], based on the first five measurement campaigns provide the most stringent upper limits available on the neutrino mass:

$$m_\beta^2 = -(0.14_{-0.15}^{+0.13}) \text{ eV}^2, \quad m_\beta < 0.45 \text{ eV (90\% C.L.)} . \quad (1.55)$$



The DUNE experiment

Under construction in the United States, the Deep Underground Neutrino Experiment (DUNE) is a neutrino physics experiment designed to investigate fundamental questions in neutrino physics. The DUNE collaboration spans tens of international scientific institutions, summing up to almost two thousand scientists and engineers. The experiment is split into three main efforts: the Long Baseline Neutrino Facility (LBNF) beam currently under construction at Fermi National Accelerator Laboratory (Fermilab), the Near Detector (ND) to be placed 600 m from the neutrino source, and the Far Detector (FD) located 1.5 km underground at Sanford Underground Research Facility (SURF) for a total baseline of ≈ 1300 km. As is standard for large-scale efforts, DUNE will adopt a staged approach with two main phases: Phase I (≈ 5 yr) and Phase II (≈ 10 yr), corresponding to a total runtime of around 15 yr. Details about what the staging implies for each part of the experiment are provided in the respective sections.

§2.1 DUNE physics

As we have shown by comparing Equation 1.15 with Equation 1.17, *if* the PMNS matrix has an imaginary part then the oscillation probability for neutrinos and antineutrinos is different. Since the PMNS matrix in the three-neutrino model is a three-by-three unitary matrix, it has nine degrees of freedom. Nevertheless, all physical observables containing it, for example the charged current

$$\mathcal{L}_{CC} \propto \bar{\ell}_\alpha U^{\alpha i} \gamma_\mu \nu_i W^\mu \quad (2.1)$$

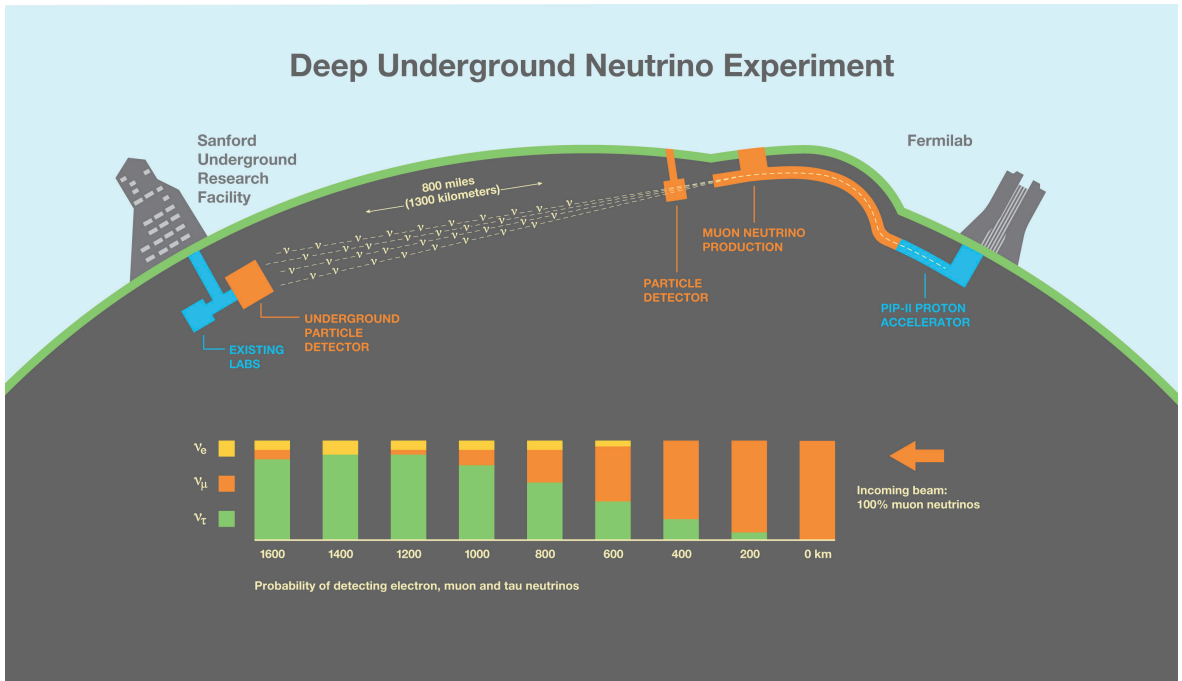


Figure 2: Poster of the DUNE experiment: neutrinos are generated at Fermilab and aimed at SURF where they are detected. The neutrino flavour oscillates as a function of the distance travelled as explained in Section 1.3.

are always constructed as the product of two spinors, hence a further five phase differences can be absorbed, leaving us with just four degrees of freedom. The PMNS matrix can therefore be parametrised using four real numbers $(\theta_{12}, \theta_{23}, \theta_{13}, \delta_{\text{CP}})$ and written as the product of three rotation matrices

$$U_{\text{PMNS}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix}}_{\text{atmospheric}} \underbrace{\begin{pmatrix} c_{13} & 0 & e^{-i\delta_{\text{CP}}} s_{13} \\ 0 & 1 & 0 \\ -e^{i\delta_{\text{CP}}} s_{13} & 0 & c_{13} \end{pmatrix}}_{\text{accelerator}} \underbrace{\begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{solar}} \quad (2.2)$$

where s_{ij} and c_{ij} are shorthands for $\sin \theta_{ij}$ and $\cos \theta_{ij}$ respectively.

In this parametrisation, the complex — and therefore CP-violating — nature of the PMNS matrix is expressed by the single parameter δ_{CP} . Neutrino oscillation experiments can be sensitive to any of these four parameters depending mainly on the physical source of neutrinos, their characteristic energies, and oscillation baseline.

It is worth pointing out that this parametrisation is correct only in the case of Dirac-type neutrino mass. Indeed the Majorana mass term of Equation 1.27 is not invariant under phase shifts, hence two additional CP-violating phases need to be introduced for Majorana neutrinos

$$U_{\text{PMNS}}^{\text{Maj}} = U_{\text{PMNS}} \text{diag}(1, e^{i\alpha}, e^{i\beta}). \quad (2.3)$$

This fact is of vital importance for experiments searching for $0\nu\beta\beta$ like those briefly explained in Section 1.8.2, however since neutrino oscillation experiments are insensitive to these two extra parameters, they will be ignored for the rest of this thesis [41].

§2.1.1 The PMNS complex phase δ_{CP}

DUNE will attempt to directly measure the vacuum oscillation asymmetry due to the complex phase δ_{CP} . To achieve this, the LBNF beam will work in both Forward Horn Current (FHC) and Reverse Horn Current (RHC) mode to produce a *mostly* neutrino or antineutrino beam respectively.

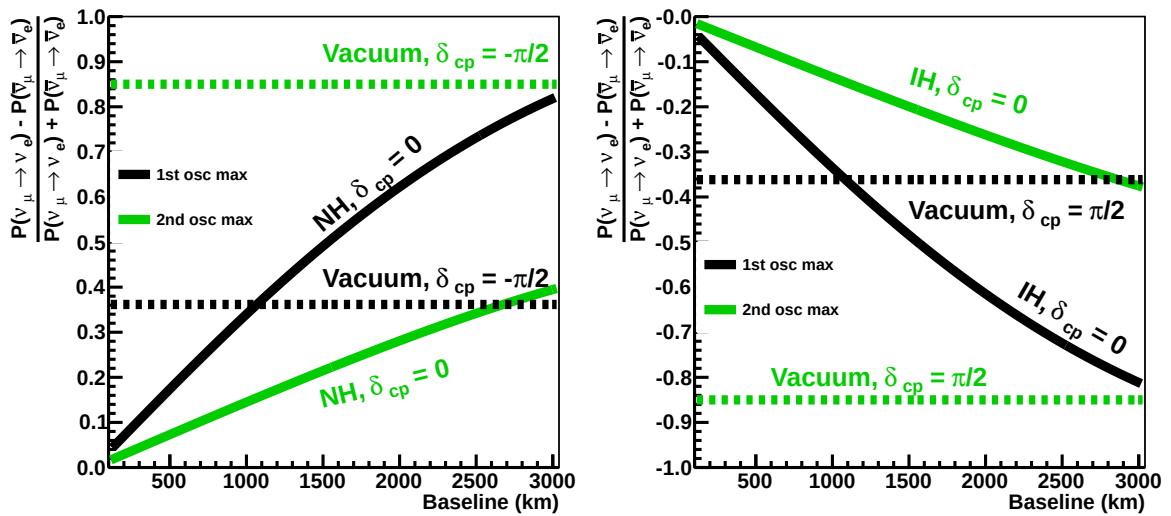


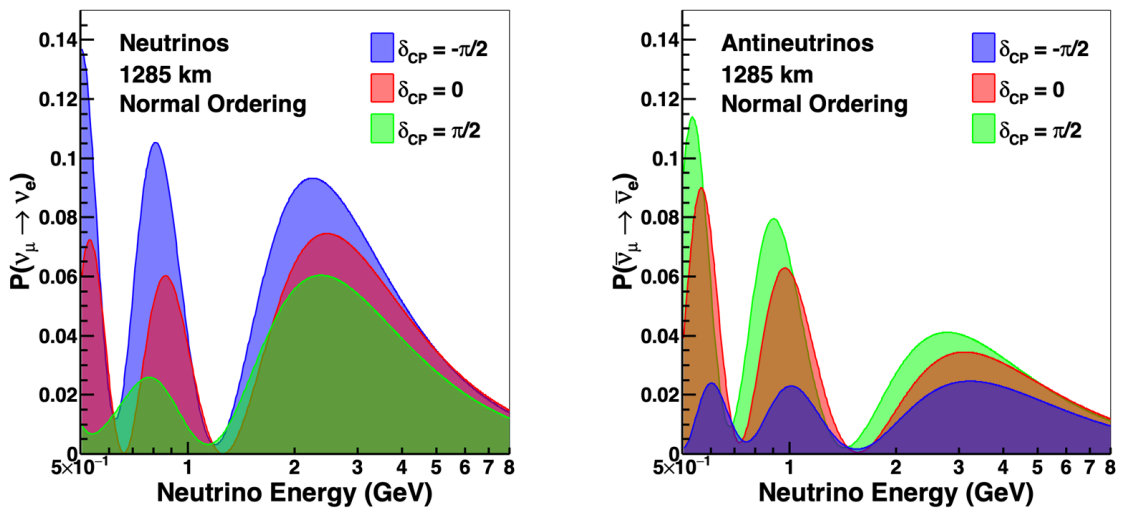
Figure 3: Asymmetry vs baseline at the first (black line) and second (green line) oscillation maxima. The solid lines show the asymmetry due to the matter effect only ($\delta_{\text{CP}} = 0$) for the normal hierarchy (left) or inverted hierarchy (right). The dashed lines show the maximum CP asymmetry in vacuum, with $\delta_{\text{CP}} = -\frac{\pi}{2}$ (left) or $\delta_{\text{CP}} = \frac{\pi}{2}$ (right). Source: [42].

As explained in Section 1.5, the observed matter-antimatter asymmetry in DUNE will be simultaneously driven by two effects, δ_{CP} and the MSW effect, making the ability to separate them of vital importance. In DUNE, the combined effect of a long baseline and wide energy range covered by the neutrino flux results in the possibility to observe multiple oscillation maxima at increasing energies at the far detector location. The oscillation peaks occur when

$$\frac{\Delta m^2 L}{4E} = (2n + 1) \frac{\pi}{2} \quad (2.4)$$

hence, the first peak ($n = 1$) is the one at higher energies, whereas the second ($n = 2$) is the one at lower energies. Among them, as shown in Figure 3, the second oscillation maximum is highly sensitive to CP-violation because the asymmetry caused by the MSW effect is much smaller than the one due to CPV for either mass ordering. This is visible also by accurately observing the theoretical DUNE-FD energy spectrum shown in Figure 4. Looking at Figure 4a it is evident that for the three peaks corresponding to the second ($n = 2$) oscillation maximum at different values of δ_{CP} the separation is larger than for the same peaks corresponding to the first ($n = 1$) oscillation maximum. On the other hand, by selecting the spectrum with $\delta_{CP} = 0$ (red) from both Figure 4a and Figure 4b, thus analysing the asymmetry from the MSW effect alone, it is clear that the change at the first oscillation maximum is larger than the one occurring at the second as is indeed also shown in Figure 3.

At baselines lower than ≈ 1000 km, the second maximum is typically too low in energy to be observable, hence experiments with this configuration can only probe matter-antimatter asymmetry at the first maximum, where CPV and MSW effects are degenerate. The coupling of a long baseline and a wide-spectrum beam allows DUNE to directly measure CPV at the second peak directly and to break the degeneracy present at the first peak. Thanks to this, DUNE will reach a sensitivity to CPV of over 5σ to 50 % δ_{CP} values after 10 years of nominal data taking as shown in Figure 5a and a δ_{CP} resolution varying from 25 % in the case of maximal CPV to around 10 % in the case of CP conservation as shown in Figure 5b.



(a) Oscillation probability at 1285 km for neutrinos assuming normal ordering.

(b) Oscillation probability at 1285 km for antineutrinos assuming normal ordering.

Figure 4: The expected oscillation probabilities spectrum for neutrinos and antineutrinos at different values of δ_{CP} ; effects from the MSW effect are included. Source: [43].

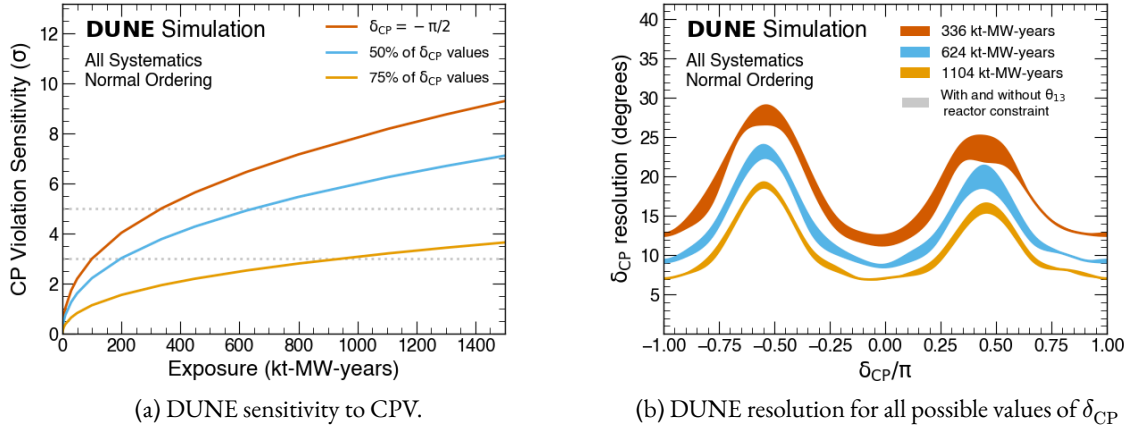


Figure 5: DUNE sensitivities and resolutions to CP-related measurements for different scenarios. An exposure of 624 kt MW yr corresponds to 10 years of data-taking at nominal values. Adapted from [30].

§2.1.2 Neutrino mass hierarchy

As briefly introduced in Section 1.4.3, the sign of the atmospheric splitting is still unknown, leaving open the question of whether m_3 is the smallest or biggest neutrino mass. Since the accurately chosen baseline and energy range of the DUNE experiment will break the degeneracy between the two matter-antimatter asymmetries, the experiment will be able to accurately measure the MSW-induced asymmetry that, as shown in Equation 1.51, depends on the sign of the squared-mass difference, allowing the determination of the neutrino mass hierarchy. Indeed, DUNE sensitivity studies outlined by DUNE Collaboration et al. [30] and cited in Figure 6b show how the experiment will be able to determine the neutrino MO assuming any possible value of δ_{CP} after just three years of data-taking.

§2.1.3 Secondary physics program

Core-collapse supernova A core-collapse supernova occurs when a massive star's iron core reaches the critical mass of 1.4 solar masses, whereby a stable configuration is no longer possible and the star reaches the end of its life [44, §2.1]. The gravitational pull towards the centre condenses all the star's mass to nuclear or super-nuclear densities of around $(10^{12} - 10^{14}) \text{ g cm}^{-3}$ creating a medium

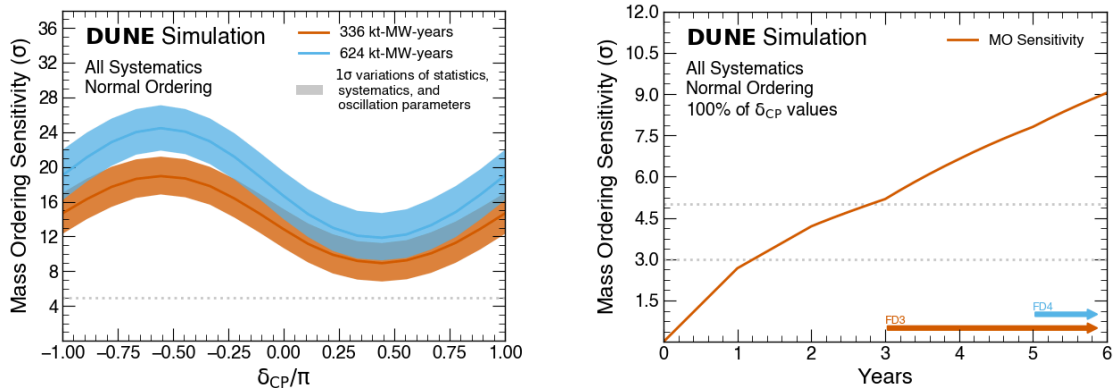


Figure 6: DUNE sensitivities to either mass ordering for all possible values of δ_{CP} . Adapted from [30].

that is opaque to all known kinds of radiation, isolating the core from the rest of the universe. The extremely high densities turn all internal mass into neutrons and neutrinos that start to push against the gravitational pull; in the time span of a few seconds neutrinos and antineutrinos of all flavours free themselves causing the collapsing outer shells to rebound and explode outwards thereby liberating around 10^{40} MJ of energy — over 99 % of all the core’s gravitational binding energy equal to around 10 % its rest mass — in the form of 10^{58} neutrinos with typical energies ≈ 10 MeV. At these low energies, the main observation channel consists of neutrino-driven inverse β -decay, also called CC absorption, of electron neutrinos by argon nuclei that expel an electron and decay into excited states of potassium

$$\nu_e + {}^{40}\text{Ar} \rightarrow e^- + {}^{40}\text{K}^* . \quad (2.5)$$

Neutrino cross sections in the low-energy nuclear bracket have never been directly measured, hence the uncertainty on this measurement is systematics-dominated; nevertheless an energy resolution of around (10 – 20) % should be achieved [44, §7]. It is worth noting that this interaction channel is not available for other experiments. Indeed most other neutrino telescopes (like Hyper-K and JUNO, outlined in Section 1.8) use water or liquid scintillators as their target material. These experiments can therefore probe supernovae neutrinos only through neutrino elastic scattering, a channel that offers better angular resolution thanks to its simple kinematics, but that is blind to neutrino flavour. DUNE will thus offer a new opportunity to measure the ν_e content of the burst, which is sensitive to interesting information about neutronisation. Further details on DUNE capabilities in observing supernova neutrinos are provided by DUNE Collaboration et al. [44].

Baryon number violation The apparent stability of the proton suggests that indeed baryon number is conserved by nature and this fact is implicit in the formulation of the SM Lagrangian. As is the case with all symmetries, the observation of any violating process—in this case proton decay or neutron-antineutron oscillations—would be evidence for Beyond the Standard Model (BSM) physics, whereas their continued non-observation calls for an answer to the question of which symmetry forbids these processes. A particularly compelling discovery channel in DUNE is the decay of a proton into a positive kaon and an antineutrino $p \rightarrow K^+ + \bar{\nu}$, for which, should the decay half-life be at the current limit of $5.9 \cdot 10^{33}$ yr, five candidate events are expected within ten years of running with all four FD modules. No observation of this signal will push the 90 % C.L. lower limit on proton lifetime to $1.3 \cdot 10^{34}$ yr [30, §2.3.2].

Other BSM searches Other BSM searches possible at DUNE include the measurement of neutrino trident production and sterile neutrino analyses. The former consists of a SM-allowed rare process in which a neutrino interacts with the electric field of a nucleus thanks to the production of a lepton-antilepton pair (generally $\mu^- \mu^+$). If the measured rate of this kind of event were to be higher than expected, it would hint towards the existence of the hypothesised Z' baryon. The latter probes the plausible existence of a fourth, completely sterile, neutrino through the effective mixing angle $\theta_{\mu e}$; further details on both fronts are provided in the DUNE TDR [30, §2.1.4].

§2.2 LBNF beam

The LBNF beam is where the journey of DUNE’s neutrinos begins. Protons will be accelerated by Fermilab’s 215 m long Linear Accelerator (LinAc), reaching energies of ≈ 800 MeV thanks to the almost completed Fermilab accelerator complex upgrade dubbed Proton Improvement Plan II

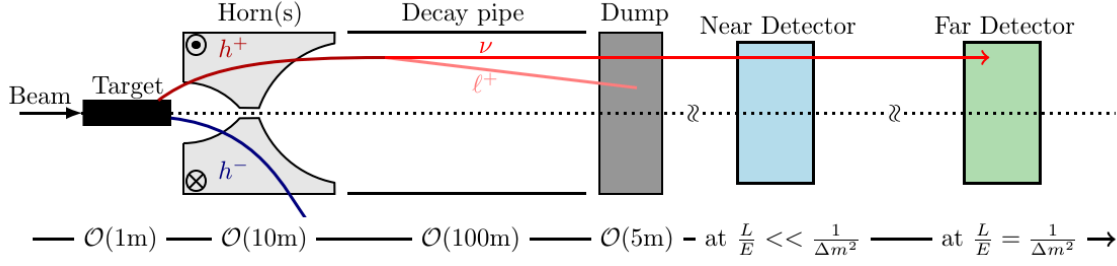


Figure 7: Standard neutrino beam setup. Source: [45]

(PIP-II). Protons are then delivered to the Booster circular accelerator, 150 m in diameter, which will further increase their energy tenfold to ≈ 8 GeV and finally sent to the much larger Main Injector ring, approximately 1 km in diameter, where the final energies of ≈ 120 GeV are reached. This accelerator system will provide DUNE with 1.2 MW of power during Phase I, doubling to 2.4 MW in Phase II [30, §1.1.2].

At 1.2 MW, the accelerator complex will deliver an expected 10^{21} Protons On Target (POT) per year to the target hall, where it will bombard a graphite target. The interaction of highly-energetic protons with the carbon nuclei creates light mesons (mainly pions and kaons), the decay products of which consist mostly of muons, and muon neutrinos. Through a series of three electromagnetic horns, the pion component of the meson beam is focused and one type of electric charge is selected: in Forward Horn Current (FHC) mode positive pions are focused and negative ones are discarded as shown in Figure 7, whereas in Reverse Horn Current (RHC) mode the opposite is true. The beam then passes through a ≈ 200 m long decay pipe, after which an absorber stops the un-decayed hadrons. This series of beam manipulations allows us to produce our neutrino or antineutrino beam depending on the horn current:

$$\text{FHC: } \pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \text{RHC: } \pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (2.6)$$

The main sources of beam contaminations come primarily [45] from the following components:

1. Production of heavier hadrons, namely $K^\pm$ and K_L^0 , whose decay products introduce both a wrong-sign neutrino and an electron flavour flux component.
2. Part of the muons themselves decaying in the decay pipe with consequent production of an electron neutrino component.

For this reason, neutrino beams will always contain some amount of all four possible neutrinos ($\nu_\mu, \bar{\nu}_\mu, \nu_e, \bar{\nu}_e$), however standard beams can generally reach a primary flux component that is one to two orders of magnitude higher than the secondary in the desired energy range as shown in Figure 8.

§2.3 Far Detector

The DUNE Far Detector (FD) is currently under construction at Sanford Underground Research Facility (SURF) in Leads, South Dakota (USA). The FD hall is located 1.5 km underground and mostly comprises two identical detector caverns — each 145 m long, 20 m wide and 28 m high — and an additional utility cavern between them. The detector caverns are built to accommodate four (two during Phase I) LArTPCs each containing at least 10 kt of LAr. The two modules to be installed as part of DUNE Phase I consist of two cryostats each 14.1 m wide, 14.0 m high and 62.0 m long

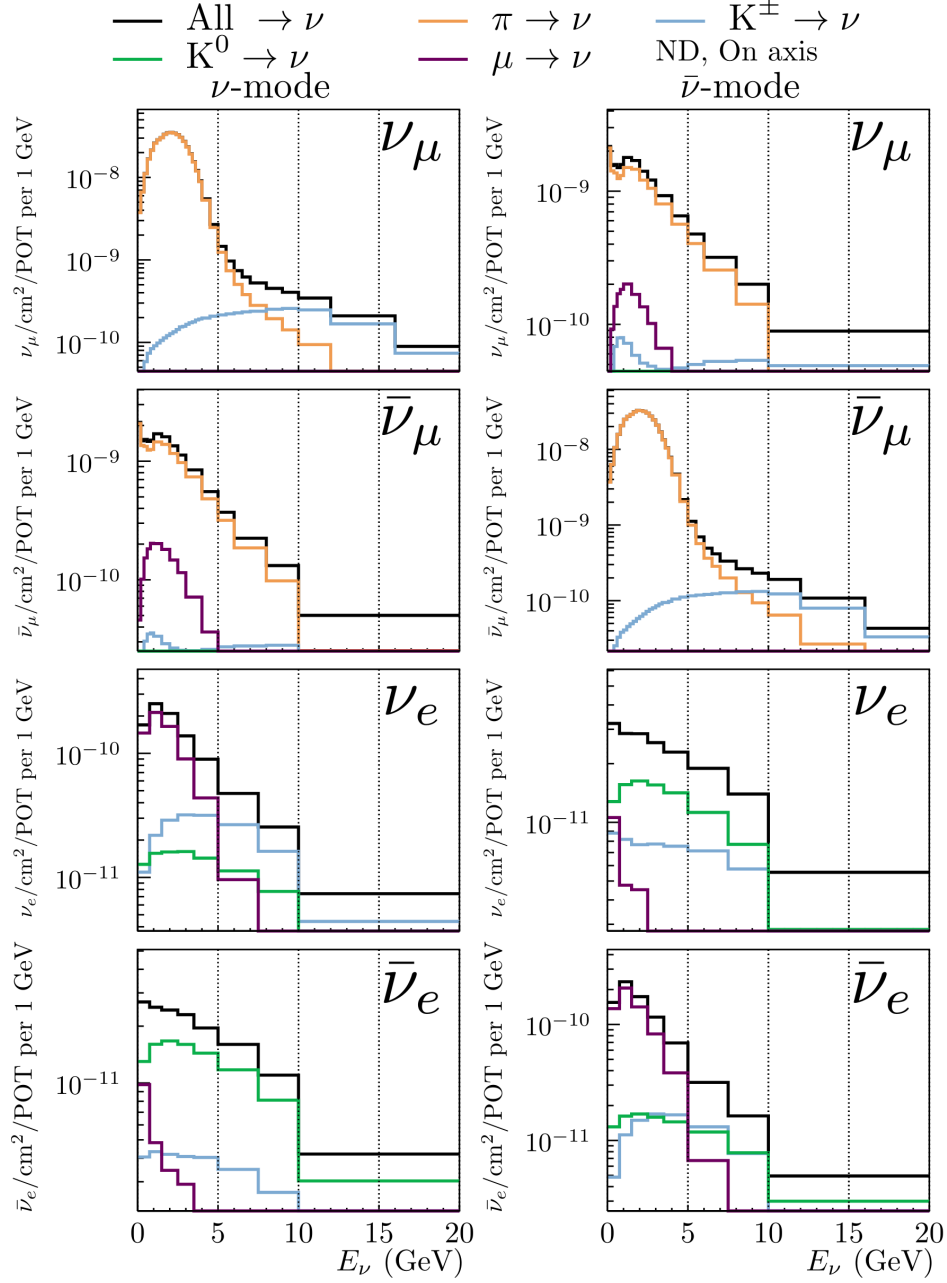


Figure 8: LBNF beam predicted neutrino fluxes at the ND for FHC (left) and RHC (right). Source: [46].

instrumented with two different detector technologies referred to as Horizontal Drift (HD) and Vertical Drift (VD).

§2.3.1 DUNE-HD

The functioning principle of the DUNE-HD module is that of a standard LArTPC, akin to ICARUS and SBND and is shown in Figure 10. A description was provided in Section 1.7.1. The DUNE-HD module will be split in four different drift volumes by three anode and two cathode planes in the sequence A-C-A-C-A. Due to the large dimensions of each DUNE FD module, both the anode and cathode planes consist of an array of smaller units, referred to as Anode Plane Assembly (APA) and Cathode Plane Assembly (CPA) respectively. The cathodes will be kept at -180 kV resulting in an internal electric field of 500 V/cm. This electric field is such that ionisation electrons produced by

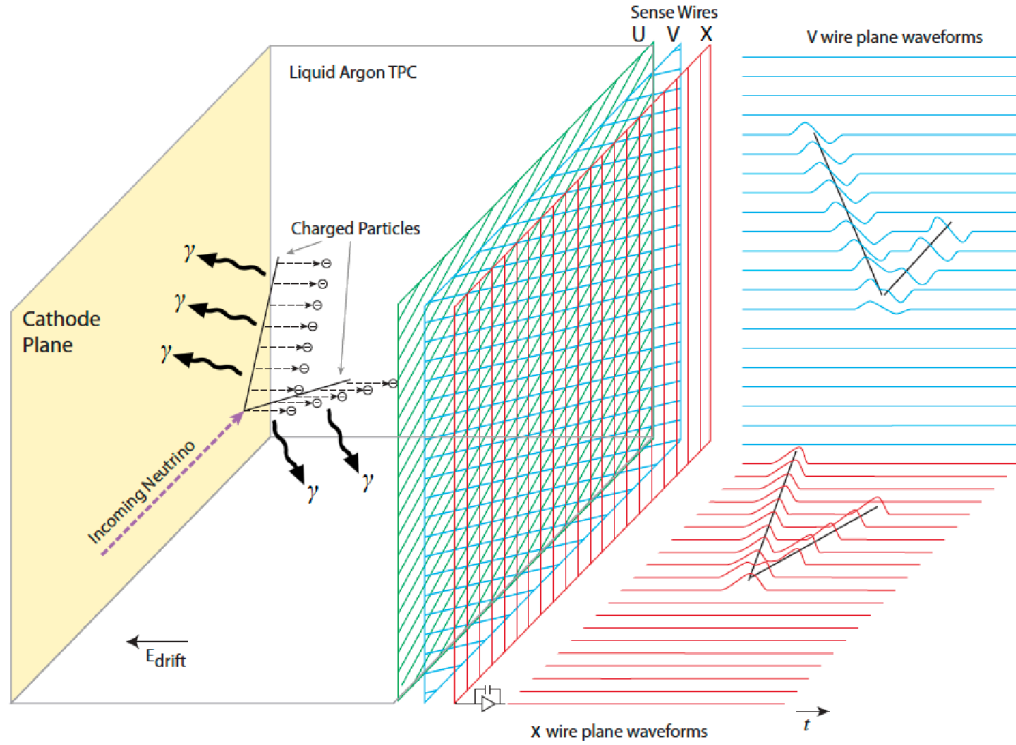


Figure 10: Functioning principle of a LArTPC. Source: [30]

charged particles passing through the detector, drift towards the anode planes for up to ≈ 3.5 m. To detect these ionisation electrons, the anode planes are each constituted by three wire planes (referred to as U, V and X) strung at angles chosen to reduce ambiguities in event reconstruction. The relative voltage between them is selected such that the U and V planes are transparent to the drifting electrons, allowing us to measure the charge induced on the wires by the passing drift charges. The X planes finally collect the electrons. Novel SiPM-based photon detectors called X-ARAPUCAs located on the anode planes will allow the detection of LAr scintillation light, providing both the event's t_0 vital for TPC reconstruction and calorimetric information [30, §3.1].

§2.3.2 DUNE-VD

The DUNE-VD module is the evolution of the Dual Phase (DP) module initially planned to be installed alongside DUNE-HD. In the DP design, ionisation electrons produced in LAr would drift vertically and then be transferred into a gaseous layer to finally deposit their charge onto a segmented anode [30, §4.1]. While providing several benefits, this design presented various challenges, some unforeseen, that became apparent during the operation of the ProtoDUNE-DP prototype and were outlined in the conclusions of an article from the DUNE Collaboration et al. [47]. Namely the Large Electron Multiplier (LEM), planned to be placed in the gas phase, showed extensive performance degradation, with its gain decreasing by almost 80 % by the end of data-taking. Additionally, turbulence of the liquid surface coupled with contamination from floating debris further undermined the long-term operation stability of ProtoDUNE-DP. Since the scaling of this Dual Phase readout technology to the dimensions needed for the FD modules would only augment these issues, a new solution now constitutes the baseline design for the second FD module. The new DUNE-VD module presents itself as a good solution, offering many of the same advantages as the DP module—for example the simpler and more cost-effective construction and installation, and the slightly larger instrumented volume—while avoiding challenges associated with the, now removed, liquid-gas phase interface. In DUNE-VD

the cryostat is completely filled with LAr and split in two drift volumes by a horizontal cathode that is kept at a voltage of ≈ 300 kV and instrumented with the same X-ARAPUCA photon detectors as DUNE-HD. The anode planes, located at the top and bottom of the cryostat are constructed by stacking two double-sided perforated anode Printed Circuit Boards (PCBs) with readout strips etched at different angles instead of the very fragile, and arduous to build, wire planes. This setup offers a detector with three readout channels (induction-1, induction-2, collection) plus one shield layer, resulting in a system whose operation almost identical to that of traditional wire-based LArTPC, but improved active volume and detector performance [48].

§2.4 Near Detector

The DUNE Near Detector (ND) serves as the experiment's control. By measuring the un-oscillated neutrino energy spectrum, the ND will be able to constrain systematic uncertainties by independently making measurements that allow to constrain the flux, cross section and detector response and partially or fully deconvolve their complicated relationships. Indeed, the FD will measure the neutrino interaction rate for different flavours as a function of the reconstructed energy and not neutrino oscillation probability directly. It is useful to formalise the measurements that are performed in the near and far detector modules in the following equations [30, Eq. A.1 and A.2]:

$$\frac{dN_x^{\text{FD}}}{dE_{\text{rec}}}(E_{\text{rec}}) = \int \Phi_{\nu_\mu}^{\text{FD}}(E_\nu) P_{\nu_\mu \rightarrow x}(E_\nu) \sigma_x^{\text{Ar}}(E_\nu) T_x^{\text{FD,Ar}}(E_\nu, E_{\text{rec}}) dE_\nu \quad (2.7)$$

$$\frac{dN_x^{\text{ND}}}{dE_{\text{rec}}}(E_{\text{rec}}) = \int \Phi_{\nu_\mu}^{\text{ND}}(E_\nu) \sigma_x^m(E_\nu) T_x^{d,m}(E_\nu, E_{\text{rec}}) dE_\nu \quad (2.8)$$

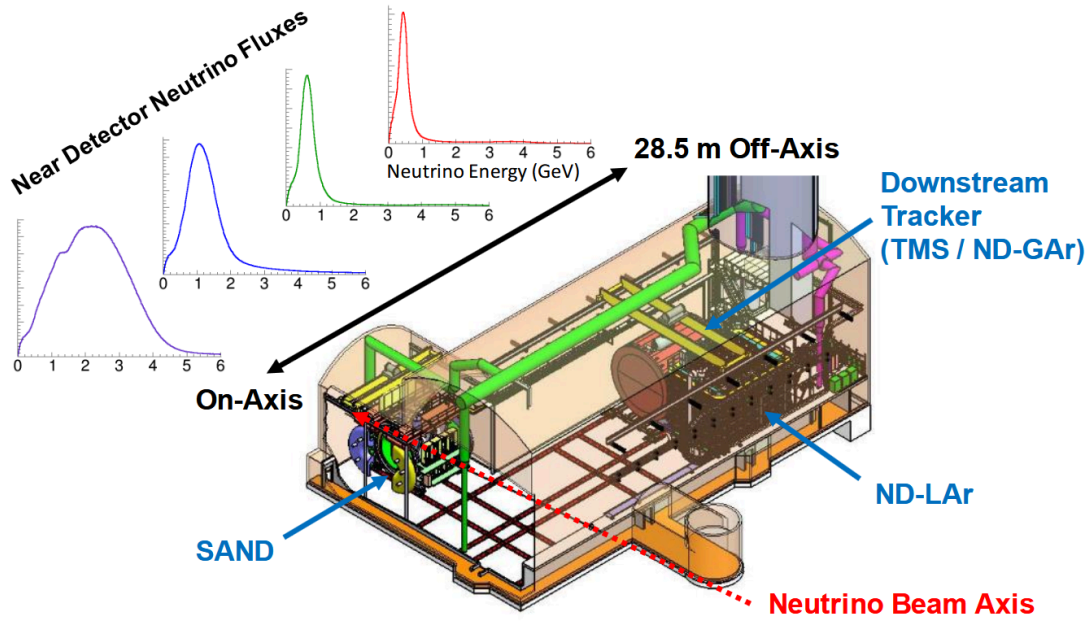
where

- the index x stands for neutrino flavour ν_e, ν_μ , d is the detector index and m is the interaction target (H, C, or Ar)
- E_ν is the true neutrino energy and E_{rec} the reconstructed neutrino energy
- $T_x^{d,m}(E_\nu, E_{\text{rec}})$ is the true-to-reconstruction transfer function
- $\sigma_x^m(E_\nu)$ is the neutrino interaction cross section
- $\Phi_x^d(E_\nu)$ is the un-oscillated neutrino flux
- $dN_x^d / dE_{\text{rec}}(E_{\text{rec}})$ is the measured differential event rate per target.

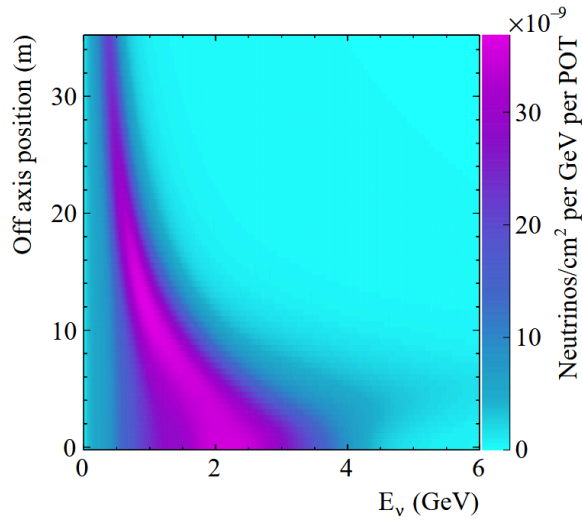
It is not possible to constrain the FD flux directly, but the near-to-flux ratio is believed to be tightly constrained by existing hadron production data. Their ratio

$$R(E_\nu) = \frac{\Phi_{\nu_\mu}^{\text{FD}}(E_\nu)}{\Phi_{\nu_\mu}^{\text{ND}}(E_\nu)} \quad (2.9)$$

can be extracted from the beam simulation. By comparing Equation 2.7 with Equation 2.8 it is possible to check that despite the usage of the same target material and technology at the ND and the FD, certain effects do not cancel trivially, mainly due to the non-diagonal nature of the true-to-reconstruction matrix $T_x^{d,m}(E_\nu, E_{\text{rec}})$ [30, §A.4]. Thus it is important that the ND suite constrains as many components as possible.



(a) Schematic of the DUNE ND hall. Source: [49].



(b) Projected neutrino spectrum as a function of off-axis position. Source: [50]

Figure 11: PRISM schematic and spectrum.

The ND will also have a rich and independent secondary physics program aimed at reinforcing available knowledge about the Standard Model (SM) and performing searches of Beyond the Standard Model (BSM) signals. The DUNE ND is split into two different parts: SAND and PRISM.

§2.4.1 DUNE-PRISM

Neutrinos cannot be observed directly, hence their properties must be deduced from the observable final state particles. However, the relationship between said observables and the incident neutrino energy is currently not known with sufficient precision to achieve DUNE physics goals [30, §A.5.1]. This results in a mismatch between reconstructed and true neutrino energy that, unlike many other systematics (see Equation 2.7 and Equation 2.8), does not cancel in the far/near ratio due to the different

detector geometries and neutrino spectrum. To more accurately measure neutrino interactions, the DUNE-PRISM detector will be built on top of a track, allowing it to move to off-beam positions. As explained in Section 2.2, the neutrino beam is generated by pion decay. The kinematics of this physical process is such that the neutrinos produced at different angles have different energy spectra as shown in Figure 11b. A linear combination of data taken at various off-axis positions by PRISM allows the production of a data-driven prediction of the oscillated neutrino spectrum at the FD with minimal interaction model dependence [49, §4]. For these reasons, PRISM will spend about half of the 29-week-long beam-year on-axis, moving to a range of off-axis positions for the remaining time [50, §4.1 and §4.4.1]. During Phase I, PRISM will be constituted by two detectors: a novel segmented LArTPC called Near Detector—Liquid Argon (ND-LAr) and the Temporary Muon Spectrometer (TMS).

ND-LAr Since oscillation experiments are interested in measuring the near/far ratio, the use of the same target nucleus and technology at the ND and FD is a simple way to reduce cross section and detector systematics. Thus, since the FD outlined in Section 2.3 will exploit the LArTPC technology, the ND needs to contain a major LArTPC component. Nevertheless the expected high event rate of $\mathcal{O}(50)$ ν interactions per spill in the planned ND-LAr volume makes the use of a standard LArTPC impossible due to the intrinsic slow detector response caused by the $\mathcal{O}(1 \text{ m/ms})$ velocity of drift electrons. To overcome this issue, ND-LAr will be segmented into 7×5 optically-separated LArTPCs based on the ArgonCube technology [30, §A.5.2] for a total target mass of $\approx 100 \text{ t}$. This technology has several upsides. Indeed, the reduced dimensions soften the cathode high voltage (from hundreds to just tens of kV) and LAr purity requirements whose importance was outlined in Section 1.7.1. Furthermore, in this design the standard wire planes are replaced with pixelated anode charge readouts with 4 mm pixel spacing and $2.5 \mu\text{s}$ time binning, allowing for a full 3D reconstruction of tracks and drastically simplifying the reconstruction. The main challenge of this detector setup consists in associating tracks and events across the different modules. However, this problem can be overcome by exploiting information from the LAr scintillation light [50, §2.1 and §2.3].

TMS The Temporary Muon Spectrometer (TMS) consists of 100 layers of steel alternating with layers of scintillator strips. The stack is subject to a magnetic field of $\approx 1 \text{ T}$ generated by copper wires coiled around the structure. The short longitudinal dimensions of ND-LAr result in an acceptance loss for muons with momentum above $\approx 0.7 \text{ GeV}$. Thus, the purpose of TMS is, as the name implies, to measure the momentum of muons, for which it is designed to reach a resolution of $\approx 5 \%$ while maintaining enough material to stop muons up to a momentum of $\approx 5 \text{ GeV}$. In the Phase II upgrade, TMS will be replaced by Near Detector — Gaseous Argon (ND-GAr), a high pressure gaseous argon TPC, with the dual objective of measuring muons uncontained in ND-LAr and of measuring neutrino interactions at energies below the reconstruction threshold of ND-LAr.

§2.5 SAND

It is essential for both the FD and PRISM’s physics objectives that the neutrino beam remains stable or, failing that, that distortions can be quickly identified and modelled. For this main reason, DUNE needs a continuous on-axis beam monitoring system, a role fulfilled by System for on-Axis Neutrino Detection (SAND). SAND’s design is based on the re-use of parts from the K-Long Experiment (KLOE), namely its superconducting magnet and Electromagnetic CALorimeter (ECAL) (Section 2.5.2). The performance of the KLOE components [51, §11], further enhanced by the addition of a low density tracker (Section 2.5.1) and of a LAr active target (GRAIN, Section 2.6), will

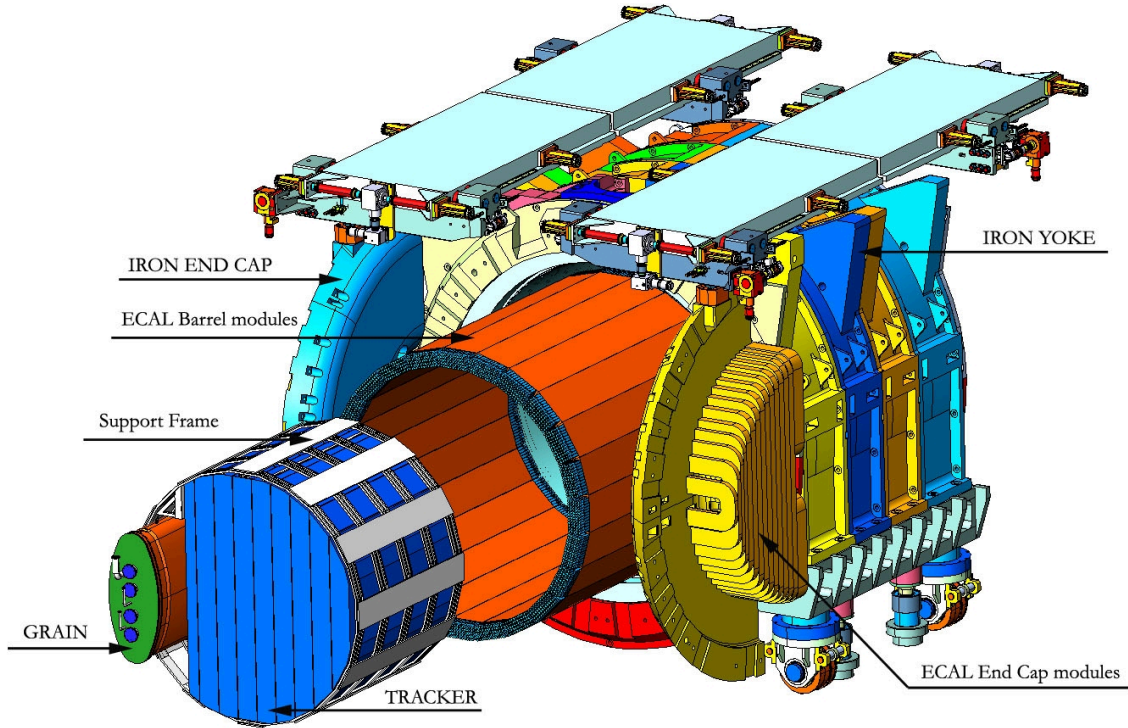


Figure 12: Exploded view of SAND. Source: [22]

allow SAND to fulfil its primary beam-monitoring role while also providing unique measurements to constrain systematic uncertainties relevant for the long-baseline analysis as well as supporting a broad physics program of precision measurements and searches for new physics.

Beam monitoring involves two main tasks:

1. Detecting a 1 % change in the neutrino event rate in near-real time.
2. Monitoring the neutrino interaction rate with a sensitivity better than 1 % over the course of a week.

To achieve this goal SAND must measure the energy of incoming neutrinos as well as provide a reconstruction of the full event hierarchy, with the primary interaction vertex and particle tracks. According to the DUNE ND Conceptual Design Report [50, Table 8.1 and §1.4.3.10], the muon momentum resolution must be $\sigma(p_\mu)/p_\mu < 10\%$ at $p_\mu \approx 5\text{ GeV}$ improving to 5 % at $p_\mu \approx 1\text{ GeV}$ equivalent to a neutrino energy resolution of $\sigma(E_\nu)/E_\nu < 15\%$ (requirement ND-C5.2) and the vertex resolution must be below 5 cm. Indeed, as shown in Figure 11b, the neutrino spectrum depends on the off-axis position: the requirement on vertex accuracy enables the distinction of interactions occurring at positions where the neutrino spectrum may vary (requirement ND-C5.3).

§2.5.1 Tracker

In order to measure muon momentum, SAND is instrumented with a light tracker ($\rho \leq 0.17\text{ g/cm}^3$) with a diffusive target composed of various materials. The target material planes can be easily swapped, allowing SAND to pursue a deep secondary physics program aimed at accurately measuring neutrino cross sections on atomic nuclei. Indeed, as previously mentioned, all DUNE analyses heavily rely on phenomenological neutrino interaction models that currently lack the desired accuracy for DUNE's

physics program. Nevertheless, favourable accuracies are reached for the simplest atom, Hydrogen, since no atomic models are needed. For this reason SAND’s target mass will include both a plastic material (composed of CH_2 chains) and graphite (pure C): statistical subtraction of measurements on graphite from those on plastic provides a strategy to obtain a model-independent measurement of a “solid hydrogen” target [52]. The solid hydrogen target is a pivotal point of SAND’s beam monitoring campaign. Indeed, since $\nu - \text{H}$ interactions do not require any further modelling beyond standard CC interactions between a neutrino and a proton, all systematics are under control and a flux measurement can be performed. For this tracker, two different technologies are under study: straw tubes and a drift chamber. The main difference between the two is the channel count: over 100k for the STT and around 50k for the drift chamber. The increased channel count in the STT produces more granular information on particle tracks augmenting the reconstruction potential. Nevertheless, working with fewer readout channels can remarkably reduce detector complexity and cost, which is why the drift chamber layout is a valid backup option.

Straw Tube Tracker (STT) A straw is a thin tube, 5 mm in diameter and up to 4 m long, filled with a Ar/CO_2 70/30 gas mixture kept at a pressure of 2 bar and a $20\text{ }\mu\text{m}$ wire placed in its centre. One STT module is a structure consisting of a layer of target material followed by four straw planes creating a grid that is orthogonal to the beam. This geometry could result in the introduction of so-called *ghost-tracks* when hits from multiple tracks are present. The phenomenon is extremely simple: imagine that two particles cross the detector in the same time bin. Since the wires provide information solely about the X or Y coordinate, in general, the detector output will be in the shape of a grid as shown in Figure 13. Without extra information it is impossible to determine whether the true tracks are the circles or crosses. However, simulations have shown that track matching with signals from the 4π ECAL is enough to allow the rejection of ghost-tracks generated by this geometry removing the need of a third stereo orientation [22, Table 6.4]. The minimal STT configuration is composed of 54 modules: 39 with a plastic target, 8 with a graphite target and 7 without target.

Drift chamber In case of technical, operational or budget limitations, the tracker is planned to utilise the drift chamber technology instead of straw tubes. This detector layout keeps the same modular detector structure and is split in 8 super-modules each containing 9 plastic modules and one carbon module. Each module consists of a target sheet followed by four aluminised Mylar cathode foils that split the drift volume into three sections. Each 12 mm wide sub-volume contains a wire plane made out of alternating field and sense wires with a periodicity of 7 mm. To provide three-dimensional

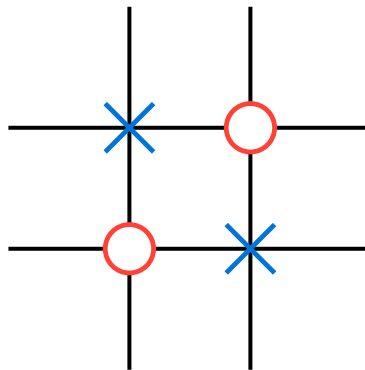


Figure 13: Ghost tracks phenomenon. Without extra information either the crosses or the circles could be reconstructed as 2-dimensional hits.

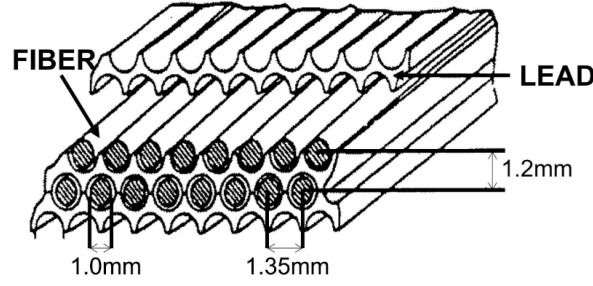


Figure 14: KLOE ECAL interleaved lead and fibers structure. Adapted from [53, Fig. 4.1]

information the wire layers are disposed at three different angles with respect to the magnetic field axis: -5° , 0° , and 5° .

§2.5.2 Calorimeter

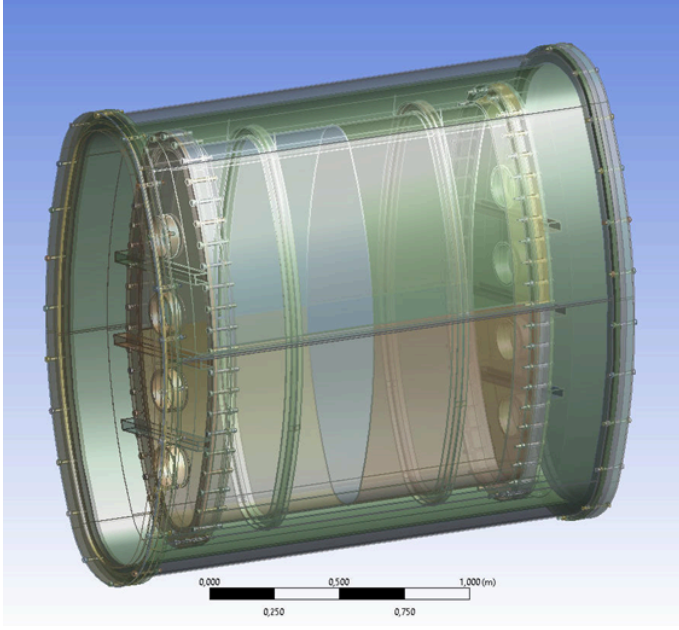
The KLOE calorimeter consists of an alternating stack of 1 mm scintillating fibre layers glued between thin 0.5 mm thick appropriately shaped lead foils. The grooves on the two sides of the foil are displaced half a pitch so that the fibres are located at the corners of quasi-equilateral triangles resulting in a uniform arrangement of fibres in the stack as shown in Figure 14. Each ECAL module is read on both sides by fine mesh PMTs for a total granularity of $\approx 4.4 \times 4.4 \text{ cm}^2$. The most stringent requirement for the ECAL in SAND is time resolution since its primary objectives are the rejection of background and accurate event timing to allow for Time of Flight (ToF) Particle IDentification (PID). ECAL timing and energy accuracies have been evaluated using KLOE data and are equal to

$$\sigma(t) = \frac{54 \text{ ps}}{\sqrt{E[\text{GeV}]}} \oplus 140 \text{ ps}, \quad \sigma(E) = \frac{5.6 \%}{\sqrt{E[\text{GeV}]}} \quad (2.10)$$

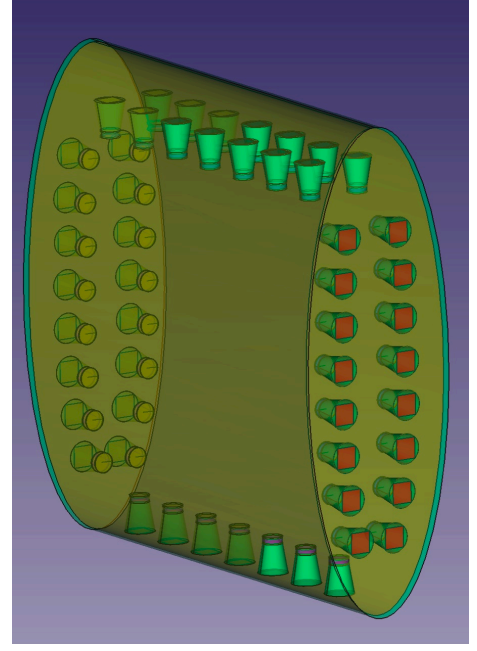
where the symbol $\oplus$ indicates a constant off-set [51, §12.3.2 and §11.7]. These accuracies are well within the physics requirements hence this re-purposing allows for large budget savings without any physics compromises.

§2.6 GRAIN

An active 1 t LAr target is designed to be installed inside the ECAL, upstream of the tracking system, with the goal of performing nuclear-effect studies on Ar and for cross-calibration with other DUNE-ND components. For what concerns the long-baseline oscillation analysis, GRanular Argon for Interactions of Neutrinos (GRAIN) will provide samples of the different flavour components of the neutrino beam, which can be used for neutrino flux evaluation. Additionally, it will help constrain background processes and systematic uncertainties in $\nu - \text{Ar}$ cross section measurements providing data to tune the neutrino interaction model through the complementary information provided by GRAIN and the tracker, which exploit different nuclear targets. Instead of detecting ionisation electrons, GRAIN will employ solely LAr scintillation light [Section 1.7.2] to reconstruct charged particle tracks produced in neutrino interactions. The synergy between the independent measurement from the large mass of ND-LAr which uses a technology closer to that of the FD and the smaller mass of GRAIN, located in a magnetic field and subject to different systematic effects, ensures an increased robustness of the ND complex analyses.



(a) GRAIN cryostat. Source: [22, Fig. 5.1]



(b) GRAIN camera placement. Source: [22, Fig. 5.14]

Figure 15: GRAIN cryostat and camera placement.

§2.6.1 Cryostat and cryogenics

The GRAIN cryostat consists of an internal vessel housed within a vacuum vessel, both of which are in the shape of an elliptical tube as shown in Figure 15a. The internal vessel's elliptical base has axes measuring 49 cm and 148 cm and a longitudinal length of 150 cm. The two bases are closed by two 30 mm thick endcaps with feedthroughs for signals and detector power. Thermal insulation relies on the vacuum present between the external and internal vessels which is maintained at about 10^{-5} mbar and ensured by a dedicated pumping system. Instead of having a forced circulation of liquid argon, a quasi-passive recirculation will be implemented [22, §5.3]. Intrinsic heat production in the detector vaporises part of the LAr; the produced gaseous argon is collected through a pipe welded at the top of the internal cryostat and liquefied in a liquid nitrogen-based condenser. The now-liquid argon is finally sent back to the cryostat by gravity through a second pipe. If the intrinsic boil-off were not to suffice in sustaining circulation, a heating element will be inserted into the circuit. The purification achieved is not at the level obtainable with forced circulation but it should be sufficient to guarantee the purity required for optical transparency, namely a concentration of oxygen molecules lower than 0.1 ppm and one of nitrogen molecules lower than 1 ppm [22, §5.3].

§2.6.2 Optical detectors

In order to image tracks and vertices from neutrino interactions, an optical system is required to collect and possibly focus argon scintillation light produced in said interactions. Whereas integral light collection would be sufficient for purely calorimetric purposes, as anticipated in the introduction to this section, GRAIN aims to also produce a geometrical reconstruction of the event topology, which makes the use of imaging devices necessary. As better described in Section 1.7.2, argon scintillation light is the result of the de-excitation of the Ar_2^* excimers produced by particle-nucleus scattering. Its spectrum has the shape of a fairly thin (6 nm FWHM) peak centred at 128 nm and a specific time signature constituted by two emissions with one characteristic short (6 ns) and one long (1.6 μs) time constants. Currently two technologies are under evaluation to detect this light: UV lenses and coded apertures.

Nevertheless, both technologies will utilise SiPM matrices to collect the impinging photons. The final SiPM matrix configuration is not decided yet, but pixel size will range from 1×1 to 4×4 mm² and single photon timing resolution is planned to reach less than 100 ps [22, §5.4]. To achieve a three-dimensional reconstruction, both designs plan to fill the side, top and bottom surfaces of GRAIN with tens of light detection apparatuses. Each technology has upsides and downsides, reason why a hybrid design is also conceivable.

Coded-aperture masks Coded-aperture imaging is an extension of the pinhole camera concept and it can be a viable solution when the light wavelength poses a challenge. The pinhole camera is capable of excellent angular resolution but is also inherently inefficient at light collection since about half the light gets absorbed by the opaque element itself. The coded-aperture mask stems from the idea that superimposing the result from various closely-located pinholes can improve collection efficiency without penalising angular resolution. This technique offers several practical advantages: the hardware is simple, passive and lightweight enabling a wide field of view and a camera design that is thin and compact. It also works across a broad range of wavelengths since imaging relies on geometric coding rather than wavelength-dependent refraction or reflection. However, despite the existence of several algorithms able to reconstruct the original source position, such as deconvolution or iterative maximum likelihood estimation, they all tend to be extremely resource-intensive, often requiring heavy use of the presently ever-so-scarce GPUs. This great computational toll coupled with the still sub-optimal light collection efficiency is the main reason why an alternative lens-based design is being studied.

Lens-based cameras The use of a lens-based imaging technology inside a cryogenic liquid would be innovative since these peculiar circumstances require the employment of specific materials compatible with the cryogenic temperatures, with a refractive index suitable for use in LAr and finally that exhibit a high enough transmittance at the wavelength of interest. The LAr scintillation wavelength sits at 128 nm hence wholly within the Void Ultra Violet (VUV) range of the electromagnetic spectrum. Just two crystals (MgF₂ and CaF₂) are known to have good transmittance at this wavelength ($\approx 60\%$) but since their use in a cryogenic environment has not yet been documented, other possibilities have been explored. The most simple and effective solution to this issue, which is hence part of the baseline design for GRAIN, is the doping of LAr with $\mathcal{O}(10$ ppm) of xenon. Indeed it has been demonstrated, that dissolving Xe in LAr can efficiently convert scintillation light from 127 nm to 174 nm through

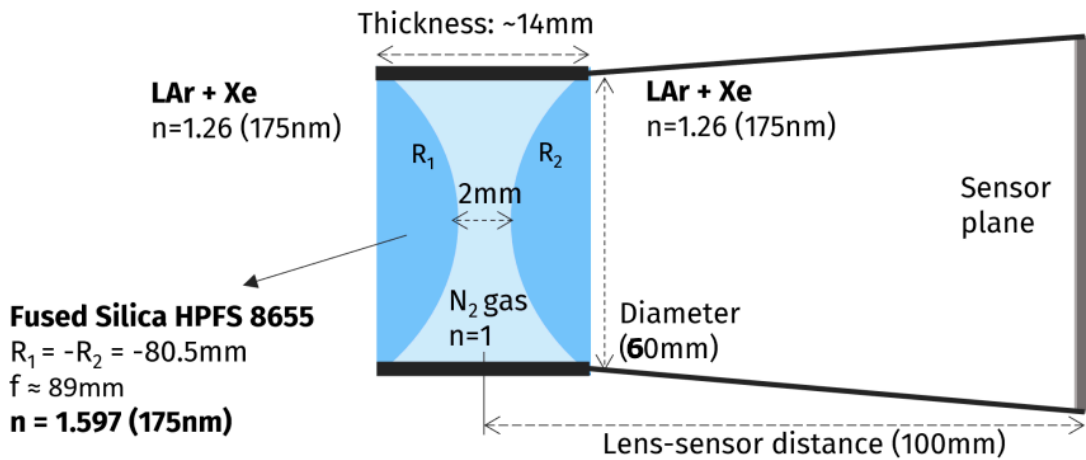


Figure 16: Schematic of a GRAIN lens, image not to scale.

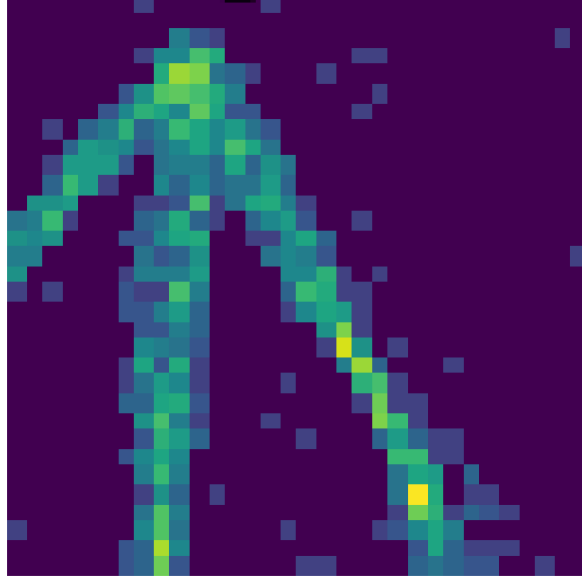


Figure 17: Simulation of a GRAIN camera.

an extremely quick energy transfer from argon to xenon excimers, whose de-excitation will produce scintillation light at this higher wavelength [27]. This energy transfer's characteristic time of ≈ 1 ns allows virtually no loss of spatial resolution whilst simultaneously moving the signal to a region where the SiPM photon detection efficiency is increased by around 10 % and where the use of much more common materials, like fused silica, becomes a valid solution to construct the lenses. To keep the lens thickness reasonable, GRAIN will exploit the “gas-lens” design: two plane-convex silica glass lenses with a refraction index very close to that of LAr form an air-tight container that is filled with a volume of N_2 gas which acts as the actual lensing material. This trick makes the lens design solid, reliable and easier to manufacture. Nevertheless, lens and camera designs are still under study and optimisation; further details on hardware developments are provided by F. L. Polleri [54].



Machine Learning

Necessary background on *Machine Learning (ML)*, a branch of applied mathematics, is introduced in this section to provide context for the original work presented in this thesis. While a deep understanding of ML is not required to follow the discussion of event reconstruction in the GRAIN detector, a more in-depth overview of its main concepts can be useful for understanding the methods employed. Section 3.1 will briefly introduce the history of machine learning, its origins and recent revival and explosion. Section 3.2 will formally introduce linear learning with the scope of providing the reader with the mathematical proofs behind these concepts. Section 3.5 will go over the two main ways of solving non-linear problems with machine learning, namely kernel methods and neural networks. Finally Section 3.6 will formally introduce Convolutional Neural Networks (CNNs), the specific technology used for this thesis’ original work.

§3.1 Brief history of machine learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) which is itself a branch of applied mathematics that was born over 80 years ago with the purpose of investigating the workings of the brain. Indeed the first article ever written about AI was published in 1943 by W. S. McCulloch and W. Pitts [55]: it consists of an attempt to build a mathematical description of the neuron, the building block of our nervous system. This definition of the neuron was the main inspiration behind the invention of the *perceptron*, the building block of the Neural Network (NN) better described in Section 3.5.2. Nevertheless it wasn’t until 1956 with the Dartmouth workshop—organised by Claude Shannon, one of the first to introduce boolean algebra as a means to convey information—that AI was formally defined as a branch of mathematics and more specifically of information theory. The first historical use of the term Machine Learning comes from an article written three years later by A. L. Samuel [56] whose objective was to train an algorithm to play checkers better than its programmer. This first AI boom ended shortly after: a book published in 1969 by M. Minsky and S. A. Papert [57] showed some limits of single layer perceptrons—at the time the best performing AI algorithm—like its inability to learn the simple XOR function. This event marks the beginning of the first “AI winter”, a period characterised by a remarkable slow-down in AI research that lasted throughout the 70s. The winter was briefly lifted in the 80s when D. E. Rumelhart, G. E. Hinton, and R. J. Williams [58] introduced *back propagation*, an algorithm to calculate gradients of highly-composed functions that is currently the foundation of most modern AI algorithms. This second AI explosion lasted only about a decade, after which *expert systems*, machine simulations of a human expert trained on a set of a few ground rules with the objective of inferring new rules, started failing to meet expectations, marking the beginning of a second AI winter. Finally, the most recent AI explosion of the 2010s and 2020s is outlined in the article “*Deep Learning*” by Y. LeCun, Y. Bengio, and G. Hinton [59]. This resurgence was caused mainly by advancements in the underlying hardware technology that allowed researchers to process increasingly large datasets. Nevertheless, breakthroughs in theory didn’t take long to arrive: the introduction of the concept of *attention* by A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin [60], and specifically its particular

implementation called a *transformer* sparked the popularity of Large Language Models (LLMs), the kind of AI the general public is most familiar with.

§3.2 Introduction to supervised statistical learning

The objective of the present section is to formally introduce ML. It is worth noting that to understand the original work presented in this thesis, it is not necessary to understand all the concepts and proofs that the following sections will provide. However, a formal understanding of what Machine Learning actually is can be of great benefit to any modern-day scientist. The contents are mostly a summary of a course attended during the first year of my MSc mainly based on the book *The Elements of Statistical Learning* by T. Hastie, R. Tibshirani, and J. Friedman [61].

ML algorithms can be split into two categories: *supervised* and *unsupervised* learning. In the former, which is the one this work will focus on, the algorithm is shown a number of inputs for which the desired output is already known and is tasked with inferring the output for some other input. In the latter, instead, there is no previous data knowledge. An example of an unsupervised ML algorithm is DBSCAN, discussed in Section 4.2.3.

Definition 3.2.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets we shall refer to as the *input* and *output* sets respectively.

It is worth noting that in theory these two sets could contain anything, however for computing reasons they are often sets of collections of real numbers, namely $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}^t$. The specific case in which $\mathcal{Y}$ denotes an interval is called *regression*, while the case when $\mathcal{Y}$ is a finite collection of numbers is called *classification*. The basic idea behind Machine Learning is to find some function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that somehow relates these two sets. To formally define this problem one must dabble some more in the realm of mathematical definitions.

Definition 3.2.2. Let Ω be a domain and $(X, Y) : \Omega \rightarrow \mathcal{X} \times \mathcal{Y}$ be a random variable with probability measure P . We shall denote with capital letters (X, Y) the random variables and with small letters (x, y) a realisation thereof.

Definition 3.2.3. Let $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow [0, +\infty)$ be a function we call the **loss function**.

Definition 3.2.4. $\forall f : \mathcal{X} \rightarrow \mathcal{Y}$ let the *expected error*, also referred to as the *expected risk*, $L(f)$ be defined as

$$L(f) := \mathbb{E}[\ell(f(X), Y)] = \int_{\mathcal{X} \times \mathcal{Y}} \ell(f(x), y) \, dP(x, y). \quad (3.1)$$

Definition 3.2.5. The *problem of learning* is to find

$$\min_{f : \mathcal{X} \rightarrow \mathcal{Y}} L(f). \quad (3.2)$$

Definition 3.2.6. The *target function* f_* is the solution to the learning problem

$$f_* := \arg \min_f L(f) \iff L(f_*) = \min_f L(f). \quad (3.3)$$

Of course the target function is an ideal and unobtainable object. Indeed, in order to find this function one would require knowing the probability measure P , an assumption which renders the learning problem redundant. This moment is when the statistical framework in which this section is embedded comes to aid.

Let $\mathcal{Z}_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ be a *training set* that contains the realisations of $\mathcal{S}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, a set of n identical and independent copies of the random variable (X, Y) . It is possible to then define what a learning algorithm is.

Definition 3.2.7. A **learning algorithm** $\mathcal{A}$ is a map that associates any training set $\mathcal{Z}_n$ with a function $\hat{f}_n : \mathcal{X} \rightarrow \mathcal{Y}$

$$\mathcal{A} : \mathcal{Z}_n \mapsto \hat{f}_n. \quad (3.4)$$

To test the validity of a learning algorithm, one must utilise convergence criteria that take into consideration the statistical nature of the problem. Indeed the function $\hat{f}_n$ depends on the training set, that is a set of realisations of random variables and hence is itself the realisation of a random variable.

Definition 3.2.8. A learning algorithm $\mathcal{A}$ is **consistent** if and only if

$$\mathbb{E}[L(\hat{f}_n) - \min L] \xrightarrow{n \rightarrow +\infty} 0. \quad (3.5)$$

If we expand the definition of this limit and hence write

$$\forall \varepsilon > 0 \quad \exists n(\varepsilon, P) \mid \text{if } n > n(\varepsilon, P) \Rightarrow \mathbb{E}[L(\hat{f}_n) - \min L] < \varepsilon \quad (3.6)$$

then it is evident how the amount of entries in the training set required for the algorithm to converge depends on the probability measure P which is unknown. It may be tempting to solve this problem by requiring that the algorithm be *uniformly* consistent and hence that

$$\sup_P \mathbb{E}[L(\hat{f}_n) - \min L] \xrightarrow{n \rightarrow +\infty} 0. \quad (3.7)$$

Unfortunately, a theorem forbids this.

Theorem 3.2.9 (No Free Lunch). *Given a finite set $\mathcal{X}$ and a finite set $\mathcal{Y}$ of real numbers, assume that $f : \mathcal{X} \rightarrow \mathcal{Y}$ is chosen at random according to a uniform distribution on the set $\mathcal{Y}^{\mathcal{X}}$ of all possible functions from $\mathcal{X}$ to $\mathcal{Y}$. For the problem of optimising f over the set $\mathcal{X}$, then no algorithm performs better than blind search.*

In the present framework this is equivalent to stating that the supremum of Equation 3.7 does not exist. Indeed it is only possible to make some assumptions on the probability distribution P and find the best algorithm out of the selected class of distributions, e.g. the best algorithm that assumes data to be linearly distributed. In practice, this also implies that **no superior algorithm exists** and that an algorithm that performs well on a specific problem will inevitably either fail at some other task or require unthinkable amounts of data to succeed.

§3.3 Linear regression: least squares

This section attempts to solve a *regression* problem; hence, let $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}$. As claimed in Definition 3.2.6 the objective of ML is the determination of

$$f_* := \arg \min_f L(f) = \arg \min_f \int_{\mathcal{X} \times \mathcal{Y}} \ell(f(x), y) dP(x, y) \quad (3.8)$$

or, at least, of another function $\hat{f}$ output of an algorithm $\mathcal{A}$ that converges to f_* when the amount of data increases. Of course, this problem is too general to solve. The objective of this section is to find the solution when certain simplifying assumptions are made. The assumptions are as follows:

1. Use the quadratic norm as loss:

$$\ell(f(x), y) = \|f(x) - y\|^2 \quad (3.9)$$

2. Use an approximation of the expected error $L(f)$, called the *empirical error*, defined as its value on the realisations in a training set:

$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^n \|f(x_i) - y_i\|^2 \quad (3.10)$$

3. Limit the problem to the search of *linear* solutions:

$$\hat{L}(w) = \frac{1}{n} \sum_{i=1}^n \|\langle w, x_i \rangle - y_i\|^2 \quad (3.11)$$

where $w \in \mathbb{R}^d$ and $\langle \cdot, \cdot \rangle$ is the standard inner product of $\mathbb{R}^d$ $\langle w, x \rangle = \sum_{j=1}^d w_j x_j$.

Some may recognise this specific problem as it is the very common **least squares**. For simpler manipulations the problem can be rewritten by defining the $n \times d$ input matrix $\hat{X}$ which has all the inputs $x_i \in \mathbb{R}^d$ as its rows and the vertical output vector $\hat{y}$ which contains all the outputs $y_i \in \mathbb{R}$ as its rows.

$$\hat{X} := \begin{pmatrix} x_{1,1} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,d} \end{pmatrix}, \quad \hat{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad (3.12)$$

The problem, written in matrix form, then becomes to find the $w \in \mathbb{R}^d$ that minimises the empirical error

$$\hat{L}(w) = \frac{1}{n} \|\hat{X}w - \hat{y}\|^2. \quad (3.13)$$

If one now assumes that a solution, referred to as $\hat{w}$, does exist, then, since the empirical error $\hat{L}(w)$ is *convex*, it is possible to find this solution simply by calculating the gradient, setting it to zero and solving the thus-obtained equation

$$\nabla \hat{L}(w) = \frac{2}{n} \hat{X}^t (\hat{X}w - \hat{y}) = 0. \quad (3.14)$$

The solution to the linear least squares problem is widely available in literature and generally involves defining the *pseudo-inverse* matrix $\hat{X}^+$, also referred to as the *Moore-Penrose inverse*. This matrix is

defined as the one such that $\hat{X}\hat{X}^+\hat{X} = \hat{X}$, but it can be shown that *if* the columns of $\hat{X}$ are linearly independent then $\hat{X}^t\hat{X}$ is invertible, the pseudo-inverse can be written as

$$\hat{X}^+ = (\hat{X}^t\hat{X})^{-1}\hat{X} \quad (3.15)$$

and it acts as a *left inverse*, hence $\hat{X}^+\hat{X} = \mathbb{1}_d$. Conversely, if the rows of $\hat{X}$ are linearly independent, then $\hat{X}\hat{X}^t$ is invertible, the pseudo-inverse can be written as

$$\hat{X}^+ = \hat{X}^t(\hat{X}\hat{X}^t)^{-1} \quad (3.16)$$

and it acts as a *right inverse* $\hat{X}\hat{X}^+ = \mathbb{1}_n$. Nevertheless it can be shown by exploiting *singular value decomposition* that both problems' solution are expressed by the same equation:

$$\hat{w} = \hat{X}^+\hat{y}. \quad (3.17)$$

§3.3.1 Regularisation, the first hyper-parameter

This brief summary of the least squares algorithm allows the formal introduction of a cardinal concept in Machine Learning (ML): hyper-parameters. Suppose a special case in which the matrix $\hat{X}^t\hat{X}$ is diagonal and invertible and call the diagonal elements σ_i . Under these assumptions the solution then becomes

$$\hat{w} = \hat{X}^t(\hat{X}^t\hat{X})^{-1}\hat{y} = \sum_{i=1}^n \frac{y_i}{\sigma_i} x_i. \quad (3.18)$$

In the reasonably common case that $\sigma_i \ll y_i$, then the solution that the least squares algorithm provides is extremely *unstable* since a small change in the inputs x_i can imply a large variation in the algorithm's output. A common and simple solution to this issue is called *Tikhonov regularisation*. The procedure consists in *slightly* modifying the pseudo-inverse by introducing a *hyper-parameter* $\lambda \in [0, 1]$

$$\hat{X}^+ = \hat{X}^t(\hat{X}^t\hat{X})^{-1} \implies \hat{X}_\lambda^+ = \hat{X}^t(\hat{X}^t\hat{X} + \lambda\mathbb{1})^{-1}. \quad (3.19)$$

It can be proved that this new algorithm, properly dubbed the *regularised least squares* basically redefines the diagonal values

$$\frac{1}{\sigma_i} \mapsto \frac{\sigma_i}{\sigma_i^2 + \lambda} \quad (3.20)$$

with the effect of stabilising the algorithm in the limit $\sigma_i \rightarrow 0$ and leaving it unchanged when $\sigma_i \rightarrow \infty$.

The subtle point is that there is not a unique best value that λ should take; this is the reason why *hyper-parameters* can be such a big headache: a choice of λ can optimise a metric, but because of Theorem 3.2.9 this will inevitably worsen some other metric. An optimal choice of hyper-parameters can greatly improve the performance of a ML algorithm just as much as a poor choice can drastically decrease its ability to learn. Moreover, more advanced ML algorithms can have upwards of tens, hundreds or even thousands of hyper-parameters that must all be optimised concurrently since the optimal value of a different hyper-parameter γ at a certain fixed λ value λ_1 can be vastly different from the optimal value of the very same γ hyper-parameter when λ is set to some other value λ_2 . Theoretical optimal values can be calculated for very simple—and hence often not very useful—use cases, but for standard

implementations the default procedure to optimise hyper-parameters is the use of a technique called *cross-validation*. This process begins with creating a set of different discrete values for each hyper-parameter, usually linearly or logarithmically distributed in a specified range, checking all possible combinations, and finally selecting the best-performing one.

The very same regularisation can be implemented by slightly modifying the empirical risk, thus defining the *regularised empirical risk*

$$\hat{L}_\lambda(w) := \frac{1}{n} \sum_{i=1}^n \|\langle w, x_i \rangle - y_i\|^2 + \lambda \|w\|^2 \quad (3.21)$$

and to select the solution to this modified minimisation problem, indicated by the symbol $\hat{w}_\lambda$, as the approximation of the solution to our original problem. The equation to solve then becomes

$$\nabla \hat{L}(w) = \frac{2}{n} \hat{X}^t (\hat{X}w - \hat{y}) + 2\lambda w = 0 \iff (\hat{X}^t \hat{X} + \lambda n \mathbb{1})w = \hat{X}^t \hat{y} \quad (3.22)$$

which, with a simple absorption of the factor n in the definition of the hyper-parameter λ , leads to the same solution as the pseudo-inverse regularisation. This different point of view of the same procedure is nevertheless useful since it explicitly shows, by also minimising over the squared norm of the solution vector w , how Tikhonov regularisation introduces a preference bias for solutions with smaller norms, a fact that, in the other framework, was not evident.

§3.4 Binary classification: soft Support Vector Machines (SVMs)

This section, similarly to the previous one, allows the introduction of other cardinal concepts of ML, namely *training* and *Stochastic Gradient Descent (SGD)*, in a simple and mathematically-sound framework. Let the problem be one of *binary classification*, hence let $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} = \{\pm 1\}$. The choice of the class labels as $+1$ and -1 is a smart one as this allows the condition that at least one solution exists to be expressed as

$$\exists w \in \mathbb{R}^d \mid \forall i = 1, \dots, n \quad y_i \langle w, x_i \rangle > 0 \quad (3.23)$$

which, in other terms, means that all points lying on the side of the hyper-plane generated by the vector w towards which w points must be part of the $+1$ class, whereas all the ones lying on the other side must be part of the -1 class. In general, the product $yf(x)$ of the output with the predicted output is called *margin* and there exists a whole category of losses, dubbed *margin-based losses*, able to quantify the classification error. Broadly—thanks to the smart class label definition implying that any correctly-classified data entry has positive margin—any function that is large when the input is negative and small when the input is positive can be used to this extent. One widely-utilised margin-based loss is the *logistic loss*:

$$\ell(x) = \ln(1 + e^{-yf(x)}) \quad (3.24)$$

also referred to as *binary cross entropy* since its multi-dimensional implementation is called Cross Entropy (CE). In a similar way to Equation 3.21, this loss is generally *regularised*, hence to solve this classification problem it is sufficient to minimise the regularised empirical error

$$\hat{w}_\lambda := \arg \min_{w \in \mathbb{R}^d} \hat{L}_\lambda(w) = \arg \min_{w \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \ln(1 + e^{-y_i \langle w, x_i \rangle}) + \lambda \|w\|^2. \quad (3.25)$$

Unfortunately though, even though the regularised empirical error is convex, and hence has one unique minimum, its gradient

$$\nabla \hat{L}_\lambda(w) = -\frac{1}{n} \sum_{i=1}^n \frac{y_i x_i}{1 + e^{y_i \langle w, x_i \rangle}} + 2\lambda w \quad (3.26)$$

is not linear and thus the equation $\nabla \hat{L}_\lambda(\hat{w}_\lambda) = 0$ cannot be analytically solved. It is, however, enough to explain why this algorithm is called *soft* SVM. Indeed, with simple equation manipulation it is possible to claim that *if* a solution $\hat{w}_\lambda$ does exist—and it must due to convexity—then the solution is such that

$$\hat{w}_\lambda = \sum_{i=1}^n \frac{1}{2\lambda n} \frac{y_i}{1 + e^{y_i \langle \hat{w}_\lambda, x_i \rangle}} x_i = \sum_{i=1}^n c_i x_i. \quad (3.27)$$

The solution to this minimisation problem can therefore be written as a linear combination of the input data. This property, often referred to as the *kernel trick*, is not unique to the soft SVM algorithm, and is in general valid for any algorithm that can be expressed as the Tikhonov-regularised minimisation problem

$$\min_{w \in \mathbb{R}^d} \sum_{i=1}^n \ell(y_i, \langle w, x_i \rangle) + \lambda \|w\|^2 \quad (3.28)$$

like the regularised least squares [Section 3.3.1]. One can notice that the c_i coefficients become very small as the margin rises, giving origin to an argument about *data quality*: an increase in data quantity does not inherently imply an increase in solution quality as the addition of a single, yet valuable, data entry can modify the predicted solution much more than the addition of large amounts of less-important data. Of course this argument is also extremely prone to backfire since the addition of a small amount of sub-optimal data can widely skew the outcome towards a bad solution.

In an adjacent class of algorithms, called *hard* SVMs, some of the c_i coefficients are zero and the data entries that instead do contribute to the solution are called *support vectors*. It is for this reason that, since with the logistic loss no coefficient c_i can be zero by definition [Equation 3.27], this algorithm takes the name of *soft* SVM.

§3.4.1 Stochastic Gradient Descent (SGD) also known as “training”

We are left with an issue: the need to find the minimum of the regularised empirical risk. A whole family of algorithms, called Stochastic Gradient Descent (SGD) algorithms, was created specifically for this purpose. The main idea is to guess a random solution $\hat{w}_0$, calculating the gradient at that location, and then iteratively updating its value by performing small gradient-informed steps according to the law

$$\hat{w}_{k+1} = \hat{w}_k - \gamma \nabla \hat{L}_\lambda(\hat{w}_k). \quad (3.29)$$

The step size γ is called the *learning rate* and, together with the amount of steps to take and the initial guess $\hat{w}_0$, is a hyper-parameter that needs tuning.

The word *training* in ML refers exactly to this: that the empirical error optimisation problem either does not have an analytical solution, is not convex, or both, and therefore that one needs to crunch the numbers to obtain an approximate minimum. A great amount of different SGD algorithms have been created, each with their own logic and heuristic merits.

§3.5 Non-linear learning

For now, the focus was on a fairly simple subclass of problems: those with a linear solution. Nevertheless, as seen with Equation 3.26 it is not rare to face non-linear problems. Two apparently different methods have been developed to generalise what outlined in the previous sections to the non-linear case: *kernel methods* [Section 3.5.1] and *neural networks* [Section 3.5.2]. For this whole section let $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}$.

§3.5.1 Kernel methods

Suppose to have a non-linear problem. If at least some information about *how* it deviates from linearity is present (e.g. exponential or polynomial data distribution) then it is plausible to first *linearise* the problem and finally solve the associated linear problem with the previous methods.

Consider the regularised least squares algorithm outlined in Section 3.3 and remember the *kernel trick* [Equation 3.27], that is, the fact that the solution to these minimisation problems can be written as a linear combination of the input data

$$\hat{w} = \sum_{i=1}^n c_i x_i \quad (3.30)$$

and that hence the predicted output for any given input is

$$f_w(x) = \langle x, w \rangle = x^t \sum_{i=1}^n x_i c_i = \sum_{i=1}^n x^t x_i c_i. \quad (3.31)$$

Let $\Phi : \mathcal{X} \rightarrow \mathbb{R}^p$ be a non-linear function, called a *feature map*, that linearises the data. It follows that the function defined as

$$f_{w,\Phi} = \langle \Phi(x), w \rangle = \sum_{i=1}^n \Phi^t(x) \Phi(x_i) c_i. \quad (3.32)$$

is a linear combination of non-linear functions that could properly describe a non-linear dataset.

Definition 3.5.1. $K(x, \bar{x}) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$ defined as $K(x, \bar{x}) := \Phi^t(x) \Phi(\bar{x})$ is called a *kernel* and defines an inner product.

Indeed, from the definition, a natural consequence is that K is both symmetric and positive-semidefinite. The only relevant elements are therefore these internal products, removing the need to calculate the linearised counterpart to each data entry $\Phi(x)$ which could require a numerically intensive or unstable calculation.

This method can be extremely general: for example let $\mathcal{X} = [0, 1)$, then the kernel

$$K(x, \bar{x}) := \sum_{j=0}^{+\infty} x^j \bar{x}^j = \frac{1}{1 - x\bar{x}} \quad (3.33)$$

allows one to perform regression or classification using a infinite-degree polynomials, a highly expressive class of functions.

§3.5.2 Neural networks

It has been shown how kernel methods adapt linear algorithms to non-linear problems by first linearising the data and then proceeding linearly: $f_{w,\Phi}(x) = \langle w, \Phi(x) \rangle$. A different equally simple way to adapt a linear algorithm to a non-linear problem is to apply the non-linearity to the whole inner product. Indeed, let σ be a non-linear function referred to as the *activation function*; then, $f_{w,\sigma} = \sigma(\langle w, x \rangle)$ is a non-linear function. Algorithms that build upon this concept are called *neural networks*, the simplest version of which, $f_{w,\sigma}$, is called *perceptron*. However, since the associated minimisation problem

$$\min_{w \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n (y_i - \sigma(\langle w, x_i \rangle))^2 \quad (3.34)$$

is not convex, this whole class of algorithms does not have closed-form solutions and must be solved numerically with SGD [Section 3.4.1]. It is worth noting that the addition of a bias b inside the non-linearity is redundant, since its effect can be absorbed by increasing the dimensions $\mathbb{R}^d \rightarrow \mathbb{R}^{d+1}$ of both the data, by padding it with a one $x \mapsto (x_1, \dots, x_n, 1)$, and the weight by adding the bias $w \mapsto (w_1, \dots, w_n, b)$. An important theorem helps explain why neural networks have nevertheless gained such widespread popularity.

Theorem 3.5.2 (of universal approximation). *Single-layer neural networks are dense in the space of continuous real-valued functions.*

$$\begin{aligned} &\forall f \in \mathcal{C}^0(\mathbb{R}), \forall \varepsilon > 0, \exists m \in \mathbb{N}, \exists \exists \text{ weights } w_j, W_j \\ &\text{such that } \left\| f - \sum_{j=1}^m w_j \sigma(W_j \cdot) \right\|_{\infty} \leq \varepsilon \end{aligned} \quad (3.35)$$

Where the weights W_j are matrices that express the internal product of Equation 3.34, the w_j are simple coefficients that superimpose the different perceptrons, and the tiny dot $\cdot$ stands for the function's missing argument. This theorem was first proved by G. Cybenko [62] in 1989 and it means that a linear combination of perceptrons can approximate *any* continuous function so long as no assumptions are made on its amount of perceptrons m .

Despite their introduction as an alternative, neural networks are actually very closely related to kernel methods. Indeed by taking Equation 3.32 and writing the feature map as $\Phi = \sigma(Wx)$ then one gets exactly a single-layer NN

$$f(x) = \langle w, \Phi(x) \rangle = \langle w, \sigma(Wx) \rangle = \sum_{j=1}^d w_j \sigma(W_j x). \quad (3.36)$$

Neural networks can be hence thought of as kernel methods that also *learn the feature map*. It is therefore easy to understand why NNs are more and more used in data analysis: they can approximate a large family of functions and require no hypothesis on the underlying data distribution.

Deep Neural Networks It is currently not well understood why, but increasing the number of *layers* of a neural network tends to generalise with fewer parameters than by increasing the network's width. The structure of a deep NN is that of a highly-nested function. If these functions are only perceptrons,

then the network is called a Multi Layer Perceptron (MLP). For example a MLP with three *hidden layers* infers the output using the function

$$f(x) = \langle w, \sigma(W^{(3)}\sigma(W^{(2)}\sigma(W^{(1)}x))) \rangle. \quad (3.37)$$

Training a deep neural network with L hidden layers consists in the minimisation of the empirical error with respect to each weight. Let $\theta = (w, W^{(1)}, W^{(2)}, \dots, W^{(L)})$, then the problem is to find

$$\hat{\theta} := \arg \min_{\theta} \hat{L}(\theta) := \arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n (y_i - f_{\theta}(x_i))^2 \quad (3.38)$$

It is important to note that the calculation of the gradient of a highly-nested function is, in general, not straightforward, but a highly-specialised algorithm, called *backpropagation* was developed specifically for this purpose.

Backpropagation First of all, a useful notation. Let

$$a_0 = x, \quad a_{\lambda} = \sigma(W^{(\lambda)}a_{\lambda-1}) = \sigma(z_{\lambda}), \quad a_{L+1} = wa_L \quad (3.39)$$

where λ indexes the hidden layers and thus runs from 1 to L . The procedure allows the calculation of the updated weights θ_k at the k -th step of SGD minimisation under the assumption that we know the weights at the previous step θ_{k-1} . It all begins with the calculation of all the layers using the θ_{k-1} weights, a step often referred to as the **forward pass**. All of the a_{λ} and z_{λ} are henceforth to be considered known values. The *loss* at some specific data entry (x, y) will be:

$$\ell(\theta) = (y - a_{L+1})^2 = (y - wa_L)^2 \implies \frac{\partial \ell(\theta)}{\partial w} = 2(wa_L - y)a_L. \quad (3.40)$$

The gradient with respect to the parameter w is thus known. By the chain rule it follows that:

$$\frac{\partial \ell(\theta)}{\partial W^{(\lambda)}} = \frac{\partial \ell(\theta)}{\partial z_{\lambda}} \frac{\partial z_{\lambda}}{\partial W^{(\lambda)}} =: \delta_{\lambda} \frac{\partial z_{\lambda}}{\partial W^{(\lambda)}} = \delta_{\lambda} a_{\lambda-1}. \quad (3.41)$$

Even though $a_{\lambda-1}$ is known, δ_{λ} is not. Nevertheless by re-applying the chain rule

$$\delta_{\lambda} = \frac{\partial \ell(\theta)}{\partial z_{\lambda}} = \frac{\partial \ell(\theta)}{\partial z_{\lambda+1}} \frac{\partial z_{\lambda+1}}{\partial z_{\lambda}} = \delta_{\lambda+1} \frac{\partial z_{\lambda+1}}{\partial z_{\lambda}} \quad (3.42)$$

notice how the quantity δ_{λ} at some specific layer λ depends on the value of that same quantity at a deeper layer $\lambda + 1$. Furthermore, since $z_{\lambda+1} = W^{(\lambda+1)}a_{\lambda} = W^{(\lambda+1)}\sigma(z_{\lambda})$, by assuming σ to be differentiable and indicating with $\odot$ the *Hadamard product* operation for which $(A \odot B)_{ij} = A_{ij}B_{ij}$, then

$$\delta_{\lambda} = \delta_{\lambda+1} W^{(\lambda+1)} \odot \sigma'(z_{\lambda}). \quad (3.43)$$

It is therefore possible to calculate the deepest delta

$$\delta_{L-1} = \frac{\partial}{\partial z_{L-1}} (y - w\sigma(z_{L-1}))^2 = 2(w\sigma(z_{L-1}) - y)w \odot \sigma'(z_{L-1}) \quad (3.44)$$

and then perform a **backward pass** following Equation 3.43 to calculate all of the δ_{λ} . After the calculation of all the a_{λ} in the forward pass, and all of the δ_{λ} in the backward pass, it is possible to

calculate the gradient of the loss with respect to every $W^{(\lambda)}$ as defined in Equation 3.41. Finally, since the gradient is a linear operator, this algorithm can be used to calculate the gradient of the empirical error, that is the average loss on the dataset, by calculating the gradient of the loss on each entry and then performing an average.

§3.6 Geometrical Deep Learning

As claimed in the introduction to Section 3.5, to introduce these methods of non-linear learning it was assumed that the input data consists of a one-dimensional collection of real numbers $\mathcal{X} \subseteq \mathbb{R}^d$, but what if data has a larger dimensionality? Suppose for example that the input data are small 32×32 grey-scale images like the one shown in Figure 17. It could be tempting to first flatten out the image into a one-dimensional vector in $\mathbb{R}^{1024}$ and then to feed it to a MLP, but this approach quickly encounters two big issues. The first one is fairly obvious: this flattening procedure destroys any kind of spatial relations between pixels, reducing the amount of information contained in the data simply by modifying its shape. The second issue has been given a fancy name: the *curse of dimensionality*. This “curse” is a recurring phenomenon that, for the purpose of this work, can be briefly described by a disproportionate need for more data as the dimensionality of the data itself increases. Indeed, when sampling a one-dimensional unit interval around 100 samples are required to achieve a precision in the order of percent. Conversely to sample a 10-dimensional unit hypercube with the same precision, one would need around $100^{10} = 10^{20}$ samples. For $\mathbb{R}^{1024}$ the amount of points grows to 10^{2048} . This is a huge number of points. For comparison, the amount of electrons in the universe is estimated to be around 10^{80} .

§3.6.1 The blueprint

Despite these complications, it is a matter of fact that data is often complex in nature and that hence high-dimensional learning has to be possible in order for Machine Learning (ML) to become and stay relevant. A way to escape the curse must be found. Luckily, a solution does exist and it exploits a concept physicists are extremely familiar with: symmetry. Indeed as the book by M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković [63] exposes, a whole category of fairly new (10-15 years old) algorithms can be grouped in this category they refer to as *Geometric Deep Learning*. All these architectures overcome the curse of dimensionality by exploiting two priors: *deformation stability* and *scale separation*.

Definition 3.6.1 (Geometric Deep Learning blueprint). Let $\mathcal{X}$ be a domain and $\mathfrak{G}$ a symmetry group over $\mathcal{X}$. Define the following building blocks:

- *Linear $\mathfrak{G}$ -equivariant layer* $B : \mathcal{X} \rightarrow \mathcal{X}$ satisfying $B(\mathfrak{g}.x) = \mathfrak{g}.B(x)$ for all $\mathfrak{g} \in \mathfrak{G}$ and $x \in \mathcal{X}$.
- *Nonlinearity* σ applied element-wise.
- *Local pooling (coarsening)* $P : \mathcal{X} \rightarrow \mathcal{X}'$ such that $\mathcal{X}' \subseteq \mathcal{X}$.
- *$\mathfrak{G}$ -invariant layer (global pooling)* $A : \mathcal{X} \rightarrow \mathcal{Y}$ satisfying $A(\mathfrak{g}.x) = A(x) \forall \mathfrak{g} \in \mathfrak{G}$ and $x \in \mathcal{X}$.

These blocks can be used to construct a $\mathfrak{G}$ -invariant function $f : \mathcal{X} \rightarrow \mathcal{Y}$ of the form

$$f = A \circ \sigma_L \circ B_L \circ P_{L-1} \circ \dots \circ P_1 \circ \sigma_1 \circ B_1 \quad (3.45)$$

Different blocks may exploit different choices of symmetry groups $\mathfrak{G}$.

The book shows how by using this blueprint and changing exclusively domains and symmetries it is possible to obtain a wide variety of ML architectures. For example by setting the domain to graphs and the symmetry to permutations the blueprint describes Graph Neural Networks (GNNs), by

changing the domain to fully-connected graphs one gets *transformers*—the architecture on which Large Language Models (LLMs) are based on—and by setting the domain to a grid and the symmetry to translation the blueprint describes Convolutional Neural Networks (CNNs) [63, §3.5].

§3.6.2 A specific case: Convolutional Neural Networks

The convolution between two functions is defined as

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t - \tau) d\tau. \quad (3.46)$$

By using this definition it is fairly simple to verify that convolution is a linear operator and that it is also translation-equivariant, or equivalently that it commutes with the translation operator. Thus, convolution satisfies the condition required to be the B layer of the geometrical learning blueprint. This operator can be trivially discretised to work on grids

$$(f * g)[n] = \sum_{m=-\infty}^{+\infty} f[n]g[n - m]. \quad (3.47)$$

It is worth noting that in most Machine Learning (ML) development suites, a convolutional block is actually implemented via the *cross-correlation* operation

$$(f \star g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t + \tau) d\tau. \quad (3.48)$$

For example, suppose to have N pictures with C_{in} colours and dimensions $H \times W$. In the *PyTorch* framework, the `torch.nn.Conv2d` actually calculates

$$\text{out}(N_i, C_{\text{out}_j}) = \text{bias}(C_{\text{out}_j}) + \sum_{k=0}^{C_{\text{in}}-1} \text{weight}(C_{\text{out}_j}, k) \star \text{input}(N_i, k) \quad (3.49)$$

where C_{out} is a user-decided parameter and $\text{bias}(C_{\text{out}_j})$ is an additive learned weight. Cross-correlation and convolution are very closely related but there is a relevant difference, namely the fact that cross-correlation is not translation-equivariant. Nevertheless since the filter g is, for ML purposes, learned, the practical difference between using cross-correlation or convolution is saving computation resources by avoiding a superfluous kernel flip. From the short introduction on geometrical deep learning it should be clear that when working with images, which are grids, specifically with the objective of extrapolating *if* something is present in the image and not *where*, CNNs offer themselves as a great tool, specifically because of the translation symmetry.

A standard CNN closely resembles the blueprint: it features a convolution, generally with $C_{\text{out}} > C_{\text{in}}$, followed by a non-linearity—often the Rectified Linear Unit (ReLU) function—and finally a *pooling* layer with the objective of reducing the $H \times W$ dimensions of the image. These three objects define what is usually referred to as a single *layer* of a CNN.

$$\text{Layer}_{\text{CNN}} = \text{Pool}(\sigma(\text{Conv}(x))). \quad (3.50)$$

Multiple convolution layers are applied sequentially to the image. The result is then passed to the *global pooling* part of the algorithm where the image, now with reduced dimensions and augmented channels (colours), is, in general, flattened and passed through an MLP who performs the final classification or regression task.

§3.6.3 ResNet

Residual neural networks represent a slight modification of the standard CNN architecture described previously. One ResNet layer has the structure

$$\text{Layer}_{\text{ResNet}} = \text{Pool}(x + \sigma(\text{Conv}(x))). \quad (3.51)$$

This architecture was first introduced in an article by K. He, X. Zhang, S. Ren, and J. Sun [64] who showed how it was possible to reduce the training error simply by converting CNN layers to ResNet layers. This technique is more and more efficient as the depth of the architecture grows and is substantially computationally free. Due to both its trivial computational cost and its simplicity of integration, ResNets have become the standard way of building a CNN.



Lens-based reconstruction

Explaining the novel work performed in the context established previously is the primary purpose of this section. The objective of this work can be summarised as: ***proving that event reconstruction in GRAIN can strongly benefit from ML-based techniques by implementing a computer vision algorithm that is able to extract the amount of visible particle tracks in each GRAIN camera and compare this result against the performance of standard reconstruction algorithms used so far.*** A few details on the chosen hyper-parameters—e.g. number of layers, loss and optimiser type and all their parameters—will be provided, mostly because they are the result of a trial and error optimisation. Indeed the objective of this thesis is a **proof of concept**, not the final iteration of a permanent reconstruction chain, in which case much more caution would be applied. In order to understand the input to GRAIN’s reconstruction, it is mandatory to understand how an optical system works with its associated challenges: this is the subject of Section 4.1. An explanation of the current GRAIN reconstruction pipeline will follow in Section 4.2. The novel work, that is, the implementation a ML-based reconstruction algorithm to extract the amount of tracks seen by each GRAIN camera will be the focus of Section 4.3. Finally Section 4.4 will briefly overview plausible evolution paths.

§4.1 Introduction

If one wants to generate a picture it is mandatory to utilise an *imaging device*. Indeed to *image* something means exactly to transfer the light generated on one side of an optical system, called the *object space*, at some height y and distance a from the imaging device to some height y' and distance a' from the imaging device in the *image space*, that is, on the other side of the optical system [65, §3.3.1] as shown in Figure 18b. The image coordinates y' and a' are connected to their object counterparts y and a by the *imaging* and *magnification* equations [65, Eq. 3.5 and 3.7]

$$\frac{1}{f} = \frac{1}{a} + \frac{1}{a'}, \quad m = \frac{a'}{a} = \frac{y'}{y}. \quad (4.1)$$

where f is the *effective focal length* (EFL), a parameter that depends on the optical system’s physical properties (see Equation 4.2 for an example). These equations represent an approximation, but they can be generalised to more complex optical systems by using a different formalism [Equation 4.5].

Various different imaging technologies exist, two of which—lenses and coded-aperture masks—were briefly introduced in Section 2.6.2. This thesis focuses on the former. Formally, a lens is a transmissive optical device that focuses or disperses a light beam by means of refraction. In general, lenses can be categorised according to different criteria, e.g.:

- By their functionality (converging or diverging).
- By their constituent material.
- By their shape.

As shown in Figure 18a, to completely define a lens, seven parameters are needed: radii of curvature R_i , centre thickness t_c , diameter D and lens material, specified by the index of refraction n , the absorption

coefficient α and the V-number V [65, §3.1.1]. To understand the basics of lensing devices it can be useful to begin by analysing the simplest one, the *thin lens*. A thin lens is defined as a lens in which the centre thickness is much smaller than either curvature radii $t_c \ll R_i$. The effective focusing distance f of this imaging device can be calculated using the so-called *lensmaker equation*

$$\frac{1}{f} = \frac{n - n_0}{n_0} \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \quad (4.2)$$

in which n_0 is the refractive index of the ambient medium.

Furthermore, for a thin lens Equation 4.1 is a good approximation. By using it, one can conclude that, thanks to its radial symmetry, a thin lens can be thought of as a device which maps any 3D point in the object space to 2D points on a plane at a distance a' from the lens³

$$\mathbf{r} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \mapsto \mathbf{r}' = \begin{pmatrix} x \\ y \end{pmatrix} = a' \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix}. \quad (4.3)$$

It is a smart move to *linearise* this equation by increasing the dimensionality of the problem. Indeed it is possible to define

$$\mathbf{r} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{r}' = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (4.4)$$

which allows the formalisation of the imaging process through a linear operator P

$$\mathbf{r}' = P(\mathbf{r})\mathbf{r} \quad \text{with} \quad P(\mathbf{r}) = \frac{1}{Z} \begin{pmatrix} a' & 0 & 0 & 0 \\ 0 & a' & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (4.5)$$

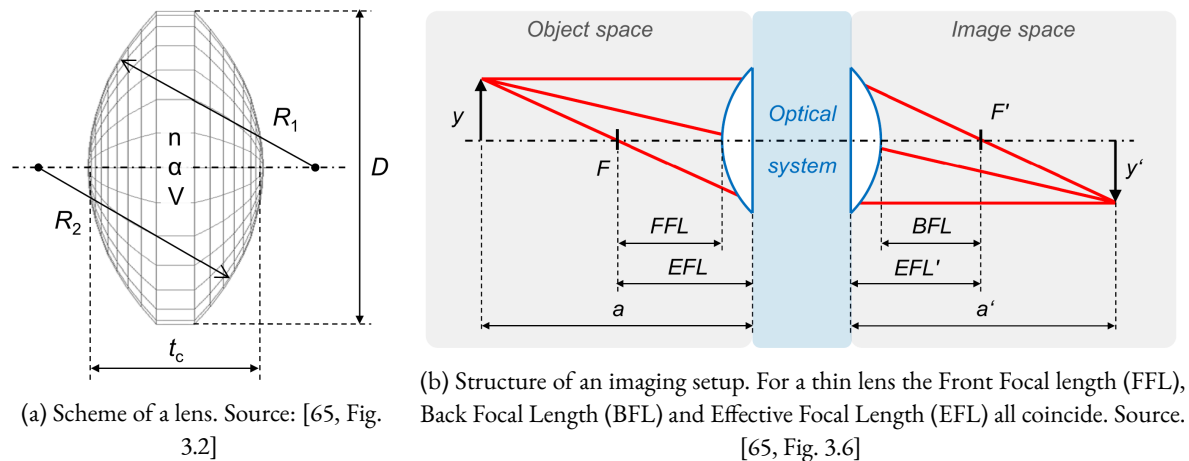


Figure 18: Schematics of a standard lens and imaging setup.

³As Figure 18b shows, the coordinates y and y' have opposite direction. Indeed lenses do also cause a *radial flip*, but since this is a known effect that can be systematically mitigated by mapping $\mathbf{r}' \mapsto -\mathbf{r}'$, this fact can be ignored in order to keep the equations simpler.

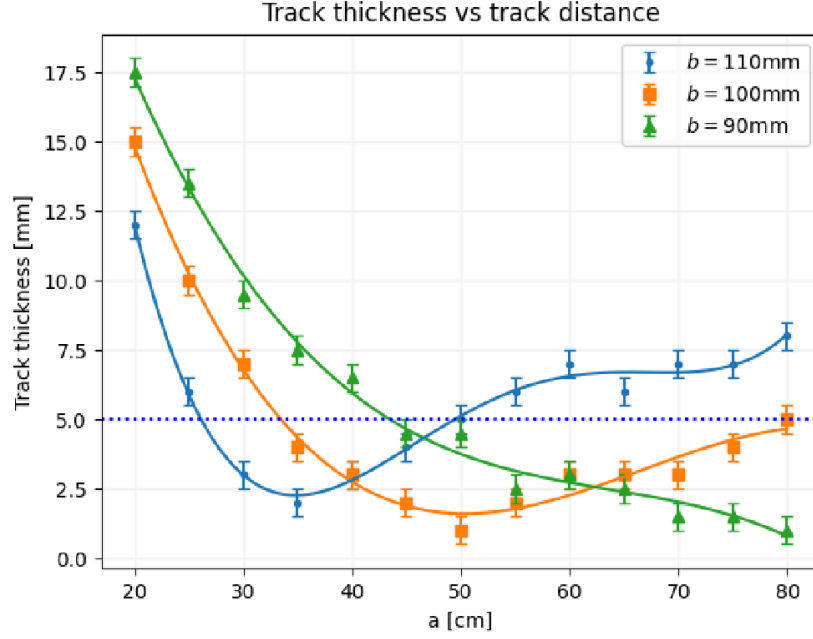


Figure 19: GRAIN lens-SiPM distance optimisation. Note, the parameter b is a' . Additionally this specific graph refers to a lens 50 mm in diameter; analogous studies were performed in order to optimise the 60 mm lens camera. Source: [66, Fig. 4.13]

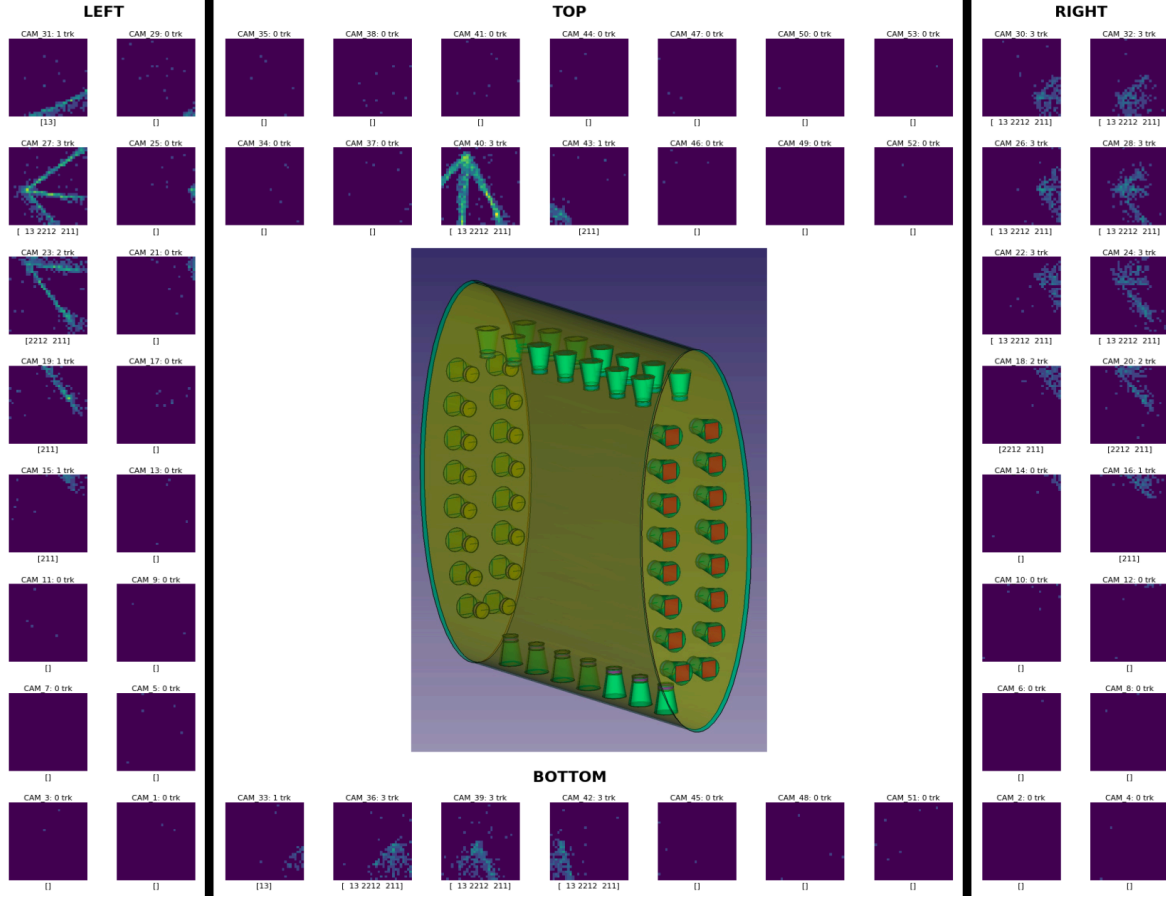
This formalism is useful because the matrix P can describe any optical system, especially more complex ones, and is useful to treat the multi-camera case as explained by M. Vicenzi [66, §5.2.2]. Indeed, from Equation 4.3 it is evident that different 3D points can be imaged to the very same 2D point. For example, both $(12, 12, 3)^t$ and $(8, 8, 2)^t$ will be imaged to $(4, 4)^t$. One way to break this degeneracy is to observe the same light source with multiple cameras. For this reason and in order to satisfy general cryogenic and mechanical requirements, the current baseline for the lens-based GRAIN design [Section 2.6] uses 53 identical cameras [Section 2.6.2.2] to detect and image the scintillation light from LAr [Section 1.7].

§4.1.1 Out-of-focus imaging

By looking at Figure 18b and imagining to move the sensor plane closer to, or further from, the optical system than a' , one can see that a point source will be imaged as a disk and that hence a thin track will be imaged as a thicker track. This deformation can be mitigated by allowing the sensor to slide back and forth in order to reach the optimal imaging lens-sensor distance. However, the implementation of such a system would drastically complicate the camera design and be overall inappropriate for the objective of this detector, that is, to image neutrino events from a beam of high intensity and frequency. Thus, it has been decided that the lens-SiPM distance a' will be fixed. This distance was optimised by M. Vicenzi [66] as shown in Figure 19 and chosen as the one that maximises the Field of View (FoV), a quantity defined as the range of source-lens distances a for which the imaged track is less than 5 mm thick on the SiPM plane. An example of the simulated GRAIN output considering the current baseline geometry and camera setup is shown in Figure 20.

§4.2 Hough transform reconstruction

The current 2-dimensional track-fitting algorithm of lens-based GRAIN is shown on a typical event in Figure 21 and it features three subsequent steps:

Figure 20: GRAIN $1\mu 1p1\pi$ event from a FHC beam-like simulation.

1. Performing the Hough transform of a single camera view.
2. Applying the Multi-Otsu threshold algorithm to classify pixels as background, low-intensity signal, and high-intensity signal.
3. Clustering high-intensity signal pixels using the DBSCAN algorithm. The number of clusters found represents the amount of visible tracks. The coordinates of the cluster's centre represent a parametric definition of the fitted line.

The fitting procedure continues with a 2D-to-3D step in which fitted lines from each camera are matched using multiple-view projective geometry to provide a 3-dimensional reconstruction of the event. Nevertheless, since, the ML step only provides a different path in determining how many tracks are visible in one camera view (as explained in Section 4.3), no additional detail will be given on this later part of the reconstruction.

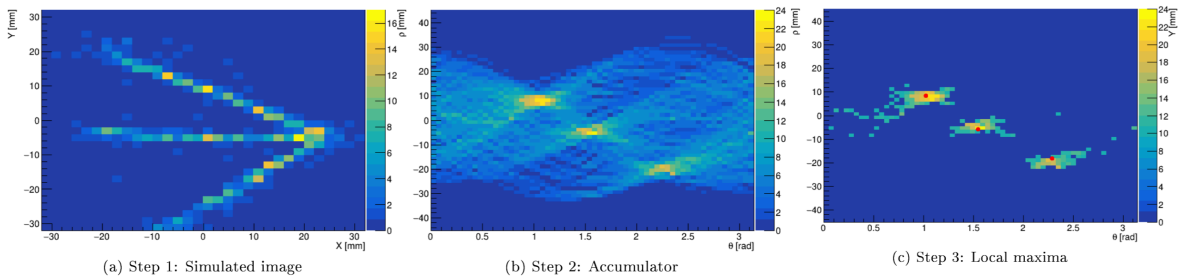


Figure 21: Example GRAIN fitting procedure. Source: [66, Fig. 5.6]

§4.2.1 Hough transform

The Hough transform is a tool developed to detect straight lines in pictures which was improved by R. O. Duda and P. E. Hart [67] who introduced its current parametrisation in 1972. Suppose to have an image, that is, a collection of 2D points $\{(x_1, y_1), \dots, (x_n, y_n)\}$. The Hough-transformed version of the picture is obtained by mapping each point in the (x, y) plane to a sinusoid in a (ρ, θ) space, formally called the *accumulator*, satisfying

$$(x_i, y_i) \mapsto x_i \cos \theta + y_i \sin \theta = \rho. \quad (4.6)$$

As shown in Figure 22, this parametrisation specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin. Thus, by restricting $\theta \in [0, \pi]$, each line in the (x, y) plane corresponds to a unique point in the (θ, ρ) plane. Conversely all the points lying on one sinusoid in the (θ, ρ) plane correspond to lines in the (x, y) plane passing through the point (x_i, y_i) . In practice both the ρ and θ variables are discretely sampled and the Hough transform is performed on a grid. An example of the Hough transform applied to a GRAIN camera with three visible tracks is shown in Figure 21. Generalisations of the Hough transform which do not require the quantisation of the accumulator [68] are briefly mentioned in Section 4.4.2.

§4.2.2 Multi-Otsu thresholding

Otsu's method is a thresholding algorithm originally developed to find a threshold able to separate an image's pixels in a low-intensity *background* and a high-intensity *signal*. The algorithm can be generalised to provide n thresholds, thus separating an image in $n + 1$ classes. In the GRAIN implementation, two thresholds are calculated, separating the accumulator in three classes: background, low-intensity signal and high-intensity signal peaks. Pixels classified as high-intensity are then selected and passed to the final step, as shown in Figure 21. An in-depth explanation of this algorithm is provided by M. Vicenzi [66, Appendix A.1].

§4.2.3 DBSCAN clustering

DBSCAN is an unsupervised Machine Learning clustering algorithm. What this means is that there is no need for a *training* phase for the algorithm to accomplish its goal. The algorithm takes as input some dataset, in this case a series of points in the accumulator space $\{(\theta_1, \rho_1), \dots, (\theta_n, \rho_n)\}$, and provides as output a spatial clustering. The algorithm loops over all points in the dataset and two hyper-parameters

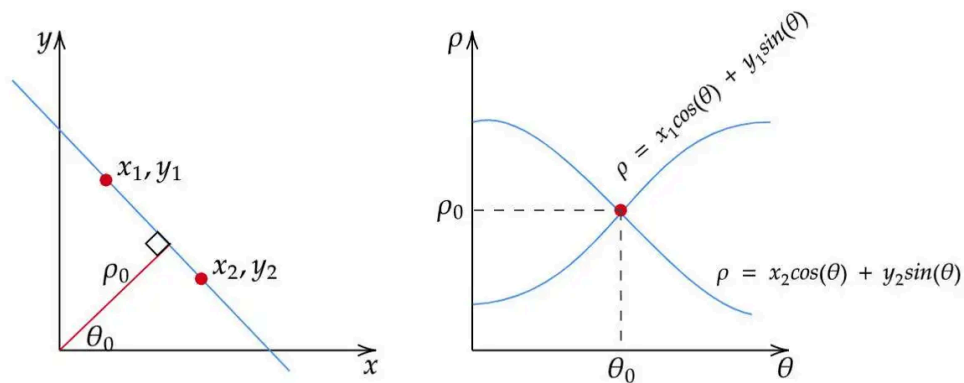


Figure 22: Simple Hough transformation. Source: [69]

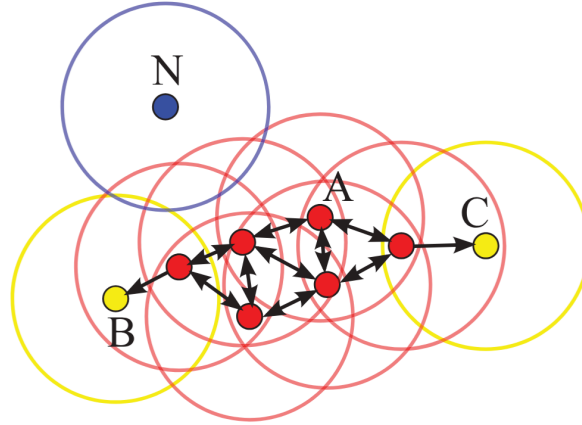


Figure 23: DBSCAN clustering. Core points are shown in red, border points in yellow and noise points in blue. A cluster is defined as all the points in the neighbourhood of a core point. Source: [70, Fig. 1]

are provided: the minimum number of neighbours $n_{\min}$ and a radius ϵ . For each point, if its circular neighbourhood with radius ϵ contains at least $n_{\min}$ points, then it is classified as a *core point* (red) and all its neighbours are added to its cluster. Instead, if that same circle contains at least one point, but less than $n_{\min}$, the point is classified as a *border point* (yellow). Border points are part of the cluster, but they do not have any effect on the classification of their neighbours. Finally if a point has no neighbours it is classified as *noise* (blue). The *amount* of clusters found by DBSCAN defines the baseline comparison for the new ML-based reconstruction outlined in Section 4.3.

§4.3 Truth definition and CNN

As explained in Section 3.6, Convolutional Neural Networks (CNNs) fall into the category of *supervised* ML, thus it is mandatory to provide the network with a training dataset. The selection of training data is a delicate process since, in case of successful learning, any bias towards a specific type of events or limitation (e.g. in terms of population of a certain class of events or general features of the events) of the dataset will be reflected by the trained architecture.

As explained in Section 1.6, the DUNE beam falls within the *intermediate* energy range of neutrino interaction classification. Hence, events in GRAIN are expected to consist of (in order of increasing characteristic energy, and thus of increasing topology complexity) Charged Current Quasi Elastic (CCQE) scattering, resonance production, or Deep Inelastic Scattering (DIS). In accordance with the exploratory nature of the present work, the study is restricted to this first class of interactions. The process consists of a positive Charged Current interaction between a muon neutrino and an Argon nucleus which is modelled as the interaction of a neutrino with a neutron with very small interference coming from other nucleons (hence the adjective “quasi-elastic”)

$$\nu_{\mu} + n \rightarrow p + \mu^{-}. \quad (4.7)$$

The expected final state has a simple topology since it consists of a single muon and proton.

This class of neutrino interactions shows a remarkably simple topology since the expected final state consists exclusively of a proton and a muon, making them the ones who best allow the measurement of neutrino flux, one of the main physics objectives of SAND [Section 2.5].

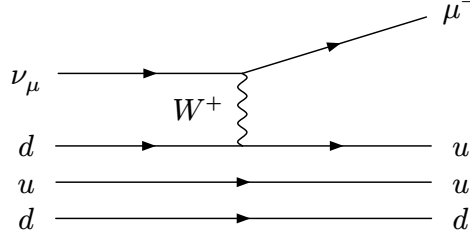


Figure 24: Feynman diagram of FHC CCQE interactions

§4.3.1 Truth propagation to the 2-dimensional level

Neutrino interactions in SAND are generated using GENIE [71], the baseline generator for neutrino interactions in DUNE, and propagated using `edep-sim` [72], a wrapper of particle physics simulation framework Geant4 [73]. Energy depositions from `edep-sim` are then fed to `OptMen` (Optical Meniscus), a custom tool built with Geant4 which is responsible for simulating VUV photons across their life-cycle: from emission, across Liquid Argon (considering effects from Rayleigh scattering and absorption) through the camera lens (considering reflection, refraction, transmission, and absorption) and finally detection from the SiPM matrix [66, §5.1.2]. Nevertheless, `OptMen` simply propagates photons from energy depositions and holds no further Monte Carlo truth information about their origin, hence the relevant information for CNN training, namely the amount of charged particle tracks that are *visible* in one specific camera, has to be projected from the `edep-sim` level. It is thus important to understand *how* this information is organised and stored inside `edep-sim`. For the purpose of this work, just two classes of `edep-sim` are relevant: `TG4Trajectory` and `TG4HitSegment`. The two hold plenty of relevant information on the event hierarchy (e.g. which detector the particle is in and which secondary particles were birthed from the considered one), but for the purpose of this study it is enough to extract the path of a particle through the detector from `TG4Trajectory`, and the fraction of deposited energy which resulted in scintillation (called `SecondaryDeposition`) from the `TG4HitSegment`.

Linear threshold The previous implementation [Section 4.2] of truth projection to the 2-dimensional level limited its scope to primary tracks that, projected on the SiPM matrix, had a length greater or equal to 10 pixels. This thesis generalises the truth generation step by considering *all* charged particles emitted in a certain neutrino interaction, even the ones exiting from secondary vertices (if they satisfy the visibility criteria explained later) and how the presence of a magnetic field can affect the particle trajectories making their tracks curve. It is worth noting that finding outliers in truth generation is a very time-consuming procedure since there is no way of performing it other than manually scanning events. The first generalisation was mainly guided by the observation of some anomalous events shown in Figure 25, in which it is clear that the old truth camera image generation was discarding information about relevant particle tracks, thereby undermining the CNN training. The second generalisation was implemented mainly because of future-proofing arguments. Indeed, due to both the high energies of the LBNF beam and the small dimension of GRAIN along the beam direction, in CCQE most events tracks are straight lines—except for small deviations derived from δ rays. This, however, is not true in beam-like events, like the one shown in Figure 26. The simplest way of approximating the length of a curved track is to split it up into shorter straight segments and summing their length. In GRAIN’s simulation framework, tracks are already naturally segmented as `TG4HitSegments`, thus truth definition begins with creating the 2-dimensional “true” image by looping over every `TG4Trajectory`, gathering all its associated `TG4HitSegments`, checking which scintillated and thus resulted in signal for GRAIN (`SecondaryDeposition` > 0) and projecting them to the SiPM plane by using Equation 4.3.

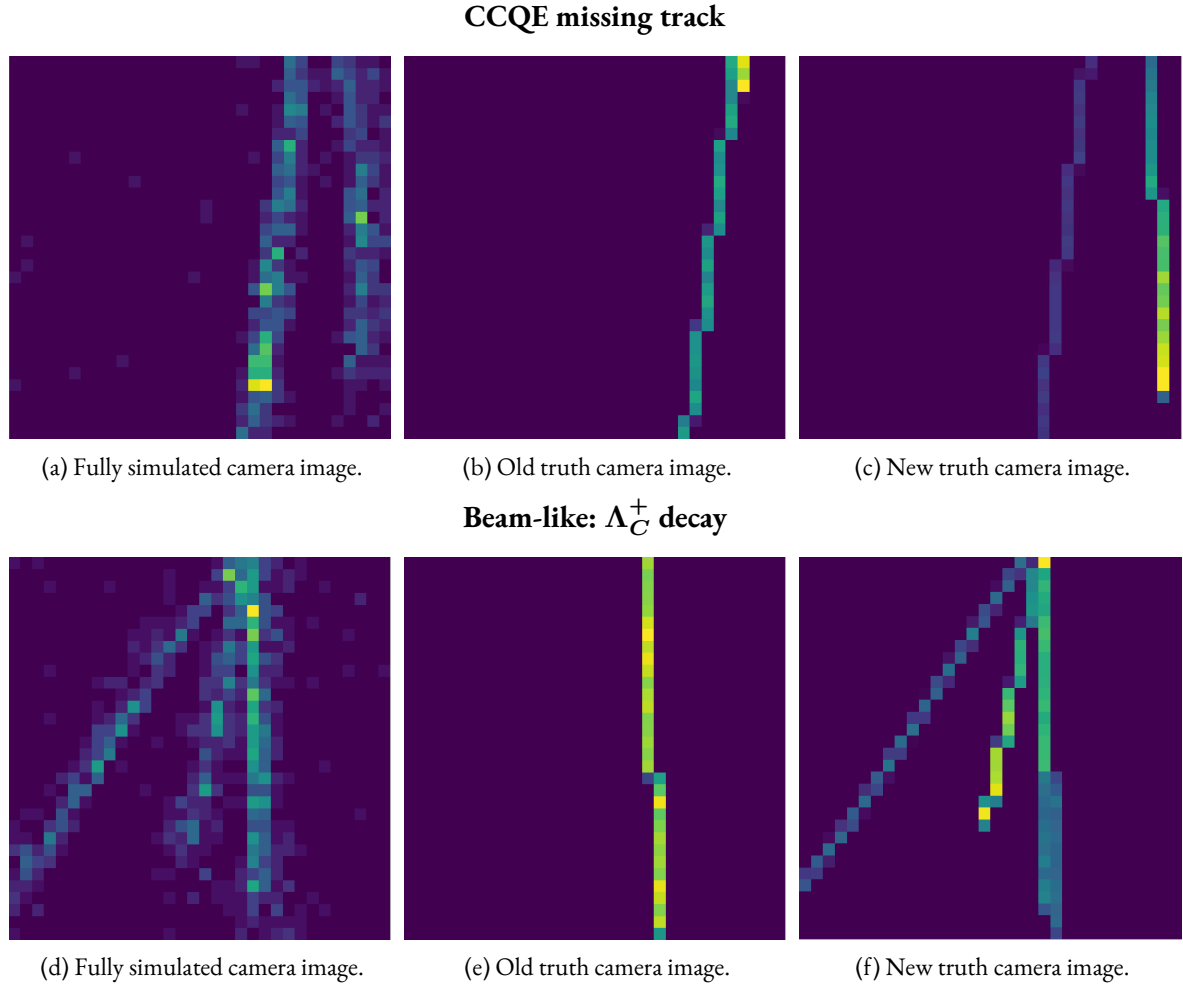


Figure 25: Old 2D truth definition, resulting in issues in the projected tracks of some events. Top: CCQE-like event with a missing track. Bottom: beam-like event where the previous approach in truth definition causes two of three visible tracks to be missing from truth information when projected to 2D.

The *uncurled visible projected* track length is then defined as the sum of the lengths of each projected TG4HitSegment and, if it is greater than 10 mm (≈ 5 pixels), the associated track is considered *linearly visible*.

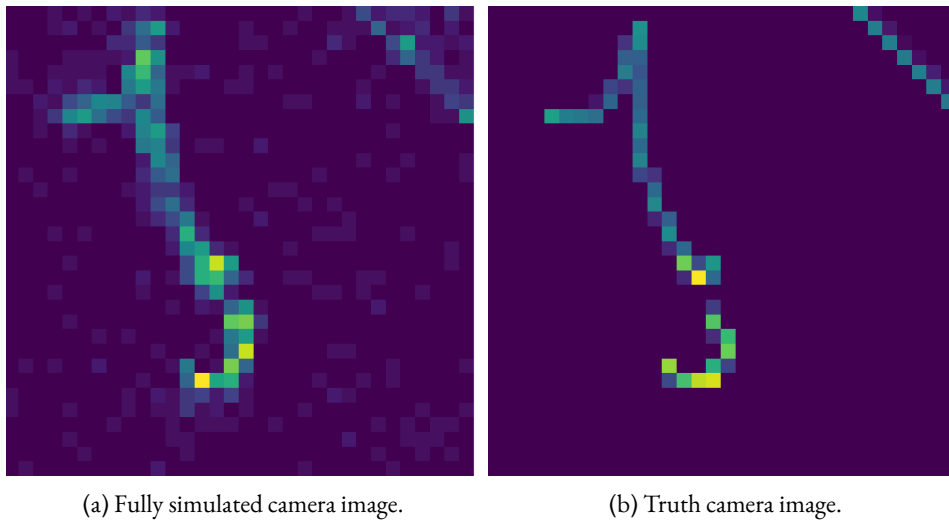


Figure 26: Example of e^+e^- pair production in GRAIN.

Angular threshold As explained in Section 4.1.1, having a fixed focus implies that often tracks appear as thick blobs. If the angle separating two tracks is not wide enough, then this smearing caused by the optical system could make it impossible to resolve two different tracks. In order to produce a reasonable visual truth, this work introduces an *angular threshold* with the aim of considering any group of tracks that are not enough angularly separated as a single track. There are two ways in which two different tracks could look like a single one due to off-focus smearing: either the two tracks come off the same interaction vertex or the tail of the first one coincides with the tip of the second one (for CCQE-like events, this happens in the case of protons elastically scattering off a LAr atom or for especially energetic ionisation electrons). Both geometrical effects are shown in Figure 27. For the purpose of this study, the direction of the tip (tail) of a track is calculated by averaging the direction of the `TG4HitSegments` that account for the first (last) 20 % of a track’s projected length. Each possible combination of track tips and tails is checked: if the two are located closer than one linear threshold, then the algorithm controls whether they are angularly separated more than $0.2 \text{ rad} \approx 11.5^\circ$. This threshold is augmented by a factor 1.5 for the tip-tail check since, as shown in Figure 27, the same angular separation clearly distinguishes the two same-vertex tracks, but not the re-interacting proton at the top. If two tracks do not pass the angular threshold, then they are flagged as to be merged.

The merging process happens *transitively*, meaning that if some track is to be merged with another and this other is to be merged with a third one, all three are treated as a single track. This problem is equivalent to generating a graph with a number of nodes equal to the number of linearly visible tracks, creating an edge between two nodes if the two tracks do not satisfy the angular threshold, and finally finding the amount of disconnected graphs. This quantity is defined as the number of visible tracks in the image.

§4.3.2 Dataset assembly

The dataset was assembled using 11k simulated ν_μ CCQE events. Since each single camera is treated independently, this equates to 583,000 data entries. Nevertheless, most views are completely empty or contain only noise. Since these camera images can be algorithmically removed by requiring that

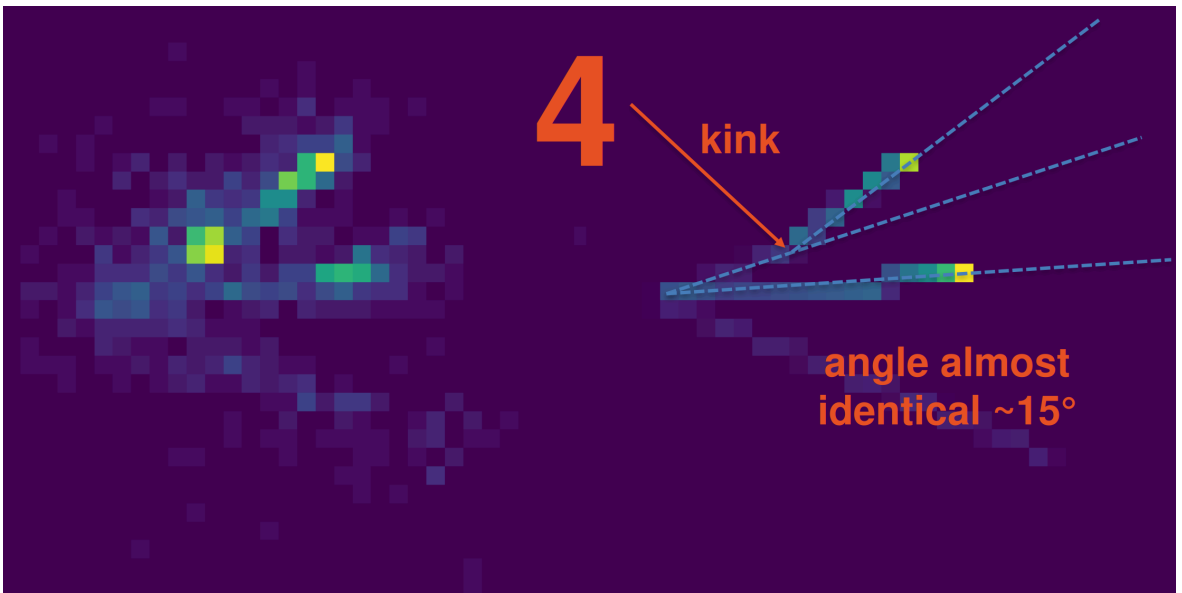


Figure 27: Event with four charged particle tracks both simulated (left) and projected (right). Even though the Monte Carlo truth features 4 `TG4Trajectories`, it is evident that the two topmost tracks are indistinguishable.

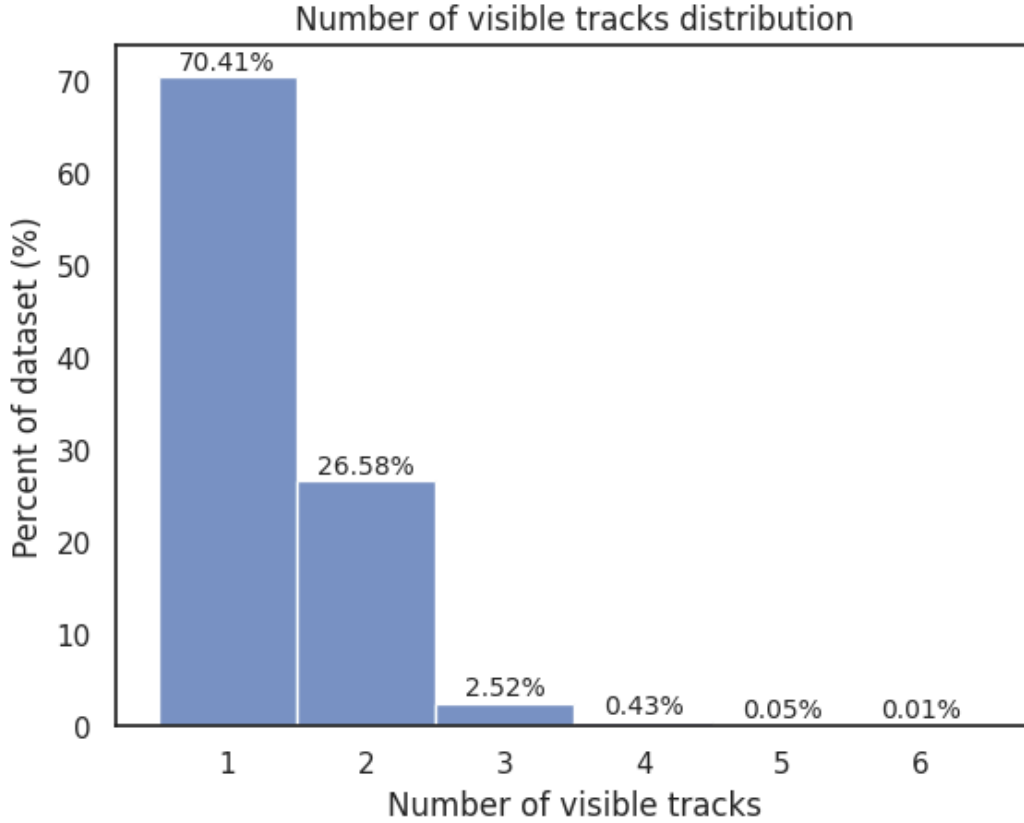


Figure 28: Visible tracks distribution in the training dataset.

enough pixels (e.g. 3 % [66, §5.2.1]) are filled, it is both unnecessary and counter-productive to flood the training dataset with meaningless information. Furthermore by looking at the class distribution histogram [Figure 28] it is possible to see that camera images with more than 3 tracks account for less than 1 % of the whole dataset, making them practically impossible to learn. Thus, it was decided to train the neural network to work only on images with one, two or three visible tracks. The resulting dataset contains:

- 99,597 cameras with one visible track.
- 37,596 cameras with two visible tracks.
- 3,565 cameras with three visible tracks.

This large class imbalance [Figure 28] is indeed the main cause of classification issues encountered. Possible solutions are discussed in Section 4.4.2.

§4.3.3 Architecture

The architecture used is a standard ResNet image classifier, a class of algorithms briefly introduced in Section 3.6.3. Let $\mathfrak{c}(N_{\text{in}}, N_{\text{out}}, s)$ be the PyTorch Conv2D function defined in Equation 3.49 with N_{in} input channels, N_{out} output channels and stride s , which expresses the number of pixels the kernel moves at each step. Let ConvBlock $\mathfrak{C}$ be defined as

$$\mathfrak{C}(N_{\text{in}}, N_{\text{out}}, s) := \mathfrak{c}(N_{\text{out}}, N_{\text{out}}, s = 1) \circ \sigma \circ \mathfrak{c}(N_{\text{in}}, N_{\text{in}}, s) \quad (4.8)$$

where N_{in} and N_{out} refer to the input and output channels respectively and σ is the ReLU function. Next, let the ResBlock $\mathfrak{R}$ be defined as

$$\mathfrak{R}(N_{\text{in}}, N_{\text{out}}, s) := \sigma \circ [\mathfrak{C}(N_{\text{in}}, N_{\text{out}}, s) + \text{Id}]. \quad (4.9)$$

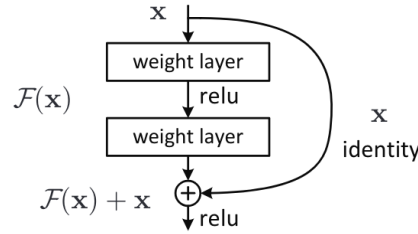
Then define the stem $\mathfrak{S}$ and the classifier tail $\mathfrak{T}$ as

$$\begin{aligned} \mathfrak{S}(N_{\text{in}}, N_{\text{out}}) &:= \sigma \circ \mathfrak{C}(N_{\text{in}}, N_{\text{out}}) \\ \mathfrak{T}(N_{\text{in}}, N_{\text{class}}) &:= \mathfrak{L}(N_{\text{in}}/2, N_{\text{class}}) \circ \sigma \circ \mathfrak{L}(N_{\text{in}}, N_{\text{in}}/2) \end{aligned} \quad (4.10)$$

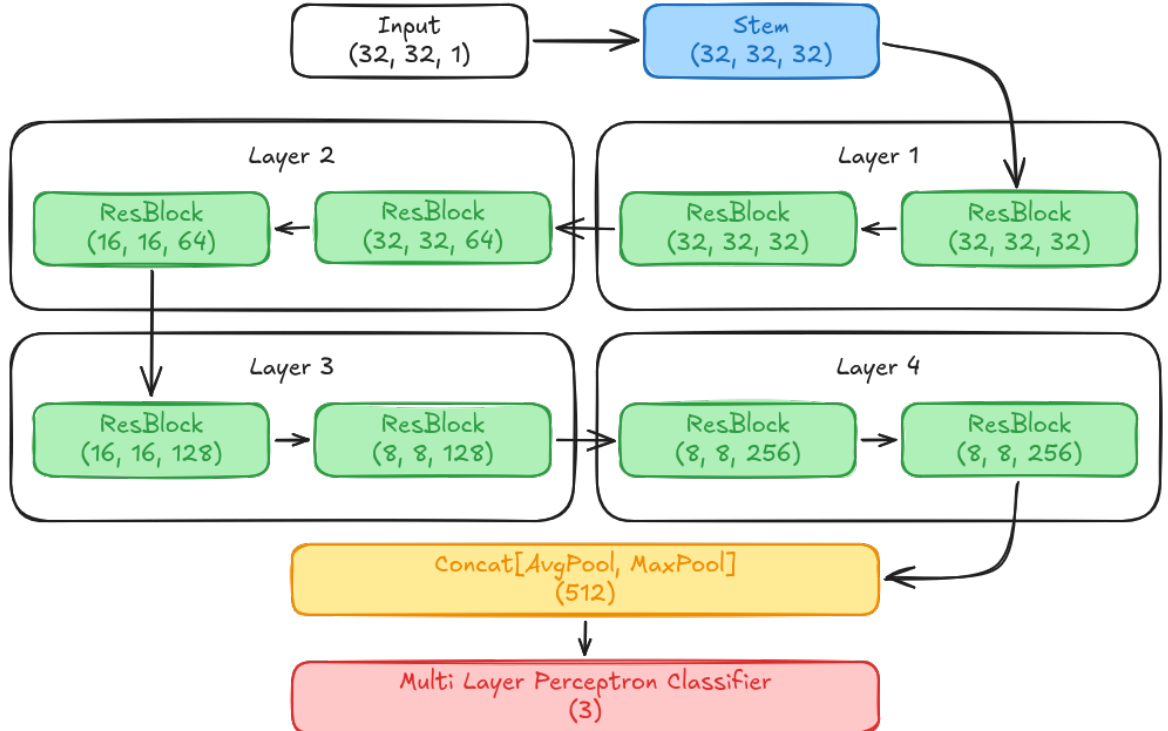
where $\mathfrak{L}(\alpha, \beta)$ is a fully connected perceptron with α input nodes and β output nodes and n_{class} is the number of classes of the classification problem. The used architecture is thus⁴

$$\begin{aligned} \text{TrackCNN} &:= \mathfrak{T}(512, 3) \circ \text{Concat}[\text{AvgPool}, \text{MaxPool}] \circ \\ &\circ [\mathfrak{R}(256, 256, 1) \circ \mathfrak{R}(128, 256, 1)] \circ [\mathfrak{R}(128, 128, 2) \circ \mathfrak{R}(64, 128, 1)] \circ \\ &\circ [\mathfrak{R}(64, 64, 2) \circ \mathfrak{R}(32, 64, 1)] \circ [\mathfrak{R}(32, 32, 1) \circ \mathfrak{R}(32, 32, 1)] \circ \mathfrak{S}(1, 32) \end{aligned} \quad (4.11)$$

where AvgPool and MaxPool extract one number per channel by averaging over it and finding its maximum respectively and Concat is simply concatenation, which consists in appending one list to



(a) Structure of each ResBlock, defined according to the introductory work. Source: [64, Fig. 2]



(b) Structure of TrackCNN, shown is the output shape of each block in the format (H, W, C) .

Figure 29: Visual representation of TrackCNN.

⁴For simplicity's sake, regularisation processes like normalisations and dropouts are omitted.

the end of the other. It is possible to see that this architecture respects the geometrical learning blueprint of Equation 3.45 where global pooling is performed by the concatenation of AvgPool and MaxPool, the non-linearity is the ReLU function, the equivariant layer is $\mathfrak{R}$ and local pooling is performed by the layers with stride 2.

The Loss: Cross Entropy (CE) The neural network was trained using the CE loss. As first stated in Section 3.4, this is simply an alternative name for the logistic loss which comes from information theory. During the training phase, the neural network is shown various images. Consider the case of a camera with 2 visible tracks: its truth label will be $(0, 1, 0)$. The purpose of the loss function is to somehow measure a distance between this true label and the output of the classifier tail $\mathfrak{T}$. It is worth pointing out that the inferred class is calculated by checking which entry of the TrackCNN output has the highest associated value, hence for classification purposes an output of $(0, 1, 0)$ is completely equivalent to, for example, $(0.3, 0.9, 0.1)$. Nevertheless, during the Stochastic Gradient Descent (SGD) procedure, asking the network to provide the perfect output is a much more stringent request than simply asking it to provide a confident answer—defined as an answer in which one entry is much bigger than the others. One method to induce the network to provide confident answers instead of perfect answers is to slightly modify the ground truth with a procedure called *label smoothing* [74, §7]. Let $\varepsilon \in [0, 1]$ be a hyper-parameter, then the ground truth can be modified according to

$$t(k|x) = \delta_{ky} \mapsto t'(k|x) = (1 - \varepsilon)\delta_{ky} + \varepsilon u(k) \quad (4.12)$$

where $k \in [0, n_{\text{class}} - 1]$ indexes the entries of the ground truth vector, δ_{ky} is the Kronecker delta and $u(k)$ a uniform distribution. This variation has the effect of smoothing the gradient profile, enabling simpler convergence towards the loss minimum.

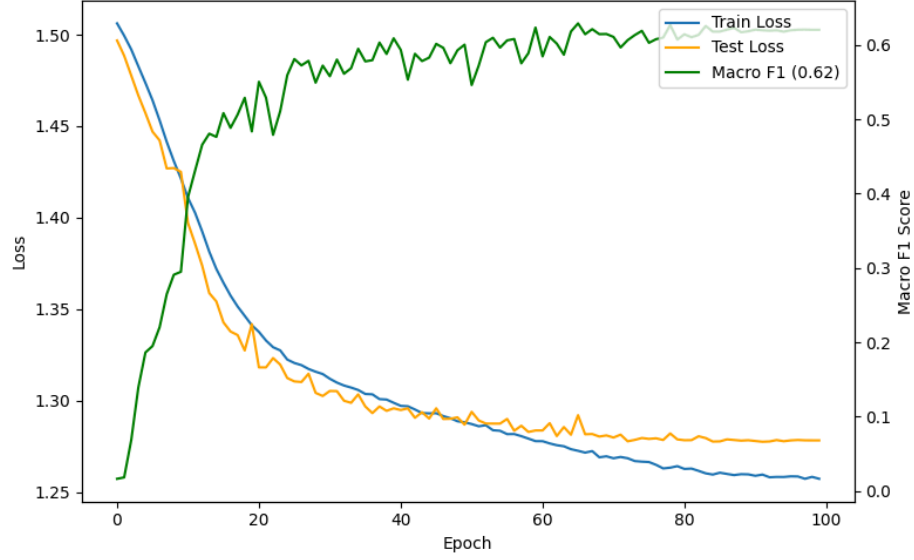
Because of the significant class imbalance shown in Figure 28, the loss was further modified by re-weighting the k -th entry of the loss output vector by $c_k^{-0.7}$ where c_k represents the number of entries in the dataset with truth label k .

The Optimiser: AdamW As anticipated in Section 3.4.1, multiple iterations and improvements on the base SGD algorithm have been developed throughout the years. A well performing version is Adaptive Momentum (Adam) [75]. Adam attempts to improve convergence speed by breaking the Markovian nature of the standard SGD algorithm: the result calculated at step $k + 1$ does not depend only on the gradient calculated at step k , instead the gradients calculated at steps $k - 1$ and $k - 2$, weighed by two hyper-parameters β_1 and β_2 , are used to calculate an optimised *learning rate*.

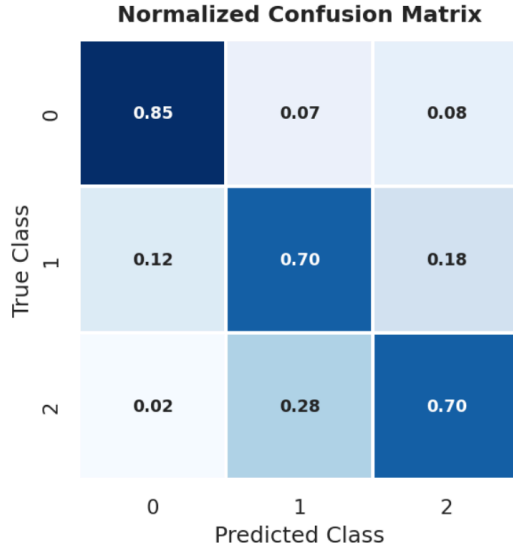
As seen in Section 3.3.1, regularising the minimisation procedure is a wise choice. Regularisation is often implemented either by regularising the model function—also referred to as L_2 regularisation—as done in Equation 3.21, or alternatively with *weight decay*, in which the minimisation step of SGD shown in Equation 3.29 is modified by introducing a hyper-parameter λ

$$\hat{w}_{k+1} = (1 - \lambda)\hat{w}_k - \gamma \nabla \hat{L}_\lambda(\hat{w}_k). \quad (4.13)$$

It can be shown [76, Appendix A] that for basic SGD weight decay by a hyper-parameter λ is equivalent to regularising the model function with hyper-parameter λ/γ . This coupling of the two hyper-parameters causes issues when Adam’s learning rate optimisation is taken into account. Indeed, the article further proves that the equivalence does not hold true for Adam [76, Prop. 2]. The authors thus introduce the AdamW algorithm, a further improvement of Adam for which weight decay and



(a) Train and test losses together with MacroF1 metric.



(b) Test normalised confusion matrix.

Test	Count	Prec.	Rec.	F1
Class 0	99597	0.95	0.85	0.90
Class 1	37596	0.77	0.70	0.73
Class 2	3565	0.15	0.70	0.24
Macro F1 score				0.62
Micro F1 score				0.84

(c) Class-wise metrics: precision, recall and F1

Figure 30: Training results and classification metrics on a class basis. Class 0 refers to cameras with 1 visible track.

learning rate are decoupled. The article further shows that if an Adaptive Momentum algorithm is used, weight decay yields substantially better generalisation performance than L_2 regularisation. Thus, for this work AdamW with weight decay has been used.

§4.3.4 Results

This architecture was trained for 100 epochs on 80 % of the dataset assembled as explained in Section 4.3.2. The remaining 20 % of the dataset was split from the rest in a way that preserves the original class proportions in order to constitute the validation set—also referred to as the test set. Training was performed at the Istituto Nazionale di Fisica Nucleare (INFN) High Performance Computing (HPC) cluster, part of the Centro Nazionale Analisi Fotogrammi (CNAF) computing centre on a single NVIDIA H100 [77] GPU.

All tracked training metrics are shown in Figure 30. Figure 30a shows the train and test losses as a function of training epoch. The smooth descent of the train loss (blue) is distinctive of good convergence towards a minimum. It is evident that the test loss (yellow) closely shows the same trend as the train loss (blue). This indicates that the network is indeed generalising (good learning) instead of memorising features—a phenomenon in jargon known as over-fitting. Indeed the plateauing behaviour of both the test and train loss after epoch 80 can be interpreted as the network reaching the maximum possible learning capability of its architecture.

The confusion matrix (Figure 30b) represents a different way of evaluating training quality. Each matrix element c_{ij} is equal to the fraction of cameras with true label i that were classified as j . A perfect classifier would have a diagonal confusion matrix.

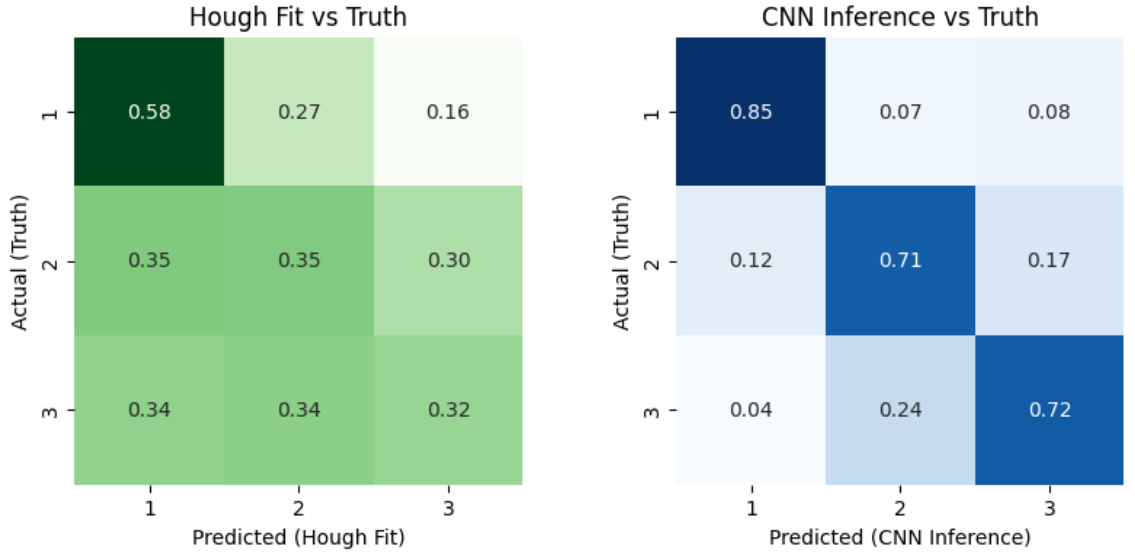
Table 30c shows some per-class classification metrics. Precision—also referred to as purity—is defined as the total number of cameras correctly classified as j divided by the number of cameras classified as j . Recall—also referred to as completeness—is defined as the number of cameras correctly classified as j divided by the number of cameras with true label j . The F1 score is defined as the harmonic mean of the two

$$F1 = 2 \frac{\text{prec} \cdot \text{rec}}{\text{prec} + \text{rec}} \quad (4.14)$$

and it allows the calculation of two global metrics: the Macro F1 score, defined as the average F1 score, and the Micro F1 score, defined as the average F1 score weighed on each class count.

From this table it is possible to see the main drawbacks of learning from an unbalanced dataset. Indeed heavily re-weighting the loss towards Class 2 (three visible tracks) introduces some bias towards it. This fact can be deduced by the low precision of Class 2: the network correctly predicts a camera image should be a part of this class only 15 % of the time. However, considering the fairly good recalls reached throughout, which are by definition also the diagonal entries of the confusion matrix [Figure 30b], it is fair to say that this result is satisfactory for the purpose of this thesis: indeed, it is clearly shown by the data that ML can boost GRAIN lens-based reconstruction in a significant way, even though more time and resources must be devolved to the cause if this path is to be pursued seriously.

Comparison with Hough transform Comparison with the standard reconstruction [Section 4.2] was performed by selecting another set of a thousand ν_μ CCQE events, applying the 3 % active pixels cut to remove empty cameras and running both the CNN inference and the Hough fit on the remaining ones. Figure 31b shows how running TrackCNN on a completely independent dataset from the one used for training [Figure 30b] produces a very similar confusion matrix and metrics, further proving that indeed the architecture learned and did not memorise features. The comparison between the two algorithms [Figure 31] requires little comment: for the task of estimating the amount of visible tracks in each GRAIN camera, TrackCNN performs better than the standard reconstruction *across all tracked metrics*. The largest improvement regards the classification efficiency of camera images for which the F1 score quadrupled. Additionally, the fact that both Macro F1, which gives equal weight to each class, and Micro F1, which is more influenced by the majority classes, approximately doubled suggests that the re-weighting accounted for class imbalance without excessively favouring the minority classes. If the re-weighting had been too aggressive, a substantially larger improvement in Macro F1 relative to Micro F1 would have been expected.



(a) Number of tracks visible per camera; Hough fit confusion matrix and statistics.

(b) Number of tracks visible per camera; CNN confusion matrix and statistics.

Hough	Count	Prec.	Rec.	F1
1 Track	8364	0.82	0.47	0.60
2 Tracks	3277	0.29	0.24	0.26
3 Tracks	296	0.03	0.22	0.06
Macro F1 score				0.31
Micro F1 score				0.45

CNN	Count	Prec.	Rec.	F1
1 Track	8364	0.94	0.85	0.90
2 Tracks	3277	0.77	0.71	0.74
3 Tracks	296	0.15	0.72	0.24
Macro F1 score				0.63
Micro F1 score				0.81

Figure 31: Confusion matrices and per-class classification performance of the Hough fit (left, green) and TrackCNN (right, blue)

§4.4 Next steps

There are two main ways in which this work could be expanded: exploiting the new truth definition in order to optimise GRAIN's camera layout and performing further reconstruction studies. Some improvements could indeed be gained by increasing the total amount of cameras, modifying their position or even their optical structure. Indeed, for example, it could be possible to modify the camera to use two lenses, an avenue that would require extensive simulations and testing. An initial study has already been performed with the objective of checking how the cameras' reconstruction potential depends on their position [Section 4.4.1]. A different path for improvements could come from further development and optimisation of the reconstruction framework. To this extent, three different paths have been taken into consideration for the present study: 2-dimensional reconstruction improvements, a 3-dimensional reformulation, or the adoption of a ML suite.

§4.4.1 GRAIN layout optimisation

The introduction of an angular threshold in the truth projection to 2D led to asking whether certain specific GRAIN cameras required angular merging more often than others. Using the same dataset as Section 4.3.4.1, it was possible to track two different camera image statistics: the percentage of non-empty camera image occurrences corresponding to each camera and the percentage of track merge occurrences corresponding to each camera. The superimposed histograms of these quantities

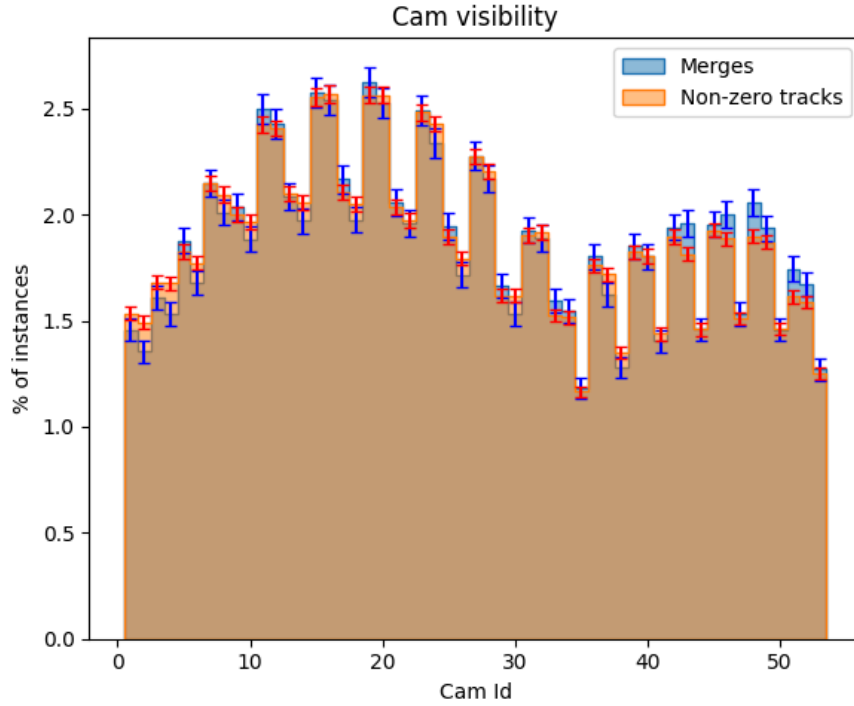


Figure 32: Histogram showing angular merging frequency per camera (blue) compared to frequency of camera usefulness (orange).

are shown in Figure 32. Since both quantities are counts, an underlying Poisson distribution was assumed and hence each bin was assigned an error equal to its square root. Both histograms were then normalised by setting their integral to 100 in order to visually compare them. The resulting saw-like profile is simply a consequence of the camera placement: particle interactions from a beam are boosted alongside its main axis, which corresponds to the z axis in GRAIN. Thus, cameras located at a lower z coordinate can only image tracks from events happening in the first half of the detector, whereas those located further downstream can image both the events happening in the second half and the longer particle tracks from events happening upstream. The overlap between the two histograms, i.e. the possibility that they originate from the same distribution, was tested by measuring the Spearman correlation factor at 0.97 with a p-value of $3 \cdot 10^{-33}$. This test result statistically implies that no single camera has worse angular resolution and that, instead, cameras that require angular merging more often are simply the ones that see tracks more often. The two-mounds structure implies that, as reasonably expected, cameras directly on the beam axis have a higher imaging rate compared to those slightly off-axis. Nevertheless, these two distributions also clearly show that in case the possibility of further increasing the amount of cameras in GRAIN were to present itself, then it would be wiser to place them on the side walls, instead of in the top or bottom, since cameras on the endcaps image tracks more often.

§4.4.2 2-dimensional improvements

One initial step towards a direct improvement could be to address the shortcomings highlighted by the present work, the main one being dataset imbalance. Indeed the comparatively small support of 3-tracks cameras makes learning Class 2 very difficult. Dataset rebalancing can usually be done in three ways: loss re-weighting (the approach used in this work), data loader weighing (which consists in

showing the network images representative of rare classes more often than those with larger support⁵), and statistics boosting by simulating more data for rarer classes. The three methods do not necessarily need to be exclusive as a combination of either or all could prove to be the best option. This should improve metrics across the board: a larger support would naturally boost learning of Class 2 and also allow a smaller loss re-weighting which could, in turn, not penalise Class 0 and Class 1 in a significant way. Additionally, some amount of improvement could be achieved by performing a proper cross-validation of all architecture hyper-parameters (e.g. number of layers, kernel sizes and strides, batch size, loss and optimiser parameters like smoothing, learning rate, and weight decay).

Another unrelated improvement could stem from changing the Hough fit to utilise the Randomised Hough Transform [68]. This modification of the Hough transform offers various benefits: namely the use of only a randomly sampled subset of pixels which boosts computational performance, greater efficiency in noisy environments, and increased accuracy due to its ability to work in an infinitely-dimensional accumulator.

§4.4.3 3-dimensional rework

It could be possible to shift the scope of this computer vision step and thus to use information from all 53 GRAIN cameras simultaneously. Indeed if the 2-dimensional fitting of particle tracks were to reach desirable accuracies, it could then be possible to complete the spatial reconstruction in GRAIN by matching the tracks fitted in different cameras and then back-projecting them to 3-dimensional space. As explained in Section 4.1, the projection of a 3-dimensional track on a 2-dimensional plane intrinsically causes some information loss which can be retrieved by imaging the same track with multiple cameras. One trivial method to implement this could be to stack the output from all cameras as different channels, modifying the stem to $\mathfrak{S}(53, n_{\text{out}})$. A different, unrelated, method could be to use all cameras to produce a 3-dimensional voxelised view of the event [66, §5.2.2] and then using the Hough transform to directly fit 3-dimensional lines [78].

§4.4.4 Adoption of a ML suite

As explained in Section 3.4, developing an ad-hoc ML algorithm for some specific task is a large endeavour which requires spending a lot of time and resources on architecture and hyper-parameter optimisation. Nevertheless, as shown by the present work, ML-based algorithms can strongly boost reconstruction performance in neutrino physics experiments, thus the larger community has been developing specialised tools with this specific objective. Here, three such tools will be briefly exposed: NuGraph [79], SPINE [80], and Pandora [81]. All these tools were developed to work in LArTPCs, but their adaptation to GRAIN could prove fruitful.

NuGraph NuGraph [79], as the name suggests, is based on the Graph Neural Network (GNN) technology. Each signal point creates a node in a graph. Said nodes are then connected and allowed to communicate in the message-passing phase. This has the effect of encoding the event information into some higher-dimensional space which is used as the input to various *decoders*, sub-modules with the objective of inferring some quality about each single node, e.g. whether it is signal or noise, or some Particle IDentification (PID) label. For example, in the ICARUS experiment, NuGraph is used to boost reconstruction of electromagnetic showers: each particle hit is classified as one of five classes,

⁵This path was initially tested, but abandoned down the line because it requires heavy usage of image transformations that proved computationally intensive for the single GPU used.

one of which is “shower-like”⁶.

SPINE SPINE [80] is a different ML suite which uses CNNs for feature extraction and GNNs for superstructure formation. The objective of SPINE is to produce an end-to-end full event reconstruction using exclusively ML algorithms. This suite has been successfully used to produce the latest physics analysis of ICARUS [26] and is currently being adapted to work on ND-LAr [83].

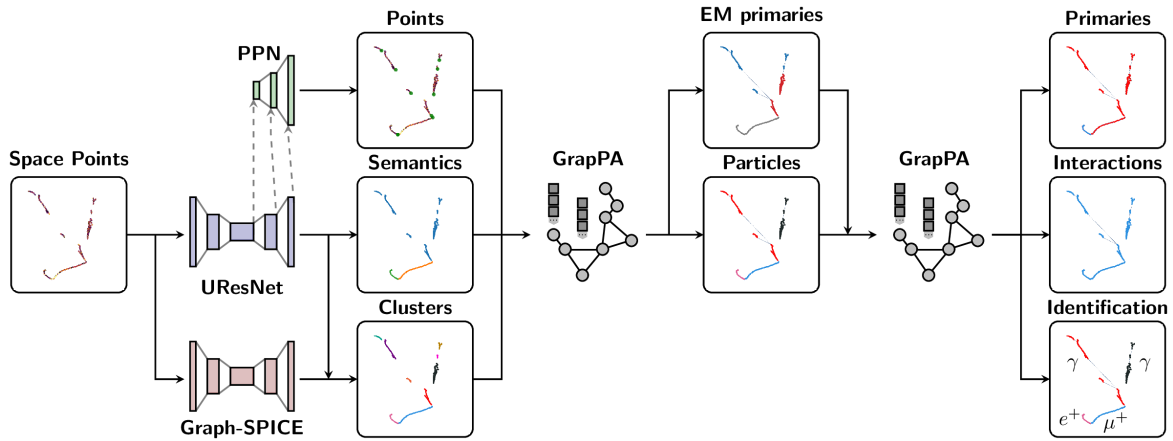


Figure 33: SPINE full end-to-end chain. Source: [80]

Pandora Pandora uses a more modular approach: by progressively selecting many tens of algorithms, it aims to carefully build up a picture of the event hierarchy providing a robust and automated pattern-recognition solution. Despite originally focusing on standard algorithms, the Pandora team has recently started implementing ML-based reconstruction algorithms to, for example, improve vertex reconstruction in the DUNE FD [84]. Pandora is currently used in both experiments belonging to the SBN program [26].



⁶Improving NuGraph’s implementation in the ICARUS reconstruction framework was the objective of my 9 weeks long Summer School at Fermilab [82].

Conclusions and outlook

The Deep Underground Neutrino Experiment is an upcoming neutrino oscillation experiment [Section 2] that aims to answer some open questions about particle physics, namely the neutrino mass ordering [Section 1.4] and whether and to what extent the CP symmetry is broken in the lepton sector [Section 1.3.2]. To achieve these goals, systematic errors must be reduced to a minimum, thus the employment of a Near Detector facility is planned [Section 2.4]. SAND acts as the experiment’s primary beam monitor; its main physics goal is to accurately measure the neutrino beam flux and energy while also reducing model uncertainties regarding neutrino cross sections thanks to its solid hydrogen target [Section 2.5]. The presence of a LAr target inside SAND, called GRAIN, will further allow the direct measure of $\nu - \text{Ar}$ cross sections, bringing an additional reduction in atomic model dependence [Section 2.6]. GRAIN will also be instrumented with novel imaging devices using a SiPM matrix to collect the light and anticipated by lenses (focus of this manuscript) or coded aperture masks with the objective of providing a full topological and calorimetric reconstruction of neutrino interactions in Liquid Argon by measuring and imaging scintillation light [Section 1.7.2] produced by charged particles in the medium. To this extent, a standard reconstruction algorithm based on the Hough transform was developed and used by the collaboration thus far for the lens-based detector imaging setup [Section 4.2]. This work aims at proving that event reconstruction in the lens-based design of GRAIN can be strongly improved by implementing Machine Learning-based algorithms [Section 4.3].

Within this context I developed a Convolutional Neural Network [Section 3.6], called TrackCNN [Section 4.3.3], in order to regress the amount of particle tracks visible in each GRAIN camera. To achieve this result I projected the Monte Carlo truth to the light detector planes for every camera [Section 4.3.1]. Furthermore, the Hough transform-based reconstruction used a definition of “track visibility” that was unsatisfactory for the present work, which I modified by also considering the observed angular separation of tracks [Section 4.3.1.2]. Extensive work was performed in finding outliers, understanding the reason behind their wrong labelling and dataset assembly [Section 4.3.2].

I trained TrackCNN on a sample of over 140k simulated camera images and assigned same task, i.e. extrapolating the number of visible particle tracks in almost 12k GRAIN cameras, to both the CNN and the standard reconstruction algorithm. I then measured the performance of each algorithm at this task by comparing the inferred amount of visible tracks with the new Monte Carlo truth. TrackCNN performs better than the Hough transform-based reconstruction on all tracked metrics [Section 4.3.4.1], almost doubling the average correct classification rate (from 41% to 76%).

The development of TrackCNN provides a quantitative example of what can be achieved by properly adapting to the physics case and exploiting these new technologies. Nevertheless, it is important to stress that TrackCNN was developed reasonably, not rigorously, that these results are preliminary, yet promising, and that studies on network optimisation, including but not limited to a rigorous cross-validation step, could further solidify this novel approach to event reconstruction in GRAIN.

Extensive work is on the horizon, with the final detector layout of GRAIN planned to be frozen by the end of 2028. From a hardware point of view, this could mean an increase in the amount of GRAIN cameras and thus a study on the best locations to place them (as anticipated in Section 4.4.1) or a modification of the imaging technology itself, for example by using two lenses instead of just one. From a reconstruction point of view, the present effort could continue either by further developing the 2-dimensional reconstruction pipeline [Section 4.4.2], by modifying GRAIN’s reconstruction pipeline

to use 3-dimensional information coming from all cameras simultaneously [Section 4.4.3], by adopting of a more solid and developed Machine Learning neutrino reconstruction suite [Section 4.4.4], or through any combination thereof.



Glossary

AI . Artificial Intelligence	37, 38
APA . Anode Plane Assembly	25
Adam . Adaptive Momentum	62, 63
BNB . Booster Neutrino Beam	15
BSM . Beyond the Standard Model	16, 23, 28
CC . Charged Current	9, 12, 23, 31, 56
CCQE . Charged Current Quasi Elastic	56–59, 64
CE . Cross Entropy	42, 62
CERN . Conseil Européen pour la Recherche Nucléaire	4
CNAF . Centro Nazionale Analisi Fotogrammi	63
CNN . Convolutional Neural Network	ix, 37, 48, 49, 56, 57, 68, 69
CP . Charge-Parity	6, 21
CPA . Cathode Plane Assembly	25
CPV . CP-violation	v, 6, 15, 20, 21
CUORE . Cryogenic Underground Observatory for Rare Events	16
CUPID . Cuore Upgrade with Particle IDentification	16
DIS . Deep Inelastic Scattering	12, 56
DUNE . Deep Underground Neutrino Experiment	v, vi, viii, ix, 1, 5, 6, 8, 9, 12–15, 19–25, 27–31, 33, 35, 56, 57, 68, 69
ECAL . Electromagnetic CALorimeter	13, 29, 31, 32
FD . Far Detector	viii, 19, 21, 23–27, 29, 32, 68
FHC . Forward Horn Current	20, 24, 25, 54, 57
FWHM . Full Width Half Maximum	14
Fermilab . Fermi National Accelerator Laboratory	v, viii, 5, 15, 19, 23
FoV . Field of View	53
GNN . Graph Neural Network	47, 67, 68
GRAIN . GRanular Argon for Interactions of Neutrinos	v, vii–ix, 12, 29, 32–35, 37, 51, 53–58, 64–67, 69
HD . Horizontal Drift	25–27
HPC . High Performance Computing	63
HPGe . High Purity Germanium	16
Hyper-K . Hyper-Kamiokande	6, 14, 15, 23
ICARUS . Imaging Cosmic And Rare Underground Signals	13, 15, 25, 67, 68
INFN . Istituto Nazionale di Fisica Nucleare	63
J-PARC . Japan Proton Accelerator Research Complex	15
JUNO . Jiangmen Underground Neutrino Observatory	15, 23
KATRIN . KARlsruhe TRItium Neutrino	8, 16, 17
KLOE . K-LONG Experiment	29, 32
LAr . Liquid Argon	viii, 13, 14, 16, 24, 26, 27, 29, 32–35, 53, 57, 59, 69
LArTPC . Liquid Argon Time Projection Chamber	viii, 13, 14, 24, 25, 27, 29, 67
LBNF . Long Baseline Neutrino Facility	1, 12, 19, 20, 23, 25, 57
LEGEND . Large Enriched Germanium Experiment for Neutrinoless $\beta\beta$ Decay	16

LLM . Large Language Model	38, 48
LNGS . Laboratori Nazionali del Gran Sasso	16
LinAc . Linear Accelerator	23
MINERvA . Main Injector Experiment to study ν -A	12
ML . Machine Learning	vii–ix, 37, 38, 40–43, 47, 48, 51, 54–56, 64, 65, 67–70
MLP . Multi Layer Perceptron	46–48
MO . mass ordering	viii, 12, 22
MSW . Mikheyev-Smirnov-Wolfenstein	8, 12, 21, 22
NC . Neutral Current	9, 10, 12
ND . Near Detector	viii, ix, 12, 19, 25, 27–30, 32, 69
ND-GAr . Near Detector — Gaseous Argon	29
ND-LAr . Near Detector—Liquid Argon	13, 29, 32, 68
NN . Neural Network	37, 45
NOvA . NuMI Off-Axis ν_e Appearance	5, 15, 16
NuMI . Neutrinos at the Main Injector	15
PCB . Printed Circuit Board	27
PID . Particle IDentification	14, 32, 67
PIP-II . Proton Improvement Plan II	24
PMNS . Pontecorvo-Maki-Nakagawa-Sakata	vi, 4, 5, 19, 20
PMT . photo-multiplier tube	14, 16, 32
POT . Protons On Target	24
PRISM . Precision Reaction Independent Spectrum Measurement	viii, 28, 29
RHC . Reverse Horn Current	20, 24, 25
ROI . Region Of Interest	16, 17
ReLU . Rectified Linear Unit	48, 60
SAND . System for on-Axis Neutrino Detection	v, viii, 28, 30, 29, 31, 32, 56, 57, 69
SBN . Short Baseline Neutrino	5, 15, 68
SBND . Short Baseline Near Detector	15, 25
SGD . Stochastic Gradient Descent	vii, 42, 43, 45, 46, 62
SM . Standard Model	v, viii, 2–4, 6, 7, 23, 28
SNO . Sudbury Neutrino Observatory	8
STT . Straw Tube Tracker	31
SURF . Sanford Underground Research Facility	v, viii, 19, 24
SVM . Support Vector Machine	vi, 42, 43
SiPM . silicon photomultiplier	14, 26, 34, 35, 53, 57, 69
T2K . Tokai to Kamioka	5, 14–16
TMS . Temporary Muon Spectrometer	29
TPC . Time Projection Chamber	13
ToF . Time of Flight	32
UV . Ultra Violet	14, 33
VD . Vertical Drift	25, 26
VEV . Void Expectation Value	7
VUV . Void Ultra Violet	14, 34, 57

Bibliography

- [1] Super-Kamiokande Collaboration et al., Evidence for Oscillation of Atmospheric Neutrinos, *Physical Review Letters* **81**, 1562 (1998).
- [2] W. Pauli, Dear radioactive ladies and gentlemen, *Phys. Today* **31N9**, 27 (1978).
- [3] J. Chadwick, The intensity distribution in the magnetic spectrum of β particles from radium (B + C), *Verh. Phys. Gesell.* **16**, 383 (1914).
- [4] F. Takahashi et al., Review of Particle Physics, *International Journal of Modern Physics a* **41**, 2630011 (2026).
- [5] A. Abbott, The singing neutrino Nobel laureate who nearly bombed Nevada, *Nature* **593**, 334 (2021).
- [6] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Detection of the free neutrino: A Confirmation, *Science* **124**, 103 (1956).
- [7] E. Fermi, Tentativo di una Teoria Dei Raggi β , *Il Nuovo Cimento* (1924-1942) **11**, 1 (1934).
- [8] E. C. G. Sudarshan and R. E. Marshak, Chirality Invariance and the Universal Fermi Interaction, *Physical Review* **109**, 1860 (1958).
- [9] R. P. Feynman and M. Gell-Mann, Theory of the Fermi Interaction, *Physical Review* **109**, 193 (1958).
- [10] C. Giunti and C. W. Kim, *Fundamentals of Neutrino Physics and Astrophysics* (Oxford University Press, 2007).
- [11] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Experimental Test of Parity Conservation in Beta Decay, *Physical Review* **105**, 1413 (1957).
- [12] M. Goldhaber, L. Grodzins, and A. W. Sunyar, Helicity of Neutrinos, *Physical Review* **109**, 1015 (1958).
- [13] M. Gell-Mann and A. Pais, Behavior of Neutral Particles under Charge Conjugation, *Physical Review* **97**, 1387 (1955).
- [14] ALEPH Collaboration et al., Determination of the number of light neutrino species, *Physics Letters B* **231**, 519 (1989).
- [15] M. Thomson, *Modern Particle Physics* (Cambridge University Press, 2013).
- [16] D. S. Parno, *Overview of Neutrino-Mass Determination*, <https://doi.org/10.48550/arXiv.2601.01672>.
- [17] SNO Collaboration et al., Measurement of Day and Night Neutrino Energy Spectra at SNO and Constraints on Neutrino Mixing Parameters, *Physical Review Letters* **89**, 11302 (2002).
- [18] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, J. P. Pinheiro, and T. Schwetz, NuFit-6.0: updated global analysis of three-flavor neutrino oscillations, *Journal of High Energy Physics* **2024**, 216 (2024).
- [19] J. A. Formaggio and G. P. Zeller, From eV to EeV: Neutrino cross sections across energy scales, *Reviews of Modern Physics* **84**, 1307 (2012).
- [20] S. Dolan, L. Pickering, P. Stowell, C. Wilkinson, and C. Wret, *CP-Violation or Nuclear Excitation: Reviewing the Role of Neutrino Interaction Model Uncertainties on Accelerator-Based Neutrino Oscillation Measurements*, <https://doi.org/10.48550/arXiv.2605.28671>.
- [21] D. Ruterbories et al., *Comparisons of Triple-Differential Cross Sections for Quasielastic-Like ν_μ -Hydrocarbon Interactions Using $E_\nu \sim 3$ GeV Versus ~ 6 GeV Beams in MINERvA*, <https://doi.org/10.48550/arXiv.2606.00745>.
- [22] DUNE Collaboration, *Deep Underground Neutrino Experiment (DUNE), Near Detector Preliminary Design Report, Volume III - SAND*, (unpublished).
- [23] W. M. Bonivento and F. Terranova, The science and technology of liquid argon detectors, *Reviews of Modern Physics* **96**, 45001 (2024).
- [24] C. Rubbia, The Liquid Argon Time Projection Chamber: A New Concept for Neutrino Detectors, (1977).

- [25] Y. Li, C. Bromberg, M. Diwan, S. Kettell, S. Martynenko, X. Qian, V. Paolone, J. Stewart, C. Thorn, and C. Zhang, Parameterization of Electron Attachment Rate Constants for Common Impurities in LArTPC Detectors, *Journal of Instrumentation* **17**, T11007 (2022).
- [26] ICARUS Collaboration, *First Search for Sterile Neutrino Oscillation Leading to ν_μ Disappearance in the Booster Neutrino Beam at ICARUS*, <https://doi.org/10.48550/arXiv.2603.22557>.
- [27] E. Segreto, Properties of Liquid Argon Scintillation Light Emission, *Physical Review D* **103**, 43001 (2021).
- [28] Hyper-Kamiokande Collaboration et al., *Hyper-Kamiokande: Neutrino Astrophysics and Status*, in *Proceedings of 39th International Cosmic Ray Conference — PoS(ICRC2025)*, Vol. 501 (SISSA Medialab, 2025), p. 1215.
- [29] Hyper-Kamiokande Collaboration et al., Sensitivity of the Hyper-Kamiokande experiment to neutrino oscillation parameters using accelerator neutrinos, *The European Physical Journal C* **86**, 170 (2026).
- [30] DUNE Collaboration et al., *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE*, <https://doi.org/10.48550/arXiv.2002.02967>.
- [31] A. Cabrera et al., Synergies and prospects for early resolution of the neutrino mass ordering, *Scientific Reports* **12**, 5393 (2022).
- [32] JUNO Collaboration et al., Measurement of reactor neutrino oscillation with the first JUNO data, *Nature* **654**, 343 (2026).
- [33] K. N. Abazajian et al., *Light Sterile Neutrinos: A White Paper*, <https://doi.org/10.48550/arXiv.1204.5379>.
- [34] The NOvA Collaboration and The T2K Collaboration, Joint neutrino oscillation analysis from the T2K and NOvA experiments, *Nature* **646**, 818 (2025).
- [35] M. Agostini, G. Benato, J. A. Detwiler, J. Menéndez, and F. Vissani, Toward the discovery of matter creation with neutrinoless $\beta\beta$ decay, *Reviews of Modern Physics* **95**, 25002 (2023).
- [36] CUPID Collaboration et al., CUPID, the Cuore upgrade with particle identification, *The European Physical Journal C* **85**, 737 (2025).
- [37] CUPID Collaboration et al., *Sensitivity of the CUPID Experiment to $0\nu\beta\beta$ Decay of ^{100}Mo* , <https://doi.org/10.48550/arXiv.2504.14369>.
- [38] LEGEND Collaboration et al., First Results on the Search for Lepton Number Violating Neutrinoless Double Beta Decay with the LEGEND-200 Experiment, *Physical Review Letters* **136**, 22701 (2026).
- [39] *LEGEND-1000 | Legend*, <https://legend-exp.org/science/legend-pathway/legend-1000>.
- [40] KATRIN Collaboration et al., Direct neutrino-mass measurement based on 259 days of KATRIN data, *Science* **388**, 180 (2025).
- [41] G. C. Branco and M. N. Rebelo, Building the full PMNS Matrix from six independent Majorana-type phases, *Physical Review D* **79**, 13001 (2009).
- [42] M. Bass et al., Baseline optimization for the measurement of CP violation, mass hierarchy, and θ_{23} octant in a long-baseline neutrino oscillation experiment, *Physical Review D* **91**, 52015 (2015).
- [43] B. Abi et al., Long-baseline neutrino oscillation physics potential of the DUNE experiment, *The European Physical Journal C* **80**, 978 (2020).
- [44] DUNE Collaboration et al., Supernova Neutrino Burst Detection with the Deep Underground Neutrino Experiment, *The European Physical Journal C* **81**, 423 (2021).
- [45] L. Zambelli, Hadroproduction experiments to constrain accelerator-based neutrino fluxes, *Journal of Physics: Conference Series* **888**, 12021 (2017).
- [46] DUNE Collaboration et al., *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics*, <https://doi.org/10.48550/arXiv.2002.03005>.

- [47] the DUNE Collaboration et al., *Operation and Performance of ProtoDUNE Dual Phase Liquid Argon Time Projection Chamber*, <https://doi.org/10.48550/arXiv.2607.15927>.
- [48] DUNE Collaboration et al., *The DUNE Far Detector Vertical Drift Technology, Technical Design Report*, <https://doi.org/10.48550/arXiv.2312.03130>.
- [49] C. Hasnip, DUNE-PRISM - a New Method to Measure Neutrino Oscillations, <http://purl.org/dc/dcmitype/Text>, 2023.
- [50] DUNE Collaboration et al., *Deep Underground Neutrino Experiment (DUNE) near Detector Conceptual Design Report*, <https://doi.org/10.48550/arXiv.2103.13910>.
- [51] M. Adinolfi et al., The KLOE electromagnetic calorimeter, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **482**, 364 (2002).
- [52] H. Duyang, B. Guo, S. R. Mishra, and R. Petti, A precise determination of (anti)neutrino fluxes with (anti)neutrino-hydrogen interactions, Physics Letters B **795**, 424 (2019).
- [53] A. Aloisio and others, The KLOE detector: technical proposal, (1993).
- [54] F. L. Polleri, Test and Characterization of the First Prototype of a Lens-Based Optical Detector for Imaging Charged Particle Tracks in the DUNE Experiment, Master's thesis, 2025.
- [55] W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics **5**, 115 (1943).
- [56] A. L. Samuel, Some Studies in Machine Learning Using the Game of Checkers, IBM Journal of Research and Development **3**, 210 (1959).
- [57] M. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry* (The MIT Press, 2017).
- [58] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, Nature **323**, 533 (1986).
- [59] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).
- [60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention Is All You Need*, <https://doi.org/10.48550/arXiv.1706.03762>.
- [61] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, New York, NY, 2009).
- [62] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems **2**, 303 (1989).
- [63] M. M. Bronstein, J. Bruna, T. Cohen, and P. Velićković, *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges* (2021).
- [64] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition, (2015).
- [65] C. Gerhard, *Lens Design Basics: Optical Design Problem-Solving in Theory and Practice* (IOP Publishing, 2020).
- [66] M. Vicenzi, A GRAIN of SAND for DUNE: Development of Simulations and Reconstruction Algorithms for the Liquid Argon Target of the SAND Detector in DUNE, PhD Thesis, 2023.
- [67] R. O. Duda and P. E. Hart, Use of the Hough transformation to detect lines and curves in pictures, Communications of the ACM **15**, 11 (1972).
- [68] L. Xu and E. Oja, Randomized Hough Transform (RHT): Basic Mechanisms, Algorithms, and Computational Complexities, CVGIP: Image Understanding **57**, 131 (1993).
- [69] 如何用霍夫变换算法实现直线检测, <https://www.infoq.cn/article/3KfCkf67yZQHuiFp0x3Q>.
- [70] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN, ACM Trans. Database Syst. **42**, (2017).
- [71] C. Andreopoulos and others, The GENIE Neutrino Monte Carlo Generator, Nucl. Instrum. Meth. A **614**, 87 (2010).

- [72] C. McGrew, *ClarkMcGrew/edep-Sim*, <https://github.com/ClarkMcGrew/edep-sim>.
- [73] 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).
- [74] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*, <https://doi.org/10.48550/arXiv.1512.00567>.
- [75] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, <https://doi.org/10.48550/arXiv.1412.6980>.
- [76] I. Loshchilov and F. Hutter, *Decoupled Weight Decay Regularization*, <https://doi.org/10.48550/arXiv.1711.05101>.
- [77] *NVIDIA H100 GPU Whitepaper*, <https://resources.nvidia.com/en-us-hopper-architecture/nvidia-h100-tensor-c>.
- [78] C. Dalitz, T. Schramke, and M. Jeltsch, Iterative Hough Transform for Line Detection in 3D Point Clouds, *Image Processing on Line* **7**, 184 (2017).
- [79] V. Hewes, A. Aurisano, G. Cerati, J. Kowalkowski, C. Lee, W.-keng Liao, D. Grzenda, K. Gumpula, and X. Zhang, Graph Neural Network for Neutrino Physics Event Reconstruction, *Physical Review D* **110**, 32008 (2024).
- [80] *DeepLearnPhysics/spine*, <https://github.com/DeepLearnPhysics/spine>.
- [81] M. collaboration et al., *The Pandora Multi-Algorithm Approach to Automated Pattern Recognition of Cosmic-Ray Muon and Neutrino Events in the MicroBooNE Detector*, <https://doi.org/10.48550/arXiv.1708.03135>.
- [82] L. Lena, A. Fava, and G. Cerati, *Improving Electromagnetic Shower Identification in ICARUS Using Machine Learning Algorithms*, <https://indico.fnal.gov/event/71834/>.
- [83] F. Drielsma, *SPINE: Progress and Plans*, <https://indico.fnal.gov/event/71860/contributions/349727/>.
- [84] D. Collaboration et al., Neutrino Interaction Vertex Reconstruction in DUNE with Pandora Deep Learning, *The European Physical Journal C* **85**, 697 (2025).