

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Scienze

Dipartimento di Fisica "Giuseppe Occhialini"

Corso di Laurea Magistrale in Fisica



Feature Extraction Neural Networks for Quantum Kernel
classifiers: low-energy background rejection in
Xenon-Doped LArTPC detectors

Master thesis

Roberto Moretti
Matricola 825617

Supervisor:

Dr. Andrea Giachero

Co-supervisor:

Dr. Daniele Guffanti

External supervisor:

Dr. Michele Grossi

Academic Year: 2021-2022

Abstract

In the last decade, the development of Machine Learning (ML) and Deep Learning (DL), and their application in several branches of physics contributed significantly to the advancements of the latter. One of the most innovative proposals in this field is Quantum Machine Learning (QML), an interplay between Machine Learning and Quantum Computing (QC), aiming to achieve an edge over traditional ML through the paradigm of quantum information. The main goal of the present thesis was to develop and characterise QML models for solving the task of supervised classification, a recurring problem in particle physics. The supervised classification consists of tagging elements of a vector space according to the value of their components, called *features*. An example of a classification algorithm is the Support Vector Machine (SVM), which is based on mapping the features into a vector space of greater dimension than the initial one, in which to define a separation hyperplane which optimizes the distance between elements of different classes and the hyperplane itself. A crucial aspect of SVM that affects the model's performance is the choice of the feature map. The latter occurs implicitly, by defining a positive, semidefinite scalar function of the features, called *kernel*. The SVM's kernel determines its expressivity, which is the capability to model the non-linear relations between variables that are useful for classification. Introducing Quantum Computing gives access to a new class of kernels, called Quantum Kernels, that are intractable with classic techniques. The SVMs that classify data using Quantum Kernels are called Quantum SVMs (QSVMs) and are examples of algorithms that are executable on modern NISQ devices (Noise Intermediate Scale Quantum devices).

The present thesis explores the potentials of QSVM within the framework of the Deep Underground Neutrino Experiment (DUNE) for the analysis of the neutrinoless double beta decay ($0\nu\beta\beta$) of the ^{136}Xe isotope. The observation of this hypothetical decay would be the first proof of lepton number violation and would confirm the Majorana theory on the equivalence between neutrinos and anti-neutrinos. DUNE is a Long-Baseline experiment designed for the measurement of neutrino flavour oscillation, the search for proton decay and the detection of supernovae and solar neutrinos. The DUNE Far Detector facility (FD) will be made of four modules, each one consisting of a 17 kton Liquid Argon Time Projection Chambers (LArTPCs), a class of detectors capable of reconstructing the track of ionizing particles in addition to their energy deposition. The $0\nu\beta\beta$ analysis proposal consists of employing a xenon-doped FD module, enriched in the 136 isotope, and searching for the decay at the expected energy peak, i.e. the Q-value of the reaction $Q_{\beta\beta}^{^{136}\text{Xe}} = 2.458 \text{ MeV}$. To increase the sensitivity of the experiment, it is necessary to implement background mitigation strategies. The latter is dominated by β emissions from the ^{42}Ar decay chain and neutron capture.

This thesis investigates the possibility of carrying out background mitigation by deve-

loping ML and QML classifiers, able to exploit the track information produced by the TPC. This goal is challenging and complicated by the sub-optimal tracking efficiency of the DUNE LArTPCs at the few-MeV energy scale, offering the opportunity to test quantum algorithms where the classical ones underperform. A necessary step to implement QSVM on limited-size quantum devices was to reduce the track information by going through a feature extraction process. This was performed by developing two Deep Learning models: a Convolutional Neural Network (CNN) and a Transformer, or Attention Neural Network (ATN). The first one achieves the task by making use of convolutional filters, while the second one is based on the innovative mechanism of self-attention. Both models were used as feature extractors for QSVM and as independent classifiers. Once the goal of feature extraction was reached, the QSVMs were trained with Quantum Kernels, varying the number of qubits and the corresponding quantum circuits. The Quantum Kernels were first chosen arbitrarily and later optimized with a meta-heuristic genetic algorithm. The data used for the training of CNN, ATN and QSVM consisted of $0\nu\beta\beta$ and β background tracks, reconstructed by an ideal LArTPC simulated using the Geant4 Monte Carlo simulation toolkit. To complete the study, the CNN was also tested for mitigating radiological background in one of the DUNE planned FD modules, called Horizontal Drift.

This work provides interesting results from QML and DL points of view as well as from a physics perspective, assessing DUNE’s potential for the search of the $0\nu\beta\beta$ decay of the 136-Xe isotope. QSVM classifiers have been implemented on IBM quantum hardware and have proven to be competitive, achieving classification accuracies equivalent to the ones of the best classic SVM of common use, based on a Gaussian Kernel: 66% and 64% with features extracted from the CNN and ATN, respectively. Further studies were made, varying parameters affecting the trace reconstruction, showing that ATN outperforms the CNN in case of high track resolution, while on the contrary, CNN prevails at low resolutions. The topological classification accuracy of $0\nu\beta\beta$ events and the main radiological background in the DUNE FD Horizontal Drift detector settles at 63% with the use of the CNN. This result adds a benchmark for DUNE’s potential in the search for the $0\nu\beta\beta$ decay and on the topological background mitigation for other low-energy channels proposed for DUNE, such as solar neutrinos and WIMPs.

The thesis activities have been carried out during an International Mobility period with the Erasmus Traineeship programme, at the *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas* (CIEMAT) of Madrid (ES), in collaboration with the CIEMAT Neutrino Physics working group and the CERN Quantum Technology Initiative group (CERN QTI).

Sommario

Nell'ultimo decennio, lo sviluppo ed il crescente utilizzo di algoritmi di Machine Learning (ML) e Deep Learning (DL) in numerose aree della fisica hanno dato forti contributi al loro avanzamento, tanto da costituirne oggi un elemento fondamentale. Tra le proposte più recenti in questo ambito si annovera il Quantum Machine Learning (QML), un insieme di strategie basate sull'utilizzo congiunto di Machine Learning e Quantum Computing (QC), con lo scopo di trarre vantaggio dal paradigma dell'informazione quantistica. La presente tesi pone come obiettivo principale lo sviluppo e la caratterizzazione di modelli di QML con applicazioni in fisica delle particelle, in particolare la classificazione supervisionata. Quest'ultima consiste nel categorizzare elementi di uno spazio vettoriale in base al valore delle loro componenti, dette *features*. Un esempio di algoritmo di classificazione è la Support Vector Machine (SVM), la quale si basa sul mappare le features di partenza in uno spazio vettoriale di dimensione maggiore di quello iniziale, in cui definire un iperpiano di separazione in grado di ottimizzare distanza che intercorre tra elementi di classi differenti e l'iperpiano stesso. Un aspetto cruciale della SVM, che ne determina la sua efficacia, è la scelta della funzione di mappatura, o *feature map*. Quest'ultima avviene in modo implicito, tramite la definizione di una funzione scalare semidefinita positiva delle features, detta *kernel*. Il kernel della SVM determina la sua espressività, ovvero la capacità di modellare relazioni non lineari tra le variabili ai fini della classificazione. In questo contesto, l'utilizzo di un processore quantistico permette di accedere in modo naturale ad una nuova classe di kernel, detti Quantum Kernels, intrattabili con tecniche classiche. Le SVM che classificano dati mediante Quantum Kernel sono denominate Quantum SVM (QSVM), e sono esempi di algoritmi implementabili su dispositivi quantistici attualmente esistenti, denominati NISQ devices (Noise Intermediate Scale Quantum devices).

Nel presente lavoro di tesi vengono esplorate le potenzialità della QSVM nell'ambito del Deep Underground Neutrino Experiment (DUNE), nell'analisi del decadimento doppio beta senza emissione di neutrini ($0\nu\beta\beta$) dell'isotopo ^{136}Xe . L'osservazione di questo ipotetico processo proverebbe per la prima volta la violazione del numero leptonico e confermerebbe la teoria di Majorana sull'equivalenza di neutrino ed anti-neutrino. DUNE è un esperimento di tipo Long-Baseline progettato per lo studio dell'oscillazione di sapore del neutrino, la ricerca del decadimento del protone e la rivelazione di neutrini di origine solare e da supernova. Il Far Detector (FD) di DUNE sarà costituito da quattro moduli Liquid Argon Time Projection Chambers (LArTPCs), ciascuno contenente 17 kton di argon liquido, in grado di ricostruire la traccia di particelle ionizzanti e misurarne l'energia. La proposta di analisi $0\nu\beta\beta$ consiste nel drogare uno dei moduli FD con xeno arricchito nell'isotopo 136 e ricercare il decadimento al picco atteso, pari al suo Q-valore $Q_{\beta\beta}^{^{136}\text{Xe}} = 2.458 \text{ MeV}$.

Per aumentare la sensitività dell'esperimento è necessario attuare strategie di riduzione

del fondo, dominato dall’emissione β proveniente dalla catena del 42-Ar e da cattura neutronica. La presente tesi esamina la possibilità di effettuare tale mitigazione costruendo classificatori ML e QML basati sull’informazione prodotta dalla TPC. Questo obiettivo risulta ambizioso, complicato dalla limitata efficienza di tracciamento di elettroni da pochi MeV nelle LArTPC di DUNE, e costituisce un banco di prova per lo sviluppo di algoritmi quantistici dove la controparte classica risulta poco performante. Un passaggio necessario per implementare la QSVM su dispositivi quantistici di dimensione limitata è la riduzione del contenuto informativo prodotto dalla TPC, tramite un processo di estrazione delle features. In questo elaborato sono stati utilizzati due modelli di Deep Learning: la Convolutional Neural Network (CNN) ed il Transformer, o Attention Neural Network (ATN). Il primo approccio è caratterizzato dall’uso di filtri convoluzionali, mentre il secondo si basa sull’innovativo meccanismo della *self-attention*. Entrambi i modelli sono stati impiegati sia come ausilio alla QSVM che come classificatori indipendenti. Una volta raggiunto l’obiettivo dell’estrazione di features, la QSVM è stata addestrata con diverse tipologie di Quantum Kernel, al variare del numero di qubit ed in funzione di diversi circuiti quantistici. I Quantum Kernel, dapprima selezionati con criteri arbitrari, sono poi stati ottimizzati tramite un algoritmo meta-euristico di tipo genetico. I dati utilizzati per addestrare i modelli CNN, ATN e QSVM corrispondono a tracce ricostruite da una LArTPC ideale, a partire da eventi $0\nu\beta\beta$ e fondo β simulati mediante l’uso del *toolkit* di simulazione Monte Carlo Geant4. A completamento dello studio, la CNN è stata adoperata per la classificazione di eventi $0\nu\beta\beta$ e fondo radiologico da una simulazione del FD Horizontal Drift, uno dei moduli pianificati per DUNE.

Il presente lavoro di tesi ha prodotto risultati di interesse nell’ambito del Quantum Machine Learning e nel Deep Learning, ed ha contribuito ad asserire le potenzialità dell’esperimento DUNE nella ricerca del $0\nu\beta\beta$ dell’isotopo 136-Xe . I classificatori quantistici QSVM sono stati implementati su hardware IBM e si sono dimostrati competitivi, ottenendo un’accuratezza di classificazione equivalente a quella del miglior SVM classico di uso comune, basato su kernel gaussiano: 66% e 64% con feature estratte rispettivamente da CNN e ATN. Ulteriori studi, condotti variando parametri che incidono sulla ricostruzione delle tracce, dimostrano che le performance della ATN in condizioni ad alta risoluzione superano quelle della CNN, mentre nel caso contrario si verifica un’inversione di tendenza. L’accuratezza nella classificazione topologica di eventi $0\nu\beta\beta$ e dei principali fondi radiologici nel detector DUNE FD Horizontal Drift si assesta al 63% con l’uso della CNN. Tale risultato costituisce un parametro di riferimento sulle potenzialità di DUNE nella ricerca del decadimento $0\nu\beta\beta$ e sulle capacità di mitigazione del fondo in altri canali di basse energie proposti per DUNE, come neutrini solari e la ricerca di WIMPs.

Le attività di tesi sono state svolte durante un periodo di Mobilità Internazionale tramite il programma Erasmus Traineeship, presso il *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas* (CIEMAT) di Madrid (ES), in collaborazione con il Neutrino Physics working group di CIEMAT ed il gruppo di CERN Quantum Technology Initiative (CERN QTI).

Contents

1 Quantum Computing and Machine Learning	7
1.1 Qubits and Quantum Information	7
1.1.1 Single and multi-qubit states	8
1.1.2 Quantum circuits	9
1.2 Quantum hardware	11
1.3 The problem of feature reduction	13
1.4 Quantum Machine Learning and QSVM	18
1.4.1 Support Vector Machine	18
1.4.2 Quantum Kernel	21
1.4.3 Pauli feature maps	23
1.4.4 Quantum Kernel Alignment	23
1.5 Automatization of quantum circuits design	25
2 $0\nu\beta\beta$ decay in xenon-doped LAr-TPCs	28
2.1 Standard Model and neutrino mass	28
2.1.1 Gauge symmetries and Higgs mechanism	30
2.1.2 Fermion masses and Yukawa coupling	30
2.1.3 Neutrino flavor oscillations	31
2.1.4 The Dirac and Majorana masses	32
2.2 Neutrinoless double beta decay	34
2.3 The DUNE experiment	39
2.3.1 DUNE liquid argon TPCs	39
2.3.2 DUNE sensitivity to ^{136}Xe $0\nu\beta\beta$	41
2.3.3 Topological background mitigation at the MeV scale	47
3 Low-energy background mitigation in LArTPCs	50
3.1 Toy model dataset description	50
3.2 CNN and ATN feature extractor	53
3.2.1 CNN architecture	54
3.2.2 ATN architecture	56
3.2.3 Performances of feature extractor networks	57
3.3 QSVM and the search for Quantum Advantage	59
3.3.1 Heuristic Quantum kernels	60
3.3.2 Genetic Quantum Kernels	68
3.3.3 QSVM test on an IBM quantum processor	74
3.4 Improving the toy model dataset resolution	76
3.4.1 Comparison between CNN and ATN	76

3.4.2	QSVM at 1 mm wire pitch	78
3.5	Background mitigation in the DUNE HD Far Detector	80
3.5.1	The background simulation	80
3.5.2	CNN multi-label and binary classification	82
4	Conclusion	88
A	Classification metrics	90
B	More on arbitrary feature maps	91
C	Genetic feature maps at 1 mm wire pitch	94

Chapter 1

Quantum Computing and Machine Learning

This chapter provides an overview of the fundamentals of Quantum Computing, an emerging field of Information Technology, Physics and Engineering, and it follows towards one of its promising applications in Artificial Intelligence models (AI), i.e. the Quantum Support Vector Machine (QSVM) [1]. The present thesis' goal consists of the development and characterization of QSVM models and their innovative use for rare decay searches in the field of neutrino physics.

1.1 Qubits and Quantum Information

The quantum bit, namely qubit, is the fundamental unit of quantum information [2] and can be considered an extension of the well-affirmed and extensively utilized classical bit concept. Bits provide a simple way to describe and store any possible piece of information. The value of a single bit consists of the logical state of one of two possible outcomes, normally represented as 0 and 1. For this reason, a sequence of n bits describes one of the 2^n possible combinations of 0s and 1s, and arbitrarily long bit-strings can be used to encode a large variety of data types, such as results of an experiment, sentences, numbers, pieces of music, etc. Qubits, instead, encode information into a quantum state, thus inheriting the properties of quantum mechanics. Quantum Computing aims to exploit these features, such as superposition, interference and entanglement, to achieve an edge over traditional classical bit systems. Reaching this goal usually means demonstrating a so-called *quantum advantage* [3], i.e. determining that an existing quantum processor performs better than any classical one for some practical tasks, by setting new benchmarks or achieving speed-ups. Lacking a more rigorous definition, sometimes the term quantum advantage is used as a synonym of *quantum primacy* or as a weaker alternative to the latter. In this second case, "primacy" indicates that a quantum computer is able to solve a problem that classical hardware could never solve given a realistic amount of time or resources. In 2019, Google's Sycamore 53-qubits processor was the first to prove Quantum primacy, by completing in a few minutes a particular *ad-hoc* engineered task that a state-of-the-art supercomputer would take approximately 10^4 years to finish, according to Google's claim [4] (despite a subsequent counter-argument brought by IBM, that lowered the

estimation down to two or three years [5]). Nowadays, addressing new proofs of Quantum advantage is still a fervent area of research, focusing on algorithms that find their applications in several scientific areas.

1.1.1 Single and multi-qubit states

A qubit is a two-level quantum system, i.e. a quantum system that can be decomposed into a linear combination of two orthonormal eigenstates, commonly represented in Dirac notation $|0\rangle$ and $|1\rangle$:

$$|\psi\rangle = a|0\rangle + b|1\rangle \quad (1.1)$$

Where a and b are complex coefficients that satisfy the normalization condition $|a|^2 + |b|^2 = 1$. Quantum observables are not affected by changes in the global phase of the quantum state. For this reason, it is possible to choose a global phase value that allows for a useful parametrization, consisting of the following:

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\theta}{2}\right)|1\rangle \quad (1.2)$$

The two remaining degrees of freedom $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$ allows to map the qubit state onto the surface of a sphere, namely the *Bloch sphere* [6].

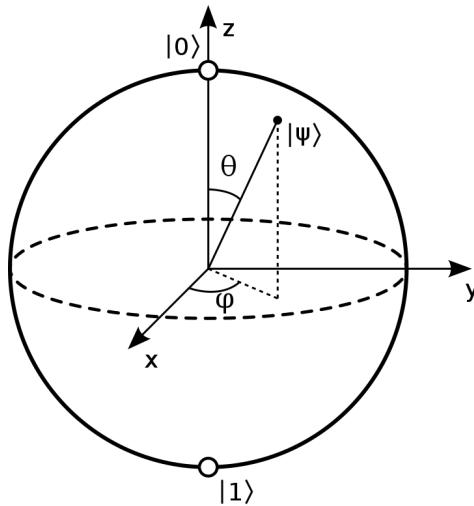


Figure 1.1: Qubit state visualization on the surface of a Bloch sphere.

North and south poles of the Bloch sphere in figure 1.1 represent respectively the physical states $|0\rangle$ and $|1\rangle$ (also called *computational basis*), while on the x and y axis lie other notable states:

$$x : \quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle); \quad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

$$y : \quad |i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle); \quad |-i\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle)$$

The single-qubit Hilbert space \mathcal{H} described so far has dimension two. By adding other qubits to the system, the number of computing bases increases exponentially, and the

overall Hilbert space becomes the tensor product of all the single qubit spaces $\mathcal{H} \otimes \mathcal{H} \dots \otimes \mathcal{H}$. As a natural consequence of quantum superposition, a n -qubit wavefunction is able to encode information of all the possible 2^n eigenstates simultaneously, unravelling the huge advantage with respect to classical computing in storing information. This can be easily shown considering n qubits all prepared in the state $|+\rangle$:

$$|\psi\rangle = |+\rangle_0 \otimes |+\rangle_1 \otimes \dots \otimes |+\rangle_n = \quad (1.3)$$

$$= \left(\frac{1}{\sqrt{2}}\right)^n [(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes \dots \otimes (|0\rangle + |1\rangle)] = \quad (1.4)$$

$$= \left(\frac{1}{\sqrt{2}}\right)^n \sum_{j \in S} |j\rangle \quad (1.5)$$

Where S is the set of all the 2^n strings of length n of 0 and 1, i.e. the computational bases of the new Hilbert space. Due to this exponential growth, the quantum state of a relatively small qubit-system (*e.g.* $n = 60$) will be described by an enormous amount of complex amplitudes (2^{60}) that could never be stored in any conceivable classical computer [7].

A limitation in quantum computing is that despite the large amount of information encoded in the quantum states, it cannot be easily extracted from the system. The process of measurement, being intrinsically destructive, will give as a result only one of the possible physical states, with a probability given by the corresponding amplitude. Nevertheless, by repeating the same experiment (for example, preparing over and over the state in equation 1.3 and measuring the outcome every time) it will be possible to estimate the output distribution probability described by the final state $|\psi\rangle$ thus compensating the information loss. In addition, until a measurement occurs, the full quantum state living in the high-dimensional Hilbert space is still accessible for computation. This allows us to explore the possibilities of quantum advantages over the classical paradigm, and design innovative quantum algorithms that operate between quantum states.

1.1.2 Quantum circuits

The elementary operations that allow to manipulate quantum information and carry it through the algorithm execution are called *quantum gates* [8], the counterpart of logic gates on a classical computer. The action of a quantum gate on qubit states can be expressed as a unitary matrix, where states are written in the vector representation $|0\rangle = (1\ 0)^T$ and $|1\rangle = (0\ 1)^T$.

Single-qubit gates act as elements of $U(2)$ (2×2 unitary matrices). The most common examples are listed below.

- X-gate (NOT):

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (1.6)$$

acts as a qubit-flip: $X|0\rangle = |1\rangle$; $X|1\rangle = |0\rangle$

- Hadamard gate:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (1.7)$$

switches from the $|0, 1\rangle$ to the $|+, -\rangle$ base: $H|0\rangle = |+\rangle$; $H|1\rangle = |-\rangle$

- Arbitrary $SU(2)$ and Pauli rotations:

$$U(\theta, \phi, \lambda) = \begin{pmatrix} e^{-i\frac{\phi+\lambda}{2}} \cos(\theta/2) & -e^{-i\frac{\phi-\lambda}{2}} \sin(\theta/2) \\ e^{i\frac{\phi-\lambda}{2}} \sin(\theta/2) & e^{i\frac{\phi+\lambda}{2}} \cos(\theta/2) \end{pmatrix} \quad (1.8)$$

General transformation that can be factored into rotations along the Bloch sphere axis (shown in figure 1.1):

$$U(\theta, \phi, \lambda) = R_z(\phi)R_y(\theta)R_z(\lambda) \quad (1.9)$$

The single rotations take the form of:

$$R_\alpha(\theta) = e^{-i\frac{\theta}{2}\sigma_\alpha} \quad \alpha = x, y, z \quad (1.10)$$

where σ_α indicates the α -Pauli matrix.

It is then possible to compose different single-qubit gates via a tensor product in order to create multi-qubit gates. Such operators act on many qubits simultaneously in such a way as to keep each qubit state evolution independent from the other. In order to harness the power of quantum mechanics for computation, it is necessary to introduce interactions between qubits with new operators that create *entangled* states, like the *CNOT* (*CX*) gate:

$$CX = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (1.11)$$

The *CX* is a conditional gate that acts on two qubits in input and represents an analogy with the classical *XOR* logic operation. The first qubit is said *control* and the second *target*. If the control qubit state is $|1\rangle$, a *X*-gate is applied to the target. For example, $CX|00\rangle = |00\rangle$ while $CX|10\rangle = |11\rangle$. A simple realization of a two-qubit entangled state is the following:

$$|\psi\rangle = CX|+0\rangle = CX\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle\right) = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (1.12)$$

where the $|+\rangle$ state is obtained by applying the *H*-gate on the first qubit prepared in $|0\rangle$. The state $|\psi\rangle$ is entangled (cannot be written as a direct product) and describes one of the so-called *Bell-pairs*.

A routine of finite coherent unitary operations that terminates with an ensemble of measurements can be considered a *quantum algorithm*. A convenient way to render the occurring set of transformations consists of the *quantum circuit* diagram, whose principal elements are *registers* (horizontal lines) that follows the state of each qubit, and quantum gates (blocks), placed on the qubit lines in sequential order. For example, the circuit diagram for the preparation of a Bell-pair is shown in figure 1.2

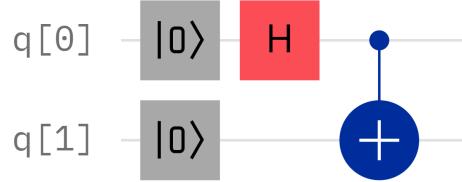


Figure 1.2: Quantum circuit diagram of a Bell state. The two qubits, initialized in the $|00\rangle$ state (grey blocks), undergo the evolution shown above. In quantum circuits, the CX -gate is represented by connecting a dot (control qubit) with a “+” symbol (target qubit). The red block in the first line corresponds to the H -gate, acting on qubit $q[0]$.

Remarkable properties of quantum algorithms are *reversibility* of all operations (inherited by gates unitarity) and the *universality*, i.e. the possibility to construct any arbitrary operation on qubits with a finite amount of gates. A quantum circuit can be seen as a parametric unitary transformation itself $U(\vec{\theta})$ (in general very large), that brings an initial state $|\psi_i\rangle$ to a final one $|\psi_f\rangle$ according to the equation $|\psi_f\rangle = U(\vec{\theta}) |\psi_i\rangle$. In most useful applications, the last step consists of measuring the final state, with the goal of estimating the probability of ending up in a query state $|\phi\rangle$, $P = |\langle\phi|\psi_f\rangle|^2$. This requires running the circuit several times in order to mitigate the intrinsic statistical fluctuations.

1.2 Quantum hardware

Up to now, we have discussed quantum computing formalism and illustrated its potential from a theoretical point of view. Well-known algorithms that provide a mathematically proven quantum advantage require the development of large-scale, fault-tolerant quantum computers to be carried out [9]. Quantum hardware has been showing steady technological progress in the latest years but is still far from providing the desired computing fidelity and power.

Among several kinds of artificial qubits (related to quantum dots, ion traps, optical photons, etc ... [10]) a prominent design for universal quantum computers is based on a particular kind of superconductive circuit, called *transmon qubit* [11], schematized in figure 1.3a. Transmons possess relatively long decay and decoherence time, allowing for fast operations, and they take advantage of the already well-advanced microelectronics and lithography techniques. This qubit type consists of an anharmonic quantum oscillator and is obtained from an LC circuit with the addition of a nonlinear inductance named Josephson junction. Such a system, in the superconductive regime (at temperatures of about few-tens of mK), presents uneven and quantized energy levels, making it possible to isolate a working subspace (consisting of ground and first-excited state), thus identifying a computational basis $|0\rangle$ and $|1\rangle$.

The main limiting factors of modern superconductive quantum computers can be summarized in three points [12]:

- **Noise:** the challenging procedure of qubit calibration, gate implementation and state readout in a cryogenic environment introduce errors that lower the computational reliability.

- **Scalability:** as the number of qubits increases, keeping the system stable and in an appropriate isolation condition from the environment becomes progressively hard.
- **Low connectivity:** while it is mathematically possible to realize entanglement between all qubits of a system, their physical disposition on the chip determines which ones can be put into entanglement (typically the nearest neighbours).

Noise, scalability and connectivity issues are also present, in different forms and importance, in quantum hardware different from the superconductive one.

At present time, quantum devices are in the so-called *Noisy Intermediate-Scale Quantum era* (NISQ-era), and as the name may suggest, they are able to execute quantum circuits with limited size, entanglement and qubit number with small error rates. One of the world-leaders quantum processors is the 127-qubit Eagle r1 developed by IBM [13], whose connectivity map is displayed in figure 1.3b.

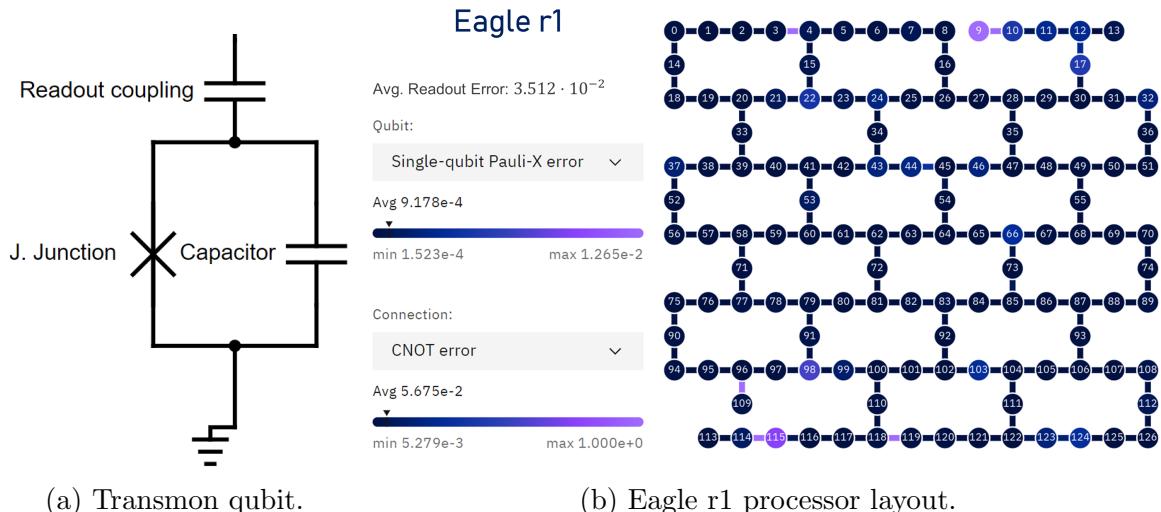


Figure 1.3: (a) superconductive transmon qubit with a Josephson Junction as a nonlinear inductive component. The qubit interacts with the environment through capacitive coupling with a microwave resonator. (b) entangling map showing control, readout and entanglement error rates of the Eagle r1 processor. The 127 qubits are far from being fully connected and this universal quantum computer is an example of a NISQ-era device.

While the experimental advancements toward realizing large-scale and high-fidelity quantum computers may take decades, it is still possible to leverage the already available resources to look for quantum advantage in classically challenging tasks, carried out by the so-called *NISQ algorithms*. Recent proposals span different fields of scientific applications, such as particle physics, machine learning, quantum chemistry and combinatorial optimization [14, 15, 16].

The core of the present thesis work consists of developing and characterizing a specific kind of near-term achievable Quantum Machine Learning classifiers (QML), named *Quantum Support Vector Machines*, and exploring its potentialities in the analysis of experimental data in Fundamental particle physics. More details on the experiment, the framework and the physics goal will be discussed in chapter 2, while the following

section describes the Support Vector Machines in their classical (SVMs) and quantum-enhanced (QSVMs) formulations.

1.3 The problem of feature reduction

For extending the classical SVM illustrated in paragraph 1.4.1 to embed quantum computing, it is necessary to develop strategies for dealing with the limited size of nowadays hardware. As a NISQ algorithm, QSVM should work with a small number of features associated with each sample in the dataset, in agreement with the number of qubits. For this reason, some techniques of feature reduction are required for analysing most datasets, like the Principal Component Analysis (PCA) [17] or more sophisticated Deep Learning approaches. This paragraph reviews two methods implemented in chapter 3, the *Convolutional Neural Network* (CNN) [18] and the *Transformer*, also called *Attention Network* (ATN) [19].

Neural Networks rely upon artificial neurons (computational nodes), shown in figure 1.4, interconnected to each other to collectively learn from an input dataset how to solve a determined task. To assemble the simplest example of a Neural Network (the

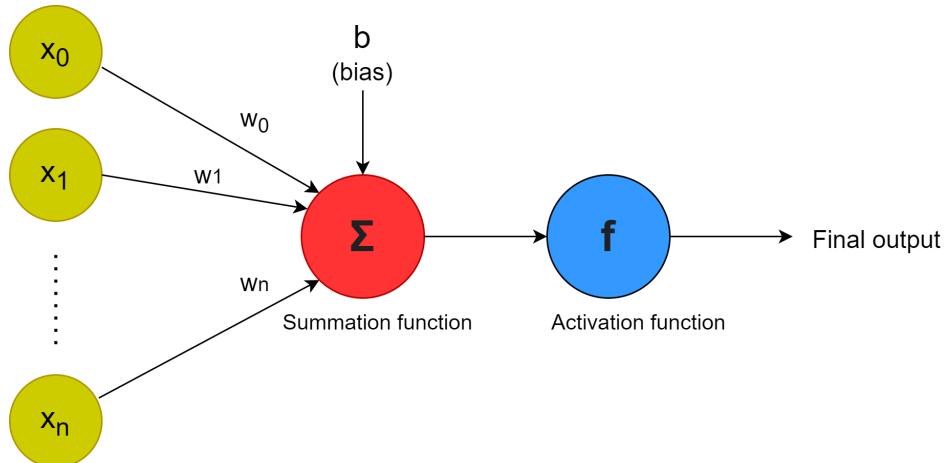


Figure 1.4: Schematic representation of an artificial neuron. The summation function Σ outputs a weighted sum of the inputs ($\sum_i^n w_i x_i + b$), where the weights w and the bias b are the trainable parameters of the model. The non-linear activation function f determines whether a neuron "activates" or not depending on the input, and allows the network to learn effectively over non-linear patterns.

feed forward type), a number of neurons are put together in consecutive *layers*, such that each layer takes as input the output of the previous one. Training a NN consists of adjusting the weights and the biases values in order to match a specific output from a given input. This is carried out through the so-called *forward* and *backward* propagation. The *forward* propagation consists of computing the neuron activations from the NN input towards the end, extracting the output and evaluating the discrepancy between the expected outcome with a *cost function*. In contrast, the *backward* propagation applies a small correction, evaluated through gradient descent, to the network weights starting from the last layer and going back to the first. The training sample is grouped into *batches*, sequentially normalized and sent to a forward-backwards propagation step. The whole procedure iterates such that the same sample flows through

the NN multiple times, each one called *epoch*, ideally until reaching convergence or a loss function stable minimum.

In this thesis, Neural Networks have been used as a powerful tool for reducing the original dataset complexity while keeping as much useful information as possible, in a way to be accessible to a small-sized quantum device. In order to achieve that, CNN and ATN were trained as binary classifiers, able to perform a *standalone* prediction, treating the second-last layer as a *feature layer*, containing a number of neurons equal to the desired feature dimension for SVM and QSVM. The last layer, consisting of a single neuron, determines the class prediction according to its neuron value. The feature layer output will be considered as the new reduced dataset, as it contains all the informative content that the neural network was able to extract from the original input.

Convolutional and Attention Neural Networks architectures present variations from the Feed Forward model, and both possess additional kinds of layers that enhance the classification performances while keeping a relatively small number of neurons and training samples, significantly speeding up the training process.

Convolutional Neural Network

CNNs are common choices in the field of visual imagery. They are characterized by Convolutional and Pooling layers.

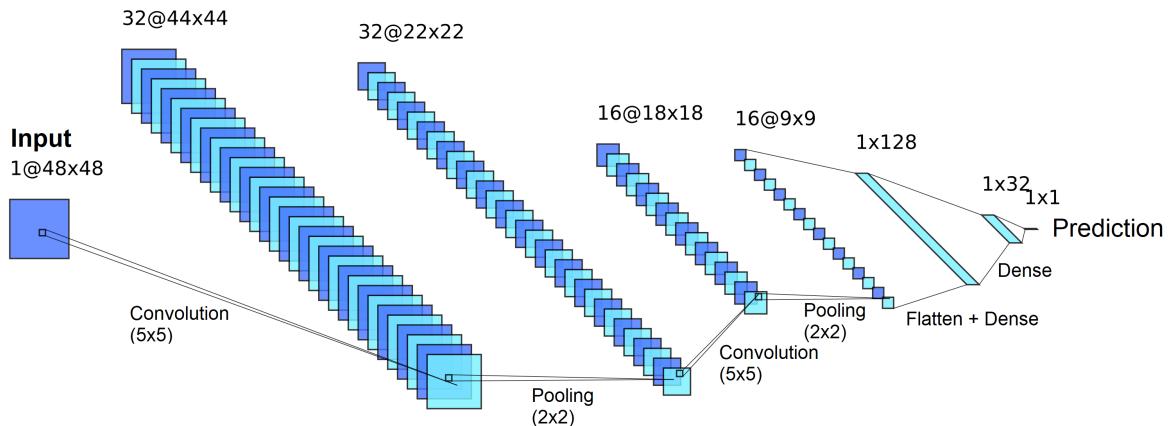


Figure 1.5: Example of CNN architecture with two convolutional and two pooling layers for analyzing images of 48×48 pixel resolution. Activation functions and batch normalization layers are omitted for simplicity.

- **Convolutional layer:** consisting of an ensemble of *filters*, i.e. trainable matrices with a given size, small with respect to the input image. Each filter serves the purpose of activating when encountering a particular local characteristic in the input (edges, colours, gradient orientations, etc..). Putting more convolutional layers in sequence allows for detecting more complex structures in the input image. The convolutional operation is illustrated in figure 1.5.
- **Pooling layer:** reduce the dimension of the tensors by taking as input a cluster of neuron outputs and computing a single value. Typical examples are *max pooling* and *average pooling*, which respectively return the maximum and the

average cluster value, as shown in figure 1.5. A pooling operation reduces the training cost of the network at the price of a partial information loss.

Transformer

Transformers are considered state-of-the-art models in the field of *Natural Language Processing*, surpassing in prediction accuracy *Recurrent* and *Long-Short Term Memory* Neural Networks (RNN, LSTM) [20]. Transformers treat each sample as a sequence of data and rely on *positional encoding* combined with a particular strategy known as *attention mechanism*, responsible for detecting relationships between different parts of the input sample (for example different words inside a phrase).

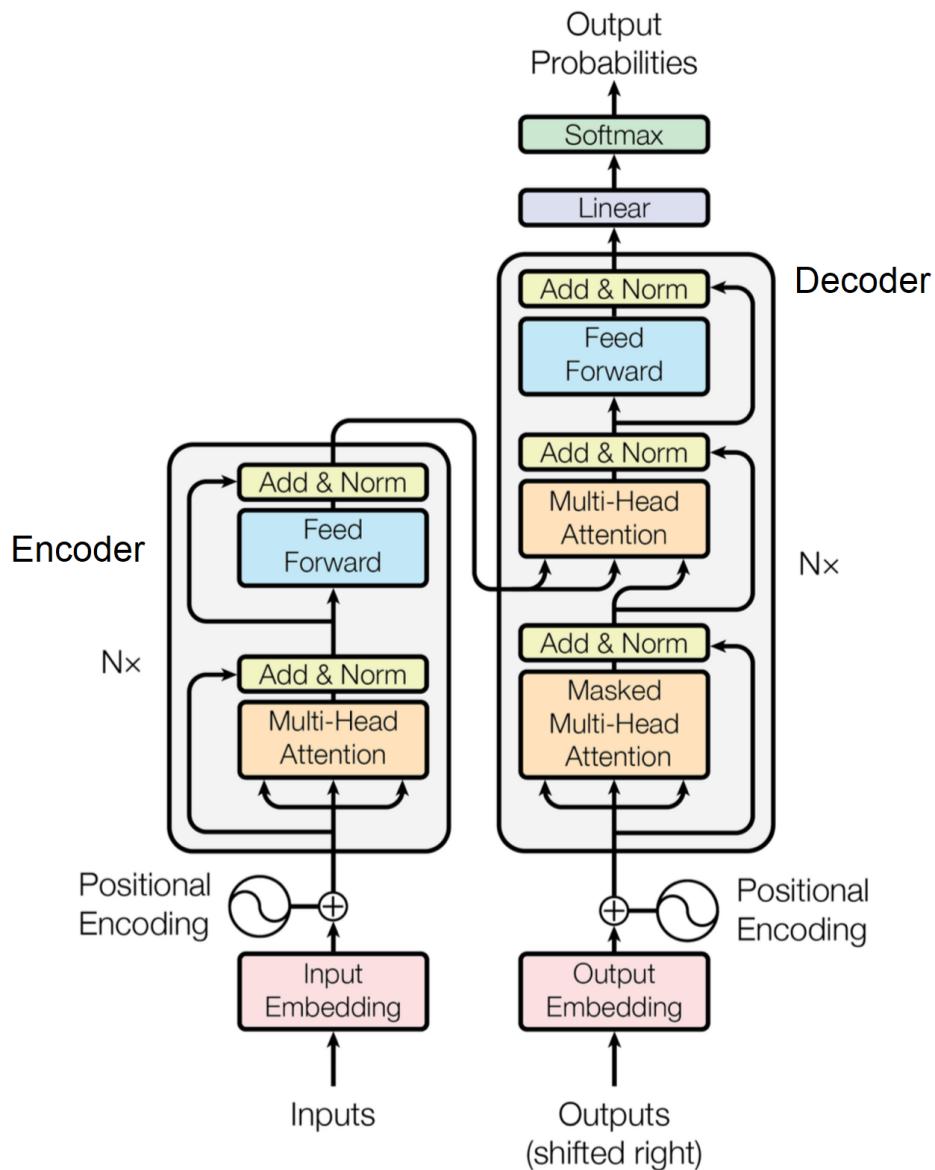


Figure 1.6: General architecture of a Transformer [19]. Many identical encoding and decoding blocks can be stacked sequentially to increase the depth of the Neural Network.

- **Embedding and positional encoding layers:** The embedding layer maps each element of a sequence to a vector of length d_{model} , living in a *embedding space*, while the positional encoding layer depends only on the length of the sequence, and assign to each element another vector of dimension d_{model} representing its relative position inside the input. The output of embedding and encoding layers are both matrices of dimension (*sequence length* \times d_{model}) and are summed together, as shown in figure 1.6
- **Attention layer:** The goal of the attention layer is to apply a weighting for every embedded sequence by assigning an *attention score*, i.e. a pairwise measure of the degree of correlation between all the elements of the original input, that takes into account *context* of the sequence itself (e.g. figure 1.7). In this way, the neural network can process the input sequence focusing only on combinations with a high attention score for predicting the output.
This effect is achieved by introducing three trainable matrices, Q (*Query*), K (*Key*), V (*Value*) combined according to the formula:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_{\text{model}}}} \right) V$$

Instead of training only a single attention function, it is beneficial to work with *multi-head attention*, i.e. several attention layers in parallel with distinct learned queries, keys and values projections into dimensions and concatenate the results as illustrated in figure 1.8.

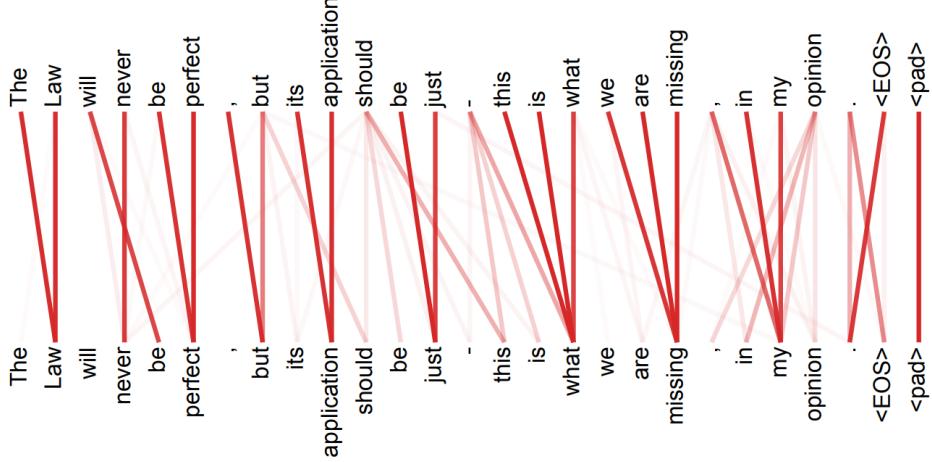


Figure 1.7: Weights from a single-head attention layer [19]. The input sample (the whole phrase) is processed word by word. The different connections' strengths correspond to the context learned by the network.

The output of multi-head attention is followed by normalization, feed-forward and normalization layers. This structure forms blocks called *encoder* and *decoder* that compose the Transformer according to the scheme in figure 1.6.

Differences between CNN and ATN

Convolutional Neural Networks and Transformers are both suitable for the task of classification. However, these models are radically different not only in architecture

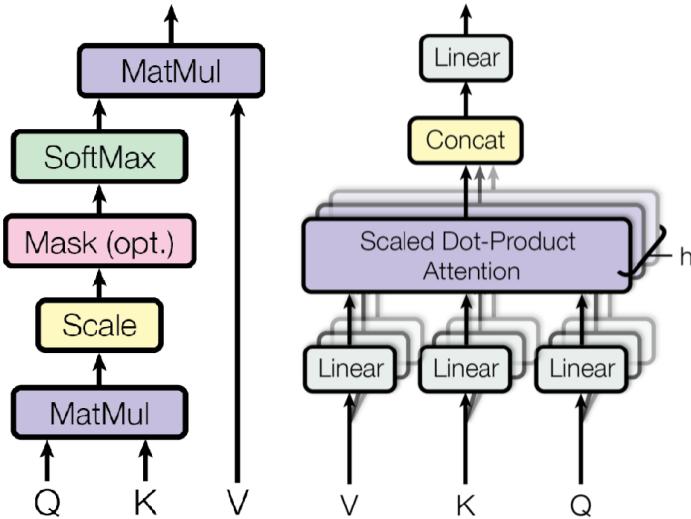


Figure 1.8: Single attention layer (left) and multi-head attention (right) [19], responsible for giving *context* to different elements of the sequential input, enhancing the network performance.

but also in how they deal with input data. While CNN extracts visual features from pictures, ATN gives context to sequential structures. For this reason, it is possible to view CNN and ATN as two different philosophies in data analysis and are expected to behave differently, whether it is convenient to consider the dataset as images or sequences. Some characteristics of both models are reported in table 1.1.

	CNN	ATN
Analysis strategy	Visual.	Sequential.
General structure	Convolutional + Pooling + Dense layers.	Encoder-Decoder Multi-Head Attention + Dense layers.
Locality	Convolutional filters collectively extract information from small portions of the image.	Attention layers analyse the entire sequence simultaneously, suitable for long-range dependencies.
Complexity	Linear with pixel/voxel number.	Quadratic with sequence length.
Adaptability	Affected by the <i>curse of dimensionality</i> : performances depend on image sparsity.	Marginally affected by input dimensionality.

Table 1.1: Qualitative differences between Convolutional Neural Networks and Transformers.

1.4 Quantum Machine Learning and QSVM

Quantum machine learning refers to algorithms that solve some of the typical tasks in classical ML while providing improvements and speedups. Different combinations in the use of quantum computing and ML for the algorithm design are summarized in figure 1.9, depending on whether the input data and/or model have a classical (C) or quantum (Q) nature. Quantum Support Vector Machine falls in both CQ and QQ categories. In this thesis, QSVM will encode classical data to make it available for quantum information processing (CQ), and then re-convert it to a classical output by measuring the quantum system. However, QSVM can also operate directly on intrinsically quantum data, without requiring encoding (QQ) [21].

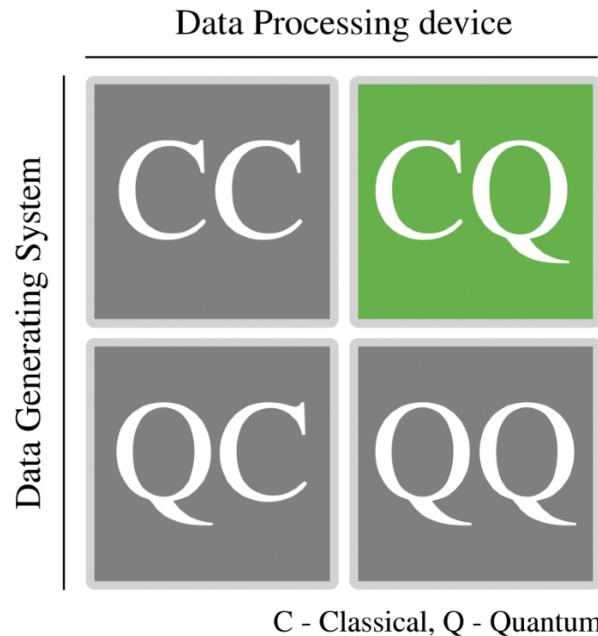


Figure 1.9: Different blending combinations of quantum computing and machine learning into Quantum Machine Learning [22]. The CC class refers to classical machine learning performed by a classical computer, running algorithms that are inspired by the formalism of quantum computing. The CQ set includes algorithms that exploit quantum advantage by solving classical problems with a quantum device (like QSVM). QC refers to the use of classical ML methods to analyse quantum systems, while QQ class analyse quantum data with a quantum computer.

1.4.1 Support Vector Machine

Classical Support Vector Machines [23] are well-known supervised learning algorithms largely employed in classification tasks, but also in regression and outlier detection. The remarkable strengths of SVM models are their effectiveness and computational efficiency when dealing with high-dimensional data and the robustness of their solutions. Moreover, SVMs are able to achieve high performances on different classification problems (binary, multilabel) and solve nontrivial relationships between dataset features. Part of their efficiency and versatility is due to the use of *kernel functions* (SVMs are an example of the so-called *kernel methods*), which will be defined in this section and

explored in detail in section 1.4.2.

As a major drawback, Support Vector Machines may not be the best choice when dealing with large, high-sampled data, as the model complexity grows quadratically with the number of samples and other models may be preferable, such as Boosted Decision Trees (BTDs) or Neural Networks (NN).

For the purpose of this thesis, which revolves around signal vs background discrimination, SVM will be solely treated in the context of a binary classifier. The problem can be formalized as follows:

Given a training dataset $S = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_l, y_l)\}$ of data points $\vec{x}_i \in X \subseteq \mathbb{R}^n$ with corresponding labels $y_i \in Y = -1; 1$, the task of classification consist of learning a function $g : X \rightarrow Y$ that correctly classifies new samples (\vec{x}, y) generated from the same underlying distribution of the training data. \vec{x}_i is the array of *features* of the i -th considered datapoint, whose distribution is unknown a priori.

In order to guarantee good generalization performances (achieving the smallest classification error on an unseen dataset), SVM finds an optimal hyperplane in the feature space that separates the two class distributions. When the training set is linearly separable, this optimal hyperplane is such that its distance from the nearest data points, called *margin*, is maximized.

In case the given dataset is not linearly separable, it is possible to define a nonlinear mapping ϕ from the initial feature space X to a new higher dimensional one. The hyperplane found in the new feature space corresponds to a non-linear decision boundary in the input space, thus achieving non-linear dataset classification. An illustrative example is shown in figure 1.10.

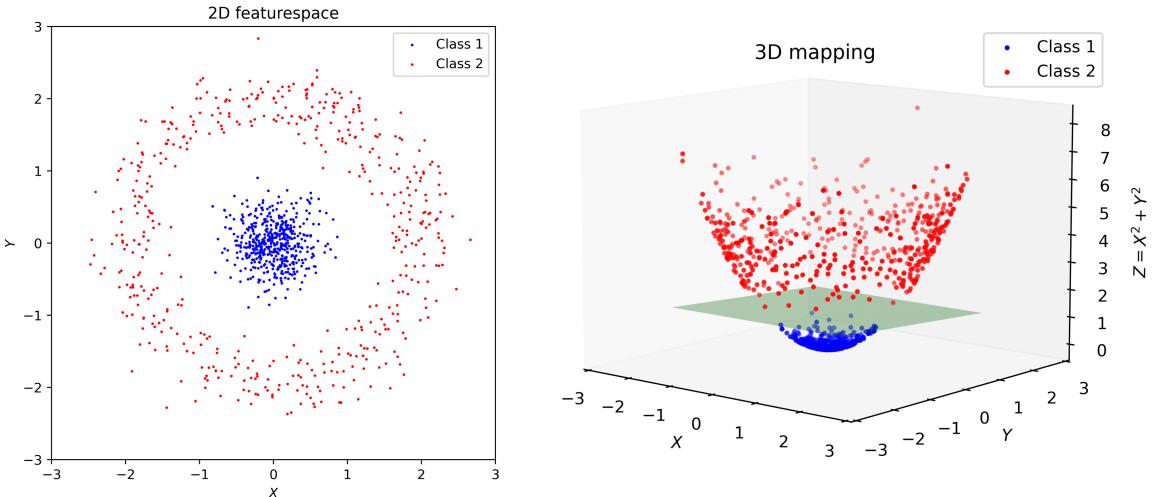


Figure 1.10: Initial non-linearly separable distribution of features (left) and mapping into an embedding space (right) that makes classes separable by a hyperplane with the function $\phi(x, y) = (x, y, x^2 + y^2)$.

Let $\phi : X \rightarrow F$ where F is the new high dimensional dataset. An hyperplane can be defined by the equation: $\vec{w}^T \cdot \phi(\vec{x}) - b = 0$, and its margin as:

$$\Gamma = \min_{1 \leq i \leq n} y_i (\vec{w}^T \cdot \phi(\vec{x}_i) - b) \quad (1.13)$$

Its worth noticing that the distance between a datapoint i and the hyperplane is

$(\vec{w}^T \cdot \phi(\vec{x}_i) - b) / ||\vec{w}||$ and the decision function corresponds to:

$$g(\vec{x}) = \text{sign} (\vec{w}^T \cdot \phi(\vec{x}_i) - b) \quad (1.14)$$

While the sign of $y_i g(\vec{x}_i)$ is positive for correctly classified data points and negative otherwise. The goal is to find the hyperplane that maximizes Γ , solving the problem:

$$\max_{\vec{w}, b} \{\Gamma\} \quad (1.15)$$

subject to $y_i (\vec{w}^T \cdot \phi(\vec{x}_i) - b) \geq \Gamma$, $i = 1, \dots, l$ and imposing $||\vec{w}|| = 1$.

This method is called *hard margin* SVM and implies that the underlying class distributions are separable, i.e. that exists a nonlinear mapping that allows identifying a hyperplane that separates not only the training data points, but also any other samples extracted from the underlying distribution. Most realistic datasets present some degree of class-overlapping, for which the hard margin technique fails (the margin will not exist for a given $\phi(\vec{x})$, or $g(\vec{x})$ will not generalize to unseen data). To deal with these cases, some error in classification must be tolerated in order to obtain a better generalization power: the *soft margin* SVM achieve these results while preserving the same philosophy of margin maximization:

$$\min_{\vec{w}, b, \xi} \left\{ -\Gamma + C \sum_{i=1}^l \xi_i \right\} \quad (1.16)$$

subject to $y_i (\vec{w}^T \cdot \phi(\vec{x}_i) - b) \geq \Gamma - \xi_i$, $i = 1, \dots, l$, $\xi_i \geq 0$ and $||\vec{w}|| = 1$. The constant C is passed as a hyperparameter before training the model and regulates the tradeoff between margin maximization and error tolerance. A low value of C allows for a smooth decision boundary, while a high value aims at classifying the training set correctly. Entries in vector $\vec{\xi}$ are *slack variables*, defined as $\xi_i = \max \{0, \Gamma - y_i \phi(\vec{x}_i) - b\} l^{-1}$, the smallest non-negative number that satisfies $y_i \phi(\vec{x}_i) - b \geq \Gamma - \xi_i$.

Computing the soft margin SVM classifier is equivalent to minimising the problem in equation 1.16. Nevertheless, a more convenient strategy requires to solve the Lagrangian dual of the above formulation [24], in order to end up with an easier and computationally cheaper problem:

$$\text{maximize } \mathcal{L}(\vec{\alpha}) = \sum_{i=1}^l \alpha_i - \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \phi(\vec{x}_i) \cdot \phi(\vec{x}_j) \quad (1.17)$$

subject to $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^l \alpha_i y_i = 0$

α_i are Lagrange multipliers that determine the separating plane equation by $\vec{w} = \sum_{i=1}^l \alpha_i y_i \vec{x}_i$

In equation 1.17 the high dimensional feature maps appear exclusively pairwise as arguments of an inner product, making it possible to solve the problem without explicitly computing $\phi(\vec{x}_i)$ for $i = 1, \dots, n$, requiring only the scalar product between them. It is then convenient to define a *kernel function*:

$$k : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R} \quad (1.18)$$

such that

$$k(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i) \cdot \phi(\vec{x}_j) \quad (1.19)$$

and rewrite equation 1.17 as:

$$\text{maximize } \mathcal{L}(\vec{\alpha}) = \sum_{i=1}^l \alpha_i - \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j k(\vec{x}_i, \vec{x}_j) \quad (1.20)$$

Providing a definition for K as an arbitrary scalar function of two input samples avoids the expensive high-dimensional computation. Also, an explicit equation for the feature map is no more necessary, as the kernel function defines it implicitly. SVMs only differ in the kernel function defined before training, and the different classification performances depend on how the classes are distributed on the implicit high-dimensional mapping.

Kernel functions, being defined as inner products (eq. 1.19), are symmetric ($K(\vec{x}_i, \vec{x}_j) = K(\vec{x}_j, \vec{x}_i)$) and positive-semidefinite, i.e. satisfy:

$$\sum_{i,j=1}^l c_i c_j K(\vec{x}_i, \vec{x}_j) \geq 0 \quad (1.21)$$

for any $c_i, c_j \in \mathbf{R}$, guaranteeing that the lagrangian in equation 1.20 has a local maximum.

Popular kernel functions are the following:

- **Linear:** $\vec{x}_i \cdot \vec{x}_j$
- **Polynomial:** $(\gamma \vec{x}_i \cdot \vec{x}_j + r)^d$
- **Gaussian:** $\exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2)$
- **Sigmoid:** $\tanh(\gamma \vec{x}_i \cdot \vec{x}_j + r)$

where γ , r and d are fixed hyperparameters, together with the penalty parameter C that constrains the lagrangian multipliers.

Training an SVM model by defining a kernel function K instead of a feature map ϕ even allows for the implicit implementation of infinite dimensional feature maps. For example, in the gaussian case for $n = 1$:

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2} = e^{-\|x_i\|^2} e^{-\|x_j\|^2} \sum_{k=0}^{+\infty} \frac{2^k x_i^k x_j^k}{k!} \quad (1.22)$$

that can be factored as a vector ϕ with infinite entries, written as Taylor expansions.

1.4.2 Quantum Kernel

After achieving an efficient reduction of the dataset feature number, a problem already addressed in section 1.3, it is finally possible to leverage few-qubits quantum processors in Machine Learning.

Integrating NISQ era QML in Support Vector Machines relies on a novel quantum definition of the kernel function [1] and its computation is carried out by running a quantum circuit. The proper training of the SVM decision function is still classical, therefore the QSVM is considered a hybrid model that requires cooperation between

both quantum and classical processors. As a side note, an alternative to the QSVM shown in this section exists for fault-tolerant quantum computers, where exponential speedup is achieved with respect to the classical SVM training [25, 26].

The idea behind the definition of Quantum Kernel function consists of promoting the high-dimensional feature map $\phi(\vec{x})$ into a quantum state $|\phi(\vec{x})\rangle\langle\phi(\vec{x})|$ living in the Hilbert space $\mathcal{H} = \mathbf{C}^{2^n \times 2^n}$. This state can be considered as an initial state $|0^n\rangle\langle 0^n|$ where some unitary evolution $U(\vec{x})$ is applied, called *quantum feature map*:

$$|\phi(\vec{x})\rangle\langle\phi(\vec{x})| = U(\vec{x})|0^n\rangle\langle 0^n|U(\vec{x})^\dagger \quad (1.23)$$

The quantum kernel function is hence defined as the square module of the inner product between quantum feature maps, i.e. the Hilbert-Schmidt product:

$$K(\vec{x}_i, \vec{x}_j) = |\langle\phi(\vec{x}_i)|\phi(\vec{x}_j)\rangle|^2 = |\langle 0^n|U(\vec{x}_i)^\dagger U(\vec{x}_j)|0^n\rangle|^2 \quad (1.24)$$

Equation 1.24 can be interpreted as the probability that an initial state $|0^n\rangle$, that undergoes the evolution described by $U(\vec{x}_i)^\dagger U(\vec{x}_j)$ will result again $|0^n\rangle$ after the measurement of all qubit states. The quantum circuit representation follows in figure 1.11.

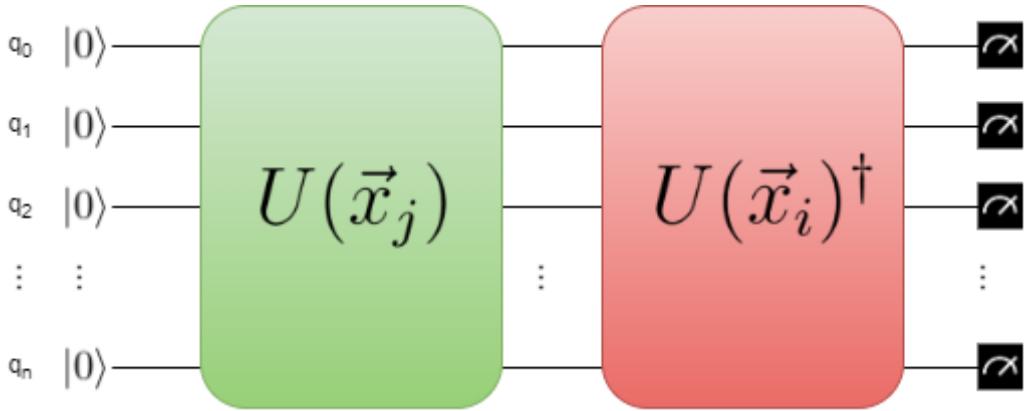


Figure 1.11: n -qubit quantum circuit diagram representing an arbitrary quantum kernel function described by an operator $U(\vec{x})$. The circuit must be executed several times in order to estimate the probability of measuring a $|0^n\rangle$ output state.

It is possible to verify that also quantum kernels are symmetric and positive-semidefinite. Moreover, quantum kernel outputs are always positive by definition and are able to quantify the similarity between the inputs: $K(\vec{x}_i, \vec{x}_j) = 1$ if $i = j$, $K(\vec{x}_i, \vec{x}_j) = 0$ if \vec{x}_i and \vec{x}_j are orthogonal in the embedding Hilbert space.

Quantum feature maps are able to spread classical data into large Hilbert spaces, which dimensions grow exponentially with the number of qubits, making it possible to easily achieve a linear class separation of training sets in that embedding. Translating into the original, classical featurespace, QSVMs are able to achieve highly non-linear decision boundaries, acquainting complex structures and correlation patterns in training data, resulting in highly expressive classifiers.

A crucial point for achieving quantum advantage with QSVM is that quantum computers are able to efficiently compute inner products of exponentially growing operators.

A generic quantum kernel function is hard to compute classically, and the search for quantum advantage in the task of classification consists of looking for hard-to-simulate kernels that achieve higher performances than classical ones [27].

Some properties that generally make quantum kernels more challenging to evaluate classically are high qubit number, degree of entanglement and depth. The implementation of such circuits is severely constrained by the capabilities of modern NISQ-era devices. Nevertheless, it is already possible to probe QSVMs at the edge of classical simulation capabilities.

1.4.3 Pauli feature maps

A typical class of commonly wielded quantum encodings takes the name of Pauli feature maps. Different kinds of Pauli feature maps are conjectured to be hard to simulate classically, and are defined as a certain number of repetitions (*reps*) of an H -gates layer followed by the operator:

$$U_{\phi(\vec{x})} = \exp \left(i \sum_{S \subseteq [n]} \phi_S(\vec{x}) \prod_{i \in S} P_i \right) \quad (1.25)$$

where P_i belongs to the set of Pauli matrices $P_i \in \{I, X, Y, Z\}$ and S is a set of possible qubit connectivities $S \in \{\binom{n}{k}\}$ combinations, $k = 1, \dots, n$. The entries of the \vec{x} vectors regulate the input parameters of the quantum gates, through the mapping:

$$\phi_S : \vec{x} \rightarrow \begin{cases} x_{i \in S} & \text{if } k = 1 \\ \prod_{i \in S} (\pi - x_i) & \text{if } k \geq 2 \end{cases} \quad (1.26)$$

The H -gate layer and the unitary operator $U_{\phi}(\vec{x})$ can be repeated several times sequentially on the circuit, in order to increase the depth of the feature map and explore new data encodings.

A simple example of a Pauli feature map is the Z -feature map, where $P_i = Z$ and $S = \{i\}$ for $i = 1, \dots, n$ ($k = 1$), represented in figure 1.12. Such encoding, due to the lack of entanglement, is not considered to be a candidate for providing a quantum advantage. A more complex example, conjectured to be hard to simulate classically is the ZZ -feature map, characterized by $k = 2$ and $P_0 = Z$, $P_{0,1} = ZZ$, shown in figure 1.13.

1.4.4 Quantum Kernel Alignment

Typical kernel circuits are rigid structures set arbitrarily and do not depend on data (although the kernel function value does), nor are trainable. Nonetheless, it is possible to extend the quantum kernels formulation by introducing a new set of variables θ that help kernels adapt to given data samples. The corresponding variational quantum feature map takes the form of:

$$\Phi(\vec{x}) = U_{\vec{\theta}}(\vec{x}) |0^n\rangle \langle 0^n| U_{\vec{\theta}}^\dagger(\vec{x}) \quad (1.27)$$

In particular, the proposed optimization strategy for the present work, called *kernel alignment* [28], consists of modifying the quantum kernel formula as follows in equation 1.28:

$$K_{\vec{\theta}}(\vec{x}_i, \vec{x}_j) = |\langle 0^n | V_{\vec{\theta}}^\dagger U^\dagger(\vec{x}_i) U(\vec{x}_j) V_{\vec{\theta}} | 0^n \rangle|^2 \quad (1.28)$$

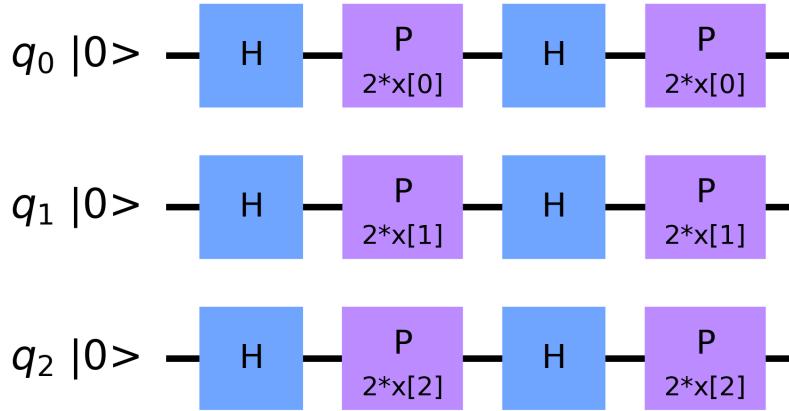


Figure 1.12: Example of a quantum circuit for the Z feature map, applied to three qubits and with a repetition number ($reps$) = 2. $x[0]$ and $x[1]$ are two features from the original dataset that parametrizes the Pauli rotation angle. The P gate, called *phase gate*, is equivalent to a R_Z Pauli rotation up to a phase factor: $P(\theta) = e^{i\theta/2}R_Z(\theta)$.

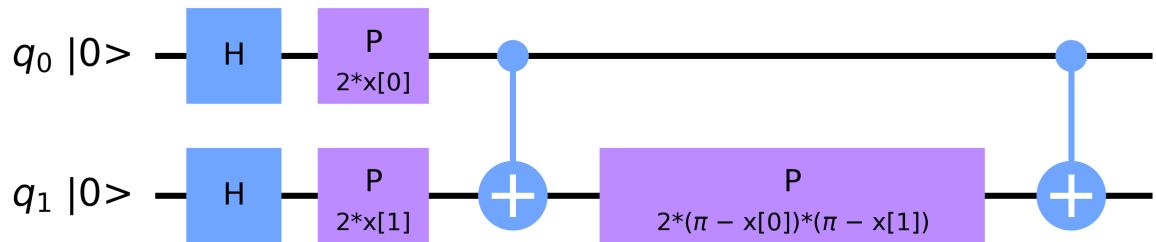


Figure 1.13: Example of a quantum circuit for the ZZ feature map, applied to two qubits and with $reps$ = 1.

In equation 1.28 the dependencies on $\vec{\theta}$ and \vec{x} are factorized, and the operator $V_{\vec{\theta}}$ is a set of rotation along an equatorial axis in the Bloch sphere representation. The general quantum kernel structure assumes now the form illustrated in figure 1.14.

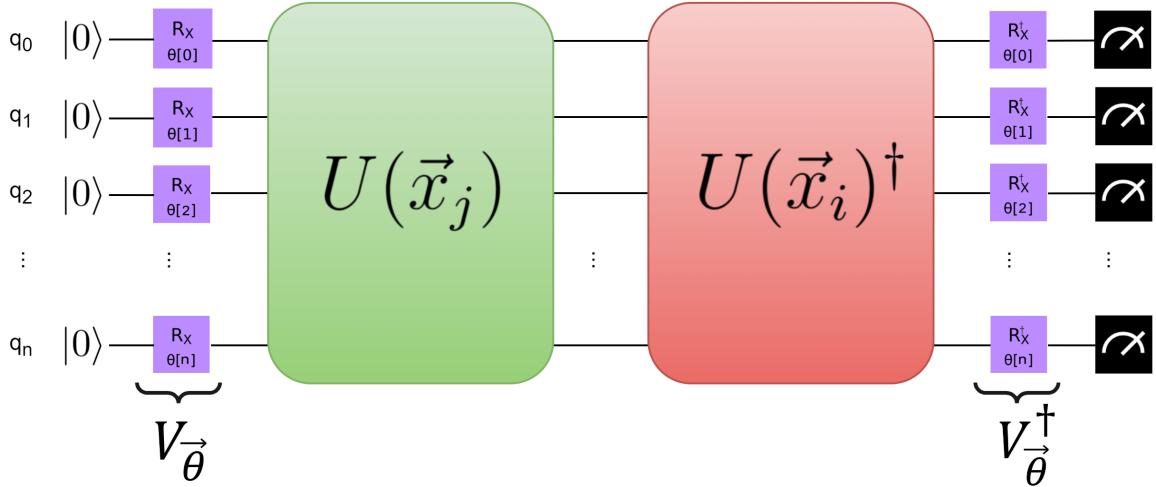


Figure 1.14: Circuit representation of a generic n -qubit quantum kernel with alignment consisting of trainable rotations along the X axes of the Bloch sphere.

The parameter vector $\vec{\theta}$ can be set for a given problem by searching the rotation angles that most facilitates the QSVM classification of a validation dataset. To avoid biases in the performance evaluation, this ensemble and the normal training set must be disjoint.

1.5 Automatization of quantum circuits design

As already hinted in section 1.4.3, the unitary operators that describe quantum kernels are normally set arbitrarily, or with *heuristics*. While in some particular cases the dataset feature distributions themselves may suggest the structure of the kernel circuit, there are no rigorous methodologies that allow us to find the best quantum feature map for a general classification problem. Part of the thesis work aimed to explore an alternative to heuristic by recurring to a *meta-heuristic Genetic Algorithm* (GA) [29]. The Genetic Algorithm is a generative model inspired by the theory of evolution of life on Earth and based on the concept of natural selection. An initial population of individuals (or *chromosomes*) undergoes iterative modifications in their genome due to the action of *genetic operators*. The evolutionary process consists of the survival of the fittest individuals, that will transmit their information to their *offspring* through the next generation.

Genetic Algorithms are a viable dataset-independent solution for the search for optimal quantum feature maps in order to avoid making assumptions *a priori* on the gates that compose the circuit and the typical *trial and error* approach of the use of heuristic methods. Moreover, GA-generated quantum kernels result from a broader exploration of the possible degrees of freedom that describe a kernel circuit.

The key elements for implementing a Genetic Algorithm are the following:

- **Chromosome representation:** the informative content that describes each

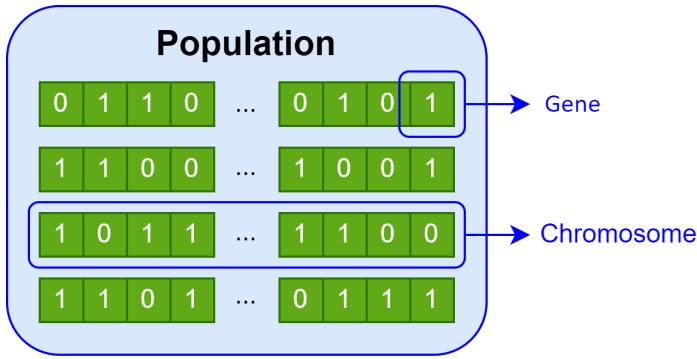


Figure 1.15: A population of bitstring-encoded individuals/chromosomes, where each one of the bits is called *gene*.

chromosome, or individual, exhaustively reported in units called *genes*, as shown in figure 1.15. In the quantum kernel example, each chromosome can be represented by a bitstring, where each bit is a gene and an ensemble of bits encode a quantum gate, its position within the circuit and its parametrization.

- **Fitness function:** a function that assigns a score to each individual in the population and represents the objective function to maximize, quantifying the degree of adaptation.
- **Parent selection:** criterion for selecting the individual that will survive to the next generation and generate new offspring, taking into account the fitness scores. Examples are *steady state*, *roulette wheel* and *rank* selection.
- **Genetic operators:** modifications of the individual's genetic pool for providing variability to the sample, as represented in figure 1.16. Genetic operators are divided into two groups:
 - **Mutation:** operates on the single individual and change genes randomly, or by *swap*, *inversion* or *scramble* criteria.
 - **Crossover:** operates on two individuals, mixing information of both in composing the new offspring. It consists of swapping chunks of genes from one individual to the other, according to different modalities, such as *n-points* or *uniform*.

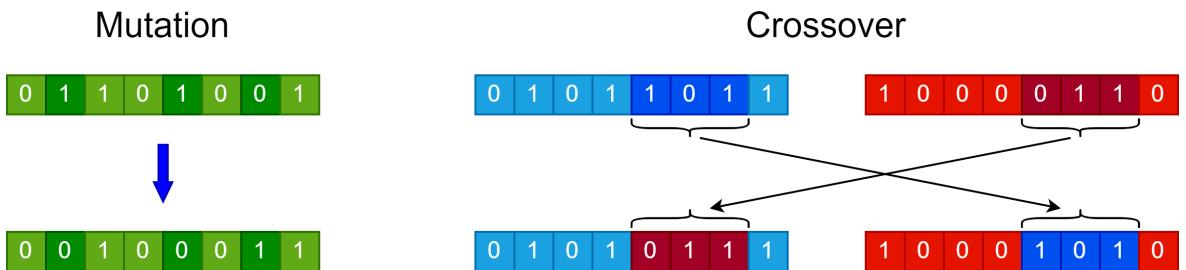


Figure 1.16: Examples of random mutation and two-points crossover operators on 1-byte chromosomes.

A GA generation routine starts by instantiating a population of a given size μ (number of individuals), and then iteratively applies genetic operations and parent selections, in

the way described in figure 1.17. For the application concerning the work in chapter 3, the population size is preserved at each generation, i.e. at every generation, where an offspring of size λ is produced, λ individual must be discarded. This method is known as the $\mu + \lambda$ Genetic Algorithm [30].

A critical aspect of GAs is that they do not guarantee convergence towards an absolute optimum.

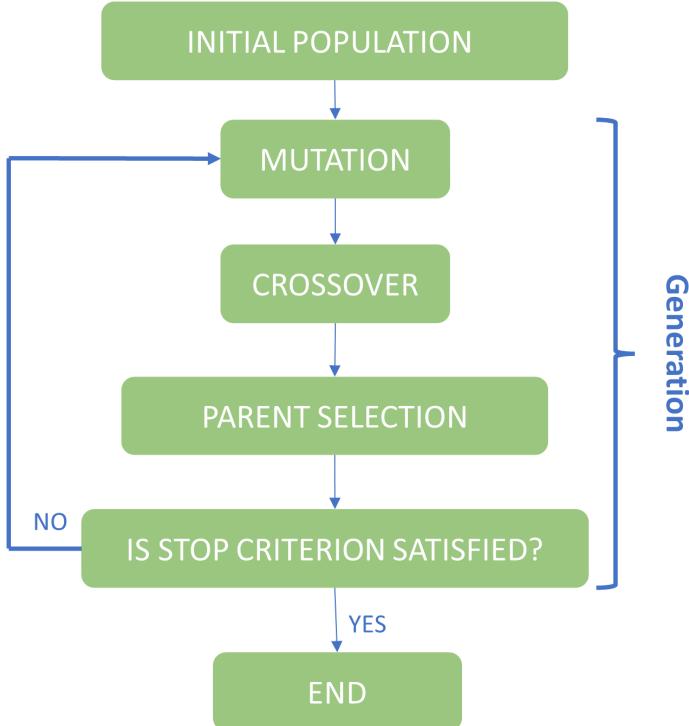


Figure 1.17: Genetic Algorithm flowchart. The stop criterion may be the reach of a fixed maximum number of generations, or a sufficiently large generation number in which no more improvements of the fitness value were encountered in new individuals.

lute or local maximum (if they exist) of the fitness function. A concrete risk is the one of reaching *premature convergence*, i.e. a condition where all individuals are identical and the genetic operators are not able to generate better offspring, resulting in a sub-optimal solution. One way to prevent the excessive shrinkage of the genetic variability over time is to previously adjust the amount of mutation and crossover, increasing the initial population μ or the offspring size λ .

When evaluation of the fitness function of every individual is computationally expensive, like for the quantum kernel case, it is crucial to find a balance between the GA initial parameters in order to achieve a sufficiently high amount of exploration within the limits of available computational resources.

This section concludes the discussion of Quantum Support Vector Machines and an overview of the strategies adopted in this work for addressing Quantum Advantage through the use of Quantum Kernels. In the present thesis, QSVM is proposed as an innovative tool for Fundamental Research in neutrino physics, in particular for achieving topology-based background mitigation in *neutrinoless double beta* ($0\nu\beta\beta$) decay searches. The next chapter will introduce the physics motivation for the study of this process and discuss the analysis of the $0\nu\beta\beta$ decay of the ^{136}Xe isotope in the framework of the DUNE experiment.

Chapter 2

$0\nu\beta\beta$ decay in xenon-doped LAr-TPCs

The present chapter introduces the physics of neutrinos, the most elusive particles known to this day, and the role they play in the modern understanding of Fundamental Physics. In particular, this chapter focuses on the neutrinoless double beta decay ($0\nu\beta\beta$), a hypothetical process that, if observed, would have important implications determining the properties of neutrinos and the understanding of matter creation. Among several experimental efforts, this chapter discusses the DUNE experiment and its $0\nu\beta\beta$ discovery potential, proposing a background mitigation strategy that exploits the topological information produced by DUNE’s Liquid Argon Time Projection Chambers (LArTPCs) facilities.

2.1 Standard Model and neutrino mass

The Standard Model (SM) of particle physics [31] embodies our current understanding of the interaction between elementary particles, the building blocks of our universe. SM comprehends weak forces, strong forces and electromagnetism, providing a wide variety of predictions in excellent agreement with experimental results [32]. Despite the great success of the theory, the Standard Model is known to be *incomplete* [33], i.e. unable to provide an explanation for several aspects:

- **Gravity:** SM does not incorporate a renormalizable quantum theory of gravity, one of the fundamental forces, and it is not compatible with the theory of General Relativity.
- **Astrophysical observations** suggesting the existence of Dark Matter and Dark Energy, not included in the SM, which together would make up about 95% of the total mass and energy in the Universe.
- **Particle mass hierarchy:** the masses of elementary particles differ by several orders of magnitude, as well as the coupling strengths of the different types of interactions. The cause is unclear in the framework of SM.
- **Matter-antimatter asymmetry:** only a small violation of the asymmetry in the production of matter and antimatter is predicted by the SM, while the

Universe seems to be composed mostly of matter, implying the existence of other violation mechanisms not included in the model.

- **Neutrino masses:** SM considers neutrinos as massless particles, however, neutrino oscillation experiments have proven the contrary [34]. While it is possible to extend the SM by taking that into account, it is unclear if neutrinos gain mass in the same way as other particles do.

Among the Standard Model particles (shown in figure 2.1), neutrinos are arguably the least understood, due to their minimal, exclusively weak interaction with matter [35]. For this reason, improving our understanding of neutrinos plays an important role in stepping further in our understanding of the Universe.

The observation of neutrino flavour oscillations confirmed that these particles have mass, but their values are still unknown. The KATRIN experiment set a mass-scale upper limit of about 0.8 eV at 90% CL [36], way smaller than any other SM particle (except for massless ones). This result is consistent with cosmological constraints imposed by the Λ CDM model [37].

Some of the most relevant experiments in neutrino physics carry out the search for the so-called *neutrinoless double beta decay* ($0\nu\beta\beta$), a hypothetical process, expected to be extremely rare, whose existence would prove that neutrinos are their own antiparticles and encourage the *seesaw mechanism* hypothesis, that would provide an explanation for such a small neutrino mass scale. Moreover, the $0\nu\beta\beta$ decay would be the first ever discovered process that violates the lepton number conservation, with possible implications in matter-antimatter asymmetry [38].

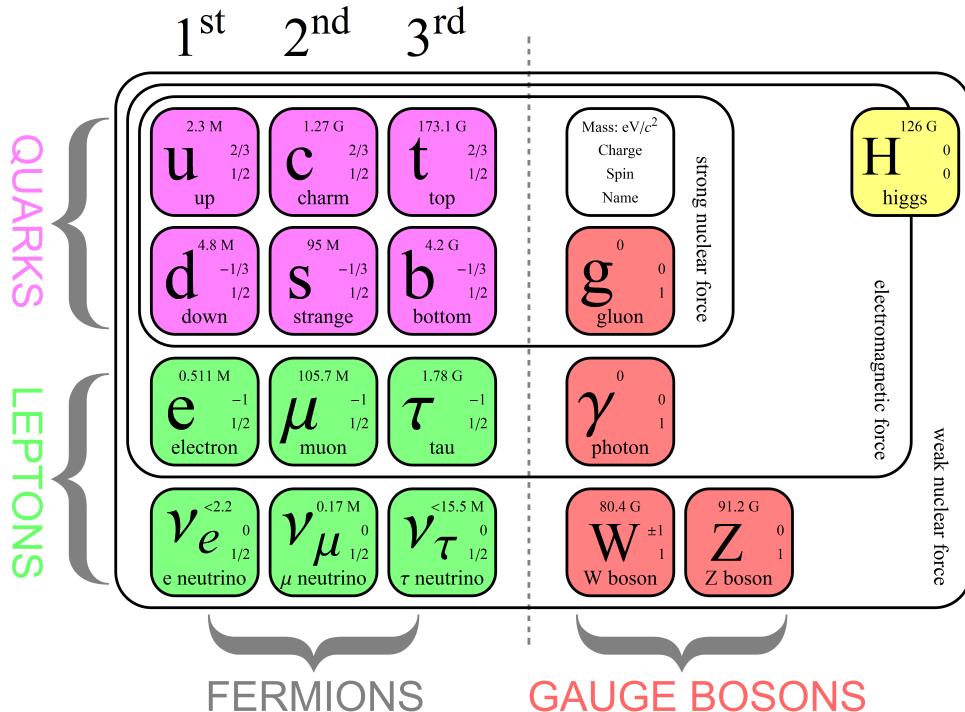


Figure 2.1: Elementary Standard Model particles grouped into quarks/leptons, fermions/gauge bosons and the Higgs [39].

2.1.1 Gauge symmetries and Higgs mechanism

The Standard Model is a gauge theory that considers particles as Lorentz invariant quantum fields and it is invariant under local transformations that belong to the symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y$$

$SU(3)_C$ corresponds to Quantum Chromodynamics (QCD), the Yang-Mills theory of strong interactions, while $SU(2)_L \times U(1)_Y$ refers to the Electroweak theory (EW) introduced by Glashow, Salam and Weinberg. From the local symmetries arise the so-called *gauge bosons*, particles that mediate interactions between fields. It is possible to infer the number of gauge bosons from the symmetry group's Lie algebra dimension, i.e. to $SU(N)$ are associated $N^2 - 1$ bosons, while for $U(N)$ there are N^2 of them.

The eight gauge bosons of QCD are gluons, massless and color-charged. The four bosons of EW coming from $SU(2)_L \times U(1)_Y$ ($W_\mu^1, W_\mu^2, W_\mu^3, B_\mu$) are mixed due to the occurrence of the *spontaneous symmetry breaking* $SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}$ [40] and the associated *Higgs mechanism* [41] into the physical bosons $W_\mu^+, W_\mu^-, Z_\mu^0$, that mediate weak interaction, and the photon A_μ , mediating electromagnetic interaction:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W^1 \mp W^2) \quad (2.1)$$

$$A_\mu = \cos(\theta_W) B_\mu + \sin(\theta_W) W^3 \quad (2.2)$$

$$Z_\mu^0 = \cos(\theta_W) W_\mu^3 - \sin(\theta_W) B_\mu \quad (2.3)$$

where θ_W is called *Weinberg's angle* ($\sin^2 \theta_W \simeq 0.23$ [42]) and quantifies the mixing between the third generator of $SU(2)_L$ (W_μ^3) and the $U(1)_Y$ one (B_μ).

The Higgs mechanism is responsible for giving mass to weak bosons, as new mass terms appear in the EW Lagrangian:

$$m_W = \frac{gv}{2} \quad m_Z = \frac{v}{2} \sqrt{g^2 + g'^2} \quad (2.4)$$

Where g and g' are the coupling strength of $SU(2)_L$ and $U(1)_Y$ respectively, and v is the *Vacuum Expectation Value* (VEV) of the Higgs boson potential. The masses of W^\pm and Z^0 bosons are related by the formula $m_W = \cos(\theta_W)m_Z$. After the spontaneous symmetry breaking, the EW Lagrangian possesses a new reduced symmetry, $U(1)_{EM}$, related to the photon that did not gain mass in the process.

2.1.2 Fermion masses and Yukawa coupling

Besides the Higgs and the gauge bosons, the Standard Model comprises quarks and leptons. Quarks are grouped into three doublets (families) of type "up" u_i (*up, charm, top*) and "down" d_i (*down, strange, bottom*) where $i = 1, 2, 3$ is the family index. Similarly, leptons are organized into three doublets of different flavours, each containing a neutral component ν_i (neutrinos) and a charged one l_i .

The mass terms for fermions are given by the interaction with the Higgs scalar field, a *Yukawa coupling*, and assume the form:

$$\mathcal{L}_Y = - \sum_{jk} \left\{ (\bar{u}'_j, \bar{d}'_j)_L \left[c_{jk}^{(d)} H d'_{kR} + c_{jk}^{(u)} \tilde{H} u'_{kR} \right] + (\bar{\nu}'_j, \bar{l}'_j)_L c_{jk}^{(l)} H l'_{kR} \right\} + h.c. \quad (2.5)$$

where:

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad \tilde{H} = -i\sigma_2 H^* = \begin{pmatrix} \phi^{0*} \\ -\phi^- \end{pmatrix} \quad (2.6)$$

And coefficients $c_{jk}^{(d)}, c_{jk}^{(u)}, c_{jk}^{(l)}$ are proportional to the respective lepton masses. Equation 2.5 preserves the $SU(2)_L \times U(1)_Y$ symmetry, and the primed sign on spinors indicates that they are expressed in the flavour basis. After the spontaneous symmetry breaking, that can be realized in the unitary gauge by expanding the Higgs field around the minimum $(0, v)^T$, i.e.

$$H = \begin{pmatrix} 0 \\ v + h \end{pmatrix}$$

the mass terms become:

$$\mathcal{L}_Y = - \left(1 + \frac{h}{v} \right) [\bar{\mathbf{d}}'_L M'_d \mathbf{d}'_R + \bar{\mathbf{u}}'_L M'_u \mathbf{u}'_R + \bar{\mathbf{l}}'_L M'_l \mathbf{l}'_R] + h.c. \quad (2.7)$$

where $(M'_\xi)_{ij} = c_{ij}^{(\xi)}$ for $\xi = d, u, l$.

Due to the arbitrariness of coefficients $c_{ij}^{(\xi)}$, the mass matrices M'_ξ are generally different from each other and non-diagonal. For this reason, the mass eigenstates, i.e. the ones that satisfy:

$$\mathcal{L}_Y = - \left(1 + \frac{h}{v} \right) [\bar{\mathbf{d}}_L M_d \mathbf{d}_R + \bar{\mathbf{u}}_L M_u \mathbf{u}_R + \bar{\mathbf{l}}_L M_l \mathbf{l}_R] + h.c. \quad (2.8)$$

where M_ξ for $\xi = d, u, l$ are diagonal, do not coincide with the flavour ones, inducing a *mixing* in the quark sector described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [43]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.9)$$

A direct consequence of quark mixing is the phenomenon of *neutral meson oscillation* [44], abundantly observed at particle accelerators, which consists of neutral mesons produced in an initial flavour state transforming into their respective antiparticle, after propagating for a certain time interval.

2.1.3 Neutrino flavor oscillations

In the Standard Model, neutrinos are considered massless, therefore no Yukawa coupling is present in equation 2.5. This allows for choosing neutrino flavours eigenstates that eliminate the mixing in the lepton sector. Nevertheless, neutrino oscillation has been observed by various experiments throughout the last decades, proving that neutrinos are not massless particles, and undergo a mixing mechanism similarly to quarks:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (2.10)$$

The neutrino mixing matrix U is called Pontecorvo-Maki-Nakagawa-Sakata (PMNS) [45]. In the standard parametrization:

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}s_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\frac{\alpha}{2}} & 0 \\ 0 & 0 & e^{i\frac{\beta}{2}} \end{pmatrix} \quad (2.11)$$

where $c_{ij} = \cos(\theta_{ij})$, $s_{ij} = \sin(\theta_{ij})$, $\theta_{ij} \in [0, \pi/2]$ are the mixing angles, $\delta \in [0, 2\pi]$ the CP-violation phase and α, β are the Majorana phases.

The oscillation probabilities are summarized by the following expression:

$$\begin{aligned} P\left(\overset{(-)}{\nu_l} \rightarrow \overset{(-)}{\nu_{l'}}\right) = & \delta_{ll'} + p_{ll'}^{12} \sin^2 \varphi_{12} + p_{ll'}^{13} \sin^2 \varphi_{13} + p_{ll'}^{23} \sin^2 \varphi_{23} \\ & \pm 8J \sin \varphi_{12} \sin \varphi_{13} \sin \varphi_{23} \sum_{l''} \epsilon_{ll'l''} \end{aligned} \quad (2.12)$$

where $\delta_{ll'} = 0$ if $l \neq l'$ (old flavour disappearance), $\delta_{ll'} = 1$ if $l = l'$ (new flavour appearance) and $\epsilon_{ll'l''}$ is the Levi-Civita tensor. The terms that define the oscillation amplitudes depend on mixing angles, and are:

$$p_{ll'}^{ii'} = -4\text{Re}(U_{li}U_{l'i'}U_{l'i}^*U_{li'}^*) \quad (2.13)$$

$$8J = \cos \theta_{13} \sin 2\theta_{13} \sin 2\theta_{12} \sin 2\theta_{23} \sin \delta \quad (2.14)$$

while the terms that modulate the frequencies depend on the differences between neutrino mass eigenvalues, the energy E and the travel distance L :

$$\sin^2 \varphi_{ij} = \sin^2 \left(\frac{\Delta m_{ij}^2}{4E} L \right)$$

The \pm sign in equation 2.12 differentiates the behaviour of neutrinos (+) and (-) behaviour, introducing a CP violation term proportional to the value of $J =$, named Jarlskog invariant. An illustration of three-flavour neutrino oscillation is shown in figure 2.2.

At the present time, not only are the values of the three neutrino mass unknown, but also the relative mass ordering between them. Neutrino oscillation experiments determined two possible scenarios, called *normal* and *inverted* orderings [47]. In both cases, two of the three neutrino masses have similar values, while the other is much lighter.

- **Normal Ordering (NO):** the lightest neutrino is ν_1 , i.e. the one with the largest mixing with the ν_e component, $\Delta m_{31}^2 > 0$.
- **Inverted Ordering (IO):** the lightest neutrino is ν_3 , i.e. the one with the smallest mixing with the ν_e component, $\Delta m_{31}^2 < 0$.

2.1.4 The Dirac and Majorana masses

An open question in neutrino physics is whether neutrinos are described by *Dirac* or *Majorana* spinors [48]. In the first case, the Yukawa coupling to the Higgs would be analogous to the other fermion ones, while in the second case it would be possible to explain the small neutrino mass scale thanks to the seesaw mechanism.

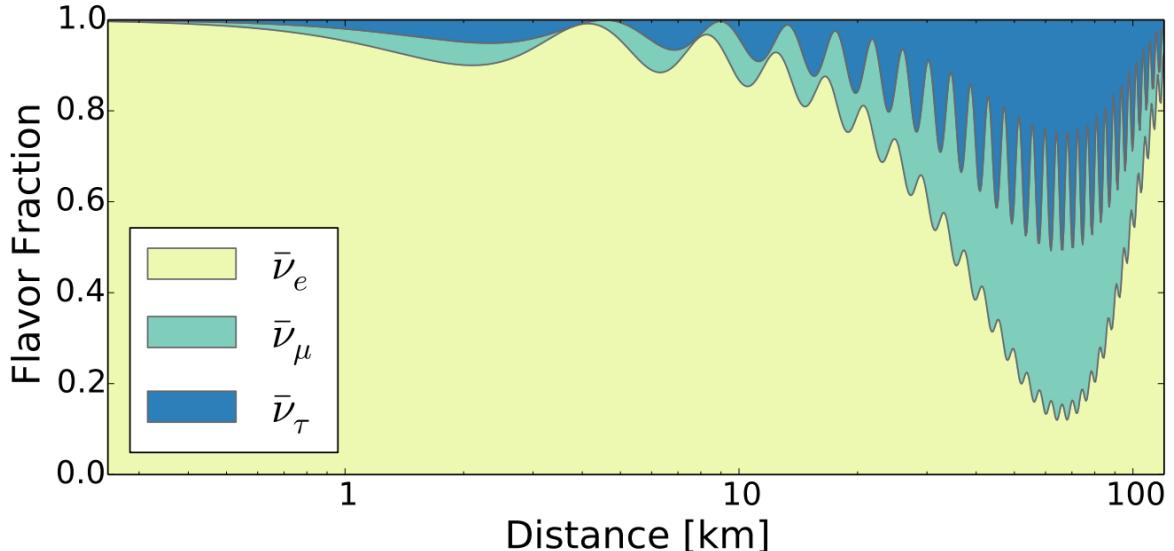


Figure 2.2: Example of neutrino three-flavour oscillations at 4 MeV energy [46]. The probabilities are expressed by the colour amount relatively to the corresponding flavour.

A Dirac spinor is a four-component field that satisfies the Dirac equation and can be written in terms of two upper components ξ representing a left-chirality part and two lower components η representing a right-chirality one. In total, a Dirac spinor is described by four degrees of freedom.

$$\psi = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

In this formulation, called Weyl basis, the chirality operator can be written as:

$$\gamma^5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}$$

Therefore the Weyl basis spinors are:

$$\begin{aligned} \text{left chirality (L): } & \begin{pmatrix} \xi \\ 0 \end{pmatrix} : \quad \gamma^5 \begin{pmatrix} \xi \\ 0 \end{pmatrix} = - \begin{pmatrix} \xi \\ 0 \end{pmatrix} \\ \text{right chirality (R): } & \begin{pmatrix} 0 \\ \eta \end{pmatrix} : \quad \gamma^5 \begin{pmatrix} 0 \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ \eta \end{pmatrix} \end{aligned} \tag{2.15}$$

A Majorana spinor, while still satisfying the Dirac equation, is described by only two degrees of freedom and is written as:

$$\psi = \begin{pmatrix} \xi \\ -i\sigma_2\xi^* \end{pmatrix}$$

A remarkable property of Majorana spinors is the invariance under charge conjugation:

$$\psi^c \equiv -i\gamma^2\psi^* = \psi$$

Meaning that the particle described by ψ is its own antiparticle.

The most general mass term for neutrino is given by the contributions of both Dirac and Majorana terms. Considering only one flavour for simplicity:

$$\begin{aligned}\mathcal{L}_{D+M} &= -\frac{1}{2}\bar{\nu}_L M_L^M (\nu_L)^c - \bar{\nu}_L M_D^D \nu_R - \frac{1}{2}(\bar{\nu})_R^c M_R^M \nu_R + h.c. \\ &= -\frac{1}{2}\bar{n}_L M^{D+M} (n_L)^c + h.c.\end{aligned}\quad (2.16)$$

where:

$$n_L = \begin{pmatrix} \nu_L \\ (\nu_R)^c \end{pmatrix} \quad M^{M+D} = \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \quad (2.17)$$

The flavour eigenstate n_L has a left chirality component ν_L , that interacts weakly, and the right chirality one n_R , the *sterile* neutrino, excluded from weak interaction that only couples to left-handed spinors. One type of the seesaw mechanism can be achieved by setting $m_L = 0$ and assuming that m_D , coming from a Dirac term, is at the electroweak (charged leptons and quark) mass scale while $m_R \gg m_D$. Hence, diagonalizing the M^{M+D} matrix in order to find the neutrino mass eigenvalues:

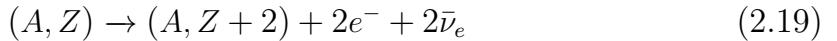
$$m_1 \simeq \frac{m_D^2}{m_R} \ll m_D \quad m_2 \simeq m_R \gg m_D \quad (2.18)$$

meaning that one neutrino is much lighter than the electroweak scale, and the other one is much heavier. Also, being the mixing angle $\theta \simeq m_D/m_R \ll 1$, flavour and mass eigenstates are almost coincident, justifying the sub-eV mass scale of weakly interacting neutrinos.

2.2 Neutrinoless double beta decay

The most proposed physics channel for investigating whether neutrinos are Dirac or Majorana particles is the search for a double beta decay without neutrino emission in those isotopes already subject to the $2\nu\beta\beta$ decay.

Beta decays are weak processes where a nucleus emits an electron (positron) and an antineutrino (neutrino), due to a nucleon conversion, increasing (decreasing) the nucleus' atomic number by a unit, in order to reach a more stable configuration. For some isotopes, the single beta decay is energetically forbidden (see figure 2.3), while the double beta may be allowed:



Conventional two-neutrino double beta decays ($2\nu\beta\beta$) is a rare second-order weak process and has been observed in several isotopes within a lifetime range between 10^{18} to 10^{21} years [50]. The neutrinoless counterpart, instead, is represented by the formula:



and is forbidden by the Standard Model. Currently, no experiment has provided evidence for the existence of this decay, with lower lifetime limits of about $10^{25} - 10^{26}$ years, indicating the extreme rarity of the process, if occurring at all.

At the Feynman diagram level, as shown in figure 2.4, the difference between $2\nu\beta\beta$ and $0\nu\beta\beta$ is that in the second case the two antineutrinos annihilate. For this reason,

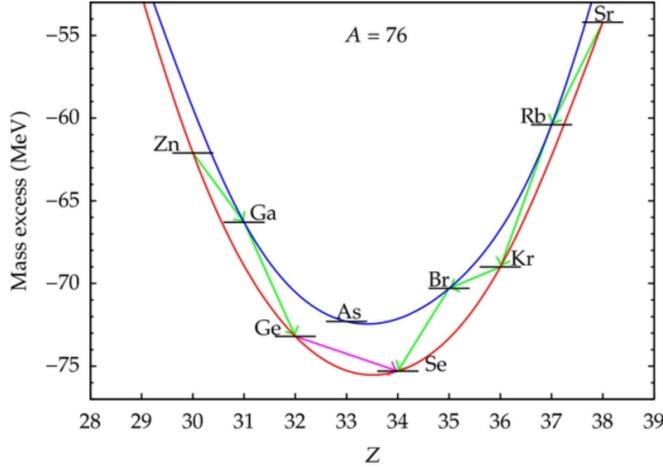


Figure 2.3: Mass difference as a function of Z for nuclei with atomic mass $A = 76$ [49]. In this example, the only way for 76-Ge to lose energy is by undergoing a double beta decay into 76-Se.

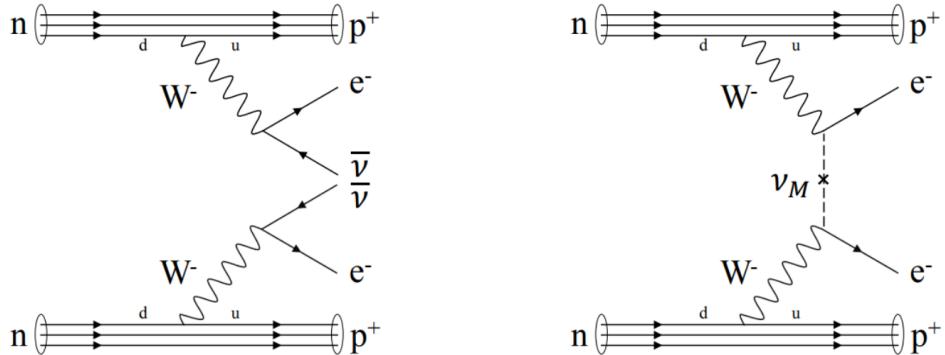


Figure 2.4: Lowest-order Feynman diagrams of conventional double beta decay (left) and neutrinoless (right) [51].

the $0\nu\beta\beta$ is allowed only if neutrino and antineutrino correspond to the same particle, confirming the Majorana nature of the neutrino.

Equation 2.20 underlines the violation of the total lepton number: $\Delta L = 2$, which is instead conserved in any other process observed so far. The $0\nu\beta\beta$ decay indeed would allow the production of two electrons without the two corresponding antimatter particles, providing new insights into the matter-antimatter asymmetry in the Universe.

The decay rate of the $0\nu\beta\beta$ is :

$$\Gamma_{\beta\beta}^{0\nu} = \frac{1}{T_{\beta\beta}^{0\nu}} = G^{0\nu}(Q, Z) ||M||^2 m_{\beta\beta}^2 \quad (2.21)$$

where $G^{0\nu}$ is the phase space factor, Q the Q-value of the $\beta\beta$ process, and M a nuclear matrix element, different for every isotope, that can be calculated using sophisticated nuclear many-body models.

The parameter $m_{\beta\beta}$ is the *effective Majorana mass*, defined as:

$$m_{\beta\beta} = \sum_i U_{ei}^2 m_i \quad (2.22)$$

m_i are the three neutrino mass eigenstates and U_{ei} are elements of the PMNS matrix defined in equation 2.11. It is possible to notice that while Majorana phases α and β have no effect on neutrino flavour oscillations (does not appear in equation 2.12) because they cancel each other when computing the square module of the PMNS matrix, they are related to $m_{\beta\beta}$ due to the U^2 dependence. This is the reason why oscillations occur regardless of the Dirac or Majorana nature of the neutrino, and the $0\nu\beta\beta$ decay does not.

The lower limit set by experiments on the neutrinoless double beta decay lifetime also set an upper limit on the $m_{\beta\beta}$ parameter. Candidates for the $0\nu\beta\beta$ search and their current exclusion limits are summarized in table 2.1. The value of the parameter $m_{\beta\beta}$

Isotope	$T_{1/2}^{0\nu} (\times 10^{25} \text{ y})$	$\langle m_{\beta\beta} \rangle (\text{eV})$	Experiment
^{48}Ca	$> 5.8 \times 10^{-3}$	$< 3.5 - 22$	ELEGANT-IV
^{76}Ge	> 8.0	$< 0.12 - 0.26$	GERDA
	> 1.9	$< 0.24 - 0.52$	MAJORANA DEMONSTRATOR
^{82}Se	$> 3.6 \times 10^{-2}$	$< 0.89 - 2.43$	
^{96}Zr	$> 9.2 \times 10^{-4}$	$< 7.2 - 19.5$	NEMO-3
^{100}Mo	$> 1.1 \times 10^{-1}$	$< 0.33 - 0.62$	NEMO-3
^{116}Cd	$> 1.0 \times 10^{-2}$	$< 1.4 - 2.5$	NEMO-3
^{128}Te	$> 1.1 \times 10^{-2}$	—	—
^{130}Te	> 1.5	$< 0.11 - 0.52$	CUORE
^{136}Xe	> 10.7	$< 0.061 - 0.165$	KamLAND-Zen
	> 1.8	$< 0.15 - 0.40$	EXO-200
^{150}Nd	$> 2.0 \times 10^{-3}$	$< 1.6 - 5.3$	NEMO-3

Table 2.1: $T_{1/2}^{0\nu}$ and $m_{\beta\beta}$ limits at 90% CL for several isotopes as reported by experiments [52].

is also related to the mass ordering mentioned in section 2.1.3. In particular, mass ordering defines the allowed regions for $m_{\beta\beta}$, which are different in cases of NO or IO, driving the sensitivity goals for $0\nu\beta\beta$ searches. Figure 2.5 shows the allowed regions in terms of $m_{\beta\beta}$ and the mass of the lightest neutrino, in the hypothesis of normal or inverted order.

In a $0\nu\beta\beta$ decay, all the energy available is distributed to the two electrons (neglecting nucleus recoil). Hence, the signature of the process corresponds to a peak at the $Q_{\beta\beta}$ -value in the spectrum of total electron kinetic energy, as represented in figure 2.6. In any realistic experiment with finite energy resolution, the tail of the $2\nu\beta\beta$ spectrum may invade the expected signal peak region. To deal with this overlap and the possible presence of other backgrounds, a common procedure is defining a region of interest (ROI) around the $Q_{\beta\beta}$ value and carrying out a counting experiment. By knowing the expected background rate inside the ROI, an excess would indicate the presence of signal $0\nu\beta\beta$ events. If the measured rate is compatible with the background-only hypothesis, the signal contribution could be overwhelmed by the statistical fluctuation

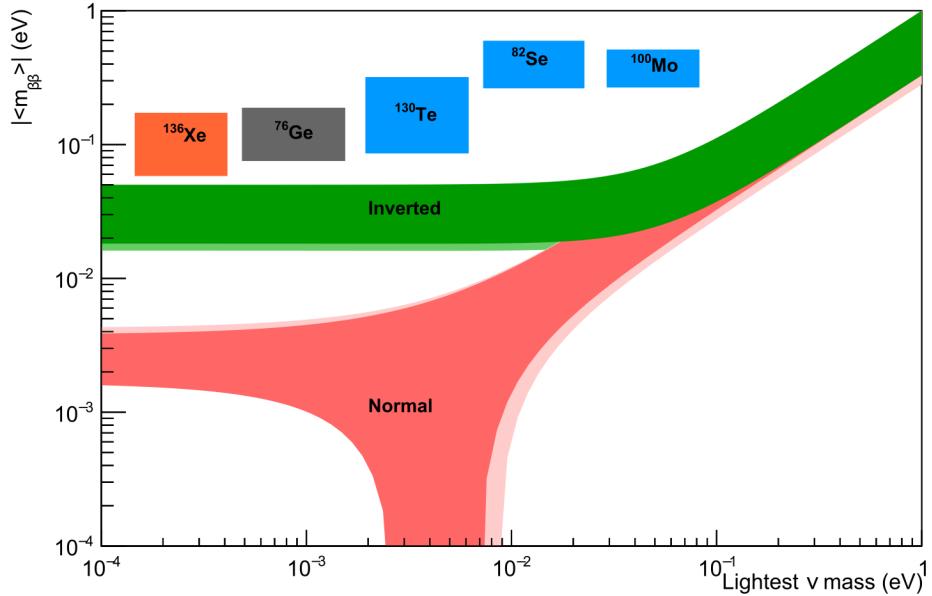


Figure 2.5: Experimental limits on $m_{\beta\beta}$ and allowed regions for $m_{\beta\beta}$ for inverted and normal ordering as a function of the smallest neutrino mass. Shaded regions represent 3σ confidence belts due to uncertainties in neutrino oscillation parameters [53].

of background events. In this background-dominated regime, it is still possible to infer a lower limit on $T_{\beta\beta}^{0\nu}$.

At one standard deviation CL, when it is possible to assume:

$$n_{0\nu\beta\beta} \leq \sqrt{n_{\text{bkg}}}$$

then:

$$T_{\beta\beta}^{0\nu} \geq \frac{N}{\sqrt{n_{\text{bkg}}}} T_m$$

where T_m is the time of the measure and N is the number of nuclei monitored by the experiment proportional to its mass M . For a ROI of width ΔE , the background rate B can be written as:

$$B = \frac{n_{\text{bkg}}}{\Delta E M T_{\text{meas}}}$$

The lower limit on the $0\nu\beta\beta$ half-life, $S_{\beta\beta}^{0\nu}$, or *sensitivity* will be:

$$S_{\beta\beta}^{0\nu} \propto \sqrt{\frac{MT_{\text{meas}}}{B\Delta E}}$$

$S_{\beta\beta}^{0\nu}$ is a quantity that expresses the potentiality of an experiment of setting an exclusion boundary over $T_{\beta\beta}^{0\nu}$.

As it could be intuitively expected, an experiment benefits from a high *exposure* (i.e. the product MT_{meas}) and the possibility to choose a narrow region of interest, determined by energy resolution.

Among various candidates for the $0\nu\beta\beta$ search, this thesis will focus on the 136-Xe isotope. The 136-Xe is one of the most appealing nuclei for its relatively high Q-value $Q_{\beta\beta} = 2.458 \text{ MeV}$, which is above most natural radiation energies, and the phase space

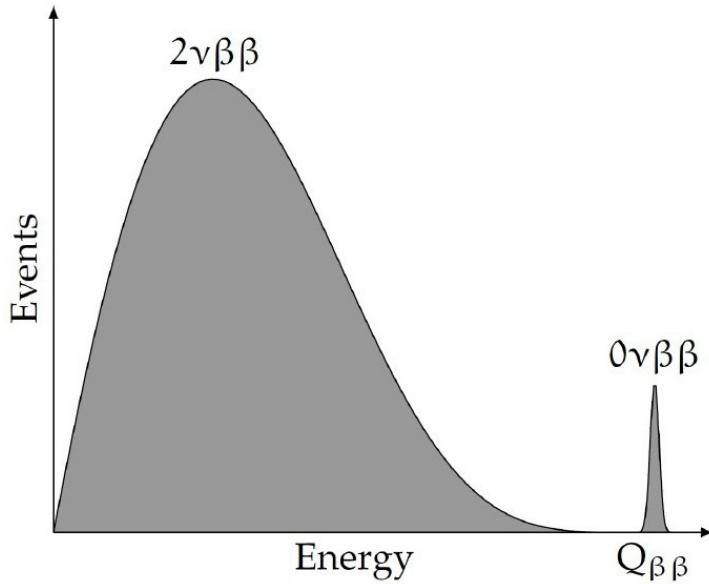


Figure 2.6: Schematic energy spectrum of a double beta decay including both 0ν and 2ν modes. The x-axis is the total kinetic energy of the two electrons, and the y-axis is the number of events. The height of the $0\nu\beta\beta$ peak is chosen arbitrarily.

factor $G^{0\nu}$ depending on Q^5 , suggesting a higher decay rate than other potential candidates with a smaller $Q_{\beta\beta}$.

One of the prominent experimental techniques for the search of the $0\nu\beta\beta$, already in use by the EXO-200 collaboration and under development by NEXT, is making use of Time Projection Chambers (TPCs) containing a large mass of 136-enriched xenon, in gaseous or liquid form. This is possible because xenon is a noble gas, with an additional advantage brought by the fact it can be enriched up to 90% isotopic concentration of 136-Xe. Currently, the most stringent limit on the half-life associated with the 136-Xe $0\nu\beta\beta$ decay corresponds to $1.07 \times 10^{26} y$ [54] and was achieved with another technique by the KamLAND-Zen experiment, working with a one-kilotonne xenon-loaded liquid scintillator.

Recent studies [55, 56, 57] suggest that similar searches could also be carried out by using monolithic xenon-doped liquid argon TPCs, or LArTPCs, discussing the fascinating prospect of extending the already wide physics reach of the DUNE experiment, illustrated in section 2.3, for including the investigation of the 136-Xe $0\nu\beta\beta$ process.

2.3 The DUNE experiment

The Deep Underground Neutrino Experiment (DUNE) is an international world-class experiment, still under construction, dedicated to answering some of the most daunting questions about neutrinos and investigating other various topics. The principal goals of DUNE are [58]:

- Solving the neutrino mass hierarchy problem, i.e. understanding mass ordering of ν_1, ν_2, ν_3 in equation 2.10 and achieving precise measurements of neutrino mixing parameters by detecting accelerator neutrinos.
- Probing the potential CP violation in neutrino flavour mixing, i.e. measuring the CP violation phase δ that appears in equation 2.14.
- Searching for the proton decay, predicted by Grand Unified Theories, which attempts to describe the unification of the SM fundamental forces.
- Studying supernovae neutrinos, emitted in the first seconds after the formation of neutron stars or black holes, potentially gaining new information about their birth and the evolution of the Universe.

DUNE is composed of three facilities: a high-intensity neutrino source, situated at the Fermi National Accelerator Laboratory (Fermilab) in Chicago, a *Near Detector* site (ND) installed in the proximity of the beam origin, and a massive *Far Detector* site (FD), located 1.5 km underground at the Sanford Underground Research Facility (SURF) in South Dakota. The components are illustrated in figure 2.7.

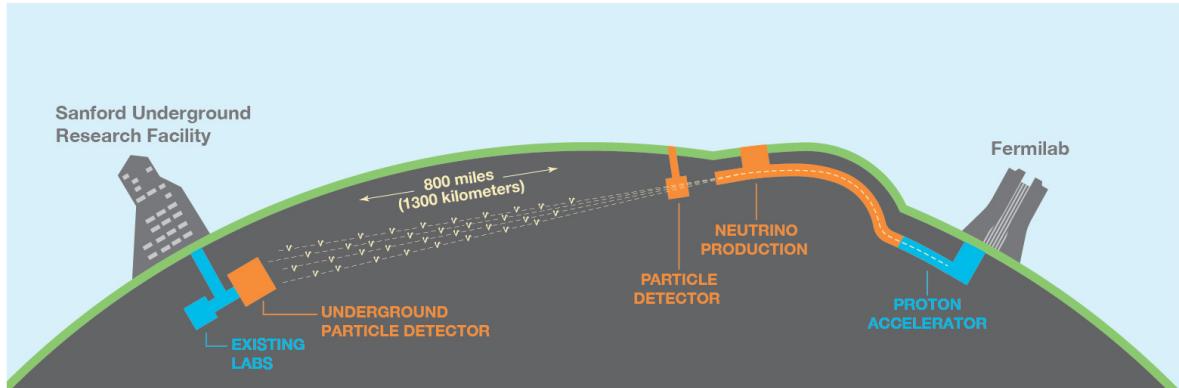


Figure 2.7: Schematic view of DUNE experiment, showing the Fermilab beamline and the ND on the right, and the FD at SURF on the left. ND and FD are distant $L = 1.3 \times 10^3\text{ km}$ from each other, classifying DUNE as a long-baseline neutrino experiment.

2.3.1 DUNE liquid argon TPCs

DUNE FDs and one ND will exploit the technique of Liquid Argon Time Projection Chamber (LArTPC), allowing for high energy resolution, and efficient tagging thanks to the 3D reconstruction of tracks, exemplified in figure 2.8. Liquid argon is a scintillating medium, meaning that two independent information carriers, ionization electrons and scintillation photons, can be leveraged for particle detection. The Far Detectors

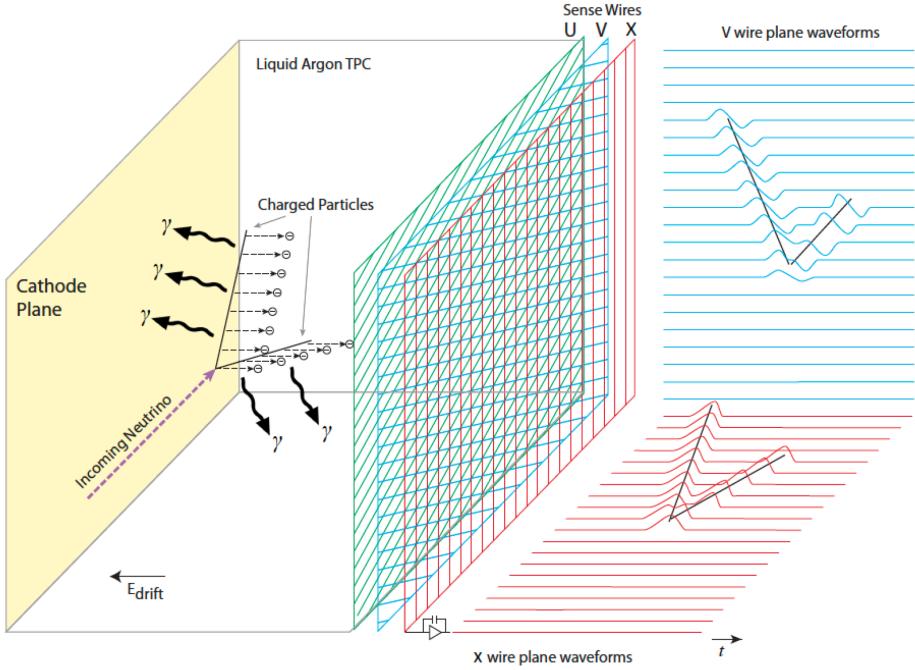


Figure 2.8: Track reconstruction by a LArTPC, from [59]. Scintillation light can be detected by photon sensors, while ionization electrons drift towards the readout planes. Each plane of the wire grid, together with temporal information, provides a 2D projection that combined with the other gives the full reconstruction.

will consist of four 10 kt fiducial mass LArTPCs modules, inside a cryostat of dimensions $15.1 \times 14.0 \times 62.0\text{ m}^3$ for a total LAr mass of 17.5 kt per module cooled to about 87 K [60]. Fiducialization consists of using the outer argon layers to shield the innermost region of the detector from external background sources. The division into four independent modules serves the purpose of designing and probing detectors based on different technologies. At the present time, the technologies for two out of the four modules have been established and will be implemented in the *Phase I* of the DUNE experiment. The remaining two modules, considered "opportunity" modules, will be left for the *Phase II* and may consist of other liquid argon-based detectors, or alternate technologies.

The first module will implement the so-called *Horizontal Drift* (HD) (Single-Phase Far Detector Module), pioneered by the ICARUS project [61]. The HD detector is divided into four drift volumes, with a maximum drift length of 3.5 m. A field cage around the active volume composed of aluminium rings ensures a uniform electric field of 0.5 kV/cm . Electrons drift horizontally towards the anodes of the cells. There are three TPC wire planes: the first two are induction planes that produce signals from the passage of ionization electrons, while the third one is the collection plane. The three wire planes are parallel and have a pitch of $d = 5\text{ mm}$. The wire orientations of the different planes form 37.5° angles, in order to allow tracking in two dimensions. The second module will adopt the *Vertical Drift* (VD) strategy. The active volume is split in the middle by the cathode, with the two anode planes placed at the top and the bottom of the detector. In this setup, ionization electrons drift upward (in the upper half) and downward (in the lower half), with a maximum drift length of 6.5 m.

Figure 2.9 shows the designs of HD and VD detectors projected for DUNE. Liquid Ar-

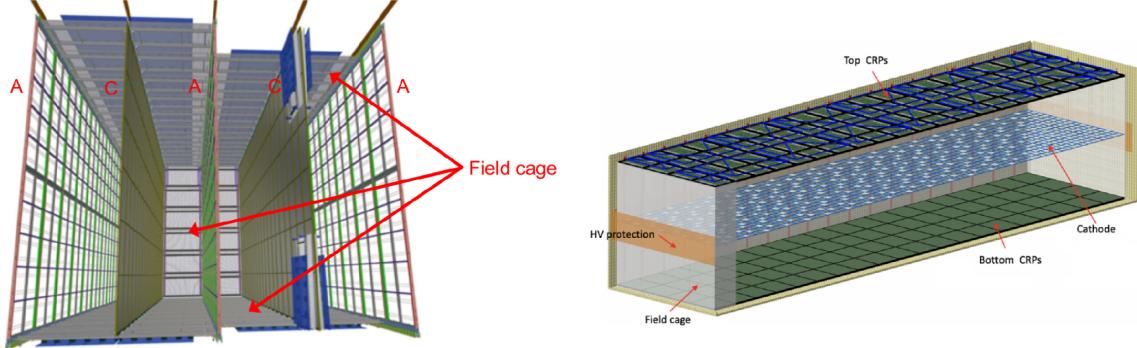


Figure 2.9: 10 kt FD modules. Horizontal Drift (left) with the alternate anode (A) and cathode (C) wire planes. Vertical Drift (right) with the central cathode and the charge readout planes on top and bottom. Image credit: [62].

gon scintillation light is produced at 127 nm. At this wavelength, most sensors are not sensitive, and it is necessary to introduce wavelength shifters (WLS). The proposed techniques are the most traditional tetraphenyl butadiene (TPB) [63], polyethylene naphthalate (PEN) [64], and xenon-doping [65]. TPB requires the installation of a coating setup that poses scalability issues to large TPCs, while PEN is a viable alternative, but it needs more characterization and it has never been operated in large-scale LAr-TPCs.

Xenon doping is a promising way to boost light detection efficiency, due to the simplicity of operation. It was proven that adding xenon at the ppm level shifts part of the scintillation light to longer wavelengths (178 nm), reducing the light attenuation (thanks to a longer Rayleigh scattering length), hence increasing the propagation distance [62].

The photon detection system for both DUNE HD and VD modules consists of *X-ARAPUCA* light traps [66]. These devices trap optical photons with the use of a dichroic filter, redirecting them towards arrays of silicon photomultipliers (SIPMs) placed inside the cage. *X-ARAPUCAs* are able to work in high-intensity electric fields and maximize the photon detection efficiency while keeping a limited number of SIPMs, reducing the overall cost of the detector.

Xenon-doping in LArTPCs for light detection purposes inspired the idea of adding the search for the neutrinoless double beta decay of the ^{136}Xe isotope in the DUNE low-energy physics programme.

2.3.2 DUNE sensitivity to ^{136}Xe $0\nu\beta\beta$

DUNE LArTPCs are optimized for detecting neutrinos produced at the Fermilab Main Injector accelerator, consisting of a wide-band energy spectrum peaking around 2.5 GeV . Nevertheless, DUNE's potentials extend well below the GeV scale, allowing for supernova neutrinos detection down to $10\text{-}20\text{ MeV}$ [67]. The low energy physics reach can further extend for studying solar neutrinos ($< 20\text{ MeV}$), which would allow increasing the sensitivity to the neutrino mass interval Δm_{21}^2 , achieving precision measurements of neutrino flux from the 8-B reaction chain and observing for the first time neutrinos from the *hep* flux [68]. Finally, by pushing to the very limit the detector

capabilities, also neutrinoless double beta decay ($Q_{\beta\beta}^{136Xe} = 2.458 \text{ MeV}$) and Weakly Interacting Massive Particles (WIMPs) dark matter ($\sim 100 \text{ keV}$ nuclear recoil) could be accessible.

In order to address the sensitivity to low-energy physics analyses, a careful background study is needed. Physics targets, along with the principal radiological backgrounds, are shown in figure 2.10.

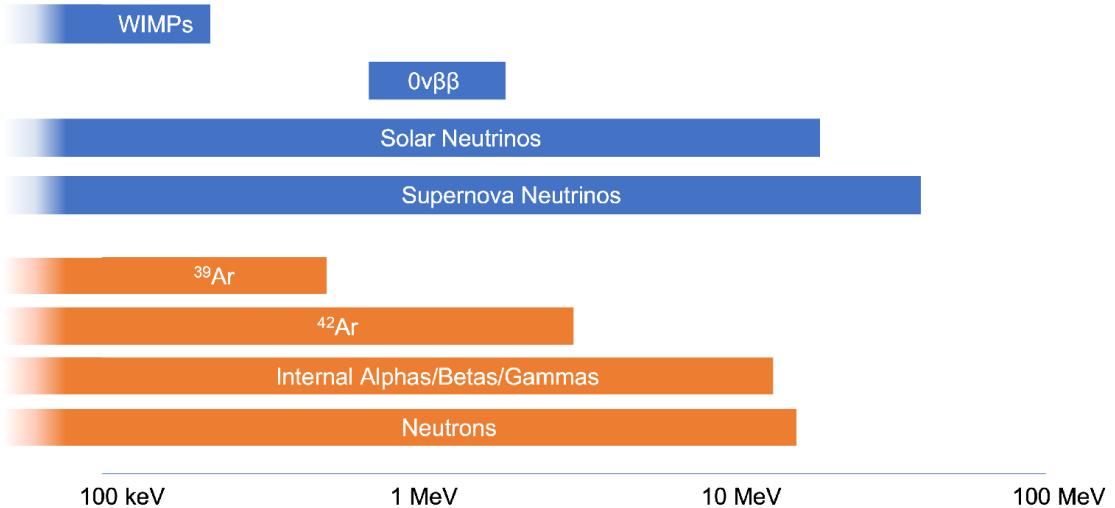


Figure 2.10: Energy range of interest for low-energy analyses and the principal backgrounds on a large scale LArTPC [57].

The $0\nu\beta\beta$ backgrounds for DUNE that are in common with the other low energy analyses can be classified as follows [55]:

- **Environmental radioactivity:** radiation from the rocks surrounding the detector, the cryostat and the field cage. Fiducialization plays a crucial role in the suppression of external radioactivity. In particular, gammas emitted by uranium and thorium are reduced by a factor bigger than 10^9 with 2.5 m. On the other hand, neutrons originating from the rock undergoing alpha capture (α, n) poses a bigger threat, as they could penetrate deep inside the bulk of liquid argon due to the mean free path on the order of meters and the low neutron capture cross section in argon.

An effective reduction of this background could be achieved by surrounding the cryostat with passive neutron-absorbing shielding. Studies have shown that a one-meter water equivalent should be enough to suppress the neutron background near the $Q_{\beta\beta}^{136Xe}$ energy region, as can be observed in figure 2.11.

- **Argon contamination:** 39-Ar is naturally present in atmospheric argon, with a half-life of 269 y, and undergoes a β decay with an endpoint at 565 keV. Despite the low energy, the high rate ($\sim 1 \text{ Bq/kg}$) may pile up with signal events, spoiling the reconstruction. In addition, contamination comes from the 222-Rn decay chain. This isotope is a noble gas deriving from 238-U that can infiltrate argon from the cryostat. In particular, the β decay of 214-Bi into 214-Po has an endpoint of 3.27 MeV. Fortunately, 214-Po decays with a half-life of 164 μs .

emitting a 7.7 MeV α particle, allowing a high-performance suppression with temporal coincidence.

Arguably the most problematic background, especially for the 136-Xe $0\nu\beta\beta$ channel, consists of 42-Ar contamination. Argon extracted from the atmosphere contains a fraction of the 42 isotope of less than 6×10^{-26} , with a lifetime of 33 y. 42-Ar decays into 42-K with $Q = 599$ keV. 42-K decays into 42-Ca with an endpoint of 3.53 MeV. Due to the long 42-K half-life of 12 h, the possibility of background suppression through coincidence is forbidden.

A way to deal with the 42-Ar background consists of employing underground extracted argon, which is depleted in the 42 and 39 isotopes. Unfortunately, the possibility of providing a sufficient amount of argon for a DUNE-scaled LArTPC is not guaranteed.

- **Cosmogenically-activated radioisotopes:** muon hadronic showers create unstable nuclei that can decay by β emission. In the underground DUNE FD the muon flux is suppressed down to 0.2 Hz. A further large reduction can be achieved by applying selection criteria: rejecting events within 2 m of a muon track occurring in a 60 s time window, and a vetoing on a coincidence with a photon detected within 32 cm distance from the candidate. Examples of unstable nuclei are 137-Xe, 39-Cl and 41-Ar.

Among all the previously cited backgrounds, the ones occurring with the highest frequency are 42-Ar and neutrons from the external cavern radiation. Figure 2.11 shows the effect of working with 42-depleted argon and the installation of a passive absorber for neutrons.

Two additional backgrounds, specifically for the $0\nu\beta\beta$ search, must be taken into consideration:

- **Solar neutrinos:** neutrino-electron elastic scattering, charged (CC) and neutral (NC) current neutrino-nucleus interactions with 40-Ar nuclei, especially from the 8-B reaction. CC and NC would mostly appear in coincidence with a cluster of low-energy hits due to the emission of de-excitation photons, exploitable for event rejection. The elastic scattering instead will produce single electrons that mimic the signal. Moreover, solar neutrino capture by 136-Xe will produce 136-Cs, β -decaying with $Q = 2.5$ MeV. This background could be also tagged due to the presence of other electrons and photons in the 136-Cs decay chain.
- **Two-neutrino double beta decay:** the $2\nu\beta\beta$ predicted by the Standard Model can be considered as an irreducible background, as it only depends on the energy resolution at the $Q_{\beta\beta}^{136Xe}$. When the two neutrinos are produced with low momentum, the event becomes indistinguishable from the signal.
For any achievable resolution, the tail of the two-neutrinos energy spectrum will overlap the peak region of the signal. This is the reason why finding evidence of the $0\nu\beta\beta$ process consists of finding a statistically significant excess of events in the signal ROI, as discussed in section 2.2.

The remaining backgrounds, after 42-Ar depletion, the use of neutron shielding and the application of all the other mitigation techniques discussed above, are shown in figure 2.13. Assuming 1% energy resolution, the dominant background would be the solar neutrino elastic scattering and 137-Xe generated by spallation. For resolution worse

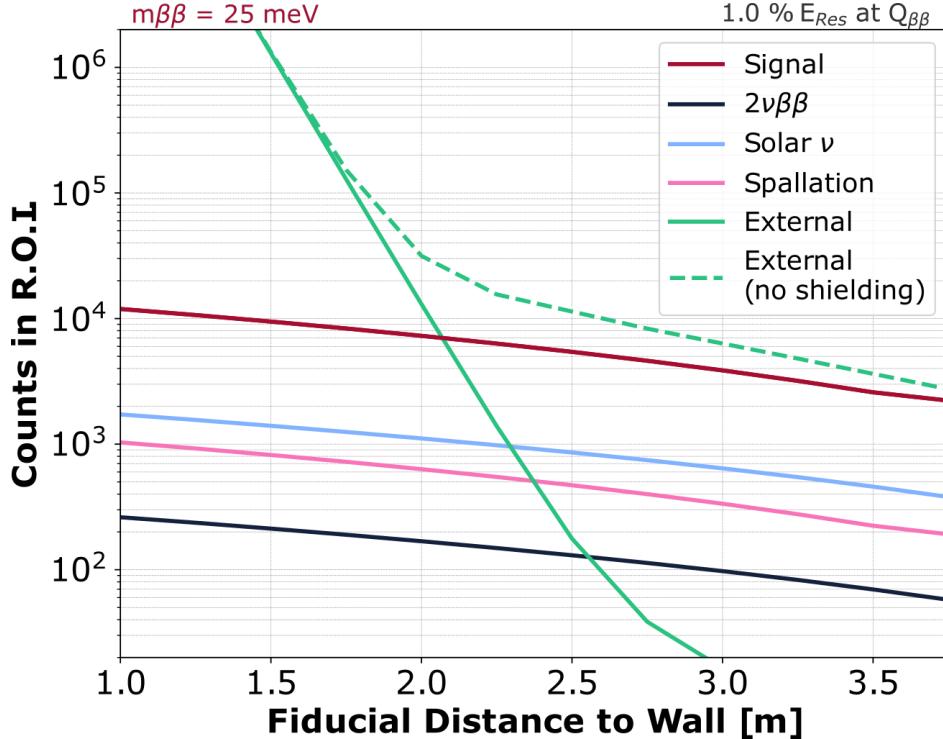


Figure 2.11: Counts in a $0\nu\beta\beta$ region of interest, assuming a 1% energy resolution at the $Q_{\beta\beta}^{136Xe}$ [55]. The ROI covers the interval $[2.41 \text{ MeV}, Q_{\beta\beta}^{136Xe} + 3\sigma]$. The effect of passive shielding, combined with a fiducialization distance from the edge around 2.5 m, is an excellent suppression of neutron background. Signal events are generated assuming $m_{\beta\beta} = 25 \text{ meV}$, corresponding to $T_{\beta\beta}^{0\nu} = 10^{27} \text{ y}$.

than 2.5%, the $2\nu\beta\beta$ tail will dominate. To summarize, DUNE’s sensitivity to $0\nu\beta\beta$ searches by considering an opportunity FD module with xenon-doping at 2%-level, enriched up to 90% of 136-Xe isotopic concentration, is potentially very high. Figure 2.14 shows the expected limits on $m_{\beta\beta}$ for different exposure time and energy resolution. Remarkably, for an energy resolution of 2% or better in a 10 year-exposure time, DUNE sensitivity surpasses the lifetime limit of 10^{29} y , corresponding to a $m_{\beta\beta}$ around 2 meV. This would be an outstanding improvement from the current best limit, set by KamLAND-Zen, of $1.07 \times 10^{26} \text{ y}$ ($m_{\beta\beta} = 61 - 165 \text{ meV}$). With a less rigid resolution constraint (6%), the sensitivity is still excellent (around $2 \times 10^{28} \text{ y}$, $m_{\beta\beta} = 6.6 \text{ meV}$) in comparison to other next-generation experiments like KamLAND2-Zen [69] and DARWIN [70], estimating to surpass the 10^{27} y scale or nEXO [71] and NEXT [72], planning to reach the 10^{28} y limit in the next decade. Such results would completely cover the allowed region for the inverted ordering shown in figure 2.14.

At the current status of DUNE, neutron shielding and depletion of 42-Ar are not confirmed. Moreover, FD LAr-TPCs design will not be optimized for the few-MeV energy scale, as already mentioned in this section. Hence, for this type of physics, it is necessary to operate at the limit of the detector’s capabilities. The energy resolution at the few-MeV scale has not been fully characterized yet, although it is known that the sole ionization signal would lead to an energy resolution far worse than the one requested for being sensitive to the $0\nu\beta\beta$ channel. Despite this, achieving a highly efficient light detection system could drastically improve the low-energy resolution and will be a crucial aspect for addressing the true feasibility for the $0\nu\beta\beta$ search, as well as for all the

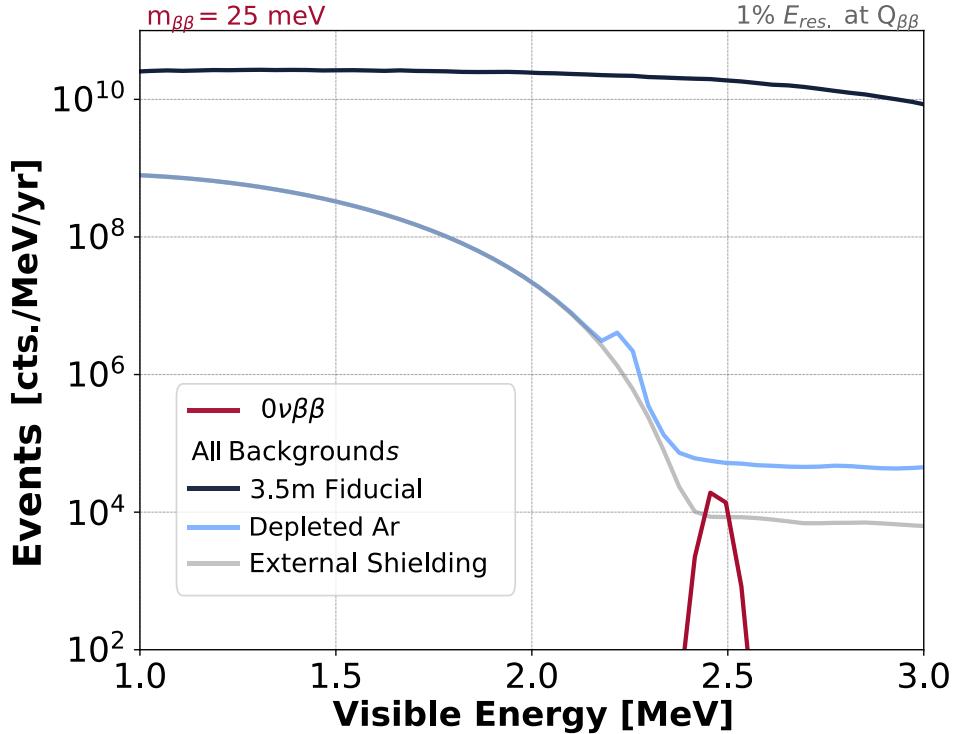


Figure 2.12: Energy distributions of all backgrounds for the $0\nu\beta\beta$ analysis, considering $m_{\beta\beta} = 25 \text{ meV}$ and a ROI of $[2.41 \text{ MeV}, Q_{\beta\beta}^{136Xe} + 3\sigma]$ for 1% energy resolution [55]. The effects of argon depletion and neutron shielding, which would suppress the dominant backgrounds, are shown.

other low-energy channels.

In addition to this, also signal-to-noise ratio in the charge readout plays an important role in determining the energy resolution, as can be seen in figure 2.15. For reference, similar experiments like MicroBooNe and ProtoDUNE report peak signal-to-noise ratios of 37.9 and 48.7 on the collection plane. A more detailed discussion about DUNE detectors and their potential in the analysis of the $0\nu\beta\beta$ decay of the 136-Xe is beyond the purpose of this thesis. The next section will introduce an additional strategy for background mitigation, apt to reduce 42-Ar and neutron contributions in the absence of argon depletion and neutron shielding. This proposal represents one of the main topics of the thesis work, and the results of the analysis will be presented in chapter 3.

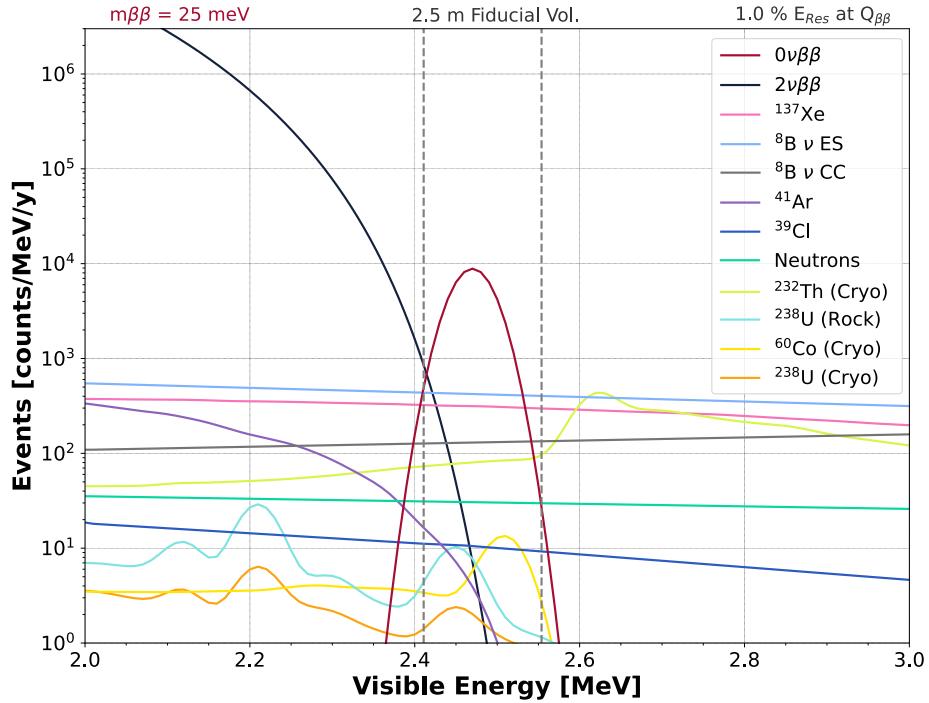


Figure 2.13: Energy distributions of all backgrounds for the $0\nu\beta\beta$ analysis after applying all the mitigation strategies discussed in this section [55]. $m_{\beta\beta} = 25$ meV, and the ROI corresponds to the interval $[2.41 \text{ MeV}, Q_{\beta\beta}^{136Xe} + 3\sigma]$, indicated by the dashed gray lines, for 1% energy resolution.

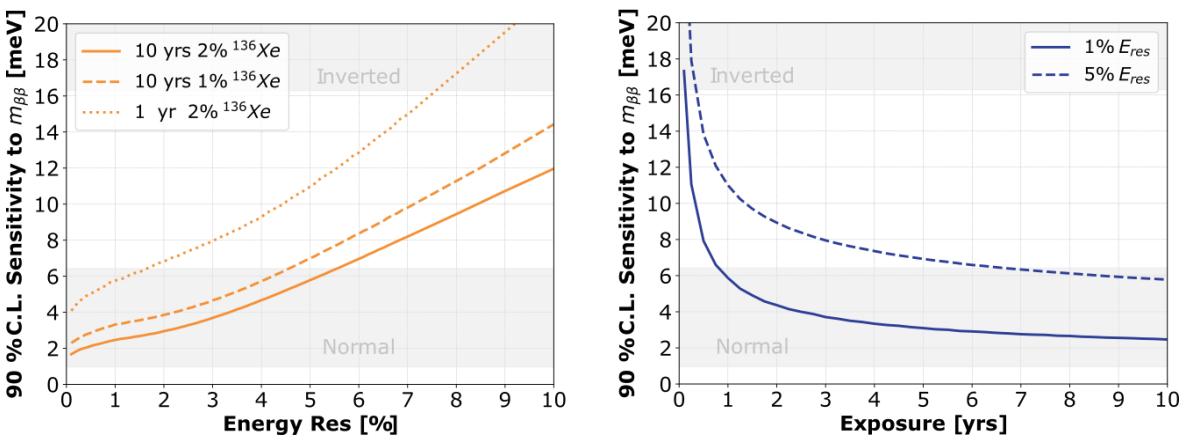


Figure 2.14: 90% CL upper limits on the $m_{\beta\beta}$ parameter, for different scans in energy resolutions (left) and exposures (right) [55]. The darker bands indicate the allowed values in the neutrino mass normal ordering and inverted ordering regimes.

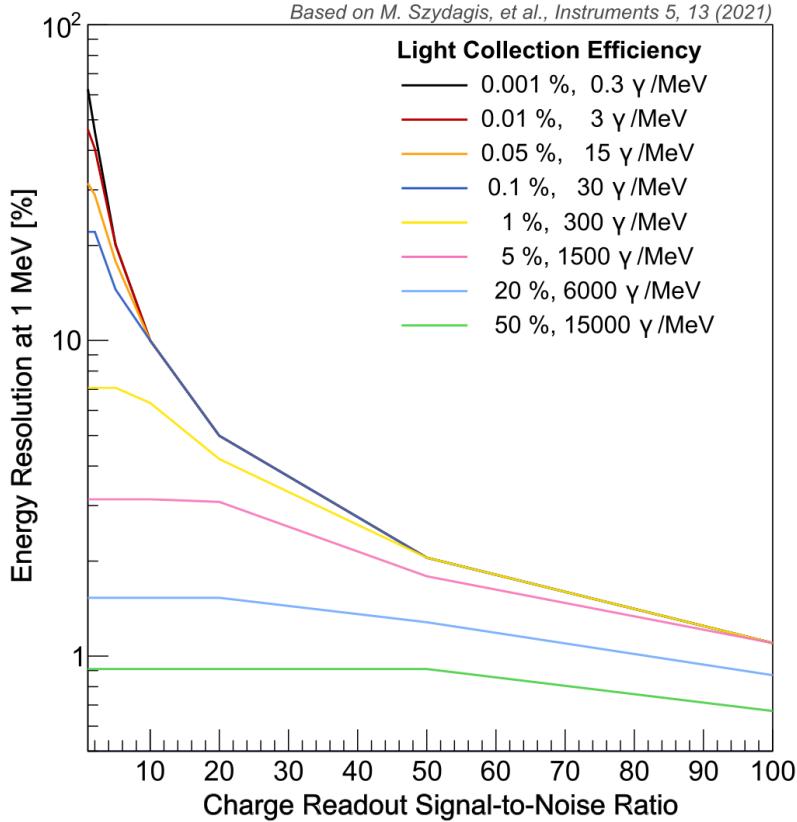


Figure 2.15: Energy resolution of a 1 MeV electron in a LArTPC for different signal-to-noise ratio and light collection efficiency [55].

2.3.3 Topological background mitigation at the MeV scale

Time Projection Chambers provide a tridimensional reconstruction of tracks and showers generated by the interaction of particles inside the detector medium. A track can be considered as the ensemble of energy depositions by ionization, occurring in a specific position inside a TPC, and associated with the same event by a reconstruction algorithm (e.g. a neutrino CC interaction or the passage of an atmospheric muon) [73]. Tracks are reconstructed by combining, *hits* measured by the TPC, i.e. charge signals collected at the wire planes. Each measured hit comes with the information on which wire and at what time a signal was produced, allowing for a full spatial reconstruction when considering at least two non-parallel wire planes.

The efficient tracking system of DUNE is expected to allow high-performance event tagging at GeV scale interactions, like in neutrino flavour identification [74]. These kinds of events produce in LArTPCs a multitude of hits that make up tracks with lengths on the order of 1 m. Considering the Far Detector wire pitch of 5 mm and the sampling time of 500 ns, the tracking of GeV-scale events provides a reconstruction in great detail of the interaction vertices, allowing for identifying the produced particles and classifying the event by topology with Visual Processing models.

At the MeV scale, the situation is far more challenging due to the small-sized tracks, almost comparable with FD LArTPCs sampling resolution. For example, an electron with the kinetic energy of $Q_{\beta\beta}^{136Xe}$ in liquid argon (density $\rho = 1.40 \text{ g/cm}^3$) possesses a CSDA range (i.e. the path length traveled as calculated with the Continuous Slowing Down Approximation [75]) of about $R_{\text{CSDA}} = 1.2 \text{ cm}$ [76]. Nevertheless, topologi-

cal background mitigation can also be attempted for the $0\nu\beta\beta$ analysis, investigating whether track differences can be leveraged to some extent, despite the experimental limitations.

Without argon depletion and neutron shielding, 42-Ar is the dominant background for the 136-Xe $0\nu\beta\beta$, and neutrons are subdominant. In a first approximation, it is possible to consider only the 42-Ar background and exploit different kinematics of the event. The problem reduces then to distinguishing a $\beta\beta$ track with a total energy of $Q_{\beta\beta}^{136Xe} = 2.458$ MeV and a single- β track, with an electron kinetic energy of $\sim Q_{\beta\beta}^{136Xe}$. In particular, three characteristics justify the different topology between β and $\beta\beta$:

- **$\beta\beta$ emission angle:** the $0\nu\beta\beta$ is a three-body decay (as shown in figure 2.16) where the nucleus recoil energy is negligible, but not its momentum. This determines an angle between the initial trajectories of the two emitted electrons, that a single β does not feature.
- **Different energy loss per unit-length:** ~ 1 MeV electrons will lose more energy far from their origin, just before stopping. The $\beta\beta$ signature will contain two distinct higher energy deposits, while the β 's only one.
- **Multiple scattering angles:** MeV electrons interacting with a medium undergo multiple scatterings against the electrons of the medium. The average scattering angles increase as the incident electron momentum decreases, therefore in $\beta\beta$ events this increase would be seen in both ends, while for β only in one.

The last two points are merely a consequence of discriminating between an event consisting of two tracks originating from the same point, and a single bigger track, without assumptions about the type of interaction.

All of these distinctive topological traits are subject to considerable fluctuations,

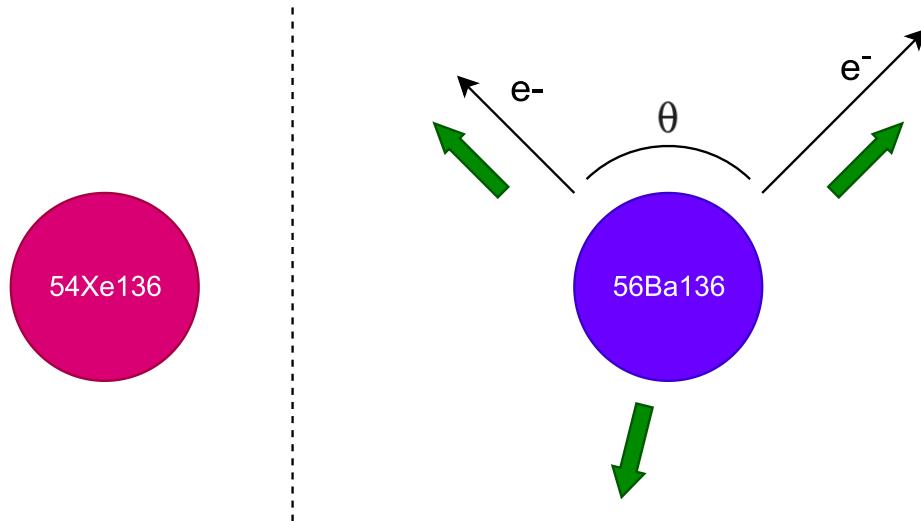


Figure 2.16: Drawing of a $0\nu\beta\beta$ decay of the 136-Xe into 136-Ba. The nuclear recoil determines an emission angle θ between the two electrons in the final state.

complicating the background rejection independently from the spatial resolution of the TPC. For example, an electron from a β decay may scatter at a great angle, causing an abrupt change in the particle trajectory that mimics the emission angle between

the electrons of the $\beta\beta$ event.

Moreover, in some cases the emission angle between the two $\beta\beta$ electrons will be close to π and will not help the classification, as can be inferred from the differential decay rate phase space factor [77]:

$$\frac{d\Gamma_{\beta\beta}^{0\nu}}{dE_1 d\cos\theta} \propto \frac{1}{16\pi^5} F(E_1, Z) F(E_2, Z) dE_1 d\cos\theta E_1 E_2 p_1 p_2 \left(1 - \frac{\vec{p}_1 \cdot \vec{p}_2}{E_1 E_2}\right) \quad (2.23)$$

where $E_1, E_2, \vec{p}_1, \vec{p}_2$ are energies and momenta of the two electrons, θ the emission angle and $F(E, Z)$ is the Fermi function that takes into account the Coulomb attraction between electrons and the nucleus.

Finally, energy loss by ionization per unit-length dE/dx fluctuates according to the Landau distribution in the moderate thickness absorber approximation [78]. The FWHM is approximately 4ξ , where:

$$\xi = 2\pi N_A r_e^2 m_e c^2 \frac{Z}{A} z^2 \left(\frac{x}{\beta^2}\right) \text{ MeV} \quad (2.24)$$

where r_e is the classical electron radius equal to $e^2/4\pi\epsilon_0 m_e c^2$, A and Z are the mass and atomic numbers, x is the considered travelled length inside the medium.

This chapter discussed the double beta decay without neutrino emissions as a BSM probe and as a relevant physic channel for improving the present understanding of neutrinos and their properties. The DUNE experiment and its potential sensitivity for the search of the $0\nu\beta\beta$ decay of the ^{136}Xe were illustrated, together with the proposal of a background mitigation strategy based on topological information reconstructed by an opportunity module DUNE Far Detector.

The next chapter will present the work carried out for the present thesis, consisting of the use of advanced deep learning models and Quantum Machine Learning, already introduced in chapter 1, for achieving a signal-to-background classification.

Chapter 3

Low-energy background mitigation in LArTPCs

This chapter describes the implementation of the Machine Learning techniques, already discussed in chapter 1, for achieving automated background mitigation in the analysis of the $0\nu\beta\beta$ decay in a DUNE-like LArTPC as reviewed in chapter 2. The present work aimed to explore the interplay between purely computational fields and experimental particle physics. The objective of this thesis, which was to investigate forefront Deep Learning models and assess quantum computing performance, is driven by a physical interest in enhancing the sensitivity of $0\nu\beta\beta$ decay experiments. In particular, the goal consisted of evaluating the benefits of the topological background rejection in the low-energy sector for the DUNE Horizontal Drift Far Detector. This chapter illustrates the methods and results achieved for the design of heuristic and meta-heuristic Quantum Kernels and applying them in the QSVM algorithm, undergoing a systematic comparison with classic SVM. This chapter also reports a detailed characterization of the Convolutional Neural Network and the Transformer in their capability to extract features and perform track identification under different resolution conditions.

3.1 Toy model dataset description

Most of the work presented in this chapter was carried out by using a two-class dataset consisting of $\beta\beta$ signal and β background, i.e. the two-electron signature of the $0\nu\beta\beta$ decay and single electron one corresponding to a β decay with energy $Q_{\beta\beta}^{136Xe}$. The tracks were generated uniformly in a volume of liquid argon using a Monte Carlo simulation developed with Geant4 toolkit [79]. While this simulation takes accurately into account the spatial and energetic distributions of the electrons and their interaction with the medium, detector effects are not included. For this simplification, the dataset is considered a toy model and can be used to set experiment-independent benchmarks. Despite the only partial fidelity to realistic data collected by a LArTPC, this simulation benefits from high versatility and control of parameters like wire pitch and time resolution, which will be exploited in section 3.4.

The Geant4 tracks consist of a collection of steps, each one characterized by its position in space and associated energy loss by ionization and *bremsstrahlung*. Information about the energy lost in the interaction and the position in cartesian coordinates allows

for reconstructing high-resolution trajectories, as shown in figure 3.1.

For approximatively taking into account the track reconstruction of a DUNE-like

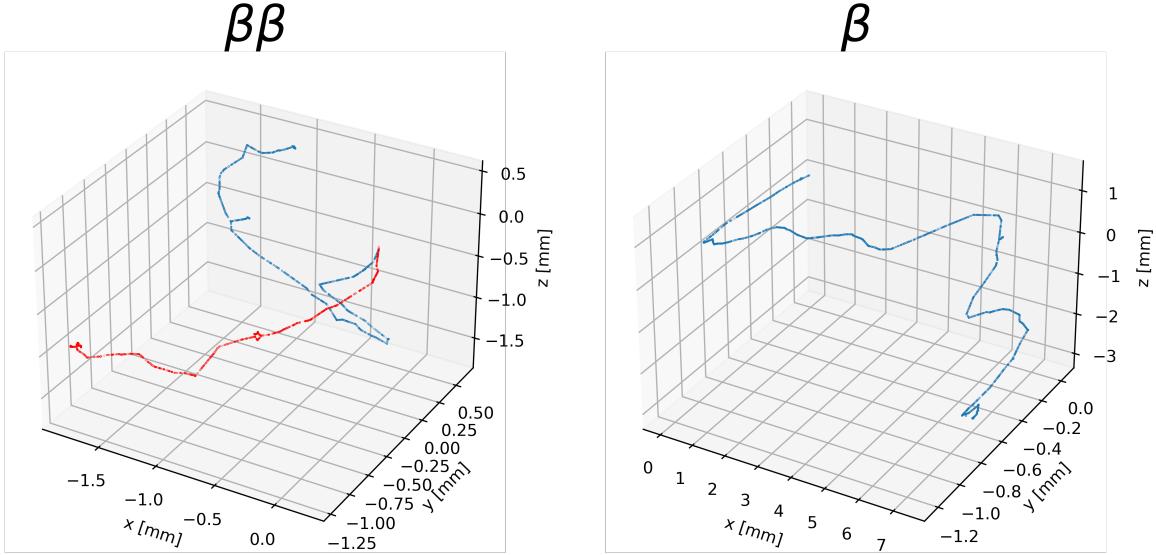


Figure 3.1: Examples of a $0\nu\beta\beta$ decay (left) and a single β decay (right) in liquid argon simulated with the geant4 toolkit. The red and blue colours in the $\beta\beta$ indicate two distinct electrons to show the emission angle introduced in section 2.3.3.

TPC, the trajectories were downsampled to a resolution of $[5 \times 5 \times 1] \text{ mm}^3$, according to the sampling time of charge signals and the wire pitches mentioned in section 2.3.1. With this procedure, the samples were converted into collections of hit-like energy depositions and the results are shown in figure 3.2.

Tracking in TPCs relies on the informative content brought by the charge signal on the induction and collection planes. At a first degree of approximation, for taking into account the conversion of energy lost in the medium into the number of charge carriers, a Poissonian fluctuation was added to the hits energies. Moreover, an energy threshold of 100 keV was applied to filter out the least energetic hits. Such a threshold value principally eliminates the energy depositions due to bremsstrahlung photons, which are not expected to be detectable in a DUNE FD. Moreover, due to the 14 cm radiation length of liquid argon, such depositions appear distant from the ionization hit cluster and would be impractical to associate them with the event. Subsequently, undesired samples were eliminated, such as events that remained empty after applying the threshold or decays occurring close to the edge of the detector where electrons escaped. The preprocessed dataset has been split according to the table 3.1.

As can be observed in figure 3.2, the downsampling to DUNE-like spatial reconstruc-

Class	Training	Validation	Test	Total
$\beta\beta$	6977	1504	1515	9996
β	7701	1641	1631	10973

Table 3.1: Dataset partition of the 20969 events into training (70%), validation (15%) and test (15%) subsets.

tion leads to a dramatic loss of information about the particles' trajectories. The two

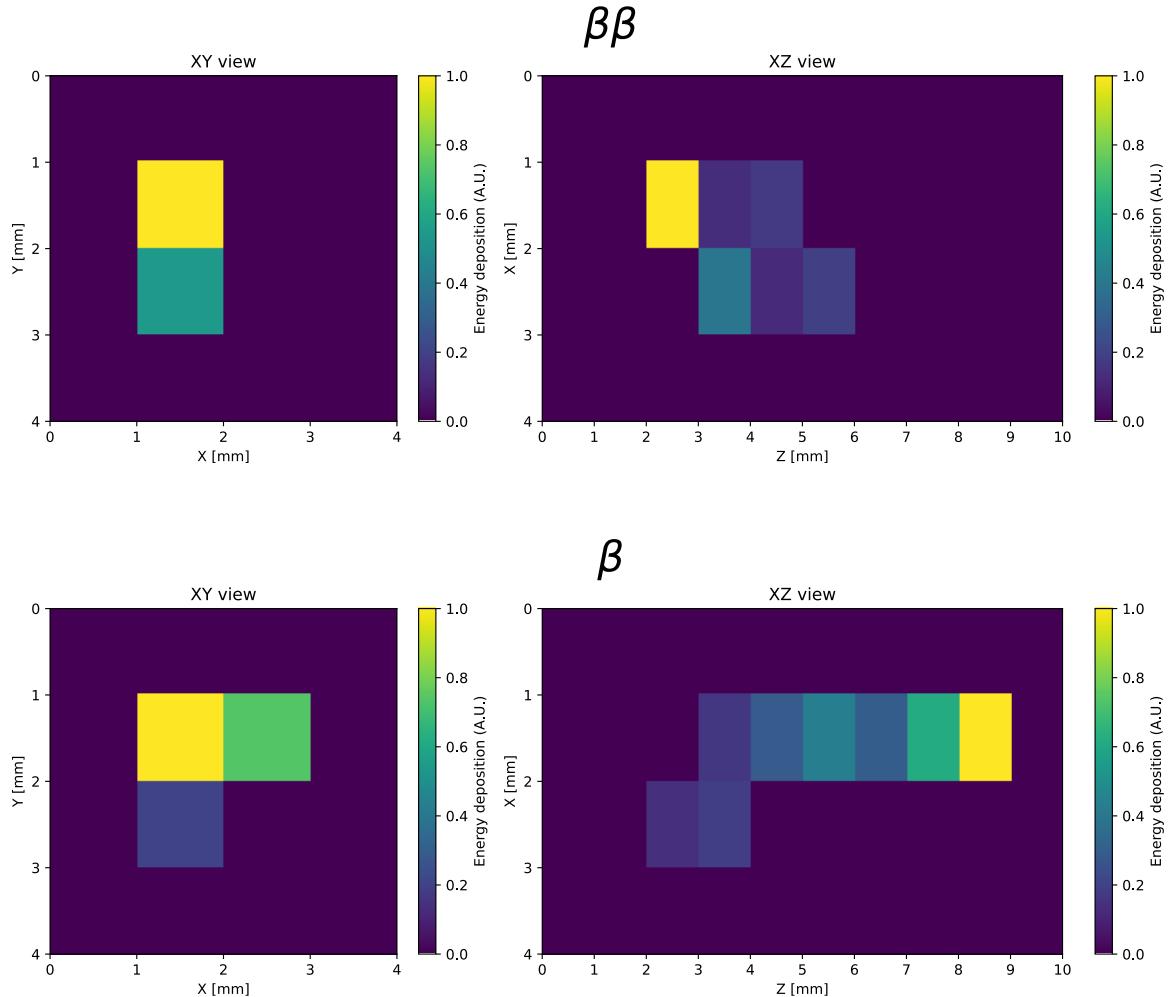


Figure 3.2: Examples of $\beta\beta$ (up) and β (down) events after downsampling to a voxel size of $[5 \times 5 \times 1]$ mm 3 . The pictures show the total 3D information projected onto the XY and XZ planes (the YZ view is omitted). Each square corresponds to a *hit*, and the colour to the corresponding energy content.

topologies become indistinguishable to the naked eye, and the feature distributions are expected to significantly overlap, making the classification problem challenging.

An interesting distribution to take into consideration is the number of hits per event. Figure 3.3 shows that, on average, background events (β) contain a higher number of hits, i.e. are more extended in space. It is possible to explain this characteristic by taking into consideration the observations about the electrons' energy loss made in section 2.3.3. Just selecting the events with a number of hits < 7 would correspond to a classification accuracy of about 58%, an efficiency (sensitivity) of 76% and a purity (specificity) of 41%. A definition of all the classification metrics used in this thesis is provided in appendix A.

For its physical intuition and its simplicity, the total number of hits can be used to

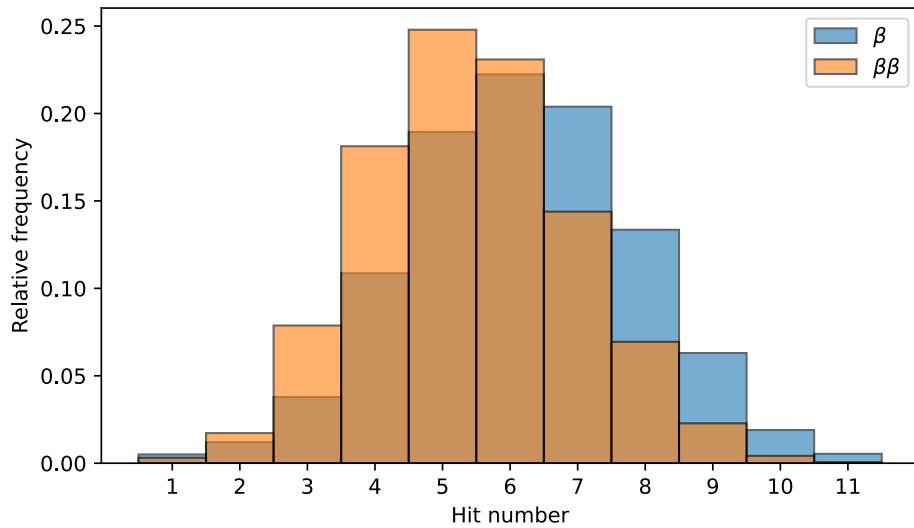


Figure 3.3: Distribution of the hit number per event in the toy-model dataset, considering the downsampling $[5 \times 5 \times 1]$ mm³. The mean of the $\beta\beta$ class is $\mu_{\beta\beta}^{\text{nhits}} = 5.44$, with $\sigma_{\beta\beta}^{\text{nhits}} = 1.56$, while for the β one $\mu_{\beta}^{\text{nhits}} = 6.17$, and $\sigma_{\beta}^{\text{nhits}} = 1.74$.

investigate its relationship with the features extracted from the CNN and the ATN, or to test a possible improvement in the SVMs performances when feeding it as an additional input. The second approach was tested, by training a classical gaussian kernel SVM with and without the hit number feature. The classification performances did not increase neither using CNN or ATN as feature extractor models, meaning that the information brought by the hit number distribution was already embedded in the deep features. A different result would have indicated an inefficiency in the feature extractor methods.

3.2 CNN and ATN feature extractor

In this thesis, Convolutional and Attention Neural Networks had the purpose to extract a reduced number of significant features from the dataset in order to make the QSVM work with few qubits and small-sized quantum circuits. This was achieved by training the networks as binary classifiers and extracting features from an intermediate layer,

as discussed in section 1.3. This procedure not only reduces the feature dimensionality of the original dataset but also increases the classification performance, as will be demonstrated at the end of this section.

CNN and ATN architectures must be complex enough to detect nontrivial correlations between hits of the same events, and at the same time contain as few learnable parameters as possible to reduce the computational resources needed, without penalizing the performance. Another fundamental aspect to consider is avoiding overfitting, i.e. the model learning nuisance structures that are specific to the training dataset, negatively affecting the generalization to unseen data.

3.2.1 CNN architecture

For the CNN approach, in order to reduce the number of training parameters, the voxelized tracks are projected on the planes XY, XZ and YZ, and treated as separate images. The proposed architecture for the Convolutional Neural Network is shown in figure 3.4 and summarized as follows:

- **Input layers:** three parallel layers take as input the projections XY, ZY, and YZ. The XY plane corresponds to information retrieved only by the wire planes ($[5 \times 5] \text{ mm}^2$) and the associated images have resolution $[4 \times 4]$. The XZ and YZ planes correspond to one of the wire planes and the temporal information ($[5 \times 1] \text{ mm}^2$). The latter allows for a more precise spatial reconstruction, and bigger-sized images are required $[4 \times 20]$.
- **Convolution blocks:** each input layer is connected to two stacks of convolution blocks, consisting of:
 - Convolutional layer, with 35 filters of size 3×3 each.
 - Batch Normalization layer
 - LeakyReLU activation function, defined as:

$$f(y) = \begin{cases} y & \text{if } y \geq 0 \\ \alpha y & \text{if } y < 0 \end{cases} \quad (3.1)$$

with $\alpha = 0.05$.

- Dropout layer, with a neuron dropout rate of 0.25.
- **Flatten and concatenate:** all tensors are reshaped and concatenated into a single array.
- **Feature layer:** a fully connected layer. The number of neurons corresponds to the number of input features to the SVMs and QSVMs.
- **Leaky ReLU activation function,** with $\alpha = 0.05$, analogously to the ones of the convolution blocks.
- **Prediction layer:** consisting only of one neuron. Its value determines the network's prediction.

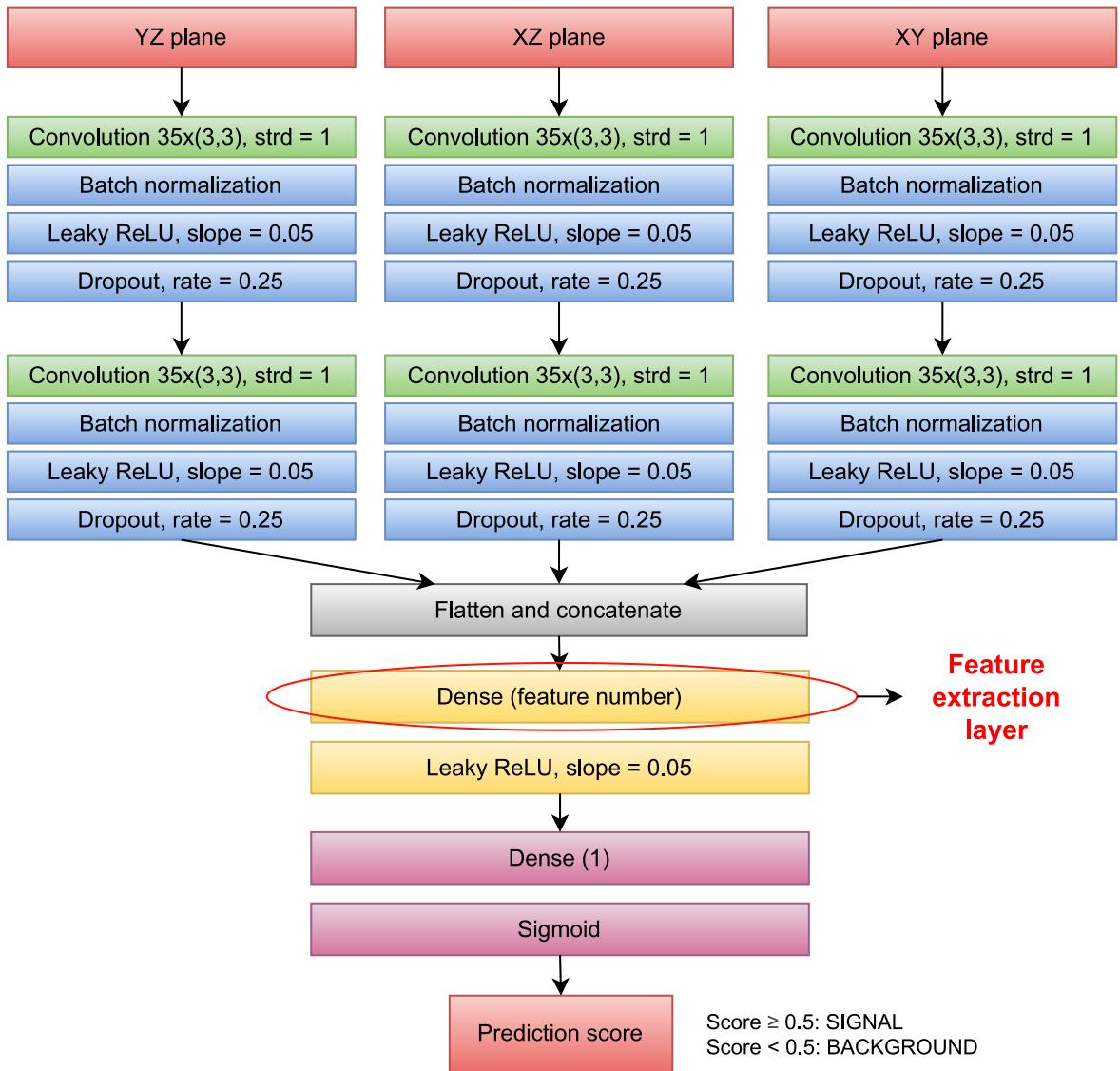


Figure 3.4: Convolutional Neural Network architecture. The two-dimensional track projections undergo independently through convolution steps and the results are merged together. Both feature extraction and binary classification are achieved with the use of the corresponding dense layers.

- **Sigmoid activation function:** nonlinear mapping of the neuron value into the final prediction score:

$$\sigma(y) = \frac{1}{1 + e^{-y}} \quad (3.2)$$

Although projecting the original samples into two-dimensional data leads to an information loss, the small size of the resulting pictures compensates by making the usage of pooling layers unnecessary.

3.2.2 ATN architecture

The Attention Network analyzes tracks as sequences of hits. For every hit associated with an event, the Transformer takes in input the $X Y Z$ coordinates and the relative energy deposit. This approach is more cost-efficient than the CNN visual one, as no data projection is needed, working with the full 3-D information.

The Transformer architecture shown in figure 1.6 is general-purpose, while the ATN employed in this chapter differs slightly in the decoding part, which takes the output from the encoding stack and returns the prediction.

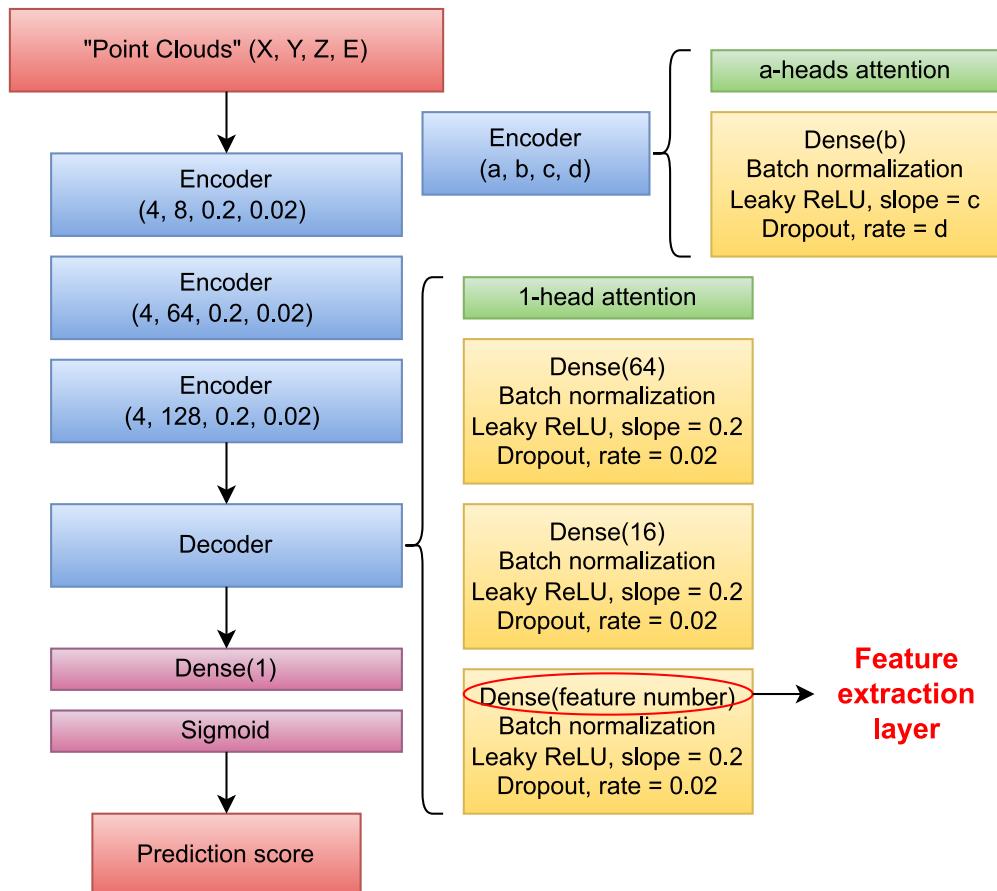


Figure 3.5: Attention Neural Network (Transformer) architecture. Variable sequences of hits are processed for extracting features and classifying samples with the mechanism of self-attention.

- **Input layer:** a layer that accepts input of variable size, in the format $[X, Y Z, E] \times$ hit number.
- **Three encoding blocks**, consisting of:
 - 4-head attention layer, where the attention mechanism described in section 1.3 takes action.
 - Dense layers with an increasing number of neurons(8, 64, 128).
 - Leaky ReLU activation function, already defined in equation 3.1, with $\alpha = 0.2$.
 - Dropout layer, with a neuron dropout rate of 0.02
- **One decoding block**, consisting of:
 - single-head attention layer.
 - three sequences of dense - batch normalization, leaky ReLU ($\alpha = 0.2$) and dropout (rate = 0.02) layers. The first two dense layers have 64 and 16 neurons respectively, while the third one is the feature extraction layer, with an arbitrary neuron number.
- **Prediction layer:** consisting only of one neuron. Its value determines the network's prediction.
- **Sigmoid activation function:** already defined in equation 3.2, that returns the final prediction score.

3.2.3 Performances of feature extractor networks

Figure 3.6 shows the score distribution of CNN and ATN on the toy model dataset. After several runs with different weight initialization, CNN slightly outperforms the ATN, and both models settle around 65% accuracy.

To contextualize these results, the latter were compared with the ones obtained from a much simpler alternative approach, avoiding Deep Learning extractors. We considered the most elementary kind of features to provide to SVM, which are the positions and the energy content of each hit. Since the highest number of hits per sample is 12 (see figure 3.3), every event is described by a total of 48 *raw* features (some of them are set to zero in case the track has less than 12 hits). By giving these features to a classical SVM with a gaussian kernel (the latter already defined in section 1.4.1), the result stands around 61% accuracy on the test set. At this point, the PCA technique for feature reduction, necessary for NISQ Quantum Machine Learning, was applied, leading to 58%-59% accuracies by selecting less than 10 principal components.

Both CNN and ATN models have a clear edge on the PCA approach in terms of classification performance, which justifies *a posteriori* the proposed methodology. Another useful result is that CNN and ATN classification accuracy does not depend on the number of features extracted, i.e. the number of neurons in the feature extraction layers, as can be seen in figure 3.7. This allows training QSVMs with different configurations of features/qubits without affecting the overall performances in background rejection.

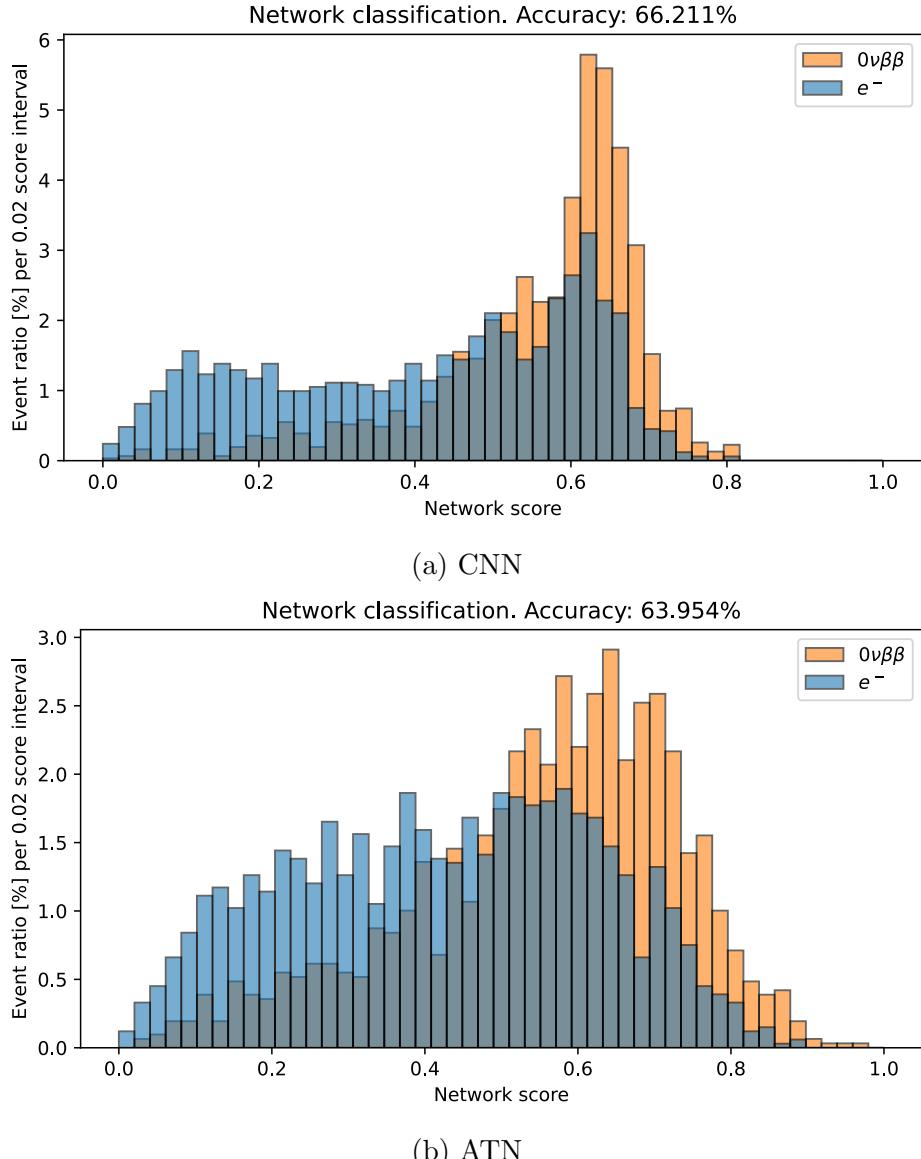


Figure 3.6: Prediction scores of (a) the Convolutional and (b) the Attention Neural Networks on a test set. The class were labeled as 1 for $\beta\beta$ and 0 for β . The considerable amount of overlap between the score distributions of different classes reflects the mediocre accuracies of 66.211% and 63.954% for CNN and ATN respectively. The slight distribution shifts towards 0 are attributed to the slightly unbalanced training dataset (β events are about 10% more than $\beta\beta$ s).

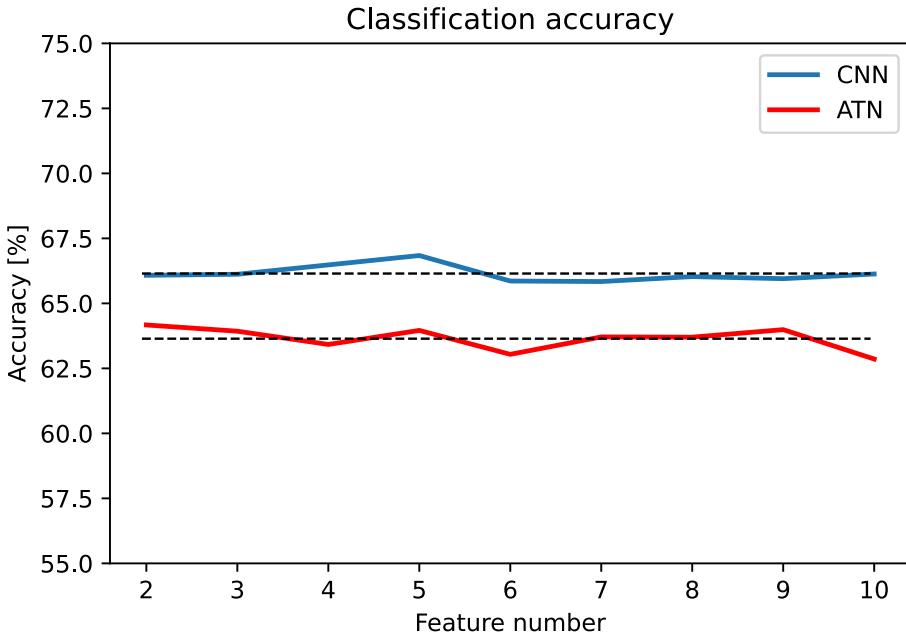


Figure 3.7: CNN and ATN accuracies for different numbers of features extracted. No significant dependence was found for both models.

3.3 QSVM and the search for Quantum Advantage

After developing and characterizing the feature extractor, Support Vector Machines were trained with several kinds of classical and quantum kernels. The goal was to design NISQ quantum kernels leading to good classification capabilities and probe the effects of using different quantum circuits.

The quantum kernels were built by using Qiskit [80], an IBM open-source software development kit, that offers the possibility to work with different backends:

- **Statevector**: ideal quantum circuit simulation, manipulating directly the probability amplitudes.
- **Qasm Simulator**: ideal or noisy quantum circuit simulation, without accessing the probability amplitudes, retrieving only the outcome of a measurement.
- **Real backend**: remote control of a real quantum processor of the IBM series.

Except for the dedicated section 3.3.3, all quantum kernels in this chapter were evaluated with the Statevector backend, as it is the fastest and allows for inspecting kernel properties in their ideal behaviour, without the effects of noise and statistical fluctuations.

Figure 3.8 summarizes the analysis workflow, and the passages that lead to SVM and QSVM predictions. Classical SVMs were trained in order to set performance benchmarks for the QSVMs. The most popular kernels were chosen: linear, polynomial and gaussian, defined in section 1.4.1. The sigmoid kernel was soon omitted in the comparisons because of the lower performances with respect to the other nonlinear ones. Another fundamental aspect, both for classical and quantum Support Vector Machines, was the choice of the hyperparameters. In equations 1.17 and in kernel formulation,

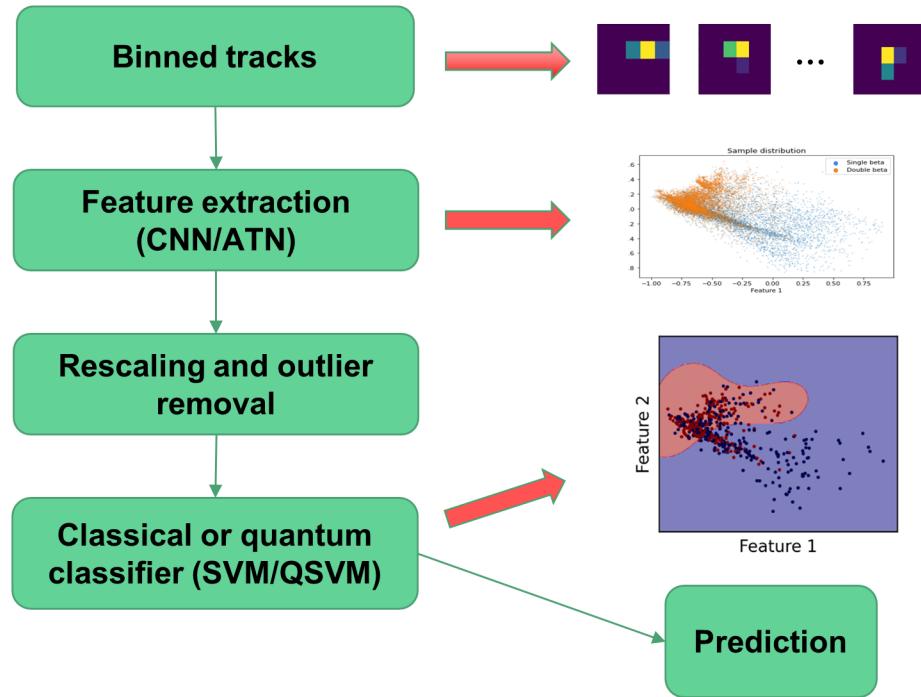


Figure 3.8: Analysis flow, from the hits reconstruction to SVM/QSVM output through a Deep Learning feature extraction. The deep features are all rescaled to fit in the interval $[-\pi/2, \pi/2]$ and the training set undergoes outlier removal.

an initial guess for some parameters must be set before training the model, affecting the overall decision boundary. QSVMs can be tuned by C , the *cost parameter*, that regulates the misclassification penalty, while in classical kernels also other parameters (γ, d, r) must be set. The hyperparameter optimization was carried out with a *grid search cross-validation* for all the results presented in this chapter.

3.3.1 Heuristic Quantum kernels

The first QSVM approach was the implementation of four arbitrary quantum kernels, generated by the feature maps:

- **Z**: first-order, non-entangling Pauli feature map, as a base approach.
 - **ZZ**: second-order Pauli feature map, with linear entanglement structure.
 - **Custom 1 (C1)**: an entangling feature map where phase rotations in some of the gates depend on more than one input feature.
 - **Custom 2 (C2)**: a non-entangling feature map with trigonometric data encoding.

The base structures, for $reps = 1$, two features and two qubits, are shown in figure 3.9. Other examples of more complex configurations are reported in appendix B.

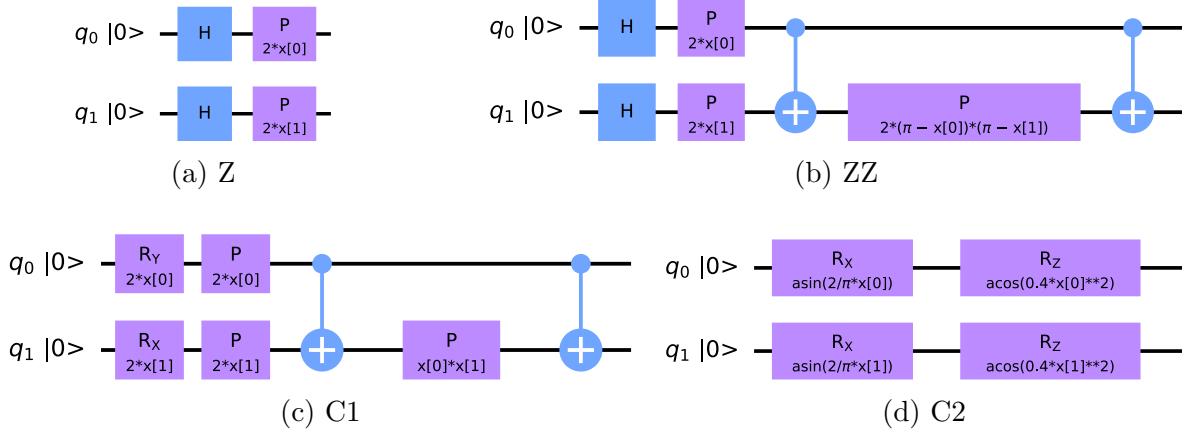


Figure 3.9: The two-qubit, one-repetition heuristic feature maps adopted.

Two features and two qubits

The first training session was carried out by using as input two features extracted from the CNN and the ATN respectively. The advantage of working with two features is the possibility to visualize SVMs' decision boundaries. On the other hand, the use of just two qubits keeps the embedding feature space relatively small, since the quantum state is described by four complex amplitudes. CNN's and ATN's distributions (see figure 3.10) are different from each other and effectively correspond to two distinct classification problems for SVM and QSVM. Feature distributions also depend on the network's random number generator seed, that was fixed for this reason.

For a better understanding of quantum data encoding, and how class separation looks

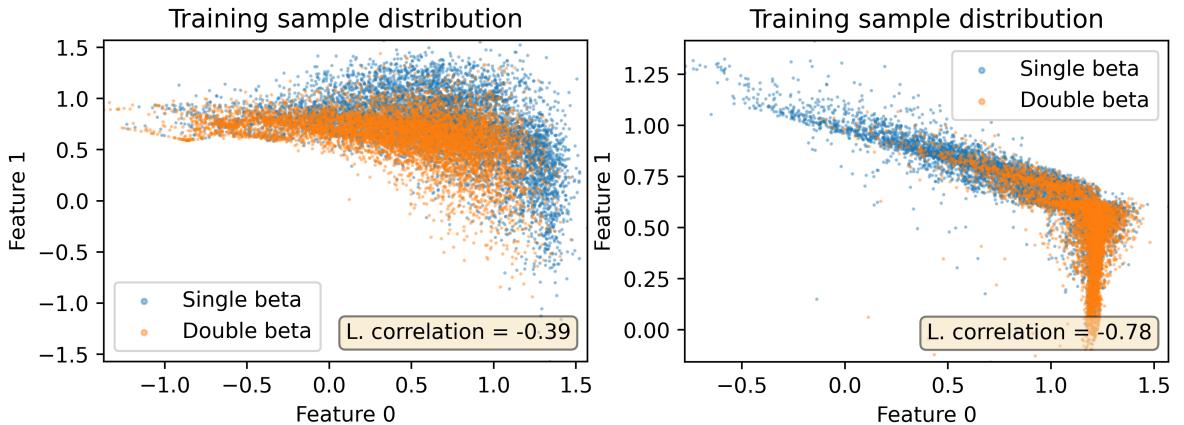


Figure 3.10: 2D feature distributions extracted from CNN (left) and ATN (right). The degree of linear correlation is also displayed.

like on the embedded Hilbert space, we recurred to the visualization of the distributions on the Bloch Spheres, represented in figure 3.11. Here, each qubit corresponds to a different sphere, and each point is the quantum state of the quantum kernel computed for a sample. Mapping multi-qubit quantum states to Bloch Sphere's angles is only possible without entanglement, i.e. qubits evolving independently from each other (Z and C2 kernels). For entangling maps, a visualization is still possible by considering the amplitudes of the single qubits separately for computing the angles. While this representation does not correctly describe the multi-qubit state, it informs approxi-

mately about the degree of class separation.

The distributions shown in figure 3.11 present considerable differences. For instance, the Z circuit maps all the points on the Bloch spheres equators, as expected when applying an H -gate on a qubit in an initial state $|0\rangle$ and then rotating along the Z axis. The other feature maps allow for better coverage of the spheres, especially the ones with entanglement. This suggests that entanglement is not a mere necessary condition when searching for quantum advantage (as already mentioned in section 1.4.3), but also plays an important role in assessing the QSVM potential. That being said, while Bloch spheres remain useful visualization tools, the final statement on the goodness of a feature map is up to the classification performance of the trained QSVM.

An additional technique implemented for optimizing the quantum circuit design was

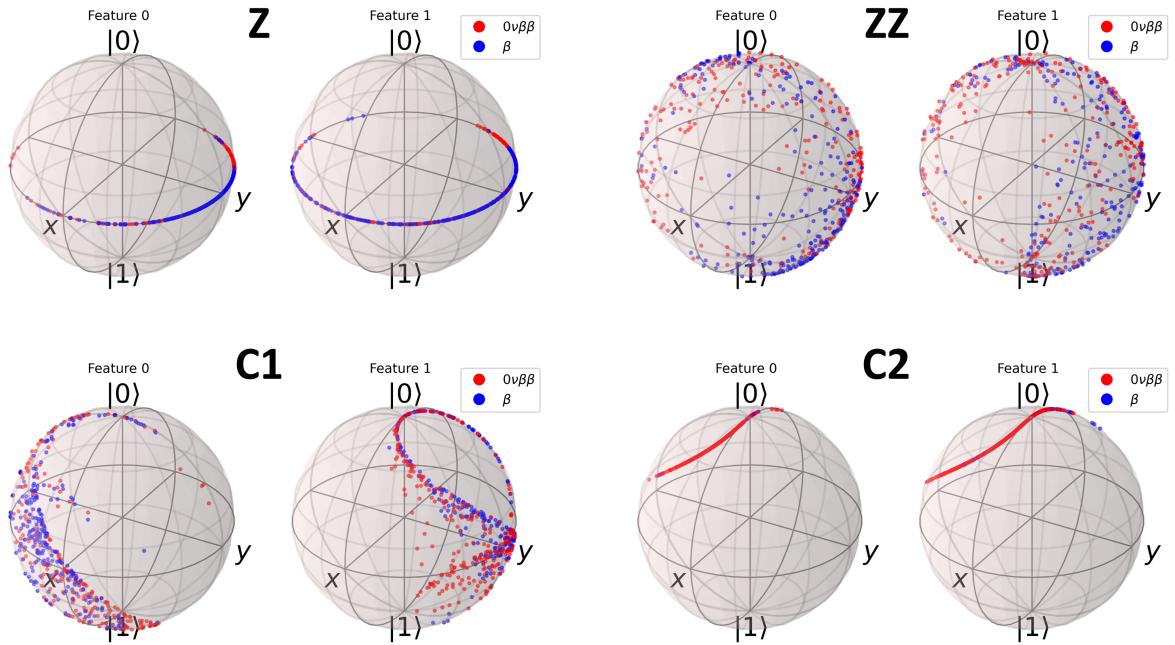


Figure 3.11: Distributions of CNN-extracted features in the Bloch sphere visualization for Z, ZZ, C1 and C2 feature maps, $reps = 1$, two qubits/features, corresponding to the circuits in figure 3.9.

the kernel alignment, presented in section 1.4.4. For each qubit, a new rotation gate R_X was inserted at the beginning of the circuit. The corresponding initial rotation angles $\vec{\theta}$ were treated similarly to hyperparameters of the QSVMs and optimized using the *Simultaneous Perturbation Stochastic Approximation* algorithm (SPSA) [81], to minimise the SVCLoss function:

$$\text{SVC Loss} = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K_{\vec{\theta}}(\vec{x}_i, \vec{x}_j) \quad (3.3)$$

that corresponds to the Lagrangian dual of the SVM problem in equation 1.20, but this time minimised as a function of $\vec{\theta}$. To compute this function it is necessary to train the QSVM and evaluate the quantum kernel function, which is expensive on a Statevector or Qasm backend. For this reason, the SPSA algorithm was introduced. SPSA is a gradient descent method developed for many-parameters optimization, requiring only two evaluations of the loss function for each iteration step independently from the

size of $\vec{\theta}$, i.e. the number of parameters to optimize. Despite being implemented for ideal circuit simulation, SPSA is particularly indicated for noisy circuits, due to its robustness to fluctuations in the loss function evaluations.

Kernel alignment was tested to enhance the QSVM performances by optimizing the SVCLoss on a 500-samples validation subset, giving modest results. An example is shown in figure 3.12 where for the C2 feature map the SVCLos decreases by about 1% after 20 steps, that corresponded to a validation accuracy increase of 0.6%, while for some other cases, no improvement was observed at all.

Finally, after preliminary studies and implementation of optimization routines, it was

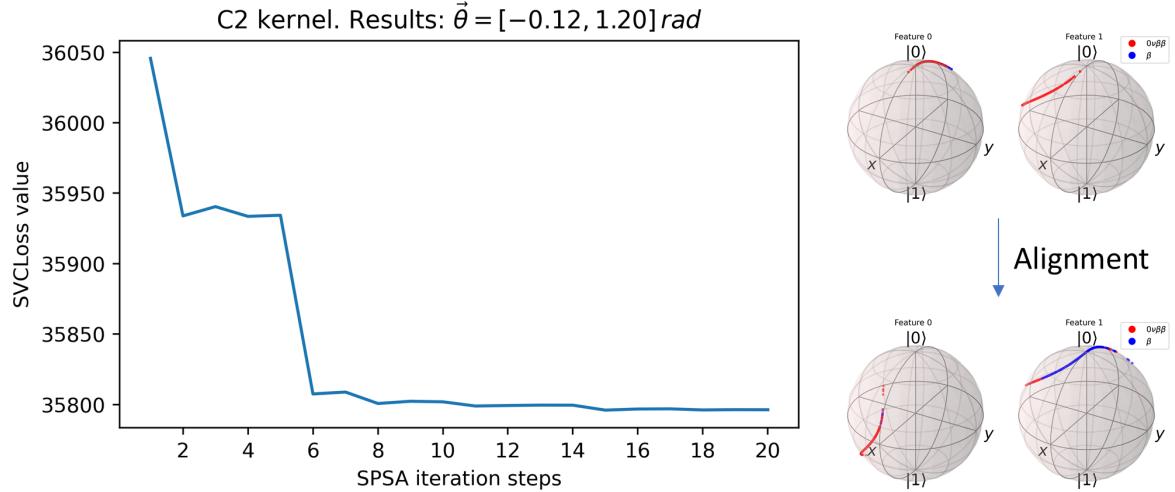


Figure 3.12: C2 feature map SVCLoss for the $\vec{\theta}$ values found at different SPSA iteration steps. The effect of the initial rotation on the Bloch sphere is also represented.

possible to effectively run SVMs and QSVMs training. The first learning curves are shown in figure 3.13, comparing linear, polynomial and gaussian kernels with Z, ZZ, C1 and C2 feature maps. Figure 3.13 shows slightly different behaviours emerge in the two feature sets. In the CNN case the entangling kernels ZZ and C1 exhibit more overfitting than the others when the training sample size is small. This can be seen from the high accuracy on the training set and the consequently low one on the validation set (< 100) samples. On the contrary, the C2 kernel is less prone to overfitting in case of a small training set size, resulting in a steeper rise of the learning curve in the leftmost part of the plots. In conclusion, ZZ and linear kernels result in being sub-optimal in exploiting the CNN features, while the gaussian and C2 obtain the best results, the latter having a validation accuracy of $65.3 \pm 0.5\%$, and $65.7 \pm 0.4\%$, respectively. In the ATN case, no significant difference between SVM and QSVM is observed, as they all rapidly converge to the same accuracy, the worse being the linear SVM with $64.8 \pm 0.6\%$ and the best being the Z-kernel QSVM with $65.9 \pm 0.6\%$ on the validation set. The samples from the test set in the CNN case were classified correctly in the 65.5% of times by the gaussian SVM, and 65.4% by the C2-kernel QSVM, while for the ATN case 63.9% and 63.8% respectively. The linear kernel is underperforming for < 500 samples, due to the hyperparameter C chosen from the grid search, set to optimize the classification at 2000 samples. The fact that the linear SVM converges to the same accuracy as the others in the ATN case suggests that the distribution in figure 3.10 (right), despite looking rather complicated, is separable at best by a linear decision boundary.

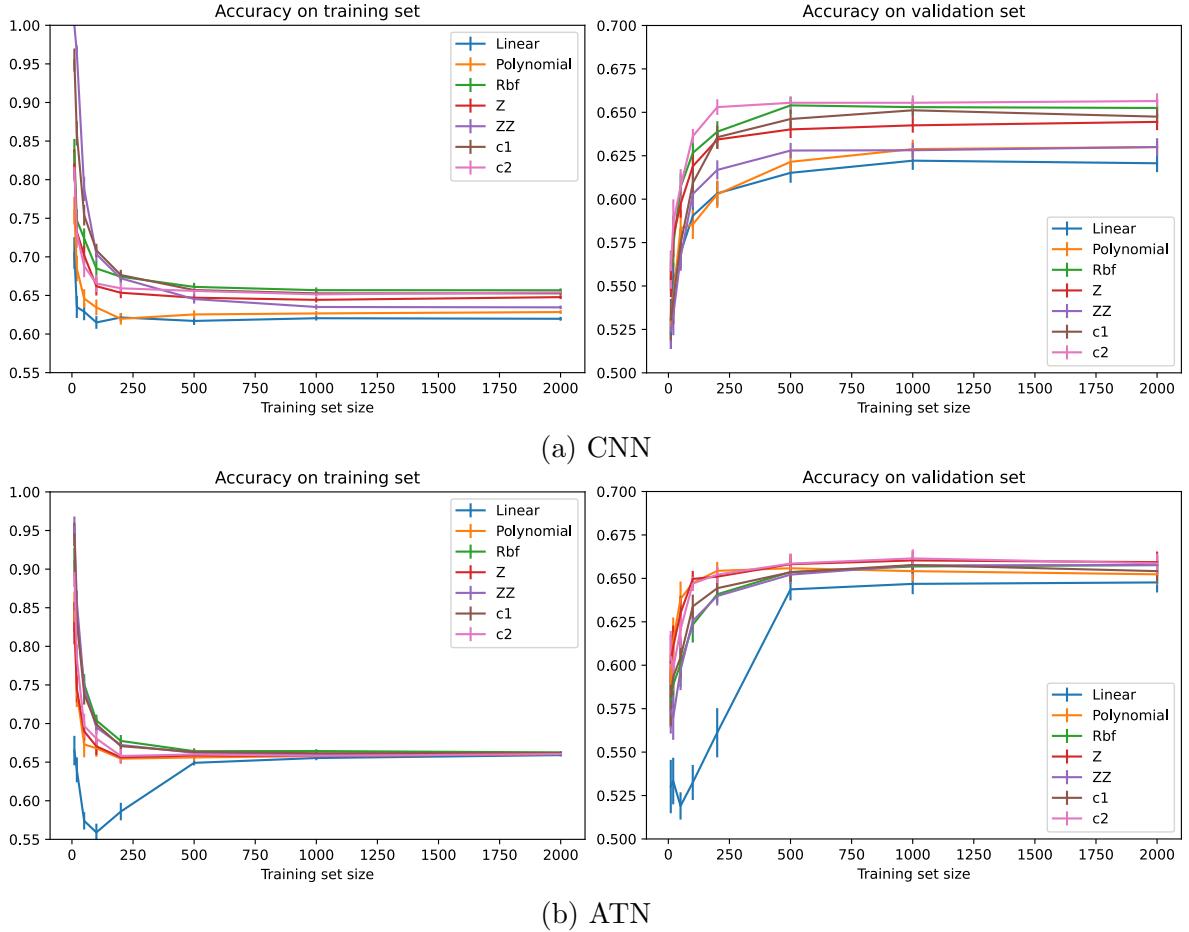


Figure 3.13: Learning curves of classical kernel SVMs (linear, polynomial, rbf/gaussian) and heuristic quantum kernels QSVMs (Z, ZZ, C1, C2), both on training and validation subsets of 300 samples, with features extracted from CNN (a) and ATN (b). Quantum kernels were optimized with alignment, have $reps=1$ and two qubits. To estimate the error bars, all trainings were repeated 20 times and the standard deviation was computed.

A qualitative difference between entangling and non-entangling feature maps is the shape of the decision boundaries, as shown in figure 3.14. The presence of entanglement seems to determine more complex boundaries, which effectively helps minimise the classification error on the specific training subset, but can also lead to missing the general dataset distribution. This behaviour matches the learning curves and suggests that entanglement enhances the classifier's expressivity, as well as the risk of overfitting. The next training session completes the case of two-feature, two-qubit heuristic feature

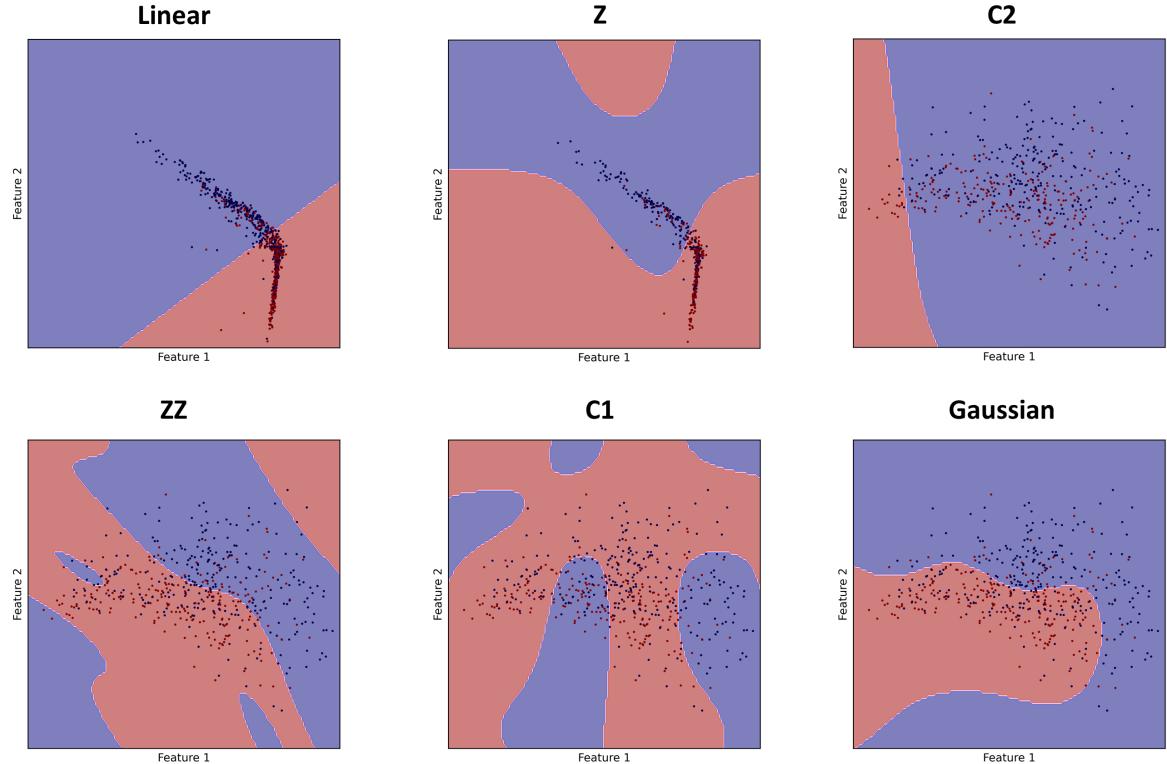


Figure 3.14: Examples of decision boundaries set by QSVMs for 500 training samples from CNN and ATN feature distributions. Not all classifiers converged to satisfying decision boundaries for these specific training subsets.

maps. It was carried out by increasing the depth of the circuits to $reps= 2$. The interest in varying the repetition number is to test circuits with a higher number of gates, more challenging to run on a NISQ device but that might lead to advantages through their complexity. As can be observed in figure 3.15, the results are very similar to the $reps = 1$ case, or even slightly worse for the CNN distribution, discouraging the idea of testing deeper feature maps for this dataset.

More features and more qubits

The following approach with heuristic feature maps aimed to exploit quantum data encoding into larger embedding spaces, by increasing the number of qubits involved and the related quantum kernel function. The quantum circuits relative to Z and ZZ for more than two qubits can be derived from equation 1.25, while C1 and C2 were extended arbitrarily as reported in appendix B. Since Z and ZZ require the number of qubits equal to the number of features, the same rule was applied for C1 and C2. The idea of training QSVMs with a different ratio between qubits and feature numbers was

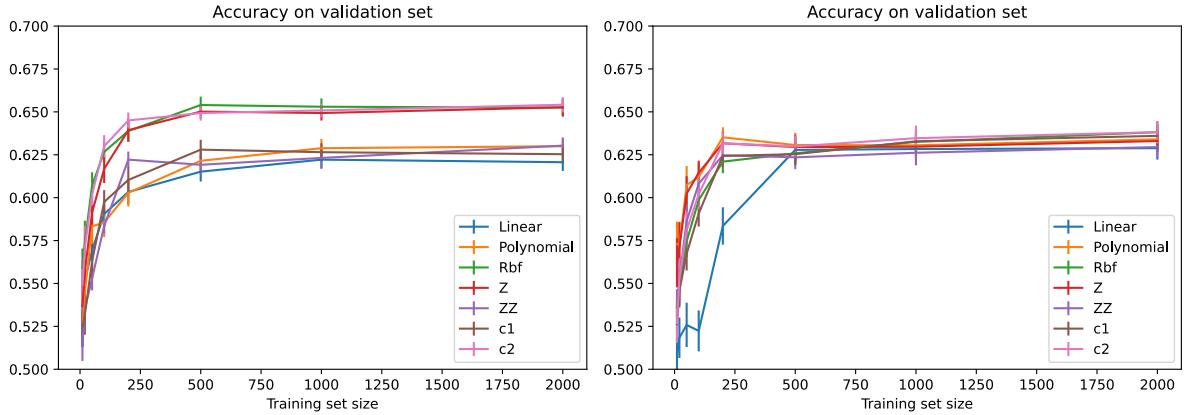


Figure 3.15: Learning curves of classical kernel SVMs and heuristic quantum kernels QSVMs with $\text{reps} = 2$ on a validation set with feature extracted from CNN (left) and ATN (right).

only explored in a later stage as described in section 3.16.

Figure 3.16 shows the learning curves obtained by incrementing the features/qubits number respectively up to three, five and ten. From these results, it is possible to conclude that the increment up to ten qubits does not affect positively the performances of the QSVMs, especially with features extracted by the CNN, where the entangling feature maps do not reach convergence within the maximum of 2000 training samples limit. This behaviour is better explained by observing the learning curve on the training set themselves (see figure 3.17), where considerable overfitting is visible, especially for the ZZ feature map.

An interesting correlation becomes evident in the ten-qubit case between the amount of overfitting and the kernel function values $K(\vec{x}_i, \vec{x}_j)$. The latter can be represented with the Gram matrix defined as:

$$G_{ij} = K(\vec{x}_i, \vec{x}_j) \quad (3.4)$$

If \vec{x}_i and \vec{x}_j are elements of the training set, the Gram matrix is symmetrical and $G_{ij} = 1$ if $i = j$. Gram matrix entries are measures of similarity between samples in the quantum embedding space, and the values go from 0 (orthogonal) to 1 (parallel). As reported in figure 3.17, the more a kernel performs well on the training set, the closer to the identity is the gram matrix. This result indicates that feature maps with great ability in spreading data in a high dimensional Hilbert space produces a highly expressive classifier and an optimal training kernel, at the risk of poor generalization. In order to achieve a satisfactory classification, a delicate balance between the choice of gates and the size of the embedding space must exist. This also depends on the shape of the initial feature distribution (this behaviour is evident for the CNN, much less for ATN), and the overlap between classes.

The studies reported in this section are related to a very specific set of kernels chosen heuristically. A general problem in QSVM is the lack of a rigorous method for designing an optimal kernel for a given dataset. In the next section, this aspect will be partially tackled by automatizing the process of kernel evaluation and optimization.

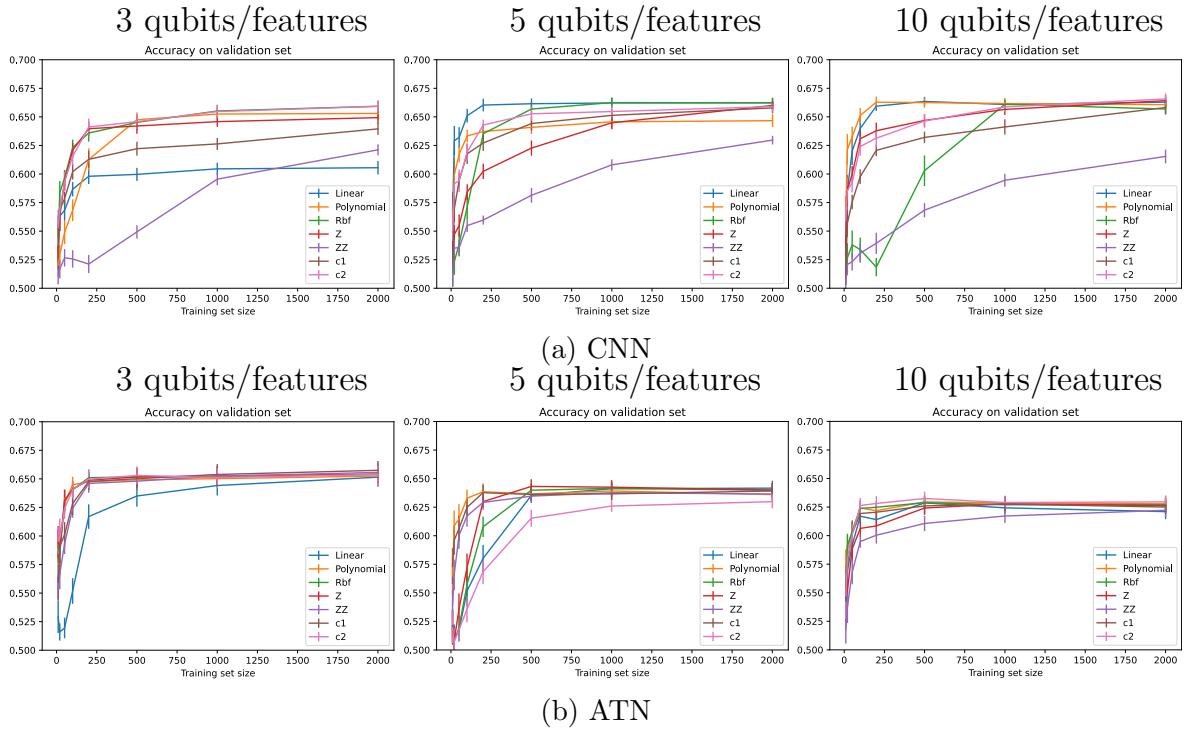


Figure 3.16: Learning curves of classical kernel SVMs and heuristic quantum kernel QSVMs on a validation set, with features extracted from CNN (a) and ATN (b), for different qubits/features number.

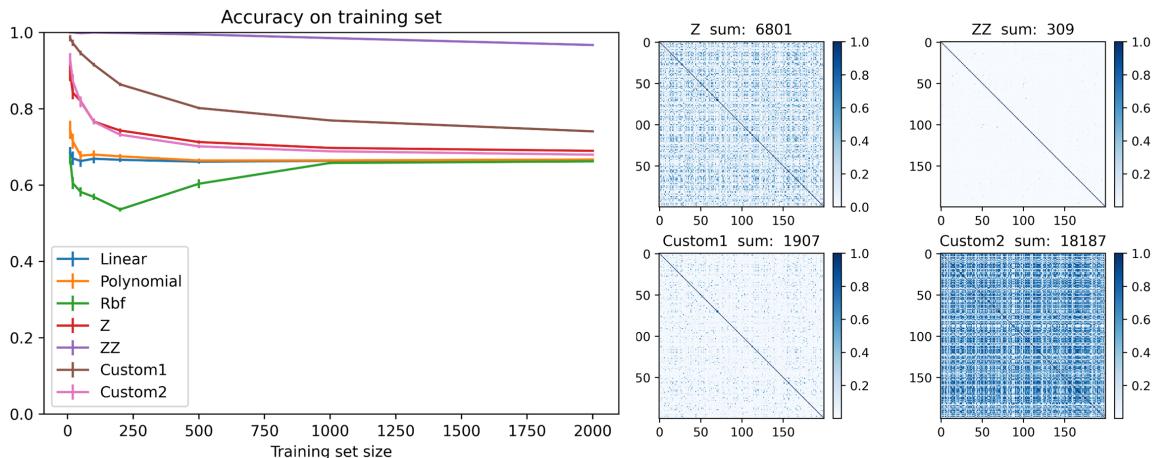


Figure 3.17: Relationship between the performance on the training set/amount of overfitting and the kernel's Gram matrix entries values. The 200 sample matrices are reported.

3.3.2 Genetic Quantum Kernels

The genetic generative algorithm (GA), introduced in section 1.5, was implemented for the automatized search of optimal kernels for the given classification task. The GA approach is meta-heuristic and was utilised to complete the search for a performance advantage in discriminating $\beta\beta$ and β events. The advantage of such a procedure over a pure heuristic approach is the following:

- **GA is dataset-independent:** the quantum kernel population automatically self-adapts through generations to the specific problem, so that the same routine may be applied to different input distributions.
- **Systematic kernel search:** the evolutionary thrust selects certain types of kernels, but genetic operators are intrinsically random, guaranteeing to explore a variety of potential candidates at each generation.

On the other hand, two drawbacks must be considered. The first one consists of the fact that in general, GA's outcome improves when increasing the population size (μ) and the number of generations. This requires the evaluation of a large number of kernels, which is computationally expensive, forcing the implementation of small-sized GA instances with modest exploration power. The second drawback is the risk of genetic variability reduction. If some genes prevail prematurely through generations, the population will be soon composed of kernels very similar to each other, slowing down or even stopping the evolutionary optimization. In order to avoid this, the genetic operator (mutation and crossover) must be tuned in order to provide sufficient kernel differentiation, but at the same time produce offspring sufficiently similar to parents (if an individual is already fit for the task, only small changes are expected to improve it).

The GA implemented for the automatized design of quantum feature maps required two levels of operation: mutation and crossover acting on feature maps transposed to bitstrings, while the fitness function was evaluated by running the quantum circuits. For this reason, it was necessary to define a mapping between the gates of quantum circuits into sequences of 0s and 1s.

Binary encoding of quantum feature maps

For simplicity of operation, the first GA implementations had strict constraints on the circuit size: two-qubit feature maps with a maximum of three gates per qubit line. The fittest kernels returned by the algorithm were then aligned to the dataset according to the procedure explained in 1.4.4. Each gate was described by six bits, for a total of 36 bits per kernel circuit. Out of the six bits, the first three determined the gate type, and the last three the rotation angle, according to the conversion table shown in 3.2. If the first three bits correspond to a non-parametric gate, like H or CX , the next three bits' values are ignored.

In a kernel's genetic pool, the six bits describing the gates are placed consecutively, and their position inside the total bitstring determines the placement inside the circuit, i.e. in which qubit to act and in which order.

Binary encoding	Gate type	Rotation angle
000	I	$2x_0$
001	H	$2x_1$
010	CX	$\frac{2}{\pi}x_0$
011	R_z	$(\frac{2}{\pi})^2 x_0^2$
100	R_x	$(\pi - 2x_0)(\pi - 2x_1)$
101	R_y	$4x_0x_1$
110	P	$\arcsin \frac{2}{\pi}x_1$
111	I	$(\frac{2}{\pi})^2 x_1^2$

Table 3.2: Conversion table between bitstrings and quantum gates. Three bits determine the type of gate, and the other three bits the rotation angle, for a total of six bits per gate. When the gate type corresponds to I , no gate is applied, allowing the GA to produce kernel circuits with less than the six-gate limit. All gate types were defined in chapter 1.

Fitness function and kernel selection

The choice of the fitness function, decisive for the success of the GA algorithm, was the following:

$$Fitness = validation\ accuracy - 10^{-4}number\ of\ gates \quad (3.5)$$

which is a simple weighted sum of the accuracy on a training set and the size of the circuit (I -gates are not counted). The -10^4 factor on the gate number serves for taking into account almost exclusively the first term of the equation, which is by far more important, as the maximum number of gates is already set at six. In this way, if the validation accuracies of two different kernels are very close to each other, the smaller circuit will be rewarded more. In order to force the use of both input features x_0 and x_1 and both qubits, the fitness score for kernels with empty qubit lines and not parametrized by both features was set to 0. Validation accuracies were calculated by training on 500 samples a QSVM for each individual, with a validation set of 1000. These values were set as a trade-off between reducing the computational cost of the fitness function evaluation and assuring classifiers with good generalization properties. The kernel population evolves in the direction of maximizing the fitness score, according to the *rank* selection technique, which follows the steps:

1. Assigning a rank to all individuals in ascending order, i.e. associating the worst kernel to rank $r = 1$, the second-worse to rank $r = 2$, etc.
2. Summing all ranks of the $\mu + \lambda$ individuals (where μ is the population size and λ is the number of the offspring individuals), obtaining a value R .
3. Selecting the individuals for the next generation with a probability proportional to their rank, i.e. a kernel with rank r will have a probability of being selected of r/R .

The rank selection rewards the fittest kernels in the population, but not proportionally to the fitness score, allowing for a balanced convergence speed.

GA settings and two-qubit results

Table 3.3 summarizes the principal parameters and characteristics set for the genetic instances: The table also reports the typology of genetic operators. The mutation is

μ	10
λ	4
Selection type	rank
Mutation type	Adaptive
Mutation probability	40% or 15%
Crossover type	two points
Crossover probability	20%
Stop criteria	Saturation after 50/200 generations, or reaching 500 generations.

Table 3.3: Setting for the genetic optimizations ran in the present chapter.

adaptive, with probabilities from 15% to 40%, meaning that the kernels with a below-average fitness score in the populations will have their genes chosen for mutation with 40% probability, while only 10% for the upper-half ones. The crossover was set to *two-points*, i.e. the bits between two specific genes will be swapped in the two parents, and the probability of choosing an individual for crossover is 20%. The stop criteria condition, called *saturation*, corresponds to ending the GA execution if after 50 or 200 generations (depending on the instance) the highest fitness score in the population does not improve.

The GA was implemented twice on the CNN feature set, and once for ATN, starting from $\mu = 10$ random kernels. Figure 3.18 reports the fitness score of the best individual through each generation, showing an improvement of 0.02-0.05, depending on the initial population, corresponding to an accuracy increment of 2%-5%. The three genetic quantum feature maps returned by the GAs, two resulting from CNN and one from ATN features, are shown in figure 3.19, and their respective learning curves in 3.20.

From these results, it is possible to draw some interesting conclusions. First of all, the genetic algorithm succeeded in the intent of automatically designing quantum circuits, and the condition of setting the max number of gates to six did not prevent the kernels to reach competitive levels of accuracy, on par with the best classical kernels in both CNN and ATN cases. Secondly, the *gen1* feature map for CNN has entanglement and performs slightly better than the heuristic entangling maps ZZ and C1 characterized in section 3.3.1. This aspect suggests that adding entanglement to a small degree (just one *CX*-gate is present, against the two in the heuristic ones) may provide benefits in classification, or at least will not harm the QSVMs performance for this dataset. An additional result, consistent with the expectations, is the simplicity of the ATN feature map, consisting only of three gates. This outcome reflects the linearity of the ATN feature distribution, already confirmed by the optimal performances reached by

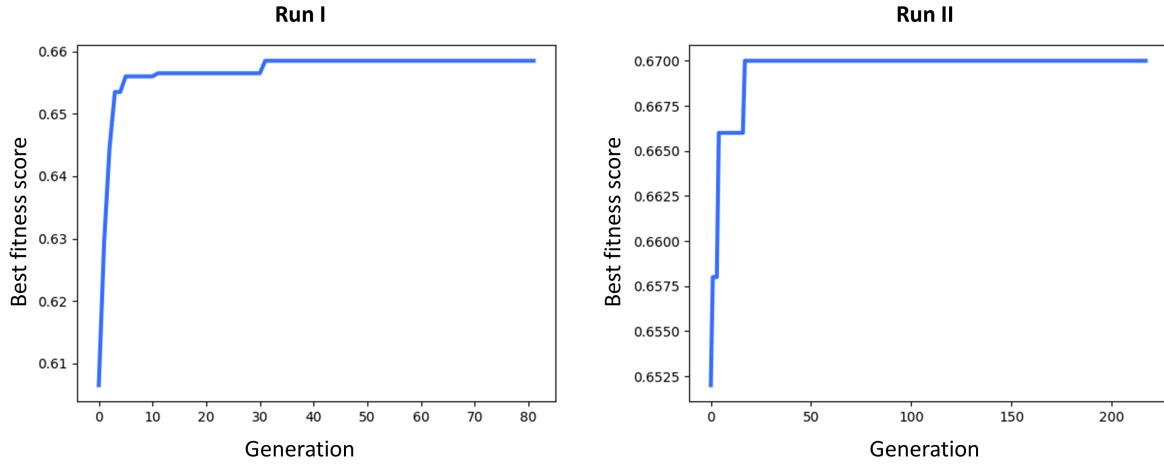


Figure 3.18: Fitness function as a function of generation number for two different genetic instances on the CNN 2D feature distribution.

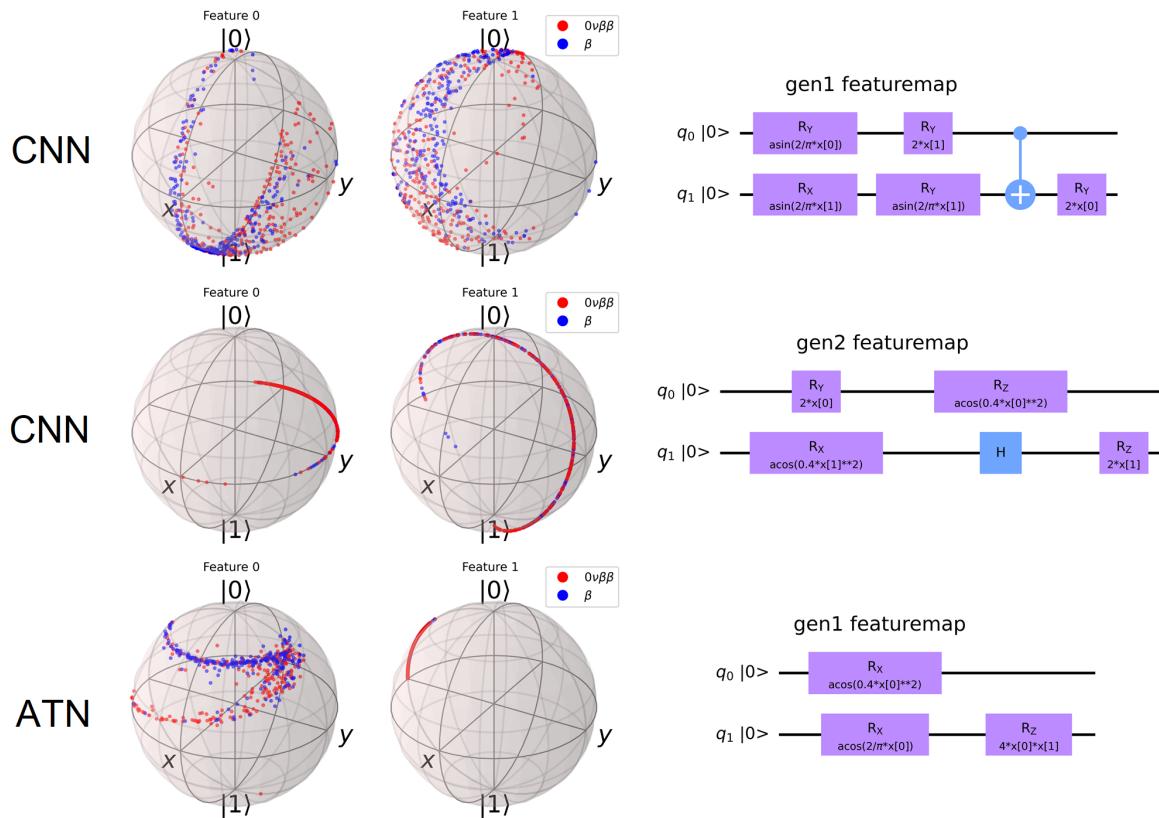


Figure 3.19: Two-qubit quantum feature maps resulting from the GA algorithm, with the corresponding distributions in the Bloch sphere visualization.

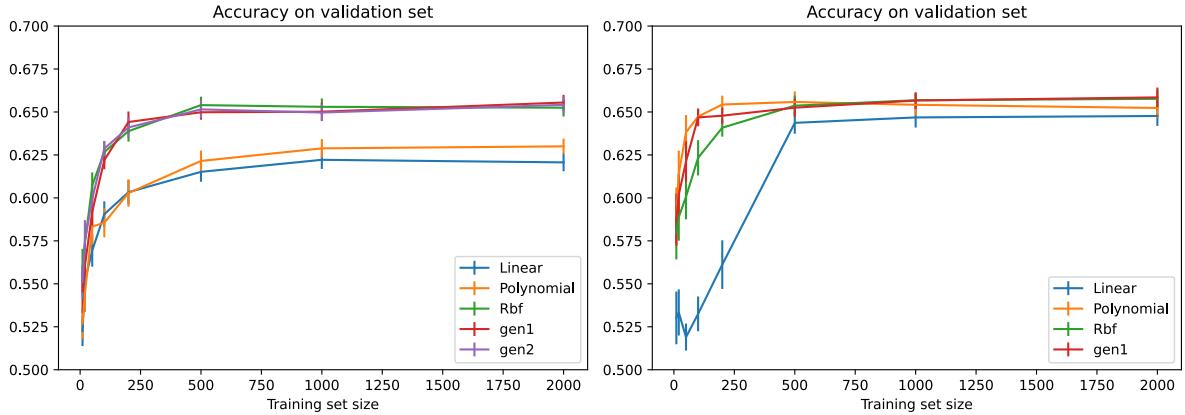


Figure 3.20: Learning curves of classical kernel SVMs and meta-heuristic quantum kernel QSVMs on a validation set, with two features extracted from CNN (left) and ATN (right).

the classic SVM with a linear kernel. Moreover, the class distribution separation on the Bloch sphere in figure 3.19 for the ATN feature map shows that only one qubit is mostly responsible for all the classification, hinting that this task could be achieved even by a single-qubit circuit.

Figure 3.21 illustrates some of the best two-qubit QSVM performances achieved so far, from the perspective of the Receiving Operator Characteristic (ROC) curve. This curve indicates the tradeoff between sensitivity and specificity in a binary classifier. The more a ROC curve differs from a diagonal line (i.e. the ROC curve of a coin-flip classifier), the better the general performance.

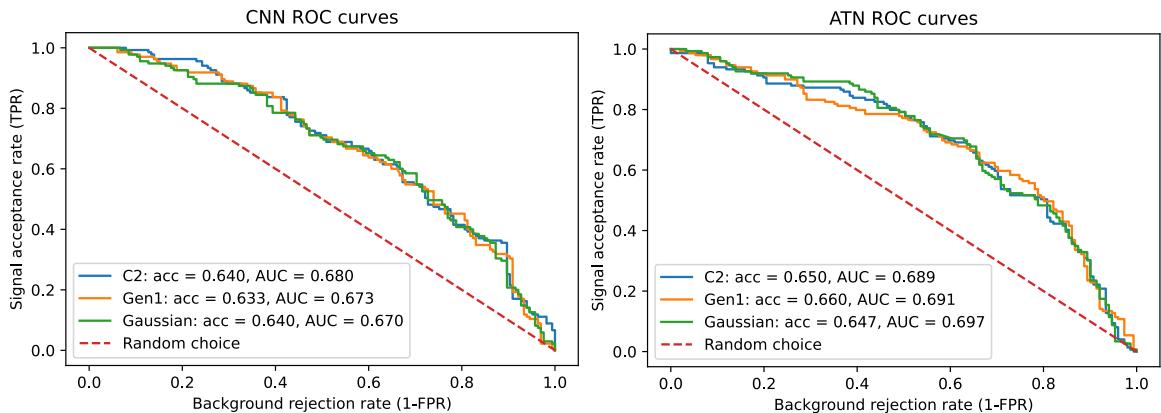


Figure 3.21: Examples of ROC curves of two-qubit QSVM classifiers and gaussian kernels for CNN and ATN extracted features.

GA for six-qubit circuits

Before introducing changes to the dataset (see section 3.4), a genetic instance with the same settings reported in table 3.3, combining the two-feature distributions of CNN and ATN, was run for a four-feature classification. The qubit number was set to six, as the max number of gates per line. Figure 3.22 shows the fitness function improvement over the generations and the learning curves. Again, the resulting circuit presents similar learning curves to the classical polynomial and gaussian counterparts. Remarkably,

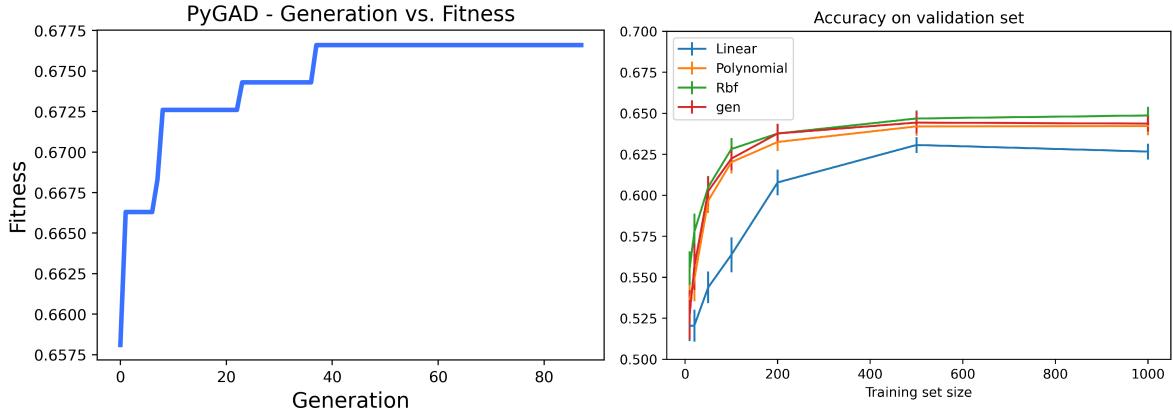


Figure 3.22: Fitness function vs generation number (left), learning curves of classical kernel SVMs and the six-qubit genetic kernel QSVM (right), with four features (two from CNN and two from ATN).

the feature map shown in figure 3.23 shows that the GA algorithm rewarded a fairly small circuit after less than 90 generations (only 24 gates out of a maximum of 36) with almost no entanglement (only one CX gate). This is a further, clear indication, that for this dataset the benefits of quantum computing were not observed, as the best-performing QSVMs were the ones with less entanglement, or with no entanglement at all, therefore easy to simulate classically. Nevertheless, the potential of entanglement in learning from distributions was observed when increasing the number of qubits on the ZZ and C1 feature maps, despite resulting in overfitting. The interpretation of these results is that if the feature distributions in a dataset have simple shapes and a large class overlap, QSVMs will not be able to provide performance improvements through their enhanced expressivity.

The genetic approach concludes the search for quantum advantage in the analysis of the toy-model dataset for $\beta\beta$ vs β classification, with resolution downsampled to $[5 \times 5 \times 1] \text{ mm}^3$. Further attempts will be made with increased resolution, as will be shown in section 3.4.

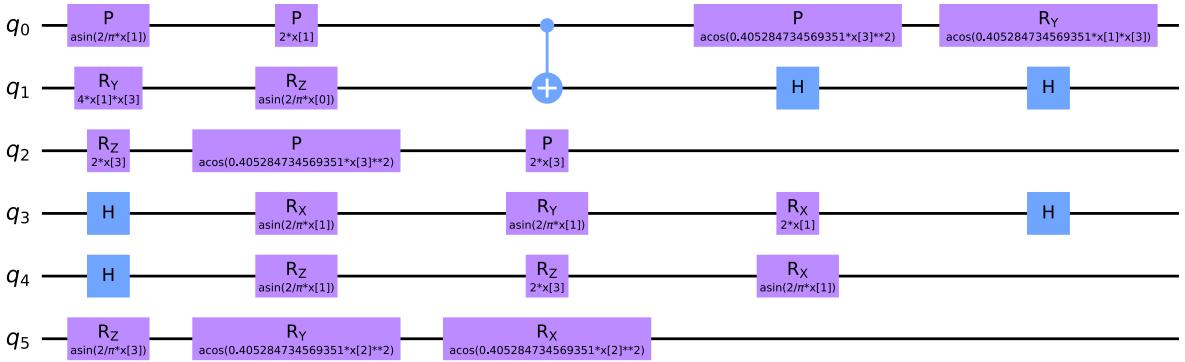


Figure 3.23: Six-qubit feature map resulting from a GA instance, on the four-feature CNN+ATN ensemble.

3.3.3 QSVM test on an IBM quantum processor

Up to this section, all the QSVM training sessions were performed using the Statevector backend, i.e. noiseless quantum circuit simulations, where probability amplitudes were accessible for speeding up the training, calculating the kernel function from the circuit's output states. In any real QSVM application, the circuit must be run several times in order to collect a significant amount of outcomes and estimate the probability of each possible final state. For ideal quantum circuits, then, quantum kernel evaluations are subject to errors due to statistical fluctuations, that go as the inverse square root of the number of *shots*, i.e. the number of circuit evaluations. On NISQ-era devices, a considerable amount of noise must be taken into account when running a circuit, affecting the outcomes.

Two-qubit kernels Z, ZZ, C1, C2 and the two genetic kernels for CNN were tested on the available backends offered by the IBM Quantum Lab on the cloud [82], including the seven-qubit IBM's processor *ibm_lagos* (see figure 3.24). Figure 3.25 shows the learning

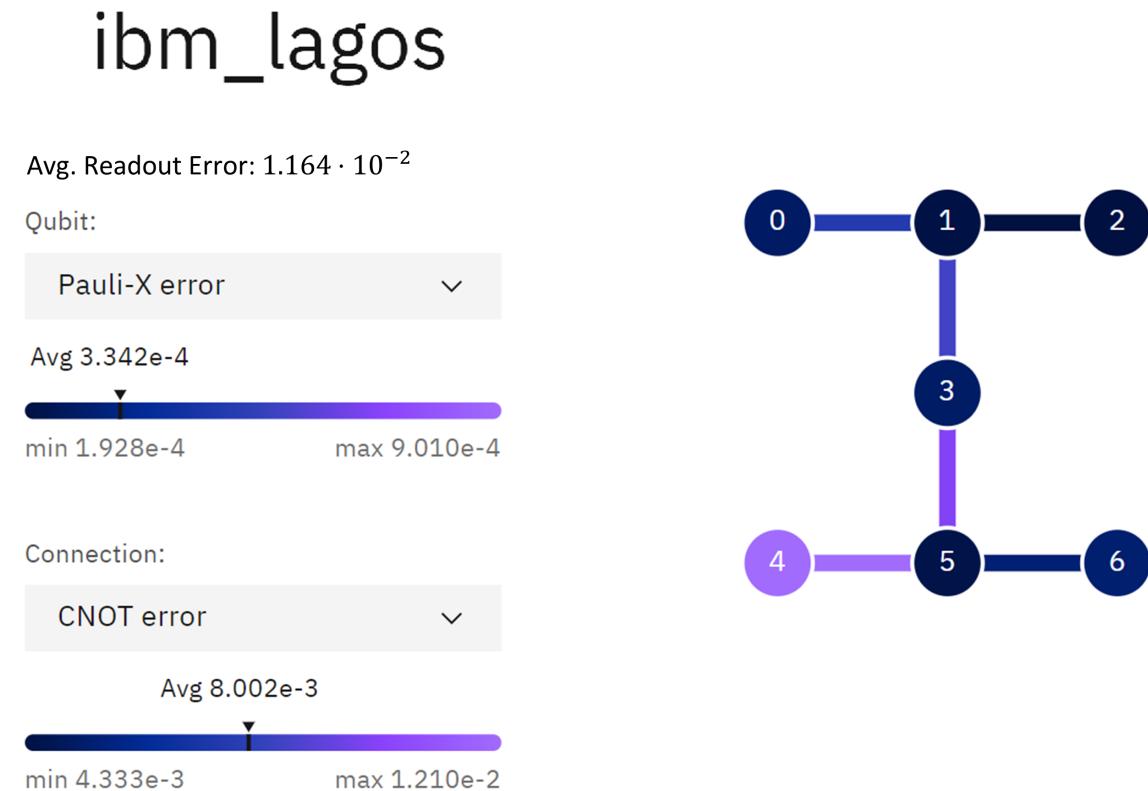


Figure 3.24: Entangling map showing control, readout and entanglement error rates of the *ibm_lagos* processor.

curve comparisons with four different backends: Statevector, ideal Qasm, *fake_lagos* (a simulation of *ibm_lagos* device included in Qiskit) and *ibm_lagos*. For this training session, due to the high computational cost of Qasm and *fake_lagos* and the limited availability of *ibm_lagos*, the training sample size was reduced to a maximum of 100, and the accuracies reported were calculated with a five-fold cross-validation. As expected, no significant accuracy loss was observed in any of the kernels when introducing noise in the circuit, confirming the NISQ nature of the QSVM algorithm.

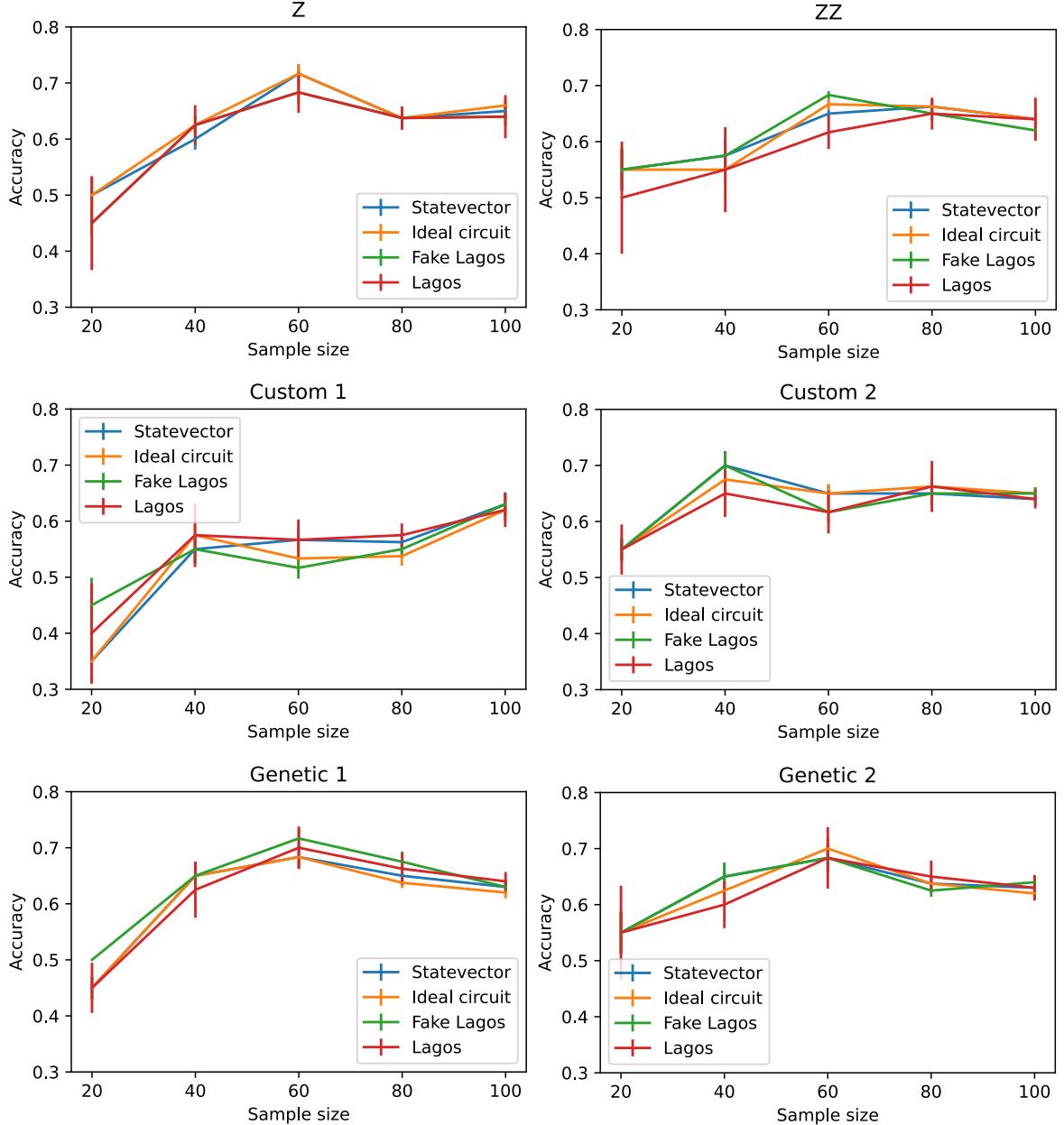


Figure 3.25: Learning curves of different two-qubit quantum circuits running on different backends, training on the CNN two-feature distribution. The number of shots for Qasm, fake_lagos and ibm_lagos was 8192.

3.4 Improving the toy model dataset resolution

This section is focused on characterizing the signal and background classification performance at different spatial resolution conditions and voxel energy thresholds. The main purpose was the investigation of the Convolutional Neural Network and the Transformer capabilities at different regimes. The track resolution tuning was performed by varying the two out of three voxel dimensions corresponding to the pitch of the wire planes. Also, in this case, no detector effects degrading the reconstruction quality were taken into account, and this approach was pursued primarily for the characterization of the two Deep Learning models. Nevertheless, the performance dependency on the wire pitch may still be taken as a benchmark for future analyses on simulations with higher fidelity. As a final step, QSVMs were trained at different resolutions and energy threshold conditions with respect to section 3.3.

3.4.1 Comparison between CNN and ATN

Figure 3.26 shows the accuracies at different wire pitches and energy thresholds. In the ATN case, the scan was in the wire pitch interval from $p = 0.1$ mm to $p = 7.5$ mm, while the CNN one was restricted to $p = 0.5$ mm to $p = 7.5$ mm due to the elevated computational cost of training on bigger sized pictures. Both CNN and ATN architectures remained unchanged from the ones shown in figures 3.4 and 3.5.

The effects of the different thresholds (10 keV, 50 keV, 100 keV) are more evident at resolutions better than $[1 \times 1 \times 1]$ mm³, where hits become numerous and smaller, containing less energy on average. From the accuracy plots in figure 3.26, it is possible to individuate three different regimes where one of the models prevails:

- **High-resolution:** below $p < 1$ mm the Transformer outperforms the CNN. This means that analyzing reconstructed events from TPC as a sequence of hits, according to the ATN paradigm, is more effective than the visual approach of CNN when the tracks are composed of a sufficient number of hits.
- **Transition region:** in the interval $1 \text{ mm} < p < 5 \text{ mm}$, Transformer and CNN performances are equivalent, within fluctuations of about $\pm 2\%$.
- **Low-resolution:** above $p > 5$ mm: The CNN outperforms the Transformer, showing that when the event is described by only a few hits, the visual approach of the CNN is preferable.

This result is in agreement with the nature of both Deep Learning models. When the resolution increases the Transformer seems to benefit from the number of hits, i.e. longer sequences for detecting internal correlations with the mechanism of self-attention. On the other side, CNN is affected negatively by the increasing sparsity of the images, as the hit number scales linearly with p , while the picture size grows quadratically. At the voxel size of $[1 \times 1 \times 1]$ mm³ (at which the ATN starts overcoming the CNN), the threshold at 50 keV and considering pictures of size 20×20 on every projection, the average ratio of nonzero pixels over the total is around 8.6%.

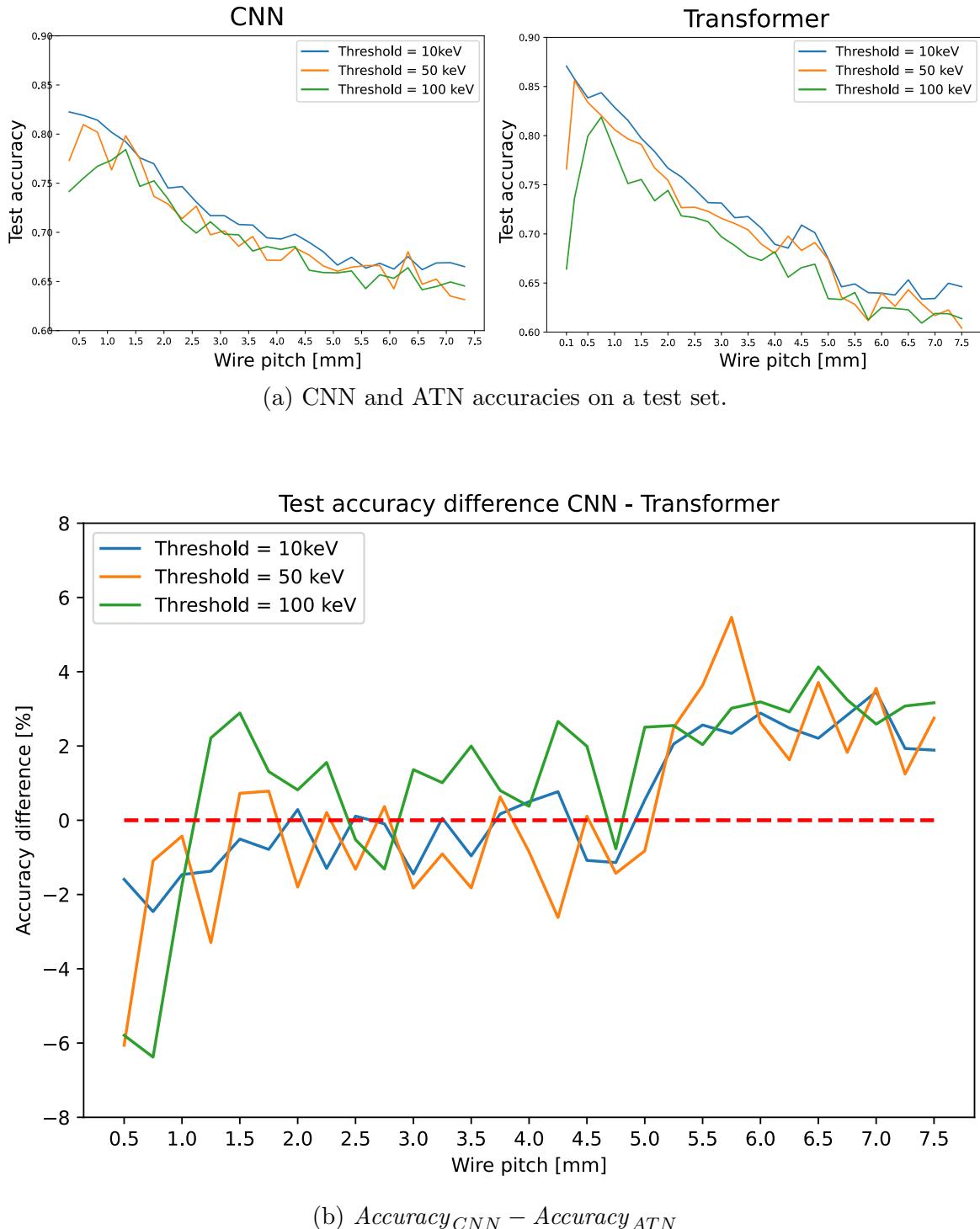


Figure 3.26: CNN and ATN accuracies (a) and comparison (b) on the toy model dataset at different hit energy thresholds and downsamplings (varying the wire pitch). The accuracy difference plot is obtained with the formula $Accuracy_{CNN} - Accuracy_{ATN}$.

3.4.2 QSVM at 1 mm wire pitch

Additional QSVM training sessions were carried out with the toy model dataset down-sampled to $[1 \times 1 \times 1]$ mm 3 and the hit energy threshold at 50 keV. In this configuration, the class overlap is significantly reduced with respect to the $[5 \times 5 \times 1]$ mm 3 case, allowing to probe the behaviours of quantum classifiers on an easier classification task, with peak accuracy at about 80%. At this resolution, the topological differences in $\beta\beta$ and β events start to be recognizable by the human eye. An example is illustrated in figure 3.27. This study was focused on the two qubits-feature case, with quantum feature

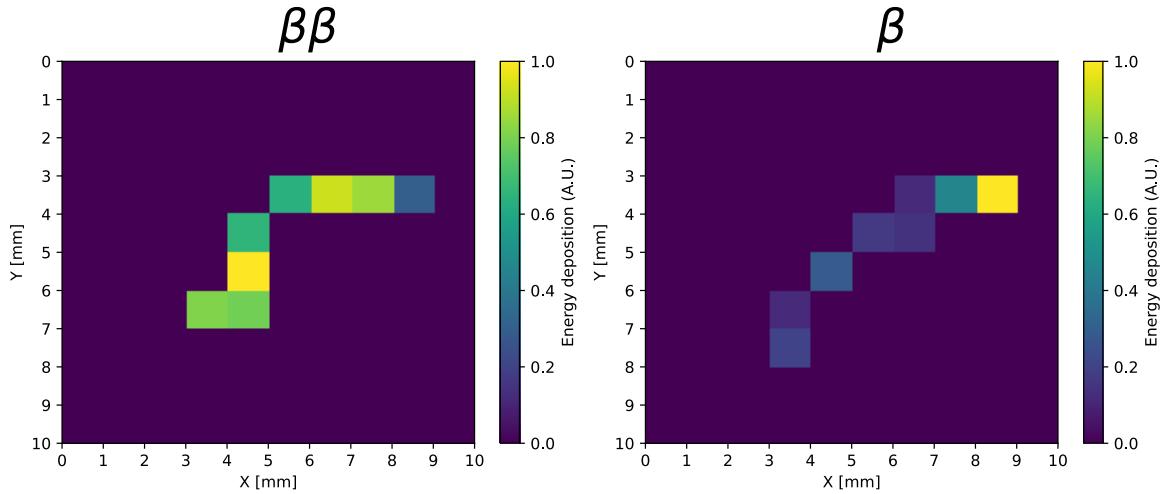


Figure 3.27: Examples of $\beta\beta$ (left) and β (right) events from the toy-model dataset after down-sampling to a voxel size of $[1 \times 1 \times 1]$ mm 3 and applying a 50 keV energy threshold, projected onto the XY plane. The colour indicates the energy of each hit.

maps Z, ZZ, C1, and C2, two genetic kernels for the CNN distribution and one for the ATN, as reported in figure 3.28. The genetic optimization settings were the same as used for the previous training sessions at a lower resolution, indicated in table 3.3. As for the $[5 \times 5 \times 1]$ mm 3 down-sampling results reported in section 3.3, the performance of all QSVMs matched the ones of the best classical SVMs (gaussian for CNN, gaussian and linear for ATN). The six-qubit genetic approach on all features combined was also carried out, as shown in figure 3.29, with analogous results. Appendix C reports the genetic feature maps generated at $[1 \times 1 \times 1]$ mm 3 .

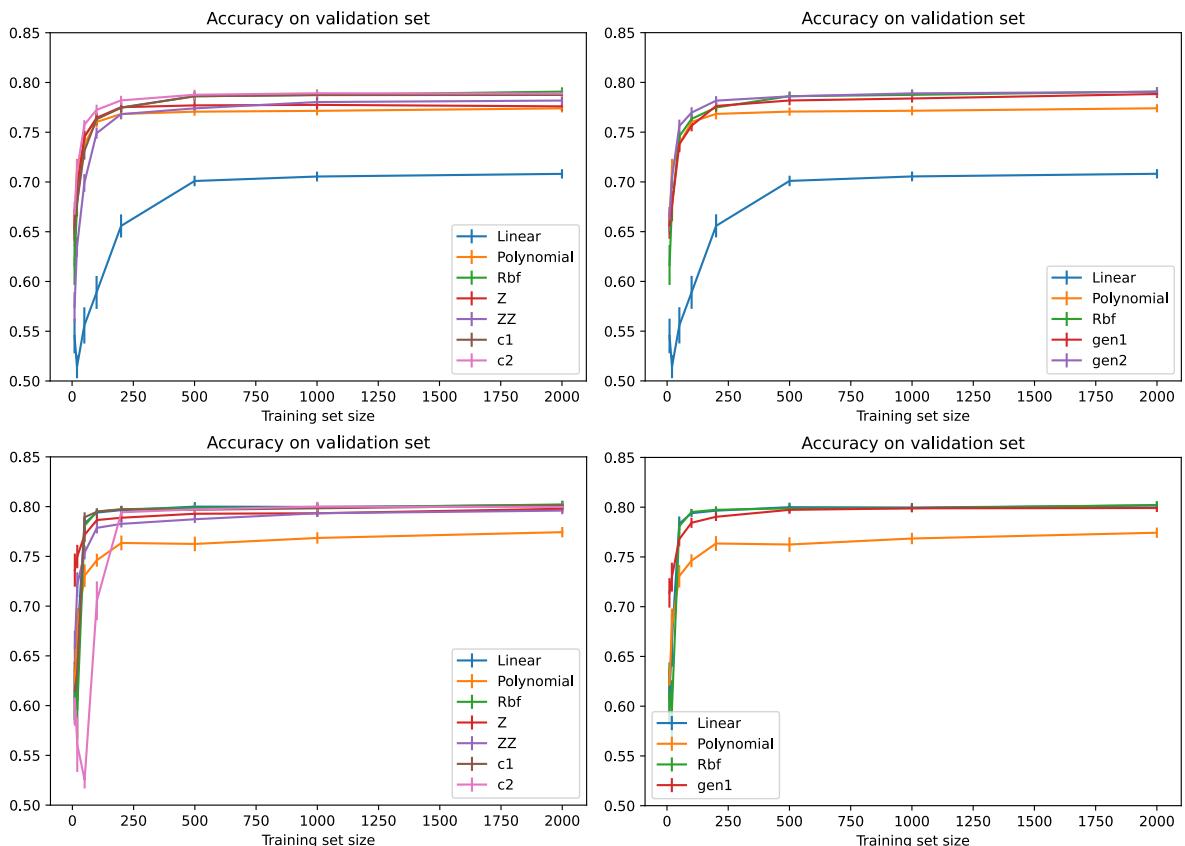


Figure 3.28: Learning curves of classical kernel SVMs, heuristic (left) and meta-heuristic (right) quantum kernel QSVMs, two features extracted from CNN (up) and ATN(down), with enhanced track reconstruction.

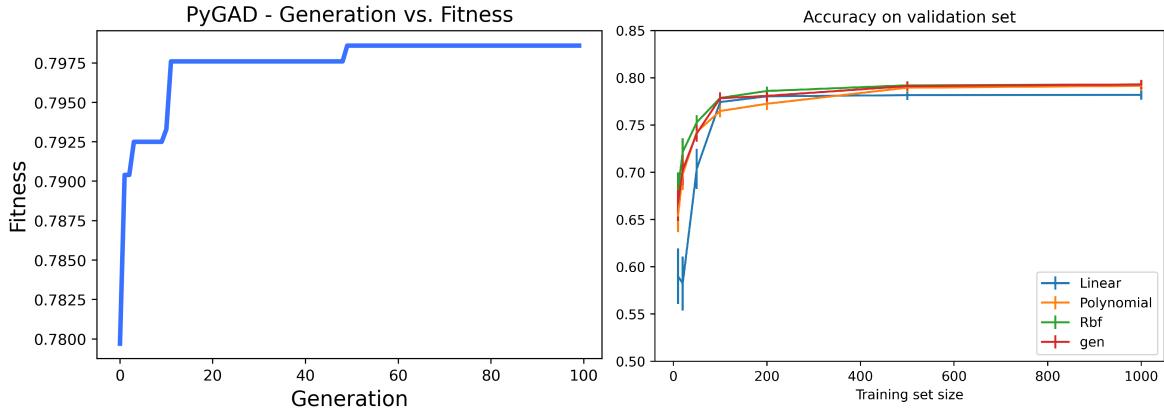


Figure 3.29: Fitness function vs generation number (left), learning curves of classical kernel SVMs and the six-qubit genetic kernel QSVM (right), with four features (two from CNN and two from ATN) and enhanced track reconstruction.

3.5 Background mitigation in the DUNE HD Far Detector

This section aims at addressing the topological background mitigation power for the $0\nu\beta\beta$ search in a DUNE FD LArTPC. For achieving this, we recurred to a state-of-the-art Monte Carlo simulation of the DUNE FD Horizontal Drift module, developed by DUNE collaboration with the use of the Liquid Argon Software toolbox (LArSoft) [83, 84].

The total active volume of an FD module is 12 m high, 14.5 m wide, and 58 m long. The readout of the ionization charge is performed by a total of 150 Anode Plane Assemblies (APAs), covering the entire anode plane, and each one of them possesses 2560 sense wires. The simulation of almost 4×10^5 channel responses is computationally expensive, requiring isolating a smaller workspace geometry, called $1 \times 2 \times 6$, referring to the APAs disposition (see figure 3.30), covering a total volume of about $[12 \times 7 \times 14] \text{ m}^3$.

3.5.1 The background simulation

The simulated radiological backgrounds, among the ones discussed in section 2.3.2, are 42-Ar, 60-Co, 40-K, 85-Kr, 39-Ar and neutrons. 42-Ar corresponds to the β topology, discussed in the previous sections of this chapter, and is the principal background in terms of event rate in the $Q_{\beta\beta}^{136Xe}$ ROI. Neutron events, despite being subdominant with respect to 42-Ar, also are a threatening background, and the topology-based rejection was attempted for both categories.

The spectrum of the radiological background, in terms of collected charge signal, is shown in figure 3.31. An important aspect to mention about the distribution shown in figure 3.31 is the absence of the $2\nu\beta\beta$ background, as the tail of its spectrum presents the same topology of the $0\nu\beta\beta$, and cannot be mitigated by the strategy proposed in this section. Moreover, the $0\nu\beta\beta$ is very spread out, despite the energy deposited is monochromatic at $Q_{\beta\beta}^{136Xe} = 2.458 \text{ MeV}$. This effect is principally determined by the recombination of ionization electrons while migrating towards the cathodes of the TPC. Depending on the electron drift length, i.e. the position where a decay occurred

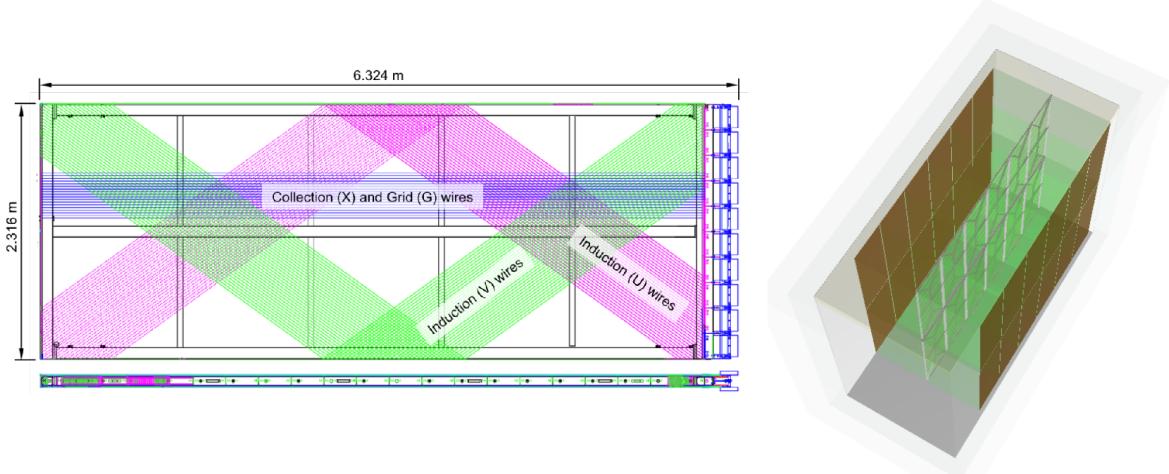


Figure 3.30: Left: scheme of an APA wire wrapping scheme [85]. U , V and X are the signal planes, while G (with wires parallel to X ones) is present to improve the pulse shape on the U plane signals. Right: DUNE FD $1 \times 2 \times 6$ workspace geometry [86]. APAs are shown in grey, while the brown panels are the Cathode Plane Assemblies.

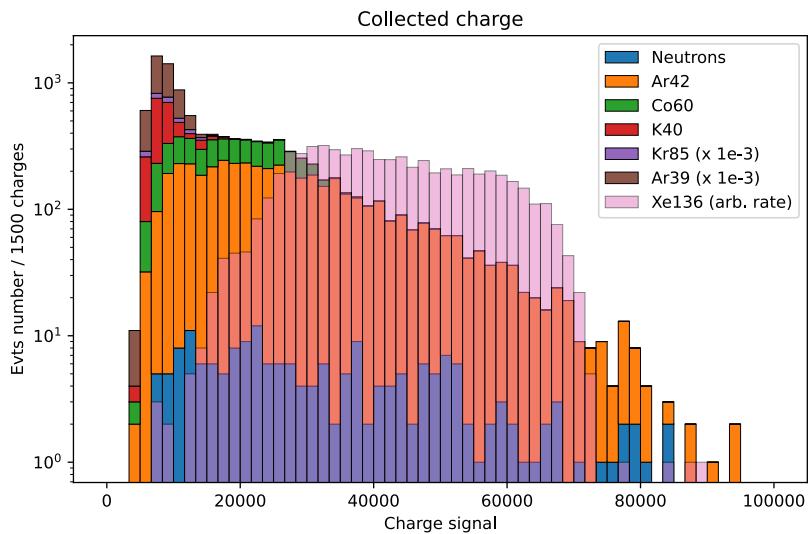


Figure 3.31: Radiological samples collected in a DUNE FD LArTPC in the $1 \times 2 \times 6$ workspace geometry, and an exposure time of 90 s, except for 85-Kr and 39-Ar, where the exposure time was reduced to 90 ms for speeding up the simulation. Background distributions are stacked, while the transparent histogram corresponds to $0\nu\beta\beta$ events, the latter set to an arbitrary rate.

inside the TPC, they have a certain probability of being recaptured by the medium before reaching the cathode, therefore not contributing to the signal. For the intent of this thesis, regarding a solely topological classification, this aspect was not taken into consideration.

3.5.2 CNN multi-label and binary classification

The dataset for the classification model contained tagged 136-Xe, 42-Ar and neutron events, where the samples consisted of hit ensembles reconstructed from the wire planes (two anodes and one collection) and the temporal label (tick). Hence, every event is described by three 2D views in the variables (*wire number, tick*).

The classification was performed by a CNN with the same layer structure as described in figure 3.4. The choice of a Convolutional Neural Network over the Transformer is motivated by the results shown in section 3.4, suggesting that when little amount of information is available, the visual approach is preferable to the sequential one. QSVM was not investigated in this phase, as the classification problem is analogous to the toy-model one discussed in the previous sections.

As a first preliminary approach, the input data was set by sampling an approximately balanced number of events for each event class, according to their natural charge distribution, and training the network as a binary classifier ($0\nu\beta\beta$ vs backgrounds). The effect of the CNN trained on such a dataset is shown in figure 3.32. As can be

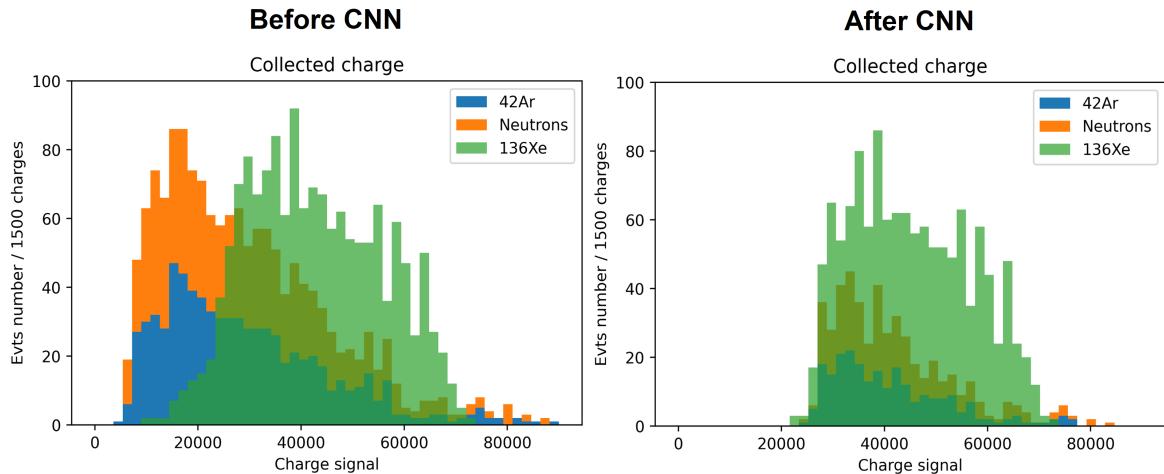


Figure 3.32: Background mitigation achieved by a CNN trained as a binary classifier ($0\nu\beta\beta$ vs backgrounds) on class distributions as they occur in the TPC.

observed from the two plots, the CNN essentially learnt how to apply a cut in the total collected charge of the event, returning a high classification accuracy (78%). This result is misleading, as it is not informative on how much the neural network actually learns from the topology of the hits, and similar accuracies can be obtained by simply applying a univariate analysis on the charge signal.

In order to force the CNN not to use the information on the amount of charge collected from the event for the prediction, the dataset was shaped in a way that every bin of the charge distribution contained fixed proportions of $0\nu\beta\beta$ and background. The next attempt consisted of training the CNN as a multi-label classifier, differentiating the two

background types. The dataset was partitioned as follows in table 3.4.

The results are described by the confusion matrix reported in figure 3.33, and the effect of background rejection can be observed in figure 3.34.

The classification outcomes are described by the confusion matrix reported in fig-

Class	Training	Validation	Test	Total
136-Xe ($0\nu\beta\beta$)	3867	823	859	5549
42-Ar (β)	3580	782	749	5111
Neutrons	3592	761	759	5112

Table 3.4: Dataset partition of the 15772 events into training (70%), validation (15%) and test (15%) subsets for the multi-label classification. The dataset is slightly unbalanced for favouring the overall event acceptance.

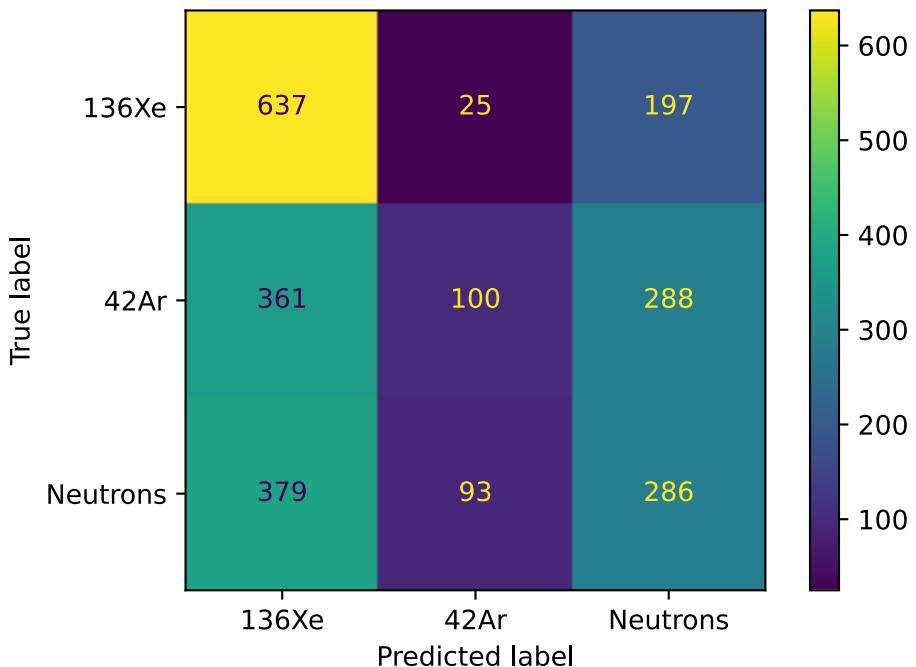


Figure 3.33: Confusion matrix for the CNN multi-label topological classification on a DUNE Far Detector Horizontal Drift dataset.

ure 3.33, and the effect of background rejection can be observed in figure 3.34. The CNN multi-label classification performances are low, correctly classifying samples with 75.0%, 13.4% and 37.7% probabilities for 136-Xe, 42-Ar and neutrons respectively. Overall, the efficiency of the selection is 75% with a purity of 50.9%, taking into account that 42-Ar events classified as neutrons and vice-versa will also be rejected.

In an additional approach, the CNN was trained as a binary classifier, again keeping the same class proportions for every bin of the charge distribution. Now 42-Ar and neutrons form a unique background class, and the dataset was organized according to the table 3.5. The classification performance is summarized by the ROC curve reported in figure 3.35, and the background rejection is shown in figure 3.36. In this case, the classification accuracy is 63.7%, with a slightly better efficiency/purity trade-off with respect to the multi-label case. In the mitigation example illustrated in figure 3.36, the efficiency 73.5% and the purity 55.2%.

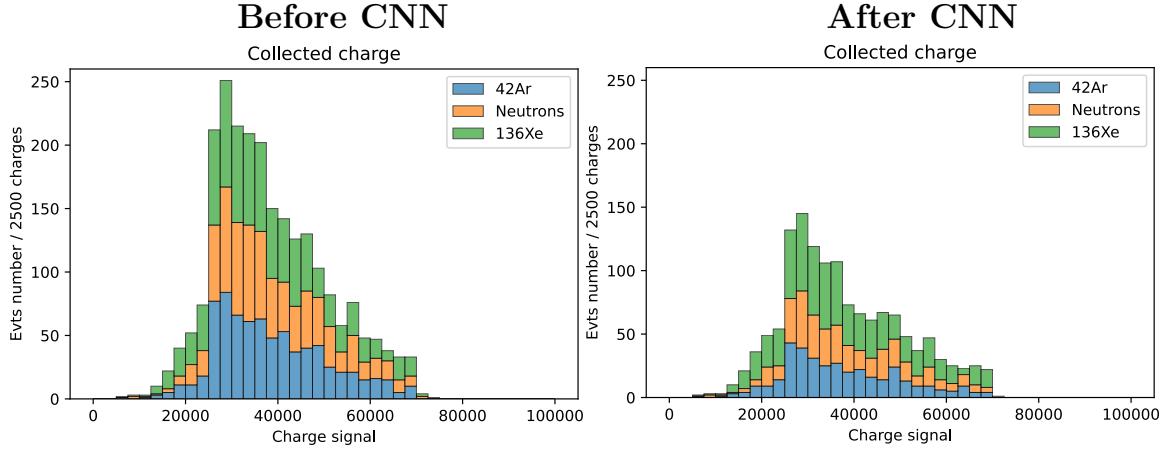


Figure 3.34: Background mitigation achieved by a CNN trained as a multi-label classifier ($136\text{-Xe } 0\nu\beta\beta$ vs 42-Ar vs neutrons) on equally distributed classes.

Class	Training	Validation	Test	Total
$136\text{-Xe } (0\nu\beta\beta)$	6403	1375	1351	9129
Background (42-Ar+neutrons)	6350	1358	1382	9090
	$(3167 + 3183)$	$(696 + 662)$	$(682 + 700)$	$(4545 + 4545)$

Table 3.5: Dataset partition of the 18219 events into training (70%), validation (15%) and test (15%) subsets for the binary classification.

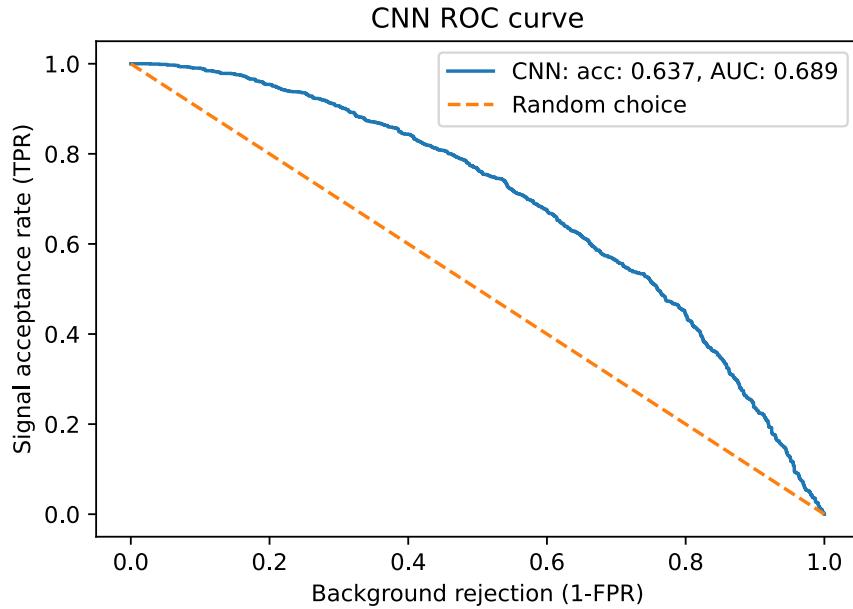


Figure 3.35: ROC curve for the CNN binary topological classification on a DUNE Far Detector Horizontal Drift dataset.

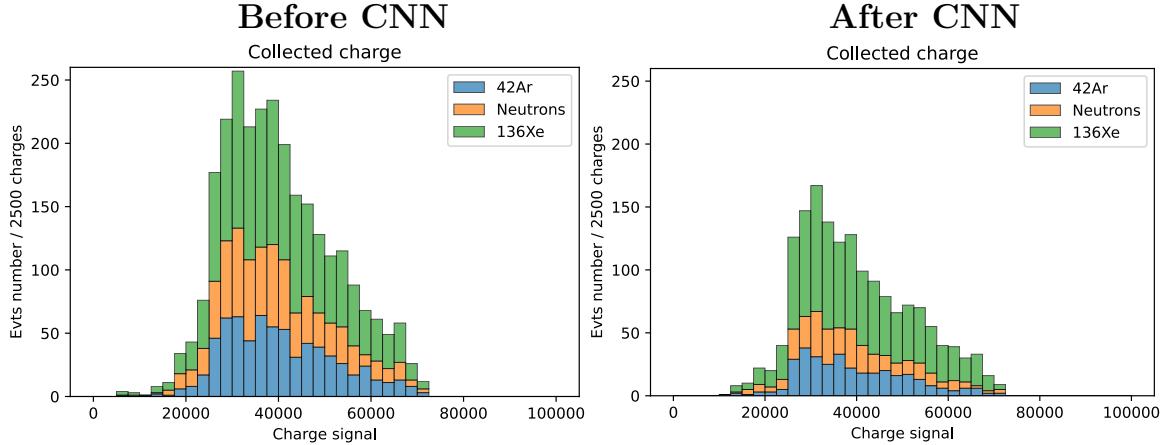


Figure 3.36: Background mitigation achieved by a CNN trained as a binary classifier ($136\text{-Xe } 0\nu\beta\beta$ vs backgrounds) on equally distributed classes.

The final attempt consisted of training the CNN as a binary classifier, again with the same charge signal distributions, but only with one background class at a time. The first dataset contained 7464 $0\nu\beta\beta$ and 7464 β samples, while the second one 5788 $0\nu\beta\beta$ and 5788 neutrons. As in the previous cases, the datasets were split into 70% training, 15% validation and 15% test. This time, it was possible to observe a great difference in performance, as shown in figures 3.37 and 3.38. When the

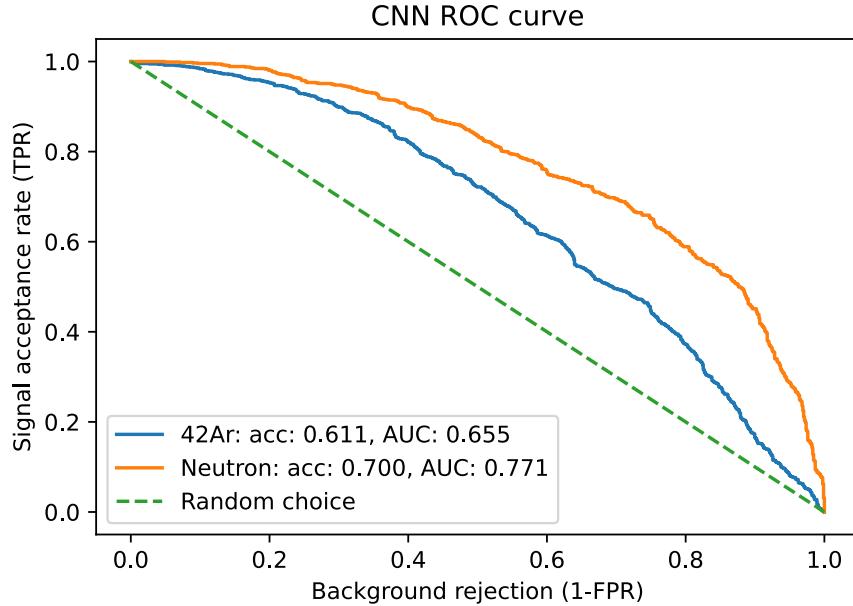


Figure 3.37: ROC curves for the CNN binary topological classifications $0\nu\beta\beta$ vs β and $0\nu\beta\beta$ vs neutrons on a DUNE Far Detector Horizontal Drift dataset.

CNN trains on the $0\nu\beta\beta/\beta$ dataset, the test accuracy amounts to 61.1%, with a corresponding 77.1% sensitivity and 45.0% specificity. On the $0\nu\beta\beta$ /neutron dataset, the performance improves significantly, reaching 70.0% accuracy, with a more balanced 68.9% sensitivity and 71.3% specificity. The track topology classification strategy results to be sensibly more effective in mitigating neutrons than β events.

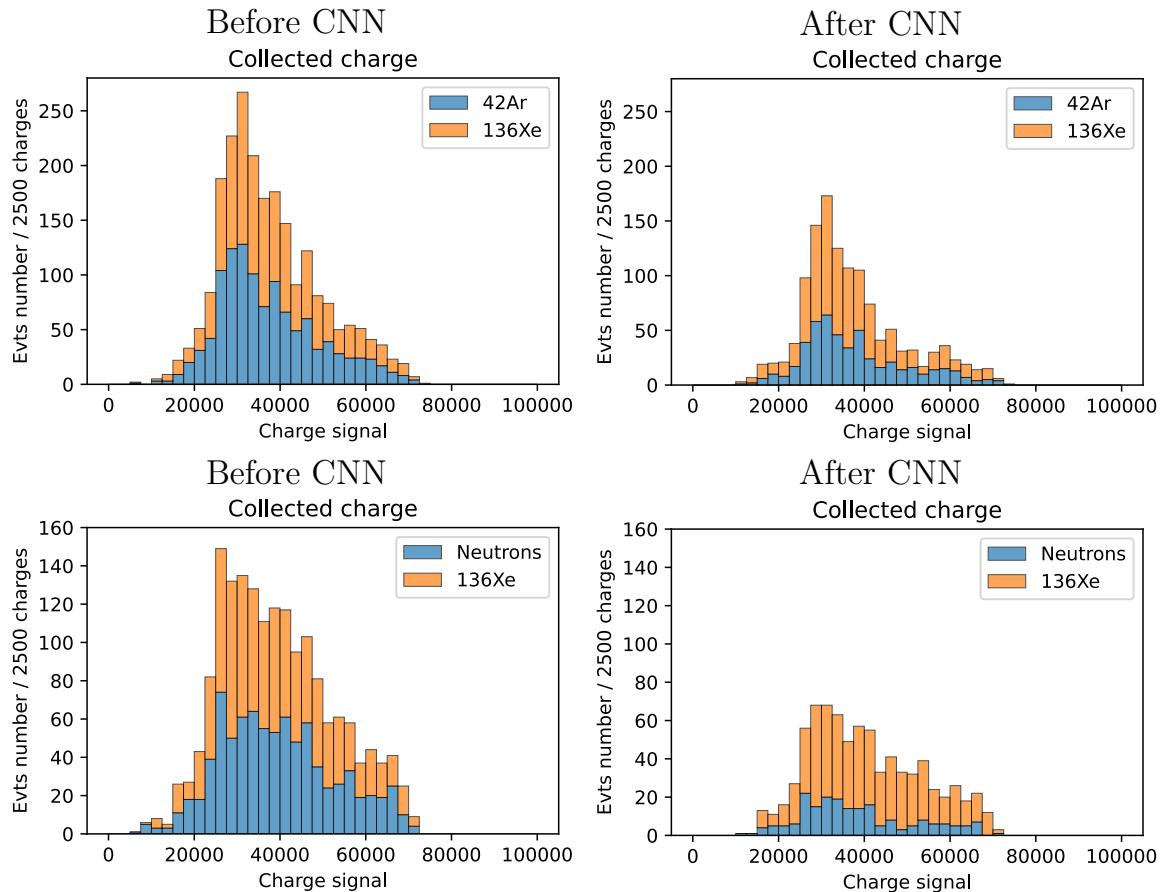


Figure 3.38: Background mitigation achieved by a CNN trained as $0\nu\beta\beta$ vs β classifier (up) and as a $0\nu\beta\beta$ vs neutrons classifier (down).

The CNN performances on the dataset from the Monte Carlo FD HD simulation are expected to be lower than the ones achieved on the toy model dataset, as it takes into account the detector effects, such as electron recombination, non-uniformities of the electric drift field, noise from electronics and the signal processing. However, the loss in accuracy only amounts to about 5%, suggesting that detector effects are not catastrophic at the MeV scale for background mitigation, and significant improvements may be achievable by designing an opportunity FD module with enhanced spatial resolution. This could be obtainable by reducing the wire pitch, i.e. the distance between TPC sense wires, compared to the one implemented for the Horizontal Drift, or recurring to different readout techniques, like pixel readout, proposed for the ND-LAr module of the Near Detector complex [87]. Such improvement would enhance the DUNE FD LArTPCs sensitivity to the search of the $0\nu\beta\beta$ decay, as well as other low energy physics channels sharing similar backgrounds, such as solar neutrinos and Dark Matter.

Chapter 4

Conclusion

Quantum Computing is experiencing rapid expansion, thanks to the technological progress marking the advent of the NISQ era and the development of quantum algorithms offering a wide range of applications in many scientific areas. In particular, Quantum enhanced Machine Learning is an attractive line of research for Particle Physics, due to the increasing usage of Artificial Intelligence in data analysis. This thesis leveraged the Quantum Computing paradigm in Neutrino Physics, proposing an automated background mitigation strategy for the analysis of the neutrinoless double beta decay of the ^{136}Xe isotope in the DUNE experiment. The idea consisted of enhancing the sensitivity to events at the MeV energy scale, learning from track information produced by DUNE's xenon-doped LArTPCs. This topology-based discrimination was achieved by implementing the Quantum Support Vector Machine, a QML classifier that encodes input data into quantum states by means of a Quantum Kernel. Quantum Kernels are scalar functions of the input features and are generally hard to compute with classical techniques, while they can be easily estimated by a quantum processor by executing a parametric quantum circuit. QSVM has a classical counterpart, the well-affirmed Support Vector Machine algorithm, which also relies on a kernel function, normally computed analytically. The thesis activities involved a thorough comparison between classical and quantum SVM in terms of signal and background classification accuracy, with the aim of addressing quantum advantage. SVM models with linear, polynomial and gaussian kernels have been trained to set benchmarks for the QSVM performances. Several Quantum Kernel circuits, up to ten qubits, have been designed and tested, first by using heuristics, and later recurring to a meta-heuristic genetic algorithm. The necessity to reduce the initial feature dimension produced by the TPC also provided the opportunity to develop and characterize Deep Learning models such as the Convolutional Neural Network and Attention Neural Network, both for classification and feature extraction.

Part of the work has been carried out analyzing a dataset consisting of $0\nu\beta\beta$ signal and β background 3D event reconstructed by an ideal LArTPC, with tunable wire pitch. The accuracy dependence on the wire axes resolution has been investigated by performing a pitch scan from 0.5 mm to 7.5 mm, comparing CNN and ATN. This study put in evidence an interesting difference between models: for resolutions worse than $[5 \times 5 \times 1] \text{ mm}^3$, the CNN is better performing than ATN, while for resolutions better than $[1 \times 1 \times 1] \text{ mm}^3$, ATN surpasses the CNN. In the region in between, no model prevails

significantly. This outcome underlines the potentialities of the innovative mechanism of attention in Deep Learning, overcoming the more conventional CNN visual approach in high-resolution track identification, and encourages adopting this strategy in other similar contexts.

When setting a DUNE-like $[5 \times 5 \times 1]$ mm³ resolution, the SVM benchmark classification accuracy approximates to 66% and 64% with features extracted with CNN and ATN respectively, while for $[1 \times 1 \times 1]$ mm³, both reach 80%. In these two scenarios, several Quantum Kernels chosen arbitrarily allowed QSVM to match the accuracy of the best classical SVM, i.e. the gaussian kernel SVM. Some of the Quantum Kernels evaluated using entanglement between qubits have shown higher expressivity than the classical ones, at the cost of generalizing less to new unseen data. A kernel optimization strategy, called kernel alignment, improved the QSVM generalization accuracy up to 0.5% and has been implemented for all Quantum Kernels. The genetic optimization algorithm succeeded in generating satisfactory Quantum feature map candidates, despite the limits imposed by the computational expensiveness. Moreover, the genetic algorithm has proven to be a powerful tool for probing the relationship between the distributions of the features and the ideal Quantum Kernel circuit structure, for example, hinting at the amount of entanglement requirement and the ideal circuit size.

The final goal of this thesis consisted of estimating the radiological background mitigation power for the $0\nu\beta\beta$ analysis with data produced by the DUNE Far Detector Horizontal Drift LArTPC simulation. In addition to β events, this classification problem involved the subdominant background due to neutron capture. A CNN has been trained, both as a three-label ($0\nu\beta\beta$, β , neutrons) and binary ($0\nu\beta\beta$, backgrounds) classifier. The second approach was the best-performing one and led to an overall accuracy of about 63% (rounding to the lower integer), corresponding to a 73% True Positive Rate and a 55% True Negative Rate. Moreover, training the CNN specifically for β rejection resulted in 61% accuracy, while doing the same for neutrons increased the performance up to 70%, indicating that the neutron background is easier to mitigate with the topological mitigation strategy. These results show that, for the characteristics of the DUNE FD Horizontal Drift module, implementing this technique may not benefit significantly the $0\nu\beta\beta$ analysis, and other kinds of mitigation strategies should be applied, such as passive neutron shielding or the use of argon depleted in the 42 isotope. However, the results achieved by this thesis contribute to probing DUNE's sensitivity to low-energy physics channels beyond the search for the neutrinoless double beta decay, such as solar neutrinos and Dark Matter, which share similar backgrounds. Moreover, the discussed techniques may interest the development of new detectors for achieving optimal background rejection capabilities in the future.

Appendix A

Classification metrics

All the models have been optimized to achieve the maximum classification accuracy on a validation dataset, while the final performances have been estimated by classifying a separate test dataset. The accuracy metric corresponds to the fraction of samples that are classified correctly. For a binary classification:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (\text{A.1})$$

where TP, TN, FP and FN respectively indicate the number of True Positive, True Negative, False Positive and False Negative samples. The Positive class corresponds to the $\beta\beta$ event, while the Negative to the background β .

Two useful additional metrics are defined as follows:

- Sensitivity, or True Positive Rate (TPR):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Quantifying the efficiency of a selection.

- Specificity, or True Negative Rate (TNR):

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Quantifying the purity of a selection.

A classifier decision function can be tuned according to the needs, increasing the sensitivity at the expense of reducing the specificity, or vice versa. When searching for rare events, it is convenient to keep a high signal acceptance, i.e. having a higher sensitivity than specificity. The capabilities of a classifier in terms of the TPR/TNR tradeoff can be represented with a ROC curve, displaying the TPR versus 1–FPR. FPR is the False Positive rate, described by the formula:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (\text{A.2})$$

The overall quality of a classifier can be determined by the area under the ROC curve, or AUC:

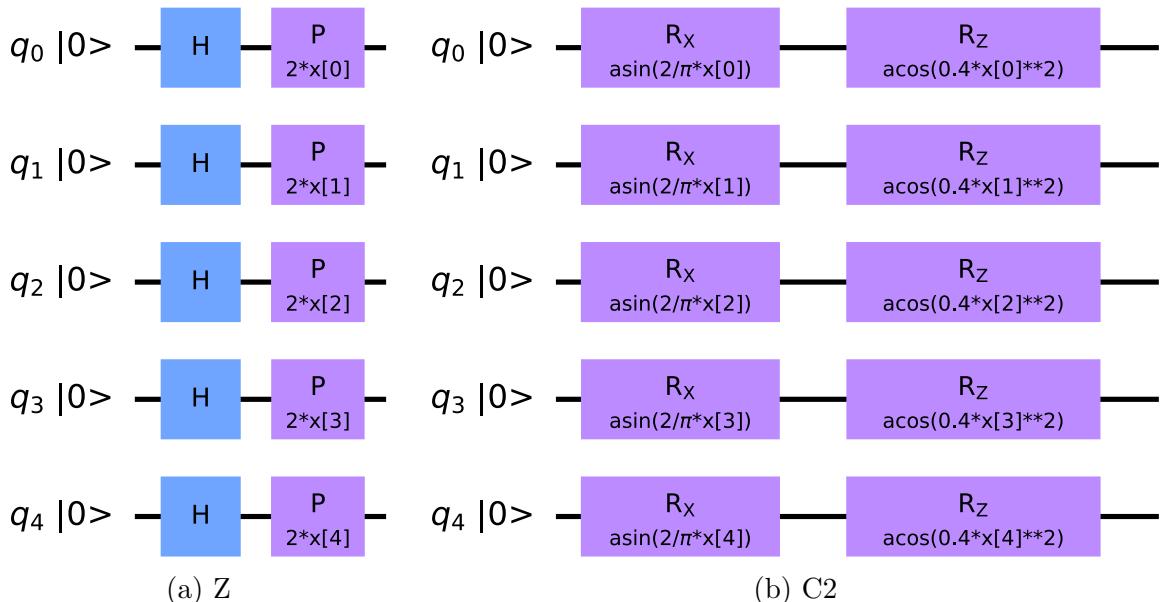
$$\text{AUC} = \int_0^1 \text{ROC}(t) dt \quad (\text{A.3})$$

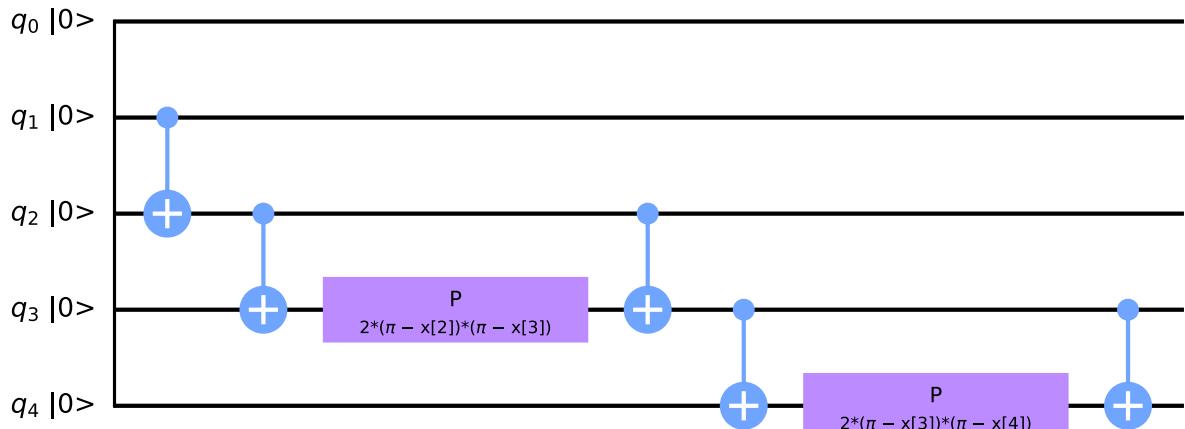
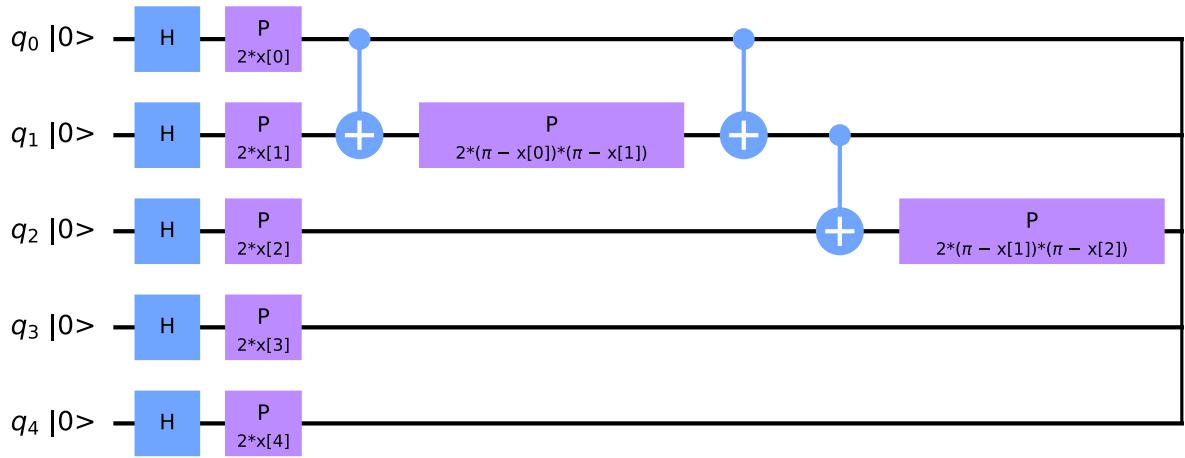
where $dt = 1 - \text{FPR}$.

Appendix B

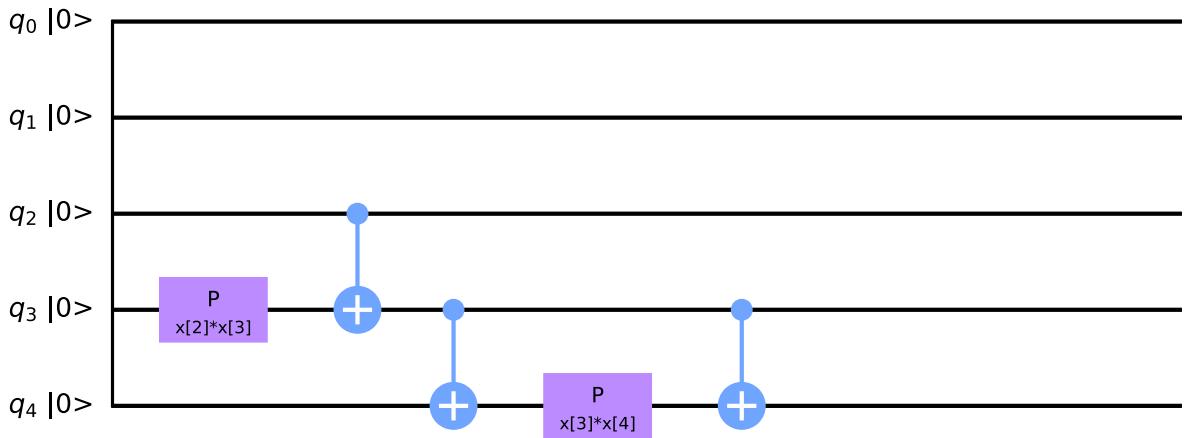
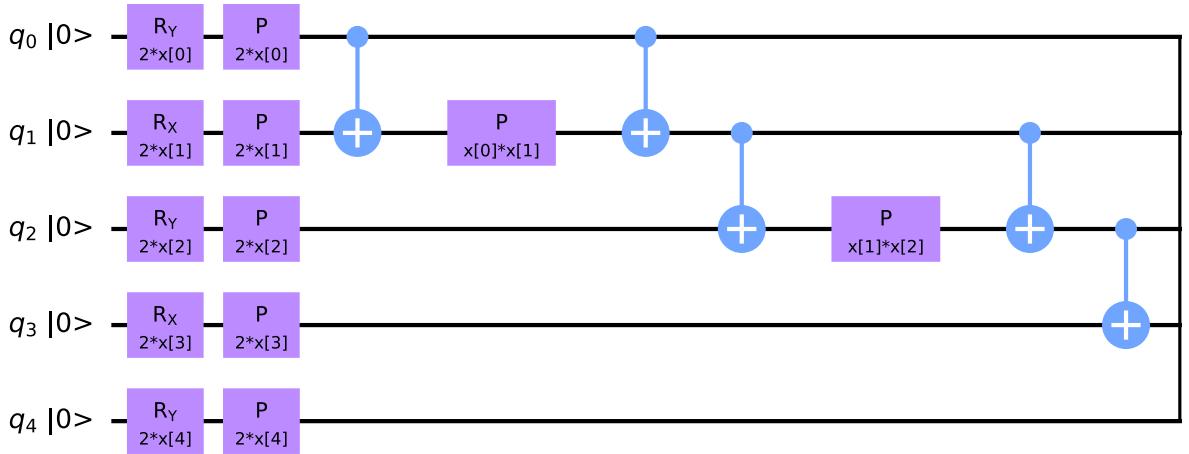
More on arbitrary feature maps

The heuristic Quantum Kernels Z, ZZ, C1 and C2 were tested in different configurations, varying the number of qubits and the repetitions. The Z and ZZ feature maps' quantum circuits were set accordingly to equation 1.25, while C1 and C2 were extended arbitrarily from the two-qubit versions shown in figure 3.9. For $reps=1$, the five-qubits circuit generalizations are represented in figure B.1. For the three-qubits case, only lines q_0, q_1, q_2 were considered.





(c) ZZ



(d) C1

Figure B.1: Five-qubits, one-repetition versions of Z, ZZ, C1 and C2 feature maps. The rules for extending to more qubits are different for every circuit and can be inferred from the pictures.

Appendix C

Genetic feature maps at 1 mm wire pitch

Figures C.1 and C.2 illustrate the feature maps produced by the genetic algorithm, referring to the learning curves of section 3.4.2. From the Bloch sphere distributions, it is possible to observe that in the ATN and one CNN cases the class separation mostly occurs just for one qubit, without affecting the performance of the classifiers.

The six-qubit feature map possesses the same number of gates as the one for the

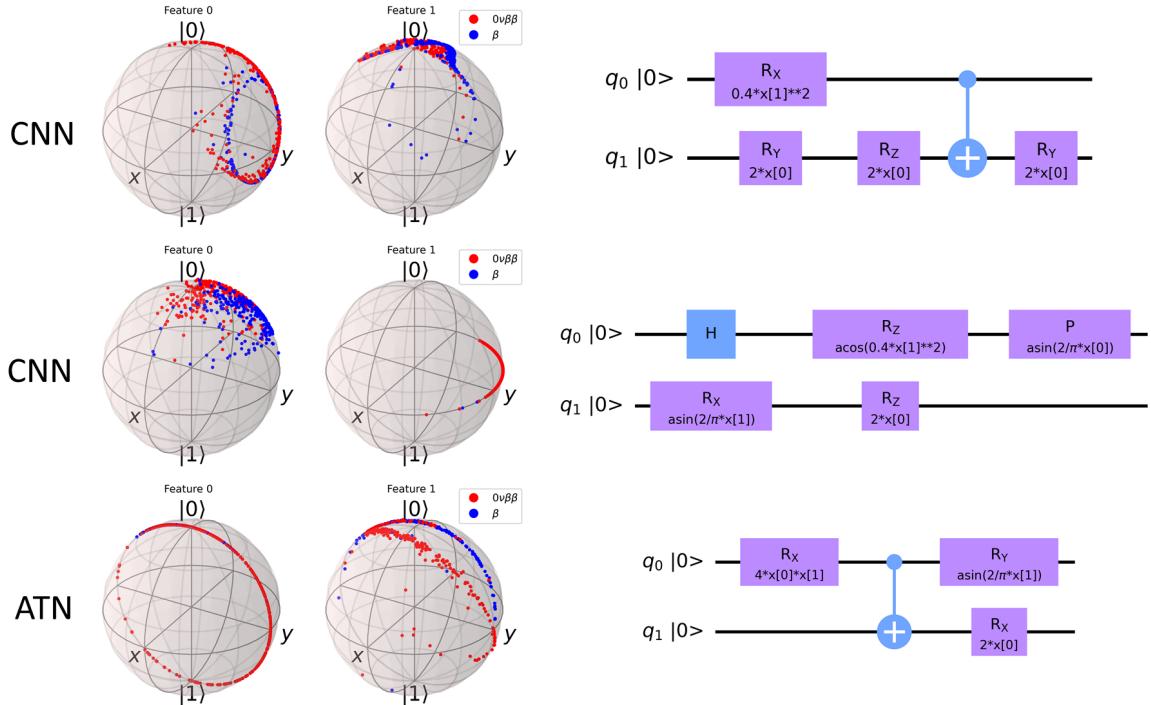


Figure C.1: Two-qubits quantum feature maps resulting from the GA algorithm, with the corresponding Bloch sphere visualization, on the toy model dataset with $[1 \times 1 \times 1]$ mm 3 voxel dimension.

$[5 \times 5 \times 5]$ mm 3 downsampled dataset. An interesting aspect is how the first qubit state does not depend on the input features, therefore won't help in the classification despite line q_0 having four gates and being entangled with q_1 . This quantum gate redundancy

may suggest that a premature convergence of the genetic algorithm occurred, or that the implemented fitness function needs improvements when approaching bigger-sized quantum circuits.

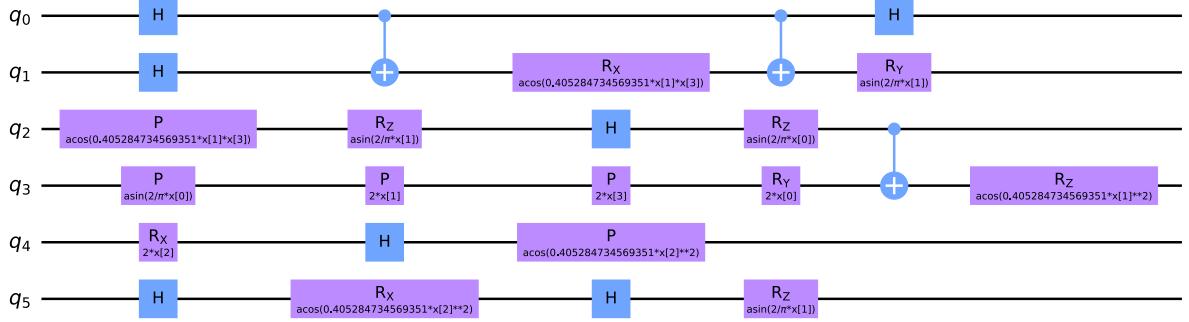


Figure C.2: Six-qubit feature map resulting from a GA instance on the four-feature CNN+ATN ensemble, on the toy model dataset with $[1 \times 1 \times 1]$ mm³ voxel dimension.

Bibliography

- [1] V. Havlíček et al. «Supervised learning with quantum-enhanced feature spaces.» In: *Nature* 567 (2019), pp. 209–212. DOI: 10.1038/s41586-019-0980-2. URL: <https://doi.org/10.1038/s41586-019-0980-2>.
- [2] Farzan Jazaeri et al. «A Review on Quantum Computing: Qubits, Cryogenic Electronics and Cryogenic MOSFET Physics». In: *arXiv: Quantum Physics* (2019). DOI: 10.48550/arXiv.1908.02656. URL: <https://doi.org/10.48550/arXiv.1908.02656>.
- [3] S. Bravyi et al. «Quantum advantage with noisy shallow circuits.» In: *Nat. Phys.* 16 (2020), pp. 1040–1045. DOI: 10.1038/s41567-020-0948-z. URL: <https://doi.org/10.1038/s41567-020-0948-z>.
- [4] F. Arute et al. «Quantum supremacy using a programmable superconducting processor.» In: *Nature* 574 (2019), pp. 505–510. DOI: 10.1038/s41586-019-1666-5. URL: <https://doi.org/10.1038/s41586-019-1666-5>.
- [5] E. Pednault et al. *On “Quantum Supremacy”*. 2019. URL: <https://www.ibm.com/blogs/research/2019/10/on-quantum-supremacy/>.
- [6] F. T. Arecchi et al. «Atomic Coherent States in Quantum Optics». In: *Phys. Rev. A* 6 (6 Dec. 1972), pp. 2211–2237. DOI: 10.1103/PhysRevA.6.2211. URL: <https://link.aps.org/doi/10.1103/PhysRevA.6.2211>.
- [7] Xin-Chuan Wu et al. «Amplitude-Aware Lossy Compression for Quantum Circuit Simulation». In: *ArXiv* abs/1811.05140 (2018).
- [8] Paul Isaac Hagouel and Ioannis Karayannidis. «Quantum computers: Registers, gates and algorithms». In: *2012 28th International Conference on Microelectronics - Proceedings, MIEL 2012* (May 2012). DOI: 10.1109/MIEL.2012.6222789. URL: <https://doi.org/10.1109/MIEL.2012.6222789>.
- [9] Kishor Bharti et al. «Noisy intermediate-scale quantum algorithms». In: *Rev. Mod. Phys.* 94 (1 Feb. 2022), p. 015004. DOI: 10.1103/RevModPhys.94.015004. URL: <https://link.aps.org/doi/10.1103/RevModPhys.94.015004>.
- [10] X. Liu and M.C. Hersam. «2D materials for quantum information science». In: *Nat Rev Mater* (2019), pp. 669–684. DOI: 10.1038/s41578-019-0136-x. URL: <https://doi.org/10.1038/s41578-019-0136-x>.
- [11] Thomas E. Roth, Ruichao Ma, and Weng C. Chew. *An Introduction to the Transmon Qubit for Electromagnetic Engineers*. 2021. DOI: 10.48550/ARXIV.2106.11352. URL: <https://arxiv.org/abs/2106.11352>.
- [12] Wen Guan et al. «Quantum machine learning in high energy physics». In: *Machine Learning: Science and Technology* 2.1 (Mar. 2021), p. 011003. DOI: 10.1088/2632-2153/abc17d. URL: <https://doi.org/10.1088/2632-2153/abc17d>.

- [13] J. Chow, O. Dial, and J. Gambetta. *IBM Quantum breaks the 100-qubit processor barrier*. 2021. URL: <https://research.ibm.com/blog/127-qubit-quantum-processor-eagle>.
- [14] Michele Grossi. «*W* Boson Polarization Studies for Vector Boson Scattering at LHC: from Classical Approaches to Quantum Computing». PhD thesis. Pavia U., 2021.
- [15] A. Kandala et al. «Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets». In: *Nature* 549 (2017), pp. 242–246. DOI: 10.1038/nature23879. URL: <https://doi.org/10.1038/nature23879>.
- [16] Zain H. Saleem et al. *Divide and Conquer for Combinatorial Optimization and Distributed Quantum Computation*. 2021. DOI: 10.48550/ARXIV.2107.07532. URL: <https://arxiv.org/abs/2107.07532>.
- [17] I. T. Jolliffe and J. Cadima. *Principal component analysis: a review and recent developments*. 2016. DOI: 10.1098/rsta.2015.0202. URL: <https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202>.
- [18] Keiron O’Shea and Ryan Nash. *An Introduction to Convolutional Neural Networks*. 2015. DOI: 10.48550/ARXIV.1511.08458. URL: <https://arxiv.org/abs/1511.08458>.
- [19] Ashish Vaswani et al. *Attention Is All You Need*. 2017. DOI: 10.48550/ARXIV.1706.03762. URL: <https://arxiv.org/abs/1706.03762>.
- [20] Shigeki Karita et al. «A Comparative Study on Transformer vs RNN in Speech Applications». In: *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, Dec. 2019. DOI: 10.1109/asru46091.2019.9003750. URL: <https://doi.org/10.1109%2Fasru46091.2019.9003750>.
- [21] Teresa Sancho-Lorente, Juan Román-Roche, and David Zueco. «Quantum kernels to learn the phases of quantum matter». In: *Phys. Rev. A* 105 (4 Apr. 2022), p. 042432. DOI: 10.1103/PhysRevA.105.042432. URL: <https://link.aps.org/doi/10.1103/PhysRevA.105.042432>.
- [22] L. Alchieri et al. «An introduction to quantum machine learning: from quantum logic to quantum deep learning». In: *Quantum Mach. Intell.* 3 (2021). DOI: 10.1007/s42484-021-00056-8. URL: <https://doi.org/10.1007/s42484-021-00056-8>.
- [23] M.A. Hearst et al. «Support vector machines». In: *IEEE Intelligent Systems and their Applications* 13.4 (1998), pp. 18–28. DOI: 10.1109/5254.708428. URL: <https://doi.org/10.1109/5254.708428>.
- [24] Bradley Efron and Trevor Hastie. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316576533. URL: <https://doi.org/10.1017/CBO9781316576533>.
- [25] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. «Quantum Support Vector Machine for Big Data Classification». In: *Phys. Rev. Lett.* 113 (13 Sept. 2014), p. 130503. DOI: 10.1103/PhysRevLett.113.130503. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.113.130503>.
- [26] Gian Gentinetta et al. *The complexity of quantum support vector machines*. 2022. DOI: 10.48550/ARXIV.2203.00031. URL: <https://arxiv.org/abs/2203.00031>.

- [27] Xinbiao Wang et al. «Towards understanding the power of quantum kernels in the NISQ era». In: *Quantum* 5 (Aug. 2021), p. 531. doi: 10.22331/q-2021-08-30-531. URL: <https://doi.org/10.22331%2Fq-2021-08-30-531>.
- [28] Jennifer R. Glick et al. *Covariant quantum kernels for data with group structure*. 2021. doi: 10.48550/ARXIV.2105.03406. URL: <https://arxiv.org/abs/2105.03406>.
- [29] S. Katoch and V. Chauhan S.S. Kumar. «A review on genetic algorithm: past, present, and future». In: *Multimed Tools Appl* 80 (2021). doi: 10.1007/s11042-020-10139-6. URL: <https://doi.org/10.1007/s11042-020-10139-6>.
- [30] Sergio Altares-López, Angela Ribeiro, and Juan José Garcia-Ripoll. «Automatic design of quantum feature maps». In: *Quantum Science and Technology* 6.4 (Aug. 2021), p. 045015. doi: 10.1088/2058-9565/ac1ab1. URL: <https://doi.org/10.1088%2F2058-9565%2Fac1ab1>.
- [31] S. F. Novaes. *Standard Model: An Introduction*. 2000. doi: 10.48550/ARXIV. HEP-PH/0001283. URL: <https://arxiv.org/abs/hep-ph/0001283>.
- [32] J. M. Butterworth. «The Standard Model: how far can it go and how can we tell?» In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2075 (Aug. 2016), p. 20150260. doi: 10.1098/rsta.2015.0260. URL: <https://doi.org/10.1098%2Frsta.2015.0260>.
- [33] Eak Paudel. «Problems of Standard Model, Review». In: *BMC Journal of Scientific Research* 4 (Mar. 2022). doi: 10.3126/bmcjsr.v4i1.42253. URL: <https://doi.org/10.3126/bmcjsr.v4i1.42253>.
- [34] S. Bilenky. «Neutrino oscillations: From a historical perspective to the present status». In: *Nuclear Physics B* 908 (2016). Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015, pp. 2–13. ISSN: 0550-3213. doi: 10.1016/j.nuclphysb.2016.01.025. URL: <https://www.sciencedirect.com/science/article/pii/S0550321316000353>.
- [35] M. Sajjad Athar, A. Fatima, and S. K. Singh. *Neutrinos and their interactions with matter*. 2022. doi: 10.48550/ARXIV.2206.13792. URL: <https://arxiv.org/abs/2206.13792>.
- [36] The KATRIN Collaboration. «Direct neutrino-mass measurement with sub-electronvolt sensitivity». In: *Nat. Phys.* 18 (2022), pp. 160–166. doi: 10.1038/s41567-021-01463-1. URL: <https://www.nature.com/articles/s41567-021-01463-1>.
- [37] F. Couchot et al. «Cosmological constraints on the neutrino mass including systematic uncertainties». In: *Astronomy & Astrophysics* 606 (Oct. 2017), A104. doi: 10.1051/0004-6361/201730927. URL: <https://doi.org/10.1051%2F0004-6361%2F201730927>.
- [38] Matteo Agostini et al. *Toward the discovery of matter creation with neutrinoless double-beta decay*. 2022. doi: 10.48550/ARXIV.2202.01787. URL: <https://arxiv.org/abs/2202.01787>.
- [39] Tasneem Rashid. «Development of pixel detector for ATLAS Inner Tracker (ITK) upgrade at HL-LHC and Searching for the Standard Model Higgs boson decay into b-quark pair with ATLAS experiment». PhD thesis. 2019.
- [40] Aron Beekman, Louk Rademaker, and Jasper van Wezel. «An introduction to spontaneous symmetry breaking». In: *SciPost Physics Lecture Notes* (Dec.

- 2019). DOI: 10.21468/scipostphyslectnotes.11. URL: <https://doi.org/10.21468%2Fscipostphyslectnotes.11>.
- [41] Peter W. Higgs. «Broken Symmetries and the Masses of Gauge Bosons». In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [42] The KATRIN Collaboration. «Measurement of the forward-backward asymmetry in $Z/\gamma^* \rightarrow \mu^+\mu^-$ decays and determination of the effective weak mixing angle.» In: *J. High Energ. Phys.* 2015 (2015). DOI: 10.1007/JHEP11(2015)190. URL: [https://link.springer.com/article/10.1007/JHEP11\(2015\)190](https://link.springer.com/article/10.1007/JHEP11(2015)190).
- [43] F. Gilman, K. Kleinknecht, and Burkhard Renk. «The cabibbo-kobayashi-maskawa quark-mixing matrix». In: *The European Physical Journal C* 15 (Mar. 2000), pp. 110–114. DOI: 10.1007/BF02683407. URL: <https://doi.org/10.1007/BF02683407>.
- [44] D. Abbaneo. *Review of experimental results on neutral B meson oscillations*. 2000. DOI: 10.48550/ARXIV.HEP-EX/0012010. URL: <https://arxiv.org/abs/hep-ex/0012010>.
- [45] C. Giganti, S. Lavignac, and M. Zito. «Neutrino oscillations: The rise of the PMNS paradigm». In: *Progress in Particle and Nuclear Physics* 98 (Jan. 2018), pp. 1–54. DOI: 10.1016/j.ppnp.2017.10.001. URL: <https://doi.org/10.1016%2Fj.ppnp.2017.10.001>.
- [46] Petr Vogel, Liangjian Wen, and Chao Zhang. «Neutrino Oscillation Studies with Reactors». In: *Nature Communications* 6 (Mar. 2015). DOI: 10.1038/ncomms7935. URL: <https://doi.org/10.1038/ncomms7935>.
- [47] Stefano Gariazzo et al. *Neutrino mass and mass ordering: No conclusive evidence for normal ordering*. 2022. DOI: 10.48550/ARXIV.2205.02195. URL: <https://arxiv.org/abs/2205.02195>.
- [48] S. Bilenky. *Neutrinos: Majorana or Dirac?* 2020. DOI: 10.48550/ARXIV.2008.02110. URL: <https://arxiv.org/abs/2008.02110>.
- [49] Andrea Giuliani and Alfredo Poves. «Neutrinoless Double-Beta Decay». In: *Advances in High Energy Physics* 2012 (Dec. 2012). DOI: 10.1155/2012/857016. URL: <https://doi.org/10.1155/2012/857016>.
- [50] Alexander Barabash. *Precise Half-Life Values for Two-Neutrino Double-Decay: 2020 review*. 2020. DOI: 10.48550/ARXIV.2009.14451. URL: <https://arxiv.org/abs/2009.14451>.
- [51] Guofu Cao. «Search for $0\nu\beta\beta$ decay with EXO-200 and nEXO». In: *PoS* HQL2018 (2018), p. 054. DOI: 10.22323/1.332.0054. URL: <https://doi.org/10.22323/1.332.0054>.
- [52] Michelle J. Dolinski, Alan W.P. Poon, and Werner Rodejohann. «Neutrinoless Double-Beta Decay: Status and Prospects». In: *Annual Review of Nuclear and Particle Science* 69.1 (Oct. 2019), pp. 219–251. DOI: 10.1146/annurev-nucl-101918-023407. URL: <https://doi.org/10.1146%2Fannurev-nucl-101918-023407>.
- [53] Anastasiia Zolotarova. «Bolometric Double Beta Decay Experiments: Review and Prospects». In: *Symmetry* 13.12 (2021). ISSN: 2073-8994. DOI: 10.3390/sym13122255. URL: <https://www.mdpi.com/2073-8994/13/12/2255>.

- [54] A. Gando et al. «Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with KamLAND-Zen». In: *Phys. Rev. Lett.* 117 (8 Aug. 2016), p. 082503. DOI: 10.1103/PhysRevLett.117.082503. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.117.082503>.
- [55] A. Mastbaum, F. Psihas, and J. Zennamo. *Xenon-Doped Liquid Argon TPCs as a Neutrinoless Double Beta Decay Platform*. 2022. DOI: 10.48550/ARXIV.2203.14700. URL: <https://arxiv.org/abs/2203.14700>.
- [56] Aneesha Avasthi et al. «Kiloton-scale xenon detectors for neutrinoless double beta decay and other new physics searches». In: *Physical Review D* 104 (Dec. 2021). DOI: 10.1103/PhysRevD.104.112007. URL: <https://doi.org/10.1103/PhysRevD.104.112007>.
- [57] A. Avasthi et al. *Low Background kTon-Scale Liquid Argon Time Projection Chambers*. 2022. DOI: 10.48550/ARXIV.2203.08821. URL: <https://arxiv.org/abs/2203.08821>.
- [58] A. Abed Abud et al. *Snowmass Neutrino Frontier: DUNE Physics Summary*. 2022. DOI: 10.48550/ARXIV.2203.06100. URL: <https://arxiv.org/abs/2203.06100>.
- [59] The DUNE collaboration. *An International Experiment for Neutrino Science*. 2020. URL: <https://www.dunescience.org/>.
- [60] B. Abi et al. *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE*. 2020. DOI: 10.48550/ARXIV.2002.02967. URL: <https://arxiv.org/abs/2002.02967>.
- [61] C. Farnese. «The ICARUS experiment». In: *EPJ Web of Conferences* 182 (Jan. 2018), p. 02042. DOI: 10.1051/epjconf/201818202042. URL: <https://doi.org/10.1051/epjconf/201818202042>.
- [62] Jose Alfonso Soto Oton. «Scintillation light detection techniques in a 750-ton liquid argon TPC for the Deep Underground Neutrino Experiment». Presented 13 Jun 2022. Dec. 2022. URL: <https://cds.cern.ch/record/2812306>.
- [63] R Francini et al. «VUV-Vis optical characterization of Tetraphenyl- butadiene films on glass and specular reflector substrates from room to liquid Argon temperature». In: *Journal of Instrumentation* 8.09 (Sept. 2013), P09006–P09006. DOI: 10.1088/1748-0221/8/09/p09006. URL: <https://doi.org/10.1088%2F1748-0221%2F8%2F09%2Fp09006>.
- [64] M. Kuźniak et al. «Polyethylene naphthalate film as a wavelength shifter in liquid argon detectors». In: *The European Physical Journal C* 79.4 (Mar. 2019). DOI: 10.1140/epjc/s10052-019-6810-8. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-019-6810-8>.
- [65] J. Soto-Oton and for DUNE Collaboration. «Impact of xenon doping in the scintillation light in a large liquid-argon TPC». PhD thesis. 2021. DOI: 10.48550/ARXIV.2109.05858. URL: <https://arxiv.org/abs/2109.05858>.
- [66] H. V. Souza. *ARAPUCA, light trapping device for the DUNE experiment*. 2021. DOI: 10.48550/ARXIV.2112.02967. URL: <https://arxiv.org/abs/2112.02967>.
- [67] B. Abi et al. «Supernova neutrino burst detection with the Deep Underground Neutrino Experiment». In: *The European Physical Journal C* 81.5 (May 2021). DOI: 10.1140/epjc/s10052-021-09166-w. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-021-09166-w>.

- [68] Francesco Capozzi et al. «DUNE as the Next-Generation Solar Neutrino Experiment». In: *Physical Review Letters* 123.13 (Sept. 2019). doi: 10.1103/physrevlett.123.131803. URL: <https://doi.org/10.1103%2Fphysrevlett.123.131803>.
- [69] R. Nakamura et al. «Research and development toward KamLAND2-Zen». In: *Journal of Physics: Conference Series* 1468 (Feb. 2020), p. 012256. doi: 10.1088/1742-6596/1468/1/012256. URL: <https://doi.org/10.1088/1742-6596/1468/1/012256>.
- [70] F. Agostini et al. *Sensitivity of the DARWIN observatory to the neutrinoless double beta decay of ^{136}Xe* . 2020. doi: 10.48550/ARXIV.2003.13407. URL: <https://arxiv.org/abs/2003.13407>.
- [71] G Adhikari et al. «nEXO: neutrinoless double beta decay search beyond 10^{28} year half-life sensitivity». In: *Journal of Physics G: Nuclear and Particle Physics* 49.1 (Dec. 2021), p. 015104. doi: 10.1088/1361-6471/ac3631. URL: <https://doi.org/10.1088%2F1361-6471%2Fac3631>.
- [72] Justo Martín-Albo. *The NEXT experiment for neutrinoless double beta decay searches*. 2019. doi: 10.48550/ARXIV.1910.07314. URL: <https://arxiv.org/abs/1910.07314>.
- [73] Michael Wallbank. «Automated Reconstruction, Signal Processing and Particle Identification in DUNE». In: *PoS EPS-HEP2017* (2017). Ed. by Paolo Checchia et al., p. 815. doi: 10.22323/1.314.0815. URL: <https://doi.org/10.22323/1.314.0815>.
- [74] B. Abi et al. «Neutrino interaction classification with a convolutional neural network in the DUNE far detector». In: *Physical Review D* 102.9 (Nov. 2020). doi: 10.1103/physrevd.102.092003. URL: <https://doi.org/10.1103%2Fphysrevd.102.092003>.
- [75] Tatsuo Tabata, Kunihiro Shinoda, and Rinsuke Ito. *An analytic formula for continuous slowing-down approximation ranges of 1-keV to 1-GeV electrons*. Dec. 1994.
- [76] M.A. Zucker M.J. Berger J.S. Coursey and J. Chang. *NIST Standard Reference Database 124*. 2017 (last updated). doi: 10.18434/T4NC7P. URL: <https://dx.doi.org/10.18434/T4NC7P>.
- [77] B. J. P. Jones. *The Physics of Neutrinoless Double Beta Decay: A Primer*. 2021. doi: 10.48550/ARXIV.2108.09364. URL: <https://arxiv.org/abs/2108.09364>.
- [78] R. L. Workman et al. «Review of Particle Physics». In: *PTEP* 2022 (2022), p. 083C01. doi: 10.1093/ptep/ptac097. URL: <https://doi.org/10.1093/ptep/ptac097>.
- [79] S. Agostinelli et al. «GEANT4— a simulation toolkit». In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. doi: 10.1016/S0168-9002(03)01368-8. URL: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [80] MD SAJID ANIS et al. *Qiskit: An Open-source Framework for Quantum Computing*. 2021. doi: 10.5281/zenodo.2573505.
- [81] James C. Spall. «An Overview of the Simultaneous Perturbation Method for Efficient Optimization». In: *Johns Hopkins Apl Technical Digest* 19 (1998), pp. 482–492.
- [82] IBM Quantum. 2021. URL: <https://quantum-computing.ibm.com/>.
- [83] The LArSoft collaboration. *LArSoft*. URL: <https://larsoft.org/>.

- [84] Eric D. Church. *LArSoft: A Software Package for Liquid Argon Time Projection Drift Chambers*. 2014. arXiv: 1311.6774.
- [85] B. Abi et al. *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics*. 2020. doi: 10.48550/ARXIV.2002.03005. URL: <https://arxiv.org/abs/2002.03005>.
- [86] T. Alion et al. *Experiment Simulation Configurations Used in DUNE CDR*. 2016. doi: 10.48550/ARXIV.1606.09550. URL: <https://arxiv.org/abs/1606.09550>.
- [87] A. Abed Abud et al. *Deep Underground Neutrino Experiment (DUNE) Near Detector Conceptual Design Report*. 2021. doi: 10.48550/ARXIV.2103.13910. URL: <https://arxiv.org/abs/2103.13910>.

Acknowledgements - Ringraziamenti

I wish to express my gratitude, first of all, to my supervisor Dr. Andrea Giachero, my co-supervisor Dr. Daniele Guffanti and my external supervisor Dr. Michele Grossi, who guided me throughout the project and made this thesis possible. Thanks to Marco Rossi, for bringing up and contributing to the Attention Network part of my thesis, and for his great coding advice. I also want to acknowledge Prof. Francesco Terranova for his great suggestions and for helping me to get started with my Erasmus.

Special thanks to Dr. Inés Gil Botella and Dr. Clara Cuesta for mentoring me during my Erasmus period, to CIEMAT Neutrino Group and to all CIEMAT PhD teams for their heartwarming hospitality. I'm looking forward to seeing you again soon!

Infine, non può mancare un ringraziamento alla mia famiglia, che da ventiquattro anni mi dà sostegno ed è una fonte inesauribile di affetto, e a tutti gli amici con cui ho condiviso il mio percorso.