UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Study of b- and c- jets identification with quantum machine learning algorithms and application to the Higgs reconstruction.

Thesis supervisor

Prof. Donatella Lucchesi

Thesis co-supervisor

Dr. Lorenzo Sestini

Candidate

Carlos Cocha

Academic Year 2021-2022

Contents

Introduction

In particle physics, machine learning algorithms have been crucial for jet classification and physics analysis. A jet is a narrow cone of hadrons and other particles produced by quark fragmentation and quark hadronization. Recently, the LHCb collaboration at CERN has already demonstrated the benefit of using Quantum Machine Learning for the b- versus $\bar{b}$-quark jet classification which potentially can perform better than classical machine learning algorithms. At the LHCb experiment, jet tagging exploits the excellent capability of the detector to precise reconstruct vertices displaced with respect to the interaction point, which allow to distinguish the long-lived heavy quarks from light quarks jets using a standard machine learning technique, called Boosted Decision Trees, trained with secondary vertex features.

In this thesis, I present for the first time a new approach for jet flavour tagging, between b- versus c-quark jets, based on Quantum Machine Learning techniques. In this technique, the data are embedded in quantum states of a quantum circuit through a quantum feature map, then the initial states get processed by a variational quantum circuit with trainable parameterized gates and finally, the measurement of the final state is mapped to a binary classification label (b- or c-quark jet). Besides, the training of the model will be carried out by using official LHCb simulated data at the center of mass energy of 13 TeV and the tagging performance is compared with standard machine learning methods like the Boosted Decision Trees. The light quark contribution is not considered here because the goal of the thesis is to study the b- versus c-quark jet separation. Finally, the optimized code will be used for the Higgs search to determine the $H \to b\bar{b}$ and $H \to c\bar{c}$ sensitivity on samples of LHCb simulated data corresponding to the integrated luminosity $6\,\mathrm{fb}^{-1}$ of the Run 2.

The results of this work for the jet flavor classification and the Higgs search are structured as follows:

- **Chapter 1**: A brief overview of the Standard Model (SM) is given. Later, the Higgs boson production mechanisms and its decay channels are discussed. Additionally, the b- and c-quark production mechanisms at colliders are presented as well as their hadronization process.
- **Chapter 2**: The LHCb experiment is presented. First, the LHC accelerator and the LHCb detector are described. Then, the jet reconstruction algorithm is described. This chapter finishes describing the current jet identification technique for jet flavor tagging between b- and c-quark jets.
- **Chapter 3**: This chapter introduces the theory about Quantum Computing. First, a theoretical overview of the qubit and entanglement concepts are presented. After, a brief introduction of key concepts of Supervised Machine Learning and the general concept of Quantum Machine Learning are treated.
- **Chapter 4**: The results of jet flavor tagging using Quantum Machine Learning models and the standard classical model are treated. An explanation of the data-set used for the analysis in terms of variables and selections is discussed as well. Then, a detailed explanation of several quantum models including their software implementation are explained. A classical model, based on Boosted Decision Trees is also presented and used as a reference. Finally, the performance of the quantum models is evaluated and compared with the classical method.

- **Chapter 5**: Here, a optimized quantum machine algorithm will be used for the Higgs to $c\bar{c}$ analysis. The Higgs sensitivity by using the algorithms discussed in Chapter 4 is calculated. The performance of the quantum model is compared with the classical method using their significance.
- **Chapter 6**: Lastly, a summary of the results of the quantum machine learning algorithms and their use for the Higgs search are presented, together with some final considerations.

Chapter 1

Theoretical introduction

1.1 Standard Model

The Standard Model is a Quantum Field Theory (QFT) mixing quantum mechanics and special relativity to describe elementary particles as quantizations of fields spread across the entire universe. Besides, the underlying foundation of the SM is built on the concepts of local gauge symmetry and spontaneous symmetry breaking. This model also provides a mechanism for mass generation through a scalar field sector but it does not include a description of the gravitational interaction, which has negligible impact at subnuclear scales.

The elementary particles can be divided in four main categories: gauge and scalar bosons, leptons and quarks. The gauge bosons are represented by four different spin-1 particles, the gluon (g) mediating the strong interaction, the photon (γ) mediating the electromagnetic interaction and the W and Z bosons mediating the weak force. The quarks, represented by six spin-1/2 fermions grouped in three families, feel the strong force and also carry an electric and weak charge, making them interacting also through the electromagnetic and weak force. Among these particles are the quark up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Three leptons interact through the weak and electromagnetic interaction and are composed of electrons (e), muons (μ) and taus (τ), grouped in three families where each one is coupled to their corresponding neutrinos, ν_e, ν_μ and ν_τ. Neutrinos are special type of leptons since they interact only through the weak interaction and do not carry any electric charge. A summary of the properties and interactions of the SM elementary particles are presented in Figure 1.1.

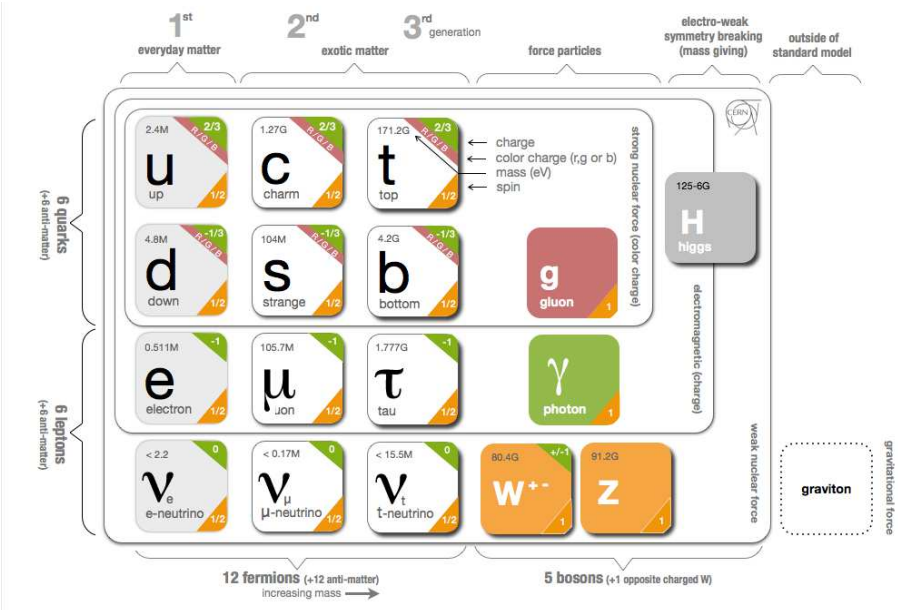


Figure 1.1: Particle content of the Standard Model [1].

The last piece is the Higgs boson which is the only spin-0 (scalar) particle which through the spontaneous symmetry breaking mechanism provides masses to the weak (W and Z) bosons, quarks and leptons. This boson differs from vector bosons as it couples only to particles carrying a mass, and not a charge as for the spin-1 bosons.

1.1.1 The Fermion and Gauge Sectors

In QFT, the requirement of gauge invariance, or symmetry of the Lagrangian density under local phase transformations of a mathematical group, introduces the presence of massless boson fields, which are interpreted as the quanta mediating the interaction. One example of a gauge QFT is the quantum electrodynamics (QED) where the Lagrangian density of the electron field (Ψ) interacting with an electromagnetic field is defined as $\mathcal{L}_{QED} = -F_{\mu\nu}F^{\mu\nu} + i\bar{\Psi}D_\mu\gamma^\mu\Psi - m_e\bar{\Psi}\Psi$, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength, A_μ is the photon field introduced by gauge invariance, $D_\mu = \partial_\mu - iqA_\mu$ is the covariant derivative, and q is the absolute value of electron electric charge. $\mathcal{L}_{QED}$ is invariant under the group local transformations $U(1)_Q$ of the electron field $\Psi'(x) = e^{iq\alpha(x)}\Psi(x)$, and photon field $A'_\mu = A_\mu - \partial_\mu\alpha(x)/q$.

The SM is a generalization of the fundamental interactions under the $SU(3)_C \bigotimes SU(2)_L \bigotimes U(1)_Y$ group. The strong interaction is studied by quantum chromodynamics (QCD) based on the $SU(3)_C$ color group. The electromagnetic and weak interactions are studied in a unified manner by the electroweak (EW) theory based on the $SU(2)_L \bigotimes U(1)_Y$ group. Experimental measurements show that only fermions with left-handed chirality interact via charged weak interactions.

The short range nature of weak interactions implies the existence of massive vector bosons. Furthermore, observed fermions like the electron are massive particles. However, explicit mass terms associated to the gauge boson and fermion fields are forbidden in the EW Lagrangian density due to $SU(2)_L \bigotimes U(1)_Y$ gauge symmetry. Therefore, the formulation needs a gauge invariant mechanism to generate these masses [2, 3].

1.1.2 The Scalar Sector

The Brout-Englert-Higgs (BEH) mechanism solves elegantly the mass-generation paradigm without violating gauge symmetry. It adds a complex scalar doublet Φ by means of a Lagrangian density $\mathcal{L}_{BEH} = D^\dagger_\mu\Phi D^\mu\Phi - V(\Phi)$, where $V(\Phi) = -\mu^2(\Phi^\dagger\Phi) + \lambda^2(\Phi^\dagger\Phi)$ and μ^2, $\lambda > 0$ [2, 3].

In the BEH mechanism, Φ has a vacuum expectation state $\langle\Phi\rangle_0$ that satisfies the $\langle\Phi^\dagger\Phi\rangle_0 = \mu^2/2\lambda = \nu^2/2$ condition, where ν is known as the vacuum expectation value. This leads to a 'spontaneous' breaking of the electroweak symmetry group $\langle\Phi\rangle_0$ (or EWSB) to the electromagnetic group $U(1)_Q$ with a massless photon, and generating the masses to the charged- and neutral-current weak bosons W and Z, respectively. Furthermore, a real scalar field named the Higgs field is introduced with its massive particle, the Higgs boson. Table 1.1 summarizes and Figure 1.2 illustrates the BEH Lagrangian terms after EWSB.

$\mathcal{L}_{BEH}$ **terms after EWSB**	**Meaning**
$\frac{1}{2}(\partial_\mu \mathrm{H})(\partial^\mu \mathrm{H})$	Higgs kinematic term
- $\frac{1}{2}\mathrm{m}_\mathrm{H}^2\mathrm{H}^2$ - $\nu\lambda_{\mathrm{HHH}}\mathrm{H}^3$ - $\frac{1}{4}\lambda_{\mathrm{HHHH}}\mathrm{H}^4$+ $\frac{1}{4}\lambda\nu^4$	Higgs potential term and self-interactions
+ $\mathrm{m}_W^2\mathrm{HW}_\mu^-\mathrm{W}^{+,\mu}$ +$\mathrm{g}_{\mathrm{HWW}}\mathrm{HW}_\mu^-\mathrm{W}^{+,\mu}$	W mass term and HWW interaction
+ $\frac{1}{2}\mathrm{m}_\mathrm{Z}^2\mathrm{Z}_\mu\mathrm{Z}^\mu$ + $\frac{1}{2}\mathrm{g}_{\mathrm{HZZ}}\mathrm{HZ}_\mu\mathrm{Z}^\mu$	Z mass term and HZZ interaction
+ $\mathrm{g}_{\mathrm{WWHH}}\mathrm{H}^2\mathrm{W}_\mu^-\mathrm{W}^{+,\mu}$	WWHH quartic interaction
+ $\frac{1}{2}\mathrm{g}_{\mathrm{ZZHH}}\mathrm{H}^2 Z_\mu Z^\mu$	ZZHH quartic interaction

Table 1.1: BEH Lagrangian terms after the spontaneous electroweak symmetry breaking (EWSB).

On the other hand, charged fermion fields acquire mass through the Higgs-fermion Yukawa interaction, where fermion masses are given by $m_f = y_f\nu/\sqrt{2}$, and y_f is the associated Yukawa coupling. Controversially, neutrino fields remain massless in the theory [2, 3].

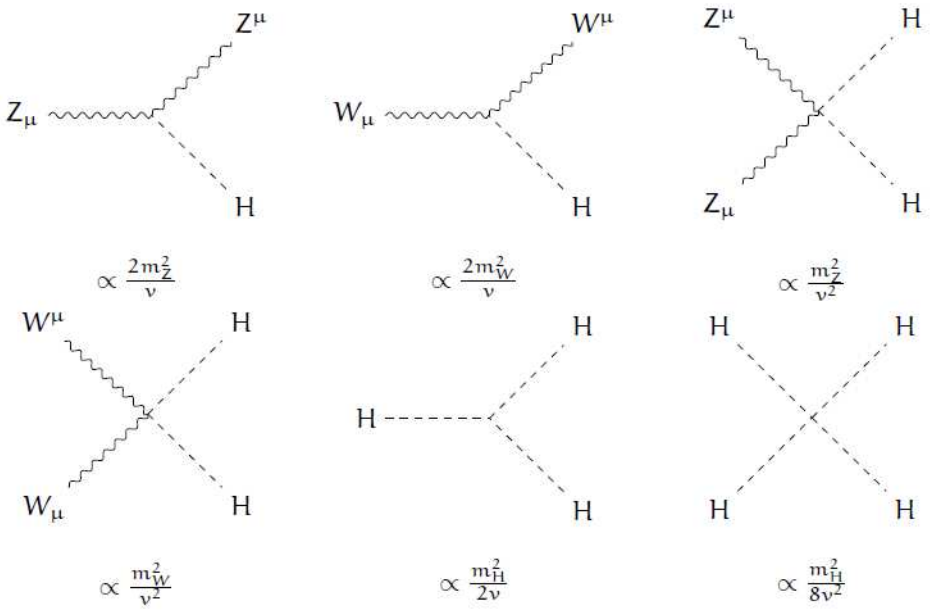


Figure 1.2: Higgs interaction terms [3].

One important prediction is the structure of the Higgs boson potential with the associated cubic (λ_{HHH}) and quartic (λ_{HHHH}) Higgs self-interactions which are responsible for the Higgs boson mass m_{H}. Furthermore, the Higgs self-couplings are connected to the Higgs boson mass and the vacuum expectation value by the relation $\lambda_{\mathrm{HHH}} = \lambda_{\mathrm{HHHH}} = m_H^2/2\nu^2$. The experimental confirmation of this relation is crucial to our understanding of the BEH mechanism and a test of the internal consistency of the SM formulation.

1.2 The Higgs Boson

On July 4th 2012, the CMS and ATLAS Collaborations of CERN announced the observation of a new resonance with mass around $m_{\mathrm{H}} = 125$ GeV and properties compatible with the SM Higgs boson [4, 5]. Since then, both collaborations have performed associated precision measurements using proton-proton collision data with 7, 8 and 13 TeV center of mass energy.

1.2.1 Production mechanisms and decay channels

The value of m_{H} was the only free parameter left to predict the Higgs boson phenomenology. After its discovery, its production mechanisms and decays can be fully studied. The Higgs boson has several production mechanisms where gluon fusion (ggF-H) is the main production mode followed by the vector-boson fusion (VBF-H), with a cross section production of approximately one order of magnitude smaller. Other subdominant production modes are the associated production with a W or a Z boson (VH) and the associated production with a top quark pair (ttH), among others. Examples of representative leading order (LO) diagrams are shown in Figure 1.3 while in the Table 1.2 different theoretical production cross section for several modes are shown.

$\sqrt{s}$	ggF-H	VBF-H	WH	ZH	ttH	total
13 TeV	$48.6^{+4.4\%}_{-7.0\%}$	$3.78^{+2.2\%}_{-2.2\%}$	$1.37^{+2.6\%}_{-2.6\%}$	$0.88^{+4.1\%}_{-3.5\%}$	$0.50^{+6.8\%}_{-9.9\%}$	55.1
14 TeV	$54.7^{+4.4\%}_{-7.0\%}$	$4.28^{+2.2\%}_{-2.2\%}$	$1.51^{+1.9\%}_{-2.0\%}$	$0.99^{+4.1\%}_{-3.7\%}$	$0.60^{+6.9\%}_{-9.8\%}$	62.1

Table 1.2: Higgs boson production cross sections (in pb) and their relative uncertainties (e.g. theory, PDF, α_S) for several Higgs boson ($m_{\mathrm{H}} = 125$ GeV) production modes at $\sqrt{s} = 13$ and 14 TeV proton-proton collisions [3]. The order of QCD (EW) calculation is N3LO (NLO) for ggF and NNLO (NLO) for the others.

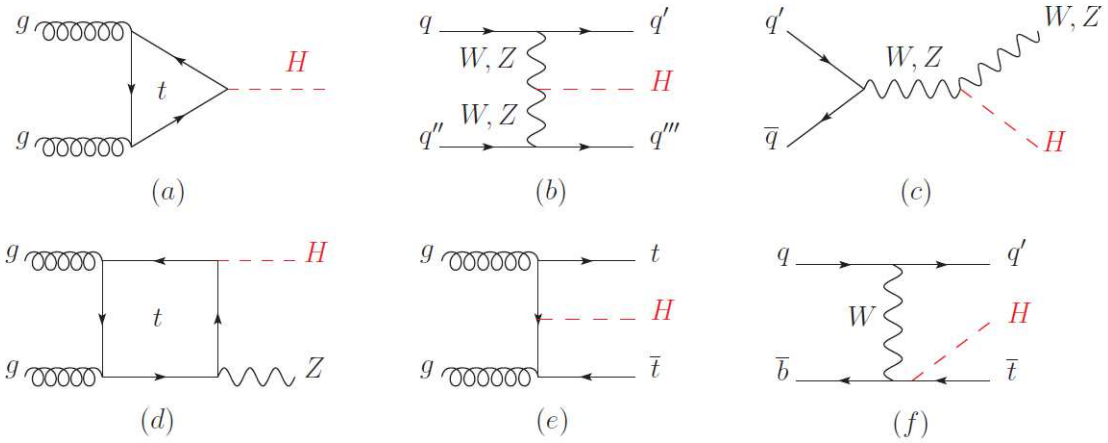


Figure 1.3: Main LO Feynman diagrams contributing to the Higgs boson production in (a) gluon fusion, (b) Vector-boson fusion, (c) associate production with a W or Z gauge boson at tree level from a quark-quark interaction, (d) associated production with a gauge boson (at loop level from a gluon-gluon interaction), (e) associated production with a pair of top quarks (there is a similar diagram for the associated production with a pair of bottom quarks), (f) production in association with a single top quark.

The main theoretical Higgs boson decays to gauge bosons and to some fermions are shortlisted in Table 1.3. The dominant Higgs boson decay channel is to a bottom quark pair ($H \rightarrow b\bar{b}$) with around 58.2% branching ratio. Note also that the Higgs boson can indirectly decay into massless particles via intermediate fermion loops: two photons ($H \rightarrow \gamma\gamma$), a Z boson and a photon ($H \rightarrow Z\gamma$), and two gluons ($H \rightarrow gg$).

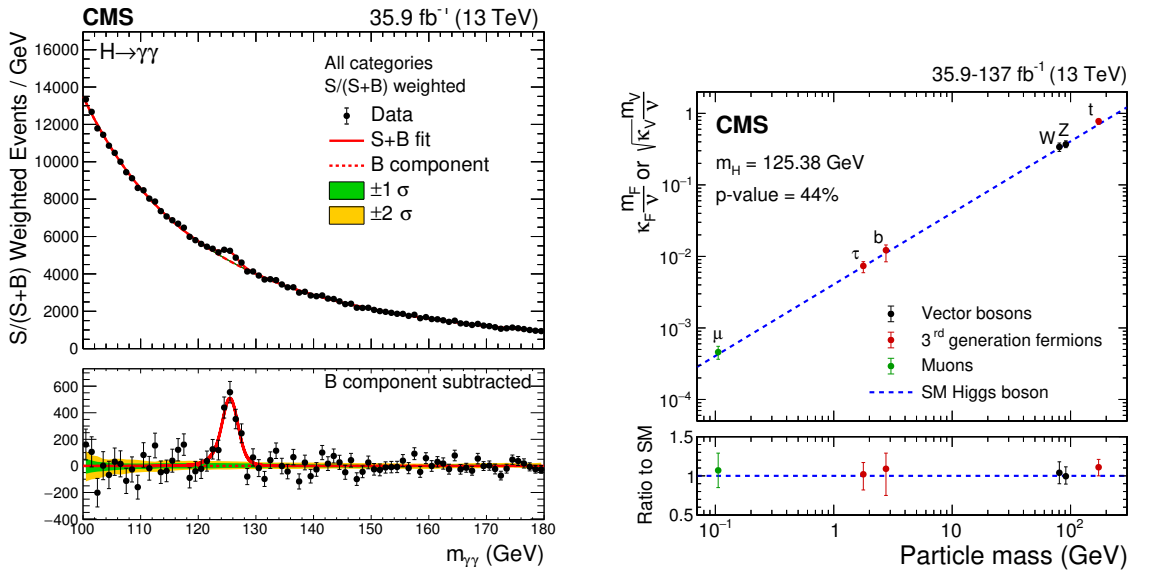
Decay mode	**Branching ratio**	**Relative uncertainty**
$H \rightarrow \gamma\gamma$	2.22×10^{-3}	$\pm 2.1\%$
$H \rightarrow ZZ$	2.62×10^{-2}	$\pm 1.5\%$
$H \rightarrow W^+W^-$	2.14×10^{-1}	$\pm 1.5\%$
$H \rightarrow \tau^+\tau^-$	6.27×10^{-2}	$\pm 1.6\%$
$H \rightarrow b\bar{b}$	5.82×10^{-1}	$^{+1.2\%}_{-1.3\%}$
$H \rightarrow c\bar{c}$	2.89×10^{-2}	$^{+5.5\%}_{-2.0\%}$
$H \rightarrow Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \rightarrow \mu^+\mu^-$	2.18×10^{-4}	$\pm 1.7\%$

Table 1.3: Higgs boson branching ratios and relative uncertainties for several Higgs boson ($m_H = 125$ GeV) decay modes [3].

1.2.2 Higgs Experimental Status

The discovery of the Higgs boson at the LHC was carried out analyzing a broad range of production mechanisms and decay channels. The most sensitive decay channels for studying the Higgs boson are the $H \rightarrow ZZ \rightarrow 4l$ ($l = e, \mu$) and $H \rightarrow \gamma\gamma$ channels where, despite the low expected rates, they benefit from small background processes to have the best mass resolution ($\sim 1-2\%$) due to well-reconstructed final state products.

The $H \rightarrow W^+W^- \rightarrow l^+\nu_l l'^-\nu_{l'}$ channel has relatively better rates, but the unmeasured (missing) energy from neutrinos has an impact on the resolution of different mass-related observables, e.g. the transverse mass resolution is $\sim 20\%$. The $H \rightarrow b\bar{b}$ ($\tau^+\tau^-$) decay channel benefits from high rates and intermediate mass resolution, 10% (15%), but is affected by overwhelming irreducible background contamination [3]. On the other hand, very challenging channels due to low very rates or large background levels are $H \rightarrow \mu^+\mu^-$ and $H \rightarrow c\bar{c}$, respectively.



(a) Di-photon mass distribution from the 2016 dataset mass measurement [14]

(b) Measurement of the Higgs couplings [22].

Figure 1.4: CMS Higgs physics results in Run 2.

In the LHC Run 1 (2009-2012), the ATLAS and CMS collaborations analyzed the pp collision datasets at $\sqrt{s} = 7$ and 8 TeV to discover the Higgs boson [6, 7]. The decay channels involving gauge bosons, $H \to ZZ \to 4l$ ($l = e, \mu$) and $H \to \gamma\gamma$, contributed with the most significant excesses. Besides, the $H \to W^+W^- \to l^+\nu_l l'^-\nu_{l'}$, $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ channels were included but contributed with lower significance. After that, studies of the spin-parity (JP) quantum numbers revealed the scalar and even-parity nature ($JP = 0^+$) of the discovered particle as predicted by the SM [8, 9]. The combined Run 1 ATLAS and CMS Higgs boson mass measurements yield to a value of $m_H = 125.09 \pm 0.24$ GeV using $H \to ZZ \to 4l$ ($l = e, \mu$) and $H \to \gamma\gamma$, taking advantage from their excellent mass resolution [10].

During the LHC Run 2 (2015-2018), ATLAS and CMS experiments had an excellent operation performance collecting high-quality data at $\sqrt{s} = 13$ TeV, around 6 times the LHC Run 1 dataset integrated luminosity. Measurements of the mass, couplings, differential and fiducial cross sections have been performed with the partial and full dataset by ATLAS [11, 12] and CMS [13, 14]. The Higgs boson mass measurement has been performed using the $H \to ZZ \to 4l$ ($l = e, \mu$) and $H \to \gamma\gamma$ decay channels. The distribution of the di-photon invariant mass in the 2016 data is presented in Figure 1.4.a. The best mass measurement yields a value of $m_H = 125.38 \pm 0.14$ GeV from the CMS combination of Run 1 and 2016 datasets [15].

Both experiments have presented the observation of the Higgs coupling to third generation fermions studying the $H \to b\bar{b}$ [16, 17] and $H \to \tau^+\tau^-$ [18, 19] decays, and the ttH production mode [20, 21]. For one side, the CMS experiment found the first evidence for the Higgs coupling to the second generation fermions in the analysis of the $H \to \mu^+\mu^-$ decay channel [22]. One the other hand, ATLAS showed the first evidence for the $H \to (\gamma/Z)^*\gamma \to ll\gamma$ decay channel [23]. Besides, Figure 1.4.b shows the Higgs coupling measurements using CMS Run-2 data.

In particular, the measurement of $H \to b\bar{b}$ decays was believed to be impossible to achieve due to the overwhelming multi-jet background and intermediate mass resolution. However, novel analysis methods based on machine learning (ML) were developed for object identification, reconstruction and signal identification, to maximize the search sensitivity. Nowadays, both experiments are developing novel methods to measure the $H \to b\bar{b}$ decay where more data are needed to measure it [24, 25].

1.3 b- and c-quark production in proton-proton collisions

At LHC, quark pairs are produced in pp collisions at $\sqrt{s} \approx 13$ TeV. The kinematic reconstruction of a pp collision requires the position and momentum of the quark and gluon content of the proton which is encoded in the Parton Distribution Functions (PDFs) [26]. The PDF $f_{a/A}(x, \mu)$ represents the probability distribution of finding the parton a inside the hadron A, carrying a fraction x of the total momentum of A. The parameter μ is related to the renormalization scale (with mass units).

The PDFs is expressed in terms of cross sections involving hadronic initial states. Using the QCD Factorization Theorem [27] for the case of quark pair production, it can be written as

$$\mathrm{d}\sigma_{pp\to Q\bar{Q}} = \sum_{n,m} \int\int \mathrm{d}x_n \mathrm{d}x_m \, f_{n/p}(x_n, \mu) f_{m/p}(x_m, \mu) \, \mathrm{d}\hat{\sigma}_{nm\to Q\bar{Q}} \tag{1.1}$$

where n and m run over all the possible partons with a non-vanishing contribution in $\mathrm{d}\hat{\sigma}_{nm\to Q\bar{Q}}$ at the parton level. Besides, the processes contributing to the b- and c- production in the Leading (LO) Order approximation, are:

- **QCD quark-fusion**: a quark and an anti-quark in the initial state annihilate each other into a gluon which produces a $Q\bar{Q}$ pair.
- **QCD gluon-fusion**: two gluons in the initial state interact with a third gluon (in the s-channel) or a quark (in the t-channel), producing a $Q\bar{Q}$ pair.
- **EW quark-antiquark annihilation**: a quark-antiquark pair in the initial state annihilates each other into a Z^0/γ which produces a $Q\bar{Q}$ pair.

The Feynman diagrams of these processes are illustrated in Figure 1.5.

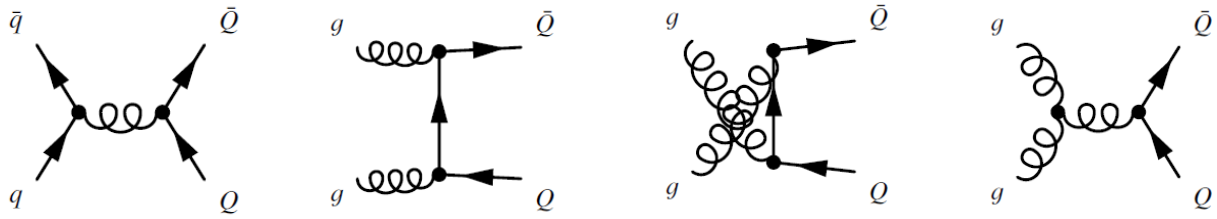


Figure 1.5: LO Feynman diagrams for heavy-quark production in pp collisions, $pp \to Q\bar{Q}$ [28].

For example, Figure 1.6 illustrates the quark-fusion process $Q\bar{Q} \to g \to b\bar{b}$ in the case of the $b\bar{b}$ production in pp collisions, weighted by the two PDFs $f_{q/p}$ and $f_{\bar{q}/p}$.

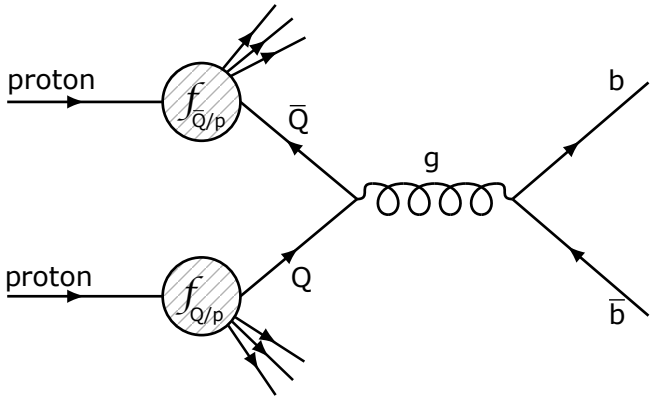


Figure 1.6: Factorization diagram of a $pp \to b\bar{b}$ event.

1.4 Hadronization of b- and c-quarks

The principle of confinement states that partons can never be observed freely in nature [29]. The partonic structure of hadrons has been widely studied but the process by which they are formed from scattered partons is not completely understood. This process is the so-called *hadronization* or *fragmentation* which is formulated in terms of *Fragmentation Functions* (FFs). These functions represent the probability density distribution for the fraction of the q-quark energy taken by the Q-hadron. The q-quark fragmentation function is defined as

$$\mathcal{D}(z) = \frac{1}{\sigma_{\text{tot}}} \frac{\mathrm{d}\sigma}{\mathrm{d}z} \tag{1.2}$$

where σ_{tot} is the total hadronic cross section, while z relates the energy of the q-quark and Q-hadron and is invariant under Lorentz boosts along the quark direction,

$$z = \frac{(E + p_L)_Q}{(E + p)_q} \tag{1.3}$$

where p_L is the momentum of the Q-hadron along the direction of the q-quark. Using the QCD Factorization Theorem [27], the perturbative and non-perturbative contribution are fold to use the FF, so that

$$\mathcal{D}_{\text{observed}}(z) = \int_0^\infty \mathcal{D}_{\text{pert.}}(t) \times \mathcal{D}_{\text{non-pert.}} \left(\frac{t}{z}\right) \frac{\mathrm{d}z}{z} \tag{1.4}$$

A consequence of the Factorization Theorem is that the non-perturbative component does not depend on the initial partonic state. Monte Carlo generators, like PHYTIA8, deal with this perturbative part using the Parton Shower model [30]. On the other hand, the non-perturbative part is treated using phenomenological schemes that are used to model the carry-over of the parton momentum and flavor to the hadrons. Finally, the Peterson model [31] proposes the following parametrization of the non-perturbative FF,

$$\mathcal{D}_{\text{non-pert.}}(z) \propto \frac{1}{z\left(1 - \frac{1}{z} - \frac{\epsilon}{1-z}\right)^2} \tag{1.5}$$

Due to their large mass, the b- and c-quark tends to keep a large fraction of its energy, which is finally transferred to their corresponding hadron. Figure 1.7 shows the FF using the Peterson model for different quarks.

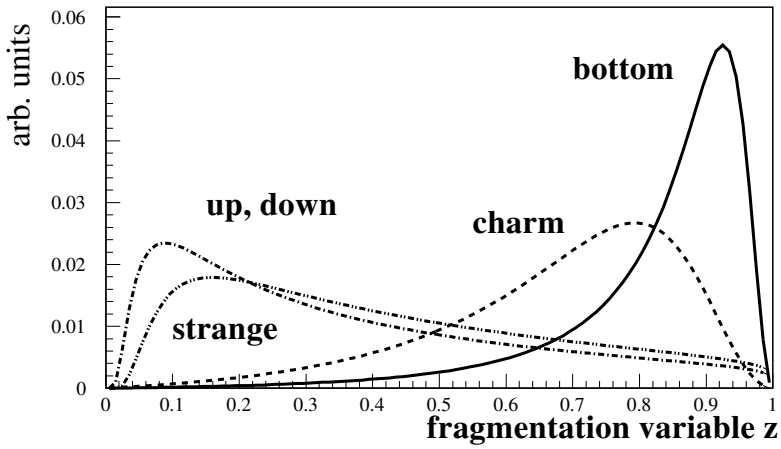


Figure 1.7: Quark Fragmentation functions as function of the z-variable for different quarks using the Peterson model [31].

1.4.1 Partons and Jets

As was mentioned before, QCD hard scattering processes create high transverse momentum quarks or gluons. Immediately after being produced, a quark or gluon fragments and hadronises, leading to a collimated spray of energetic hadrons jointly known as a jet in the partonic picture. The probability of creating a set of jets can be described by the jet production cross section, which is an average of elementary perturbative QCD quark, antiquark, and gluon processes, weighted by the parton distribution functions.

Perturbative QCD calculations have colored partons in the final state, but only the colorless hadrons that are ultimately produced are observed experimentally. Thus, to describe what is observed in a detector, all outgoing colored partons must first undergo parton showering and then combination of the produced partons into hadrons (hadronization).

As the parton, which was generated in a hard scatter, get out of the interaction, the strong coupling constant will increase with its separation. This rises the probability for QCD radiation, which is mainly shallow-angled with respect to the originating parton. Thus, one parton will radiate gluons, which will in turn radiate $q\bar{q}$ pairs, with each new parton nearly collinear with its parent. This can be described by convolving the spinors with fragmentation functions $P_{ji}(x/z, Q^2)$, in a similar manner to the evolution of parton density functions. This is described by the Dokshizer Gribov Lipatov Altarelli Parisi equations (DGLAP) type equation [32]:

$$\frac{\partial}{\partial \ln Q^2} D_i^h(x, Q^2) = \sum_j \int_x^1 \frac{dz}{z} \frac{\alpha_S}{4\pi} P_{ji}\left(\frac{x}{z}, Q^2\right) D_j^h(z, Q^2) \tag{1.6}$$

At each stage, parton showering produces partons of successively lower energy. So, the quarks and gluons in the shower are all roughly collinear. Eventually, the momentum transfer in the splittings falls below 1 GeV, the value of the strong coupling becomes large, and the strong interaction effects of QCD take over, combining quarks and antiquarks into mesons and baryons giving as a result a jet of hadrons.

Phenomenological models must then be applied to describe the length of time when showering occurs, and also the combination of colored partons into bound states of colorless hadrons, which is inherently not-perturbative. Figure 1.8 illustrates the evolution of the particles produced during the fragmentation and the hadronization of quarks.

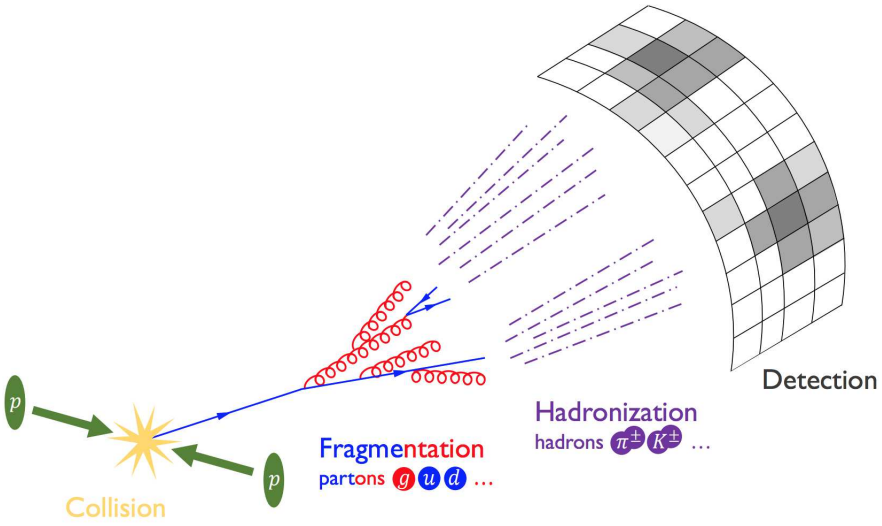


Figure 1.8: Schematic view of a hadron collision event showing the evolution of the produced particles. The parton shower phase is dominated by radiations of quarks and gluons in the initial and the final state. During the hadronization, colored partons are bound into colorless hadrons which are lately detected [33].

Chapter 2

The LHCb experiment and jet identification

2.1 The Large Hadron Collider

The European Organization for Nuclear Research, known as CERN, is a research organization operating the largest particle physics laboratory in the world [34]. Established in 1954, the organization is based in Geneva on the Franco-Swiss border. CERN's main function is to provide the particle accelerators and other infrastructure needed for high-energy physics research, as a result, numerous experiments have been constructed at CERN through international collaborations.

Many activities at CERN spans around its flagship experimental project, the Large Hadron Collider (LHC), and associated experiments. The LHC represents a large-scale, worldwide scientific cooperation project. The LHC tunnel is located 100 meters underground and consists of a 27 km circumference circular tunnel where several experiments (CMS, ATLAS, LHCb, ALICE) are located along the collider. Each of them operates with a different physical perspective and uses different detection technologies, see Figure 2.1.

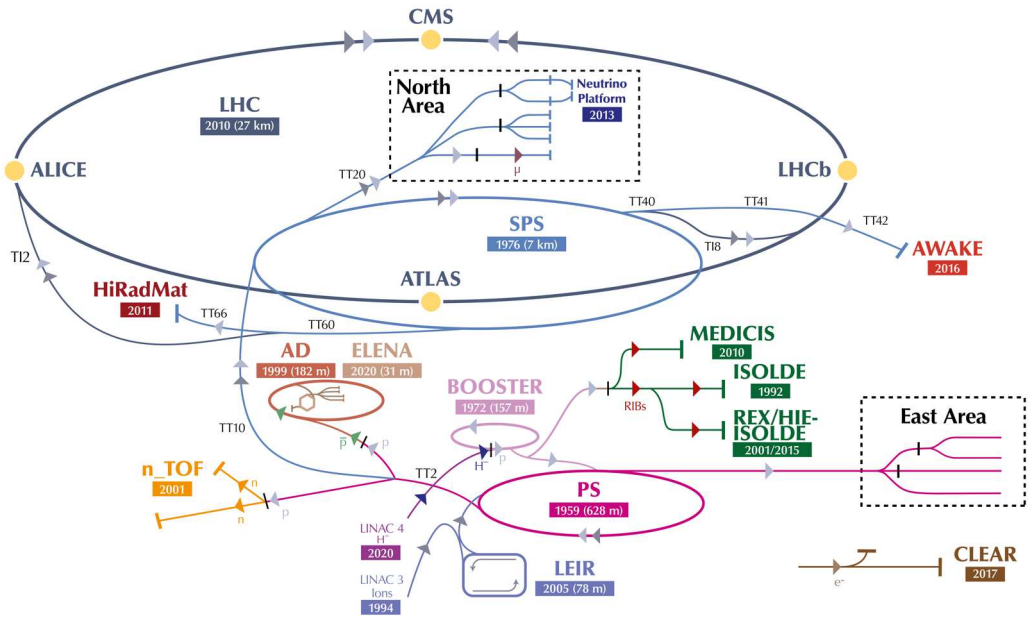


Figure 2.1: Schematic representation of accelerator facilities at CERN [35].

At the LHC, the acceleration of protons starts by stripping the electrons on hydrogen atoms originally contained in a hydrogen bottle. Next, those protons are accelerated successively by the LINAC 4, Booster, PS and by the Super Proton Synchrotron (SPS) to achieve an energy up to 450 GeV. Finally,

the protons are injected into the two LHC rings (one in clockwise and other in counter-clockwise direction) where each beam can reach a final energy of $\approx 7\,\text{TeV}$, consequently acquiring a center of mass energy $\sqrt{s} \approx 14\,\text{TeV}$. Once the beams are stable, they are separated in bunches with around 10^{11} protons each spaced by around 25 ns each other, reaching a collision frequency of 40 MHz.

Four main experiments are located around the LHC collision points: ATLAS (A Toroidal LHC ApparatuS)[7] and CMS (Compact Muon Solenoid)[6] are the two general-purpose detectors, designed with a cylindrical structure around the collision point, that concentrate on measurements of SM processes and BSM physics searches at the TeV scale. ALICE (A Lead Ions Collision Experiment)[36] is dedicated to the study of phases of matter where quark and gluons are free, e.g. quark-gluon plasma. Finally, the LHCb detector [37] which was initially designed to the study of b- and c-quarks properties in the forward region of the collision and by means of different upgrades is turning into a general purpose detector as well.

A key parameter to describe the collider operation is the instantaneous luminosity ($\mathcal{L}$),

$$\mathcal{L} = \frac{N_b^2 n_b f \gamma}{4\pi\epsilon\beta\sqrt{1+\left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2}} \tag{2.1}$$

which depends on the beam properties listed in Table 2.1. The LHC operation was originally designed to achieve a luminosity peak of $\mathcal{L} = 2 \times 10^{34}\,\text{cm}^2\,\text{s}^{-1}$. An increase in the luminosity will imply an increase in the production rate of physics processes, and thus potential new physics discoveries. However, high luminosity also implies an increase in the number of pp interactions per bunch crossing or pile-up (UP), which poses challenges to the experiments for reconstruction and radiation harness.

Variable	**LHC design experimental parameters**	**Design or nominal values**
f	revolution frequency	11.245 kHz
n_b	number of proton bunches per beam	2808
N_b	number of proton per bunch	1.15×10^{11}
β	optical beta function at the IP	55 cm
σ_z	RMS bunch length	7.55 cm
σ^*	transverse RMS beam size	16.7 μm
γ	relativistic gamma factor	7461
ϵ	normalized transverse beam emittance	3.75 μm.rad
θ_c	crossing angle at the IP	285 μrad

Table 2.1: Instantaneous luminosity parameters for the LHC proton-proton collisions [38].

The expected number of produced events of a particular process during the LHC collisions (without taking into account the detector acceptance) is predicted as $N_{\text{exp}} = \sigma \cdot \text{L}$, where σ is the theoretical production cross section, and L is the integrated luminosity defined as the integral of the instantaneous luminosity $\mathcal{L}$ over the LHC operation time.

2.2 The LHCb detector

LHCb [39] is a single-arm forward spectrometer with a pseudo-rapidity[1] η coverage between 1.8 and 4.9. The LHCb detector is composed of several sub-detectors systems, see Figure 2.2. The LHCb coordinates system involves a z-axis parallel to the beam direction, the y-axis parallel and opposite to the gravitational acceleration and the x-axis orthogonal to the yz-plane, forming a right-handed system. The main sub-detectors are:

[1]The pseudo-rapidity is defined as $\eta = -\ln\tan\frac{\theta}{2}$, where θ is the polar angle formed by the particle momentum and the beam axis.

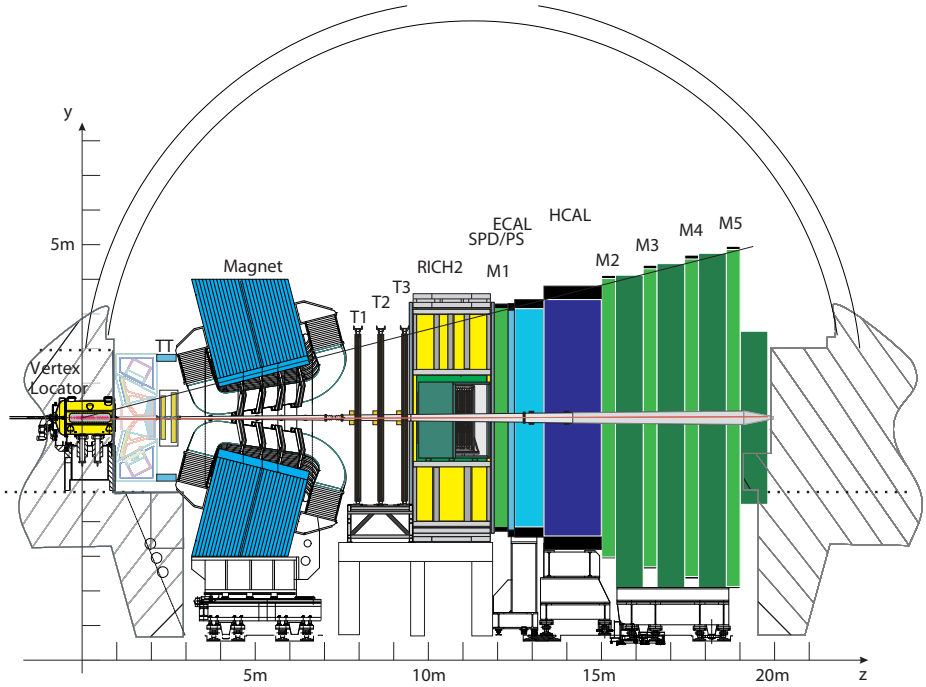


Figure 2.2: Schematic representation of the LHCb detector illustrating each sub-system. The z-axis is parallel to the beam direction while the y-axis is parallel to the gravitational acceleration [39].

- **VErtex LOcator** (VELO) is a system located around the interaction point.
- **Tracking system** is a system of four tracking stations. One station called Trigger Tracker (TT) located upstream of the magnet, while the other three (T1, T2 and T3) are located downstream of the magnet.
- **Ring Imaging Cherenkov** (RICH) detectors where one is located upstream and one downstream of the magnet. These are used for efficient particle identification, in particular for pions versus kaons separation.
- **Calorimeter system** made of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL) and an hadronic calorimeter (HCAL).
- **Muon detection system** which is located in the outer part of the detector.

The optimal LHCb luminosity has a maximum of 2.5 proton-proton interactions per bunch crossing. In order to achieve this value, the luminosity provided by LHC is reduced using a *luminosity levelling* technique [40] where the two beams do not collide head-on but instead the beam axes are tilted in such a way a larger interaction area is obtained. It allows to reach a constant luminosity of around $4.5 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Table 2.2 summarizes the integrated luminosity for the Run 2 campaign with $\sqrt{s} = 13$ TeV.

year	2015	2016	2017	2018	**Total**
$\mathcal{L}_{int}$ (fb^{-1})	0.33	1.67	1.81	2.19	6.00

Table 2.2: Run-II integrated luminosity at LHCb with $\sqrt{s} = 13$ TeV [39].

2.2.1 The tracking system

The LHCb detector includes a high-precision tracking system. It consists of a VELO, surrounding the pp interaction point, which provides high precision track measurements close to the interaction region. It has also four tracking stations, one is a large-area silicon-strip called TT, which is a upstream detector located after the dipole magnet, and three tracking stations (T1, T2, T3) placed downstream of the magnet. Silicon micro-strips are used in the Inner Tracker (IT) while straw-tubes are employed in the Outer Tracker (OT).

Vertex Locator

The VELO [41] is a sub-system designed for the measurement of the distance r (from 8 mm to 42 mm) and the direction ϕ (from 15 mrad to 390 mrad) of the tracks coordinates close to the interaction point. This is useful in the secondary vertex identification associated to b- and c- hadrons decays. The VELO is built as a series of silicon modules, each providing a measure of the r and ϕ coordinates using the R and ϕ sensors, see Figure 2.3.

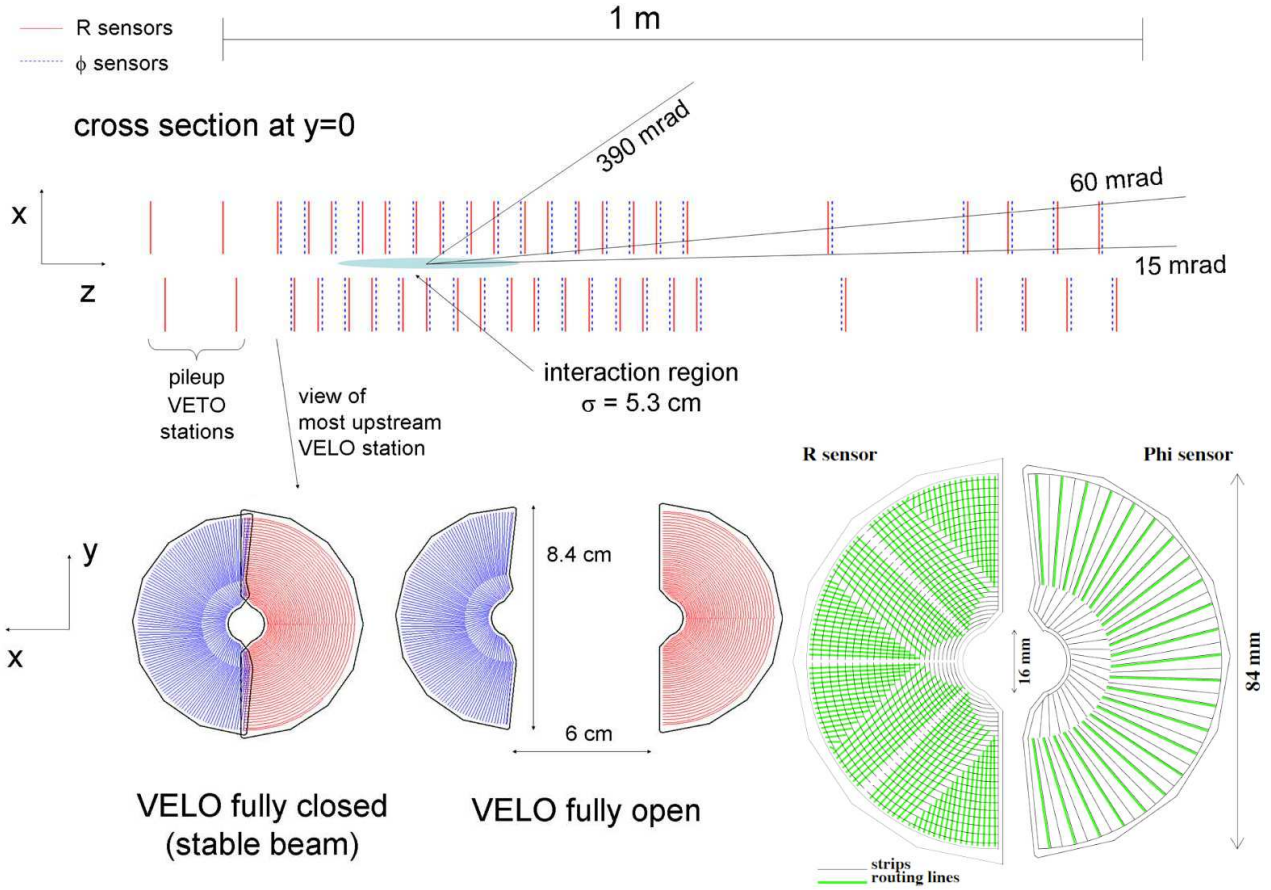
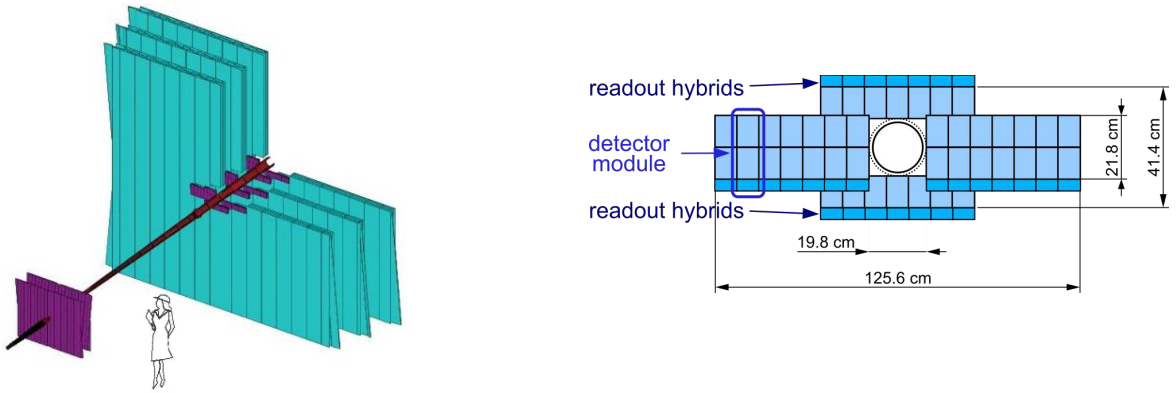


Figure 2.3: Schematic layout of the VELO system [39, 41].

The VELO system is built with 21 stations of these semi-circular sensors, placed parallel to the beam axis along 1 m. Besides, the sensors are mounted in a vessel that maintains the vacuum separated from the machine vacuum. These sensors operates at 8 mm far from the beam axis. This distance is smaller than the beam width during the injection phase, that's why, sensors are retractable and the VELO can stay in the open configuration until the beam conditions are stable and reaching an individual hit sensor resolution around 7 μm.

Tracking stations

The tracking system has four planar stations perpendicular to the beam axis. One is the TT which is 150 cm wide and 130 cm high located upstream of the dipole magnet to cover the full geometrical acceptance of the experiment. T1-3 stations are located downstream the magnet where the inner part is 120 cm wide and 30 cm high, and together with the TT forms the Inner Tracker (IT), see Figure 2.4. Besides, the IT is made of 200 μm wide silicon micro-strip sensors, displaced in four layers each with small angle stereo-view: x-layers with strips aligned with the vertical (y) axis, u-layers with strips tilted $-5°$ and m v-layers with strips tilted $+5°$.



(a) Tracking system, IT (violet) and OT (cyan).

(b) Layout of x-layer of a IT station.

Figure 2.4: Layout of a Inner Tracker (IT) and layout of the x-layer [39].

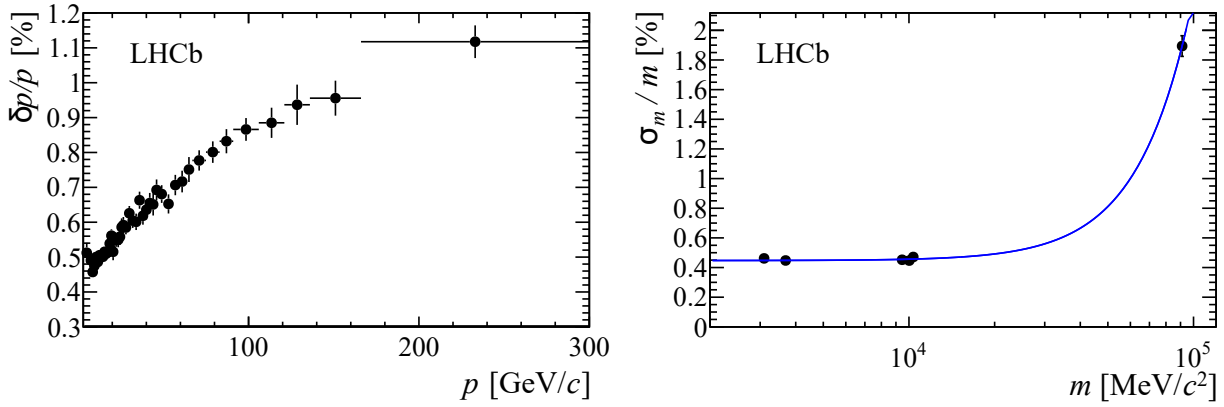
The external regions of the T1-3 stations are called Outer Tracker (OT), see Figure 2.4.a. It is a drift tube detector with a geometrical acceptance of 300 mrad and 250 mrad on the the horizontal and vertical plane, respectively. It consists of an array of straw-tube modules, each one with two layers of 4.9 mm diameter drift-tubes encapsulating a mixture of Ar (70%), CO_2 (28.5%) and O_2 (1.5%). The final drift distance resolution is around 200 μm.

2.2.2 Tracking performance

The VELO together with the tracking stations (TT and T1-3) measure the trajectory and momentum of charged particles by measuring the curvature induced by the magnet. Commonly, pattern recognition algorithms are used to reconstruct the track trajectory through all the tracking sub-detectors. The tracks are reconstructed via a Kalman fit [42] and are classified in four categories:

- **Long tracks** with hits in the VELO and in all the T stations.
- **Downstream tracks** with hits only in the VELO and the TT station.
- **Upstream tracks** with hits in all the T stations but not in the VELO.
- **VELO tracks** with hits only in the VELO.

The track momentum resolution ($\delta p/p$) has been measured using data samples of the $J/\psi \to \mu^+\mu^-$ decay. Figure 2.5.a shows its behavior as a function of the momentum of the two muon final state. On the other hand, the invariant mass resolution (σ_m/m) was determined using six resonances in the dimuon data sample, J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ mesons and the Z^0-boson. Figure 2.5.b shows its behavior as a function of the invariant mass of the dimuon resonance.



(a) The relative momentum resolution as a function of the momentum for long tracks, obtained from J/ψ decays.

(b) The relative mass resolution as a function of the mass of the dimuon resonance.

Figure 2.5: Relative momentum and relative mass resolution at LHCb [43].

The Primary Vertex (PV), where particles are generated in the pp interaction point, is reconstructed from the detected tracks and it involves the minimization of

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \chi^2_{\mathrm{IP},i} \cdot W_{T,i} \tag{2.2}$$

where $\chi^2_{\mathrm{IP},i}$ is the χ^2 of the track impact parameter with respect to the PV, and $W_{T,i}$ denotes to the corresponding weight. The PV reconstruction efficiency and resolution depend on the tracks multiplicity [44]. The average efficiency runs from 90.0% to 97.5%, with a probability of reconstructing a false PV around 1%. The resolutions on the x, y and z coordinates are 8 μm, 10 μm and 50 μm, respectively.

2.2.3 Magnet

A dipole magnet is located between the TT and the T1-3 stations. It allows the measurement of the momentum of charged particles, by measuring the curvature induced by a magnetic field parallel to the y-axis. Therefore, the trajectories are bent on the xz-plane. The integrated magnetic field for 10 m long tracks is around 4 Tm, while the residual magnetic field inside the two RICHs is ~ 2 Tm. The magnet covers a forward acceptance of $\pm$250 mrad vertically and $\pm$300 mrad horizontally.

2.2.4 Ring Imaging Cherenkov detectors

The excellent particle identification (PID) performance is due to the two Ring Imaging Cherenkov (RICH) detectors, RICH1 and RICH2, which discriminate charged hadrons (π, K, p). Besides, the RICH system can identify charged leptons (e, μ) to complement the information coming from the downstream calorimeters and muon system.

Particles crossing the RICH radiator travel at velocities larger than the speed of light, in the same medium, emit Cherenkov radiation in a cone with axis parallel to the particle velocity. The cone angle θ_c is related to the particle velocity v by

$$\cos\theta_c = \frac{c/n}{v} = \frac{c_m}{v} = \frac{\sqrt{m^2c^2 + p^2}}{np} \tag{2.3}$$

where n is the refractive index of the RICH active medium, m is the particle mass and p its momentum, c and c_m are the speed of light in the vacuum and in the active medium, respectively. By measuring the angle of the Cherenkov cone θ_c of a charged particle and estimating the p from the tracking system, an estimation of m can be achieved, allowing its mass identification.

The active media of RICH1, located upstream of the magnet, is silica aerogel and C_4F_{10} used to distinguish particles with a momentum range between 1 and 60 GeV/c. It covers an angular acceptance range from $\pm$25 mrad up to $\pm$300 mrad horizontally and up to $\pm$250 mrad vertically. On the other hand, RICH2, located downstream of the magnet, between the T3 station and the muon station M1, distinguishes particles with momentum between 15 and 100 GeV/c. It covers an angular acceptance from $\pm$15 mrad up to $\pm$120 mrad horizontally and up to $\pm$100 mrad vertically. Aditionally, it uses CF_4 as active medium.

2.2.5 Calorimeters

The calorimeter is a very important system which selects hadrons, electrons and photons with significant transverse momentum. It also provides an excellent PID for electrons, photons and hadrons. Besides, it measures the energy of neutral particles such as photons, π^0 and other neutral hadrons.

The LHCb calorimeter is made of a Scintillating Pad Detector (SPD), a Pre-Shower detector (PS), an Electromagnetic Calorimeter (ECAL) and an Hadronic Calorimeter (HCAL). Figure 2.6 shows the structure and granularity of HCAL and ECAL.

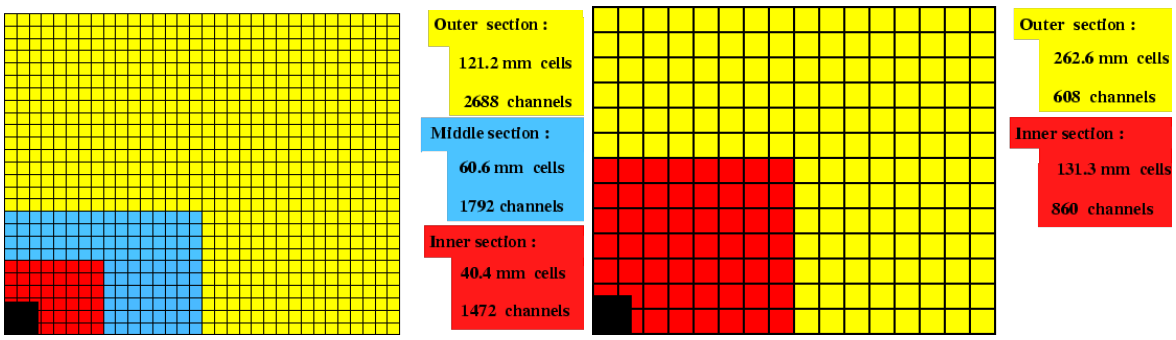


Figure 2.6: (Left) Structure and granularity of the SPD/PS and ECAL systems and (right) HCAL [39].

The SPD/PS system is used to identify electrons and photons since, in the SPD, photons do not produce any signal before triggering a shower, while the electrons do. On the other hand, photons passing the lead layer of the SPD produce showers on the PS. Hadrons do not produce showers on the PS, therefore they can be separated from electrons. Additionally, it covers and angular acceptance range from ±30 mrad to ±300 mrad horizontally and ±250 mrad vertically.

In the ECAL, electrons and photons deposit their energy. ECAL is composed by 66 layers of alternating 4 mm lead tiles and scintillators coupled to PMT via WLS fibers, resulting in a total length of 42 cm (25 radiation lengths). The granularity of the ECAL allows of a good energy resolution of

$$\left(\frac{\sigma_E}{E}\right)_{\text{ECAL}} = \frac{10\%}{\sqrt{E}} \oplus 1\%$$

Where E is the energy of the particle (in GeV). It has the same geometrical acceptance as the SPD/PS system. Finally, the HCAL collects the residual energy of charged and neutral hadrons. Its structure is very similar to ECAL made of 5.6 radiation lengths of 1 cm iron tiles and scintillators. Its energy resolution is

$$\left(\frac{\sigma_E}{E}\right)_{\text{HCAL}} = \frac{69\%}{\sqrt{E}} \oplus 9\%$$

where E is the energy of the hadron (in GeV).

2.2.6 Muon System

The muon system allows fast muon detection. It is the outermost detector from the interaction point composed of five stations M1-5 of rectangular shape, see Figure 2.7. The detector has 1380 chambers that cover an angular acceptance range from ±20 mrad to ±306 mrad horizontally and from ±16 mrad to ±258 mrad vertically. The M1 station is located upstream of the RICH2 detector to improve the transverse momentum measurement in the trigger. The M2-5 stations are located upstream of the calorimeters, they have 80 cm (20 interaction lengths) interleaved thick iron absorbers used to select muons whose minimum momentum to pass over all the five stations is approximately 6 GeV/c. The muon system is built with Multi-Wire Proportional Chambers (MWPCs), except for the most inner part where triple-GEM detectors are used. MWPCs are filled with a gas mixture of $Ar/CO_2/CF_4$, in a 40:55:5 proportion achieving a time resolution of around 5 ns. The GEM detectors are filled with a "faster" gas mixture of $Ar/CO_2/CF_4$ in a 45:15:40 proportion allowing a time resolution of 3 ns.

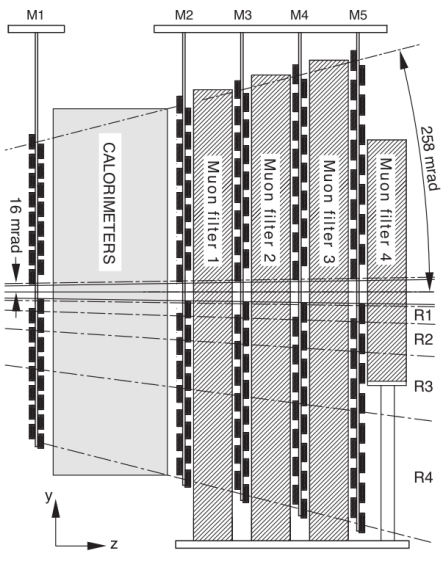


Figure 2.7: LHCb muon system in the yz plane [39].

2.3 Trigger

LHCb operates with a bunch crossing rate around 40 MHz which represents lower average luminosity than the maximum designed luminosity of LHC. At the beginning of the Run 2 data taking (2015), the storage system was able to write and save events at a rate of 12.5 kHz. The purpose of the trigger system is to reduce the rate from MHz to kHz selecting just a small fraction of interesting events. The LHCb trigger system operates in two stages, Level 0 (L0) which is an hardware trigger stage operating synchronously with the bunch crossing rate and High Level Trigger (HLT) which is a software trigger operating on a processor farm. Figure 2.8 shows the data flow on the LHCb trigger.

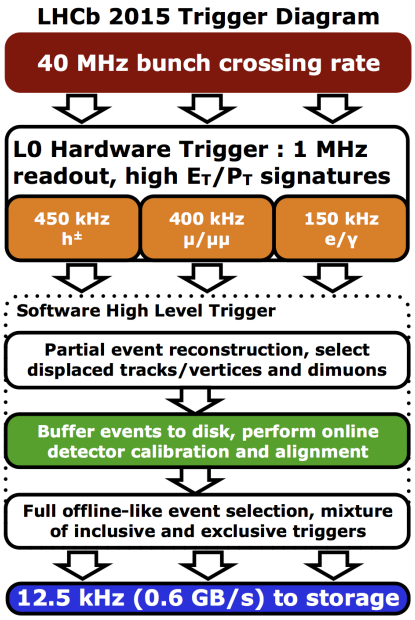


Figure 2.8: Schematic representation of the data flow in the LHCb trigger system (Run 2 configuration)

2.3.1 Level 0 trigger

The L0 trigger is designed to detect particles with high transverse momentum or high transverse energy. The L0 Decision Unit (L0-DU) is connected to the calorimeters and the muon system in order to choose a likely event, reducing the event rate from 40 MHz to 1 MHz. Then, multiple interactions events are rejected applying a cut on the number of hits inside the SPD, called Global Event Cut (GEC), where an event is accepted if one of the following conditions is satisfied:

- **L0-Muon**: a particle is identified as a muon if it has a $p_T > 1.76$ GeV/c.
- **L0-Dimuon**: two particles which p_T product $p_T^1 \cdot p_T^2 > 1.6^2$ GeV^2/c^2 are identified as 2 muons.
- **L0-Hadron**: is a particle which releases a cluster in the HCAL and has $E_T > 3.7$ GeV.
- **L0-Photon**: is a particle which emits a cluster in the ECAL and has $E_T > 3$ GeV.
- **L0-Electron**: is a particle with an $E_T > 3$ GeV producing a cluster in the ECAL and hits in the PS and the SPD as well.

2.3.2 High Level Trigger

After an event is accepted by the L0 trigger, it is processed by the HLT which runs on a large computing facility called Event Filter Farm. HLT reduces the event rate from 1 MHz to 12.5 kHz in two stages:

- **HLT1** which uses the information from the VELO and the T1-3 stations to partially reconstruct the events selecting tracks with high p_T and high displacement from the interaction point.
- **HLT2** which reconstructs and selects an event using an inclusive or exclusive algorithm to select all particles or a specific decay mode, respectively.

2.4 Jet reconstruction

The LHCb detector has been initially designed to study b- and c- hadrons physics. However, jets can be studied efficiently by means of the unique forward geometry and excellent tracking performance of the detector, covering a complementary phase space with respect to CMS and ATLAS.

As was described in the previous chapter, jets are produced as the result of the hadronization and fragmentation of partons (quarks and gluons) produced in the hard-scattering processes at the interaction point. Jets is manifested in the detector as collimated sprays of charged particles in the tracking system and as concentrated clusters of energy in the calorimeters. Despite the LHCb detector has excellent tracking performance, its calorimeters are not equally well performing in the forward region. Therefore, contrary to CMS and ATLAS which exploit calorimeters for jet reconstruction, the LHCb experiment employs both calorimeter clusters and tracking information to efficiently reconstruct jets.

At the LHCb, the jets reconstruction algorithm workflow is

1. **Particle Flow** which provides a list of particles at the input of the clustering algorithm.
2. **Anti-k_t algorithm** which clusters particles providing a list of output jets.
3. **E-recombination scheme** where four-momentum of the jet as a function of the particles momenta is calculated.
4. **Jet Energy Correction** which applies a final correction to the jet energy based on Monte Carlo simulations.

2.4.1 The Particle Flow algorithm

This algorithm chooses particles based on different criteria. Charged particles are selected depending on the category they belong (long, downstream, upstream or VELO tracks), and if they fulfil the requirements described in Table 2.3. Moreover, a mass/type information can be associated to particles ($p/\bar{p}$, $\pi^{\pm}$, $\mu^{\pm}$, $e^{\pm}$ and $K^{\pm}$) using the information coming from the RICH detectors and calorimeters.

Variable	Description	long	downstream	upstream	VELO tracks
p_T	Transverse momentum [MeV/c]	-	-	> 100	-
χ^2	Kalman fit χ^2	< 5	< 1.5	< 1.5	< 10
P_{ghost}	Wrong reconstruction probability	< 0.5	-	-	-
$\frac{\sigma(q/p)}{q/p}$	Momentum resolution	> 10	> 10	> 2	-

Table 2.3: Particle Flow algorithm requirements to select charged particles.

Then, isolated neutral particles show clusters in calorimeters but can not be associated to tracks. In the ECAL, these are mainly photons or π^0s that decay in two photons which discrimination depends on the shape of the calorimeter clusters. Therefore, a likelihood for the photon or π^0 hypothesis is computed. After, π^0s are divided in two categories: *merged* π^0 where the two photons are almost collinear and show as a single cluster, and *resolved* π^0 where the two photons are detected as two separate clusters. Table 2.4 shows the applied requirements for each particle category. In the HCAL, there are no specific particle identification requirement, just different $\chi^2_{\text{track-cluster}}$ requirements are applied for different energy thresholds, $\chi^2_{\text{track-cluster}} > 25$ for $E < 10$ GeV and $\chi^2_{\text{track-cluster}} > 15$ for $E > 10$ GeV.

Variable	Description	merged π^0	resolved π^0	photons
E_T	transverse energy [MeV]	-	-	> 200
PhotonID	photon hypothesis likelihood	-	> -2	-
$\chi^2_{\text{track-cluster}}$	track - cluster probability	> 25	> 25	> 25

Table 2.4: Cluster requirements for neutral particle selection.

Finally, non-isolated neutral particles show clusters and are associated to tracks. Therefore, the selection is as follows:

1. ECAL and HCAL $\chi^2_{\text{track-cluster}}$ requirenments:
   - ECAL: $\chi^2_{\text{track-cluster}} < 25$
   - HCAL: $E < 10$ GeV a $\chi^2_{\text{track-cluster}} > 25$; $E > 10$ GeV a $\chi^2_{\text{track-cluster}} > 15$.
2. Tracks are not shared between ECAL and HCAL clusters.
3. The expected energy E_{exp} released in the calorimeters is estimated using an empirical E/p parametrization, where E is the cluster energy released by charged particles and p the track momentum. If E_{exp} is 1.8 times larger than the measured energy E_{meas} the cluster group is discarded.
4. If $E_{\text{meas}} > 1.8\ E_{\text{exp}}$, then $E_{\text{sub}} = E_{\text{meas}} - E_{\text{exp}}$.
5. Finally, if $E_T > 2$ GeV then E_{sub} is selected as a non-isolated neutral particle (used as input of the anti-k_T algorithm).

2.4.2 The anti-k_t algorithm

It follows the next steps:

1. For each pair of particles the distance d_{ij} is calculated

$$d_{ij} = \min(k_{t,i}^{-2}, k_{t,j}^{-2})\frac{\Delta_{ij}^2}{R^2} \qquad \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{2.4}$$

   where $k_{t,i}$, y_i and ϕ_i are the transverse momentum, the rapidity and the azimuthal angle of the particle i, respectively. R is a tunable parameter of the algorithm called *radius*.
2. Then, a similar distance is calculated between each particle and the beam axis

$$d_{iB} = k_{t,i}^{-2} \tag{2.5}$$

3. After, the smallest distance between d_{ij} or d_{iB} is selected. If the smallest distance is d_{ij}, the particles i and j are combined into a single particle (summing their four-momenta) and removed from the list of input particles. If the smallest distance is d_{iB}, the particle i is defined as a jet and removed from the list of particles.
4. The algorithm restarts from step 2 until any particles is remaining to be clustered.

The anti-k_t algorithm normally clusters soft particles with hard particles instead of composed jets only with soft components. Then, the parameter R can be tuned in order to optimize the jet energy resolution depending on the experimental environment, the detector properties or the process under study. At LHCb the optimal ranges from $0.5 \leq R \leq 0.7$.

2.4.3 E-recombination scheme

The four-momenta $(E_{\text{jet}}, \boldsymbol{p}_{\text{jet}})$ of the reconstructed jets after the clustering process can be estimated as

$$E_{\text{jet}} = \sum_i E_i \qquad \boldsymbol{p}_{\text{jet}} = \sum_i \boldsymbol{p}_i \tag{2.6}$$

Monte Carlo jet simulation samples are used to validate the reconstructed jet energies. The main difference between reconstructed jets (jet_{reco}) and MC jets (jet_{MC}) is that the last ones are clustered by the anti-k_T algorithm using all the stable truth-level particles (true kinematical values) while for jet_{reco} just reconstructed particles are used where invisible particles (e.g. neutrinos) are removed from the list to unbias the reconstructed energy. To associate a jet_{MC} with a jet_{reco}, the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between them in the η-ϕ plane should be smaller than 0.4. If more than one jet fulfil this requirement, the jet with the closest distance is chosen. Also, jets reconstructed with charged particles are in agreement with MC simulation as expected. On the other hand, neutral particles show worse performance due to the lack of tracking information (only calorimeter data is used).

2.4.4 Jet Energy Correction

Later, a correction is applied to the energy of the reconstructed jets $E^{\text{jet}}_{\text{reco}}$ to match with the energy of one of the associated MC jets $E^{\text{jet}}_{\text{MC}}$. This is done using a k^{MC} factor, called *Jet Energy Correction*

$$E^{\text{jet}}_{\text{MC}} = k^{\text{MC}}\, E^{\text{jet}}_{\text{reco}} \tag{2.7}$$

Inside the k^{MC} factor the integrated effect of pile-up, noise and non-uniformity of the detector are considered. k^{MC} is calculated using MC simulations of *b*-, *c*-, *light*- quarks and *gluons* jets. It is non-uniform with respect to the η and ϕ jet axis as well as with the fraction of charged particles in the jet (cpf). Moreover, it depends on the jet flavor. In addition, other correction is applied to suppress differences between the jet energy real data and MC simulations using the *Jet Energy Scale* factor.

2.4.5 Jet identification efficiency

Finally, some requirements are applied to reject jets originated from noise or high energetic leptons:

- Tracks pointing to the PV ($n^{\text{PV}}_{\text{tracks}}$) ≥ 2,
- p_T fraction carried by a single Particle Flow particle (mpf) < 0.8,
- p_T carried by a track (mpt) > 1.2 GeV/c,
- fraction of charged particles inside the jet (cpf) > 0.1.

The jet identification efficiency ϵ_{jet} is evaluated

$$\epsilon_{jet} = \frac{N[\text{reconstructed jets}]}{N[\text{true jets}]} \tag{2.8}$$

It has been tested using $Z^0 \to \mu\mu$ and MC events at $\sqrt{s} = 7$ TeV. The jet identification efficiency increases with respect to the jet p_T, see Figure 2.9.

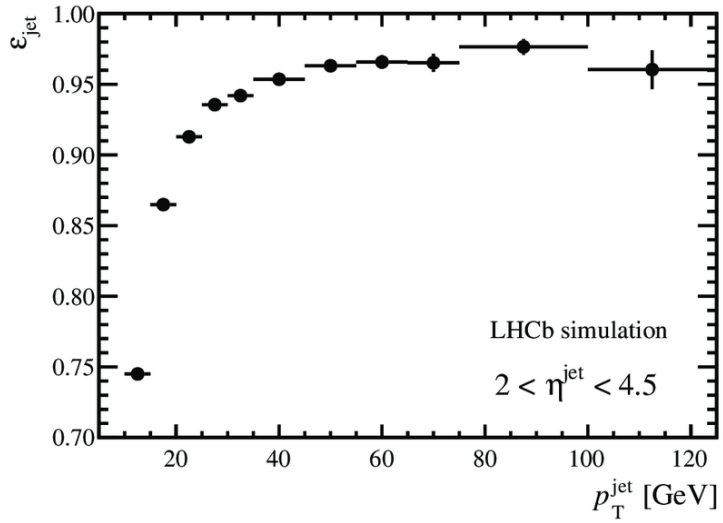


Figure 2.9: Jet identification efficiency as function of the jet p_T

2.5 Heavy-Flavor jet identification

The identification of jets originated from the heavy quarks hadronization is crucial to study Standard Model (SM) and to search New Physics (NP) processes. The measurement of the Higgs decay to heavy quarks relies in the ability to efficiently classify *b*-, *c*- and *light* jets. This task is done using the Jet-flavor tagging algorithms which determine the flavor of the heavy-quark that originated the tagged jet, identifying *b*-, *c*- and *light*-quark jets with low misidentification rates.

This technique is called *Secondary Vertex tagging* which exploits the fact that heavy (b- and c-) hadrons travel a sizable distance before decaying, producing a Secondary Vertex (SV) detached from the Primary Vertex (PV). This algorithm identifies heavy (b- and c-) events from *light*-partons events and also can separate b- from c-events.

2.5.1 Secondary Vertex tagging

Heavy (b- and c-) hadrons produced from proton-proton collisions at $\sqrt{s} = 13$ TeV at LHCb travel an average distance $\langle d \rangle$

$$\langle d \rangle = \beta\gamma c\tau \approx 7 \text{ mm} \qquad \beta = \frac{v}{c} \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{2.9}$$

where c is the speed of light, v is the speed of the hadron and τ is the average lifetime of the hadron. Then, b- and c- events contain a SV with decay products of the heavy quark. This tagging algorithm [45] works as follows

1. Tracks selection:

   - long tracks,
   - $p > 5$ GeV/c and $p_T > 0.5$ GeV/c,
   - Track fit χ^2/ndof < 3,
   - Variation of the χ^2 of the PV fit when the track is removed (χ^2_{IP}) > 16,
   - $P_{\text{ghost}} < 0.3$.

2. Tracks are used to build 2-body SVs in the 3D space. Fits are used to determine the SV position. Two tracks linked to a SV are mixed to form a particle whose flight direction vector points from the PV to the SV and its four-momentum is the sum of four-momenta tracks assuming the π mass. These 2-body particles requires:

   - Distance of closest approach (DOCA) between the two tracks < 0.2 mm,
   - SV χ^2 fit < 10,
   - invariant mass $400 < m < 5279.4$ MeV/c^2 (B^0 mass).

3. If $\Delta R = \sqrt{(\eta_{\text{jet}} - \eta_{\text{SV}})^2 + (\phi_{\text{jet}} - \phi_{\text{SV}})^2} < 0.5$, a 2-body particles is assigned to a jet. 2-body particles inside the same jet sharing at least one track are associated together, creating n-body particles that do not share tracks. The resulting n-body particles are called tag_{SV} whose position is the average of the 2-body SV positions, weighted with the χ^2 of the SV position fit. The tag_{SV} flight direction points from the PV to its position and the tag_{SV} four-momentum is the sum of the four-momenta tracks, assuming the mass.

4. Light quark jets are reduced requiring tag_{SV}s with

   - $p_T > 2$ GeV/c,
   - jet z-position < 200 mm,
   - flight distance/$p < 1.5$ mm/(GeV/c),
   - the flight distance χ^2 obtained from the PV fit must be $> 5\ \sigma$,
   - a tag_{SV} formed by two tracks is rejected if the invariant mass is compatible with K_S,
   - tag_{SV} must contain at most one track with $\Delta R > 0.5$,
   - to remove kaons or hyperons, the corrected mass $M_{\text{corr}} = \sqrt{M_{SV}^2 + p_T^{\text{miss}}} + p_T^{\text{miss}} > 600$ MeV/c^2 is required, where M_{SV} is the SV mass and p_T^{miss} represents the missing component of the momentum with respect to the particle flight direction.

5. If more than one tag_{SV} per jet pass the conditions, the one with the higher p_T is selected. A jet coming from an heavy quark is tagged if at least one tag_{SV} is chosen by this algorithm.

6. Finally, two Boosted Decision Trees (BDT) trained on simulated samples of heavy and light-parton jets are used to identify b- and c- jets:

- One `BDT(`$bc|udsg$`)` is used to separate heavy (b-, c-) jets from *light*-parton jets.
- Other, `BDT(`$b|c$`)` is trained to separate b- jets from c- jets.

2.5.2 SV-tagging performance

The performance of the SV flavor tagging algorithm is measured calculating the b-(c-) tagging efficiency $\epsilon_{b(c)}$ and the light jets misidentification ϵ_q. The b-(c-) tagging efficiency $\epsilon_{b(c)}$ is defined as the ratio between the number of tagged b-(c-) jets $N^{\text{tag}}_{b(c)}$ over the total number of reconstructed jets $N^{\text{tot}}_{b(c)}$

$$\epsilon_{b(c)} = \frac{N^{\text{tag}}_{b(c)}}{N^{\text{tot}}_{b(c)}} \qquad \sigma_{\epsilon_{b(c)}} = \sqrt{\frac{\epsilon_{b(c)}(1+\epsilon_{b(c)})}{N^{\text{tot}}_{b(c)}}} \tag{2.10}$$

while the light jets misidentification is defined as the ratio between the number of tagged *light*- jets N^{tag}_q over the total number of *light*- jets N^{tot}_q

$$\epsilon_q = \frac{N^{\text{tag}}_q}{N^{\text{tot}}_q} \tag{2.11}$$

The results from the tagging performance evaluated on MC samples of $pp \to q\bar{q}$ di-jets events of b-, c- and *light*- quarks at $\sqrt{s} = 13$ TeV generated with `PYTHIA8` [46] are

- $\epsilon_{b(c)}$ increases as p_T increases,
- ϵ_b efficiency is larger than ϵ_c,
- both $\epsilon_{b(c)}$ behave similarly with respect to p_T, while ϵ_q is almost constant.

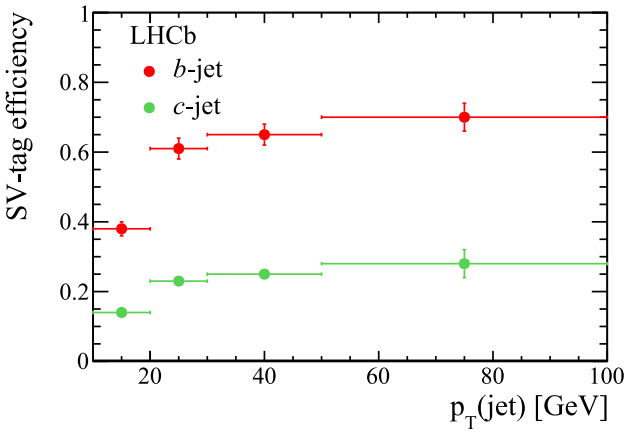


Figure 2.10: SV-tagging efficiency for b-jets and c-jets as a function of p_T, measured on simulated data for $2.2 < \eta < 4.2$ [47].

2.6 Development of new tagging algorithms

Recent results from the LHCb collaboration showed that there still room from improvement for the heavy jet classification [48]. Basically, novel machine learning methods should exploit the complex structure of jets and all the available experimental variables, looking to unveil correlations between particles that could enhance the identification process. Some modern Machine Learning techniques have been already implemented by the CMS[49, 50] and ATLAS [51] Collaboration. On the other hand, LHCb still lacks of high-performance algorithms for the classification of b- and c- events that potentially could give a boost to the $H \to b\bar{b}$ and the $H \to c\bar{c}$ search. Therefore, the main purpose of this thesis is to explore the new-born field of Quantum Machine Learning to work out in a quantum tagger for the b- vs c- classification, comparing the results with 'classical' Machine Learning models. The performance of the new tagger will be compared in terms of the following metrics:

- **Tagging efficiency** ϵ_{eff} calculated as the ratio between the number of tagged events N^{tag} over the total number of events N^{tot}

$$\epsilon_{\text{eff}} = \frac{N^{\text{tag}}}{N^{\text{tot}}} \tag{2.12}$$

- **Mistag efficiency** ω calculated as the ratio between the number of not interested tagged events $N^{\text{tag}}_{\text{wrong}}$ over the total number of not interested events N^{tag}

$$\omega = \frac{N^{\text{tag}}_{\text{not\_int}}}{N_{\text{not\_int}}} \tag{2.13}$$

2.7 Data-set selection

Data-set samples produced from official LHCb Monte Carlo simulations of di-jet produced from pp collisions at $\sqrt{s} = 13$ TeV (Run 2 conditions) were used either for the training of the machine learning models or for the Higgs Search. $b\bar{b}$, $c\bar{c}$, H$\rightarrow b\bar{b}$ and H$\rightarrow c\bar{c}$ dijet productions are done using `PYTHIA8` generator that also estimate the jet fragmentation. `PYTHIA8` includes production processes at LO QCD and LO Z^0/γ decaying into a $b\bar{b}$ pair and $c\bar{c}$, see the Feynman diagrams in Figure 1.5, as well as all the production mechanism for the Higgs production, see Figure 1.3. Events are generated with two opposite polarization of the magnet. `EvtGen` [52] simulates the b- and c- hadron decays and `Geant4` [53] simulates the jet particle interactions with the detector. Finally, the simulated signals are digitalized and the jet reconstruction is done applying the ParticleFlow algorithm and the anti-k_T clustering algorithm as was described previously. In this work the background produced by the light flavour jets is considered as negligible, since light jets with (fake) SV are very low (<0.5%). Although, another discriminator could be also implemented to remove this small contribution in future analyses.

The data-set is generated in different `HDF` files stored in the LHCb Padova cloud and split according to the momentum exchanged $\hat{p}$ in the hard collision, the polarization of the magnet (up or down) and the production mechanism (QCD or Z^0/γ). Table 2.5 lists all the files together with their total number of events N_{gen}.

Filename	**Processes**	**Magnet**	$\hat{\mathbf{p}}$ [GeV/c]	$\mathbf{N}_{\text{events}}$
`Dijet_bb_pt10_15_dw.hdf`	QCD	down	[10,15]	409 662
`Dijet_bb_pt10_15_up.hdf`	QCD	up	[10,15]	390 508
`Dijet_bb_pt15_20_dw.hdf`	QCD	down	[15,20]	791 488
`Dijet_bb_pt15_20_up.hdf`	QCD	up	[15,20]	799 134
`Dijet_bb_pt20_50_dw.hdf`	QCD	down	[20,50]	880 412
`Dijet_bb_pt20_50_up.hdf`	QCD	up	[20,50]	919 768
`Dijet_bb_pt50_dw.hdf`	QCD	down	> 50	961 646
`Dijet_bb_pt50_up.hdf`	QCD	up	> 50	999 540
`Dijet_cc_pt10_15_dw.hdf`	QCD	down	[10,15]	438 622
`Dijet_cc_pt10_15_up.hdf`	QCD	up	[10,15]	496 842
`Dijet_cc_pt15_20_dw.hdf`	QCD	down	[15,20]	772 780
`Dijet_cc_pt15_20_up.hdf`	QCD	up	[15,20]	812 122
`Dijet_cc_pt20_50_dw.hdf`	QCD	down	[20,50]	907 660
`Dijet_cc_pt20_50_up.hdf`	QCD	up	[20,50]	839 298
`Dijet_cc_pt50_dw.hdf`	QCD	down	> 50	823 954
`Dijet_cc_pt50_up.hdf`	QCD	up	> 50	893 384
`Dijet_Higgsbb_tot.hdf`	all	both	all	1 846 208
`Dijet_Higgscc_tot.hdf`	all	both	all	1 689 526

Table 2.5: Data-sets used for the Higgs search.

The total number of events increases as the exchanged momentum $\hat{p}$ increases for all the QCD files, either for $b\bar{b}$ and $c\bar{c}$. For the H$\rightarrow$ $b\bar{b}$ and H$\rightarrow$ $c\bar{c}$ files, the total number of events is one order of magnitude larger than its background counterparts. Besides, the background contribution coming from the light jets are assumed to be negligible.

Chapter 3

Quantum Machine Learning

The following chapter will cover an overview of theoretical formulations behind Quantum Machine Learning and its building block concepts: qubit and entanglement.

3.1 Introduction to Quantum Computing

Quantum computing born in 1980 when Paul Benioff proposed the quantum version of the Turing machine [54]. Later, Richard Feynman suggested a quantum computer will be capable to simulate complex processes a normal computer could not do [55]. The research in this field caused more interest after Peter Shor demonstrated that a quantum algorithm can solve the factorization problem in polynomial time [56] in 1994. Two years later, Lov Grover scripted a quantum algorithm which performs an unstructured search in a database of size n using only $\mathcal{O}(\sqrt{n})$ evaluations [57], much faster than a classical algorithm $\mathcal{O}(n)$. In 1998 Isaac Chuang, Neil Gershenfeld and Mark Kubinec built the first two-qubit quantum computer able to perform computations [58]. In recent years, the interested and the investment in quantum computing research has increased considerably due to its great potential.

3.2 The Qubit

In the Classical Information theory, the basic information unit is the *bit* which represents a logic state taking only two possible values: 0 (False, off, etc.) or 1 (True, on, etc.). Bits are concatenated into strings to represent a larger number where a n-bits string can represent a maximum of 2^n different states, e.g., count integers from 0 to 1000 requires at least 10 bits, $2^{10} = 1024$.

On the other hand, in Quantum Computing the elementary information unit is the *qubit* which is a two-levels quantum system that can be measured as two orthonormal states $|0\rangle$ and $|1\rangle$. These states are the *computational basis* of the 2-dimensional Hilbert space $\mathcal{H}$ where all the possible states of the qubit are defined. Then, the state $|\psi\rangle$ of a qubit can be written as

$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle \tag{3.1}$$

where $\alpha_0, \alpha_1 \in \mathbb{C}$ and satisfy the normalization condition $|\alpha_0|^2 + |\alpha_1|^2 = 1$. $\mathcal{H}$ is isomorphic to the $\mathbb{C}^2$ vector space, so

$$|0\rangle \in \mathcal{H} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2$$

$$|1\rangle \in \mathcal{H} \rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}^2$$

This isomorphism allows the representation of the quantum states as complex (normalized) vectors and linear operators as complex matrices.

3.2.1 Bloch sphere

The 1-qubit state can be represented graphically using the Bloch sphere by rewriting the equation (3.1) as:

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle \tag{3.2}$$

where θ, $\phi \in \mathbb{R}$ with $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$. Therefore, the state $|\psi\rangle$ can be visualized as a $\mathbb{R}^3$ vector $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\phi)$ pointing from the origin to the surface of a sphere, see Figure 3.1.

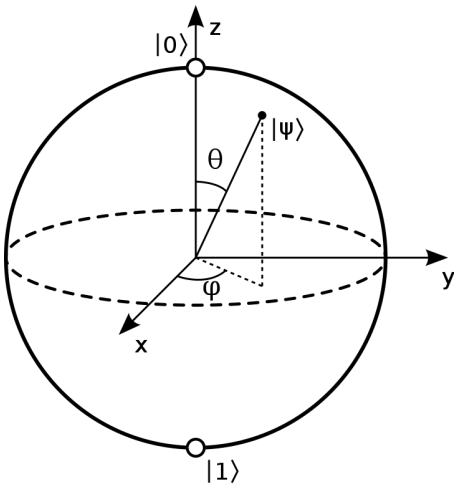


Figure 3.1: Bloch sphere representation as a generic pure state $|\psi\rangle$ [59].

3.2.2 Density matrix and mixed states

The density matrix $\rho_{\text{pure}} = |\psi\rangle\langle\psi|$ of the pure state $|\psi\rangle$ represents a projector operator:

$$\rho_{\text{pure}} = |\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha_0|^2 & \alpha_0\alpha_1^* \\ \alpha_0^*\alpha_1 & |\alpha_1|^2 \end{pmatrix} \tag{3.3}$$

which has the following properties:

- $\rho_{\text{pure}} = \rho_{\text{pure}}^\dagger$.
- $\forall\,|\phi\rangle : \langle\phi|\,\rho_{\text{pure}}\,|\phi\rangle \geq 0$.
- $\operatorname{Tr}\rho_{\text{pure}} = 1$.
- $\rho_{\text{pure}}^2 = \rho_{\text{pure}}$.

The density matrix framework allows us to define a *mixed state* ρ_{mix}

$$\rho_{\text{mix}} = \sum_i p_i\,|\psi_i\rangle\langle\psi_i| \tag{3.4}$$

where p_i represents the probabilities associated to a set of pure state $|\psi_i\rangle$. ρ_{mix} does not satisfy the same properties of ρ_{pure}, in particular:

- $\rho_{\text{mix}}^2 \neq \rho_{\text{mix}}$ and
- $\operatorname{Tr}\rho_{\text{mix}}^2 < \operatorname{Tr}\rho_{\text{mix}} = 1$.

3.2.3 Multi-qubit states

Multi-qubit quantum systems are *tensor products* between Hilbert spaces, e.g., a n-qubit quantum system is represented by the tensor product of n single qubits:

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_n\rangle \qquad |\psi_1\rangle, |\psi_2\rangle, ... |\psi_n\rangle \in \mathcal{H} \qquad |\psi\rangle \in \bigotimes_{i=1}^{n}\mathcal{H} \tag{3.5}$$

A completely generic n-qubits state is represented by 2^n complex coefficients:

$$|\psi\rangle = \alpha_1 |0...00\rangle + \alpha_2 |0...01\rangle + \alpha_3 |0...10\rangle + ... + \alpha_{2^n} |1...11\rangle \qquad \alpha_i \in \mathbb{C} \tag{3.6}$$

Finally, pure and mixed density matrices can be generalized using a n-qubits state $|\psi\rangle$, then

$$|\psi\rangle = \sum_{i=1}^{2^n} \alpha_i |i\rangle \qquad \rho_{\text{pure}} = |\psi\rangle \langle\psi| \qquad \rho_{\text{mix}} = \sum_{i,j} \alpha_{ij} |i\rangle \langle j| \qquad \alpha_i, \alpha_{ij} \in \mathbb{C} \tag{3.7}$$

3.2.4 Quantum measurements

For the seek of simplicity, let us assume the generic 1-qubit state represented in equation (3.1). For this case, the two possible results of a measurement in the computational basis (0 or 1) are associated to the two projectors on the eigenspaces

$$P_0 = |0\rangle \langle 0| \qquad P_1 = |1\rangle \langle 1| \tag{3.8}$$

Therefore, the probabilities associated to the two measurement are:

$$p(0) = \langle\psi | P_0 | \psi\rangle = |\alpha_0|^2 \qquad p(1) = \langle\psi | P_1 | \psi\rangle = |\alpha_1|^2 \tag{3.9}$$

Then, the quantum state collapses into the eigenstate associated to the measurement outcome. For a 0 result, the quantum state after the measurement is

$$|\psi\rangle \to \frac{P_0 |\psi\rangle}{\sqrt{\langle\psi | P_0 | \psi\rangle}} = |0\rangle \tag{3.10}$$

Note that the final quantum state was destroyed. This kind of measurements in a qubit system are called *projective measurements* on the computational basis.

For an n-qubits system, a computational basis measurement is obtained by sampling a binary string of length n from the quantum state distribution. Then, each the 2^n possible outcomes is associated to a projector defined in the eigenspace on which the state collapses after the measurement.

3.3 Entanglement

Another fundamental concept in quantum computing, which no classical counterpart, is the *entanglement* between quantum systems. For example, a composite system made of A and B quantum systems (e.g. two qubits or two groups of qubits), is called *separable* if it can be written as a classical probability distribution p_i over uncorrelated states of the two subsystems ρ^A and ρ^B:

$$\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B \tag{3.11}$$

When a state is non-separable, it is called an entangled state, e.g., the Bell states:

$$|\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B \pm |1\rangle_A \otimes |1\rangle_B) \tag{3.12}$$

$$|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |1\rangle_B \pm |1\rangle_A \otimes |0\rangle_B) \tag{3.13}$$

which are known as *maximally entangled* states, since the measurement on the subsystem B is exactly determined by a measurement on the subsystem A, due to both subsystem are maximally correlated. This effect is purely quantum and cannot be obtained by classical operations, as Bell proved in 1964 [60]. This phenomenon shows the incompatibility of locality and realism in Quantum Mechanics and represents a unique feature in the Quantum Information theory.

3.4 Quantum circuits

Classical computers operate using bit-strings by means of *logical function* $f : \{0,1\}^n \to \{0,1\}^m$ that gets a n-bits string as input and returns a m-bits string as output. Every logical function can be decomposed in elementary logical function, called *logical gates*, such as the AND gate, the OR gate, the NOT gate, the XOR gate, etc. Elementary logical gates can be combined in *logical circuits* implemented electronically. In analogy with classical computers, qubits can be manipulated by mean of *quantum gates* represented by unitary operators[1] acting on the quantum state.

3.4.1 1-qubit gate

It is described by 2×2 complex unitary matrices, e.g., the *Pauli operators*:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.14}$$

The X gate is also know as the NOT gate because it flips the computational basis of the qubit: $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$. Other widely used quantum gates are the *Hadamard gate* H, the *phase gate* S and the $\pi/8$ gate T:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \qquad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix} \tag{3.15}$$

Besides, the operators that rotates, an angle θ, the state vector around the x, y and z axis on the Bloch sphere are the *rotation gates*:

$$R_x(\theta) = e^{-i\frac{\theta}{2}X} = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}X = \begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \tag{3.16}$$

$$R_y(\theta) = e^{-i\frac{\theta}{2}Y} = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}Y = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \tag{3.17}$$

$$R_z(\theta) = e^{-i\frac{\theta}{2}Z} = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}Z = \begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{pmatrix} \tag{3.18}$$

3.4.2 Controlled gates

A controlled gate does controlled operations such as "If A is true, then do B". For example, let U be an arbitrary 1-qubit unitary operation, then *controlled-U* operation is a 2-qubits operation with a *control qubit* and a *target qubit*. If the control qubit is set to $|1\rangle$ then U is applied to the target qubit or in the other way around. Figure 3.2 (left) represents a controlled-U operation where the control qubit (on the top) is wired to the controlled gate (on the bottom).

Controlled-U Controlled-NOT Controlled-Z

Figure 3.2: Controlled gates. The top wire represents the control qubit while the bottom wire is the target qubit. Next to the circuit symbol the matrix representation is shown.

The Controlled-NOT (CNOT) gate and the Controlled-Z (CZ) gate are the most common controlled gates. Their matrix representations and circuit symbols are shown in Figure 3.2.

[1] An operator O on an Hilbert space is called *unitary* if $OO^\dagger = O^\dagger O = \mathbb{1}$, where $\mathbb{1}$ is the identity operator.

3.4.3 Measurement gates

These type of gates are used to do projective measurements on the computational basis, represented as"meter" symbols, see Figure 3.3.

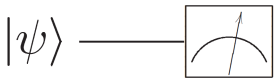


Figure 3.3: Measurement gate.

The *principle of deferred measurement* [61] demonstrated that measurements performed at an intermediate stage of a quantum circuit can always be moved to the end of the circuit without loss of generality. Besides, according to the *principle of implicit measurement* [61], any unmeasured qubit at the end of the circuit may be assumed to be measured, or in other words, measuring an unmeasured qubit at the end of the circuit does not affect the probability distribution of the other measurements.

3.5 Machine Learning with quantum circuits

3.5.1 Introduction to Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) based in the computer learning from data to make predictions or decisions without being explicitly programmed to do so. ML automatizes the process to make predictions with known and unknown information. It is a very powerful tool in tasks where the data patterns are very complex and difficult to understand. Therefore, this approach involves a very general and agnostic mathematical model where the data is adapted to each case. Normally, the ML model is considered as a black box that has learned the patterns of the system in order to produce reliable predictions.

ML is classified in three main branches depending of the prediction problem:

- **Supervised learning**. This model performs the training phase pairing the input data with the expected predictions of the model.
- **Unsupervised learning**. In this model, the expected predictions are not used during the training, but instead, it does self-learning using the patters produced from the input data.
- **Reinforcement learning**. This model is trained using a rewards/punishment mechanism where the optimization is done by trial and error.

This thesis will use supervised learning for the training and classification of b- and c- jets.

3.5.2 Supervised learning

The main structure of a generic supervised machine learning algorithm (Figure 3.4) is the following:

1. **Problem definition**. Given a training data-set $\mathbb{D} = \{(x_1, y_1), ..., (x_M, y_M)\}$ with M couples of training inputs $x_i \in X$ (input domain) and expected outputs $y_i \in Y$ (output domain). The relationship between the x_i and the y_i is still unknown. This relationship will be approximated using a ML model, see the top-left of Figure 3.4. For this thesis, a N-dimensional real vectors, knows as *feature vectors*, will be used as input data.

2. **Select a model family**. It is a function

$$y = f(x, \theta) \tag{3.19}$$

with $x \in \mathbb{X}$, $y \in \mathbb{Y}$ and $\theta = \{\theta_1, ..., \theta_D\}$ is a set of real parameters that will be optimized during the training, see Figure 3.4 (upper-right). More complex models also depend on hyper-parameters that are not included explicitly in the notation.

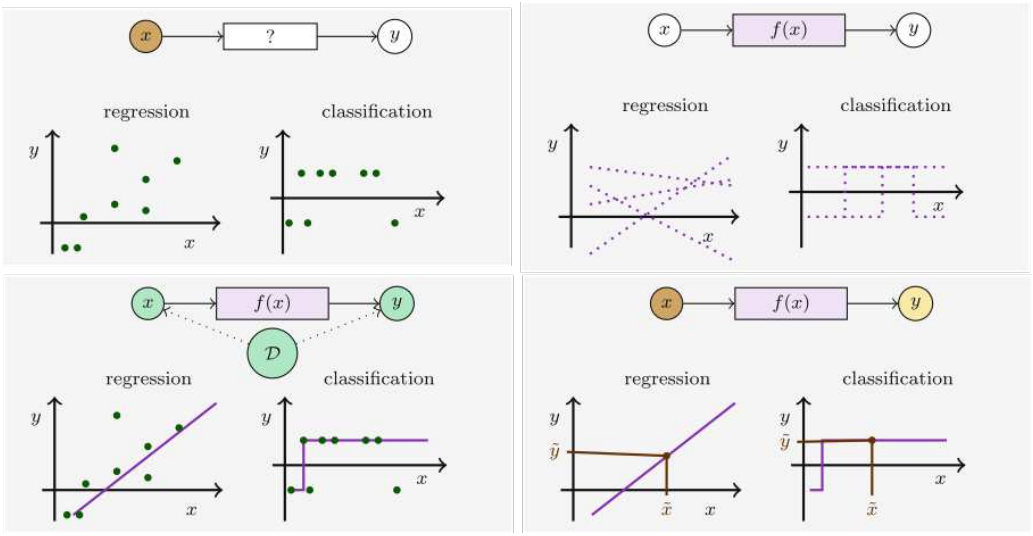


Figure 3.4: Supervised learning algorithm employs four main stages. First, the train data-set is selected (top left). Later, a model family is chosen (top right). Afterward, the model is trained (bottom left). Finally, the model is employed on new data to do predictions (bottom right) [62].

3. **Training the model**. The model is trained by fitting the θ parameters and the hyper-parameters to the training data $\mathbb{D}$ using a specific model function. Later, a test data-set can be used to validate the performance of the model on new data, see Figure 3.4 (lower-left).

4. **Employment of the model**. After the training, the optimized model can be used to do predictions, see Figure 3.4 (lower-right).

3.5.3 Training a ML model

The training phase is the most crucial step in a ML algorithm. The goal of the training is to obtain the best predictions using a set of parameters θ. These parameters are the result of the optimization of an objective function $L(\theta)$, called *loss function*. This function quantifies the quality of the predictions given the current set of parameters θ. Therefore, the training phase can be considered as an optimization problem. The choice of a certain loss function implies different training strategies and results. This thesis will use the *mean squared error loss*,

$$L^{MSE}(\theta) = \frac{1}{M}\sum_{i=1}^{M}(f(x^i,\theta) - y^i)^2 \tag{3.20}$$

and the *cross-entropy loss*,

$$L^{CE}(\theta) = \sum_{i=1}^{M} -y^i \log p^i - (1-y^i)\log(1-p^i) \tag{3.21}$$

where $p^i = f(x^i,\theta)$ are the outputs of the model in terms of prediction probabilities. The loss function $L(\theta)$ has to be minimized with respect to the model parameters θ in order to obtain the optimal parameters $\bar{\theta}$,

$$\bar{\theta} = \arg\min_{\theta} L(\theta) \tag{3.22}$$

However, complex ML tasks often fall in hard optimization problems since the loss function can be non-convex. Therefore, popular methods based on iterative searches are used such as the *gradient descent method*, see Figure 3.5. Commonly, gradient descent algorithms are local optimization methods that iteratively updates the parameters of the loss function $L(\theta)$ towards the direction of the steepest descent

$$\theta_{t+1} = \theta_t - \lambda \nabla L(\theta_t) \tag{3.23}$$

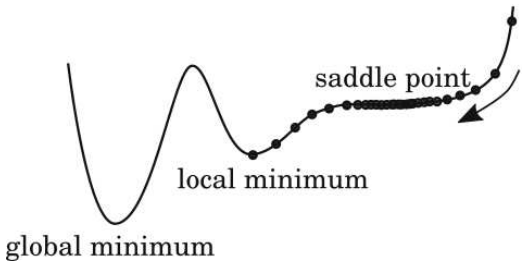


Figure 3.5: Illustration of a gradient descent method. Several minima points are observed implying a difficult and slow convergence at saddle points [62].

where λ is a tunable hyper-parameter called *learning rate* and t is the iteration index. Figure 3.5 illustrates an example the converge of a loss function in a 1-parameter model. Since the gradient of the loss function is used during the optimization, the method can get stuck on a local minimum and it convergence rate can get negligible low on saddle points, where ∇L tends to 0. To deal with these effects, several optimized methods have been derived starting from the gradient descent method. In this thesis, the ADAMAX [63] optimizer has been used.

3.6 Parameterized quantum circuits

Parametrized Quantum Circuits (PQC), also known as *Variational Quantum Circuits* (VQC), are normally build as a combination of fixed quantum gates (e.g., *CNOT*, *CZ*, *H*, etc) and variational quantum gates of tunable parameters (e.g., *RX*, *RY*, *RZ*, etc). This kinf of VQC can be employed as a Quantum Machine Learning (QML) model. Besides, VQCs have already showed that are able to produce non-trivial results not efficiently simulated by classical hardware, even at low circuit complexity.

The main constituents of a QML model based on a VQC are illustrated in Figure 3.6, where classical steps are highlighted in light blue while the quantum step in pink. These are:

1. **Pre-processing**. A feature vector x is sampled from the training data-set P_D and then is mapped, following a classical pre-processing scheme, by a function ϕ where $x \to \phi(x)$. This step can include feature importance selection and normalization or scaling techniques.
2. **Parameterized Quantum Circuit.** The pre-processed data is mapped to the parameters of a quantum encoder circuit $U_{\phi(\mathbf{x})}$ which embeds the data into an n-qubits state. A VQC, with tunable parameters θ and represented by U_θ, performs quantum operations over the n-qubits state of the model. This is followed by the measurements of the expectation values $\{\langle M_k \rangle_{x,\theta}\}_{k=1}^K$ over the final quantum states. Note that all the components of the PQC circuit are parameterized and optimized continuously during the training phase.
3. **Post-processing**. Finally, the measured expectation values $\{\langle M_k \rangle_{x,\theta}\}_{k=1}^K$ are mapped to a prediction model through a post-processing function $y = f(\{\langle M_k \rangle_{x,\theta}\}_{k=1}^K)$.

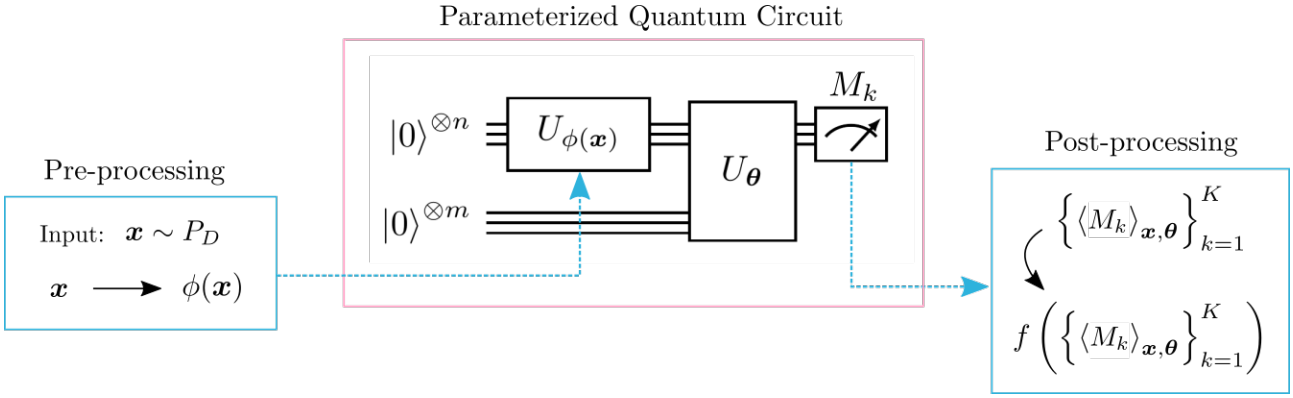


Figure 3.6: Parameterized Quantum Circuit and its main components [64].

3.6.1 PQC Data encoding

This is the first step of the QML model based on PQC after the data pre-processing. It maps a vector of classical data x into an n-qubits quantum state $x \to |x\rangle$. This task can be done in different ways. In this thesis, the *Amplitude Encoding*, *Angle Encoding* and *Quantum Approximate Optimization Algorithm Encoding* techniques were used.

- **Amplitude Encoding**. It embeds up to 2^n features into the amplitudes of the state vector of an n-qubits system. The main advantage of this method is the exponential scaling to large number of features. In the case of features smaller than 2^n, the the remaining amplitudes are fed with constant values (normally 0). Assuming exactly 2^n features, the amplitude encoding mapping is

$$|x\rangle = \sum_{i=1}^{2^n} x_i \, |n_i\rangle \tag{3.24}$$

  where x_i refers to the i^{th} feature and $|n_i\rangle$ is the i^{th} vector of the computational basis. Note how the x vector must be normalized, i.e., $\sum_i |x_i|^2 = 1$.

- **Angle Encoding**. This technique encodes up to n data features into an n-qubits state trough angle rotations around one of the three axes, i.e. with $i = 0, \dots n-1$ the i^{th} qubit state is

$$|\psi\rangle_i = R_x(x_{i+1}) R_y(x_{i+2}) R_z(x_{i+3}) \, |0\rangle_i \tag{3.25}$$

  where R_x, R_y and R_z are the rotation gates, see Eqn. (3.16) - Eqn. (3.18). Then, the feature vectors x are represented by the full n-qubits state $|x\rangle$ built as the tensor product of the single $|\psi\rangle_i$ states.

$$|x\rangle = \bigotimes_{i=0}^{n-1} |\psi\rangle_i \tag{3.26}$$

  It is important to take advantage that the rotational gates are periodic unitary operator with respect to the parameter θ. Therefore, $R_i(\theta) = R_i(\theta + 2\pi)$. Then, each feature vector must be scaled (pre-processed) to the $[0, 2\pi]$ range. However, more restricted ranges could improve even more the training performance.

- **QAOA Encoding**. It was introduced by Farhi et al. [65]. It encodes N features into $n > N$ qubits, using a trainable quantum circuit. A single quantum layer applies two circuits ("Hamiltonians") where the first encodes the features, and the second is a variational ansatz inspired by a 1-dimensional Ising model. The feature-encoding circuit associates features with the angles of R_X rotations. Its initial state $|\psi\rangle_0$ will be the uniform superposition over computational basis states:

$$|\psi\rangle_0 = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle \tag{3.27}$$

  When there are fewer features than qubits, the feature-encoding rotation is substituted by a Hadamard gate.

3.6.2 Variational circuits

They are the core of PQC circuits where, similarly to the universal approximation theorem in neural networks [64], there always exists a quantum circuit able to act as a target function within an arbitrary small error. Nevertheless, this circuit may get deeper exponentially and impractical to implement on current quantum hardware. That's why, circuit templates with fixed structure of gates (*circuit ansatz*) are used. Note, despite the fact that the dimension of the Hilbert space grows exponentially with the number of qubits, the number of free parameters to be optimized scales as a polynomial of the qubit count.

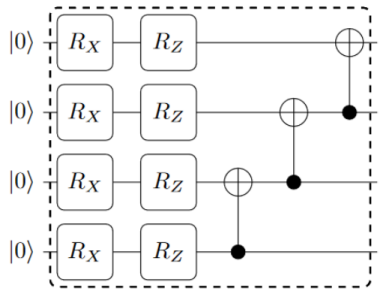


Figure 3.7: Variational circuit ansatz example in a PQC model. The angle parameter of each rotational gate (R_x and R_z) can be tuned during the training.

Figure 3.7 illustrates the structure of an example of a variational ansatz for QML model. The first stage of this circuit has tunable rotational gates applied to the single qubits and a second stage of controlled gates to entangle the qubits. This ansatz can be repeated multiple times growing the number of parameters and the complexity of the model.

3.6.3 Gradients of PQC

They are important operations during the implementation of a gradient descent algorithm on a QML model. Eqn (3.23) needs the evaluation of the partial derivatives of the loss function with respect to the tunable parameters θ_i of each rotational gate. These parameters can be evaluated numerically using the finite difference method:

$$\frac{\partial L}{\partial \theta_i} \approx \frac{L(\theta + \epsilon \hat{e}_i) - L(\theta - \epsilon \hat{e}_i)}{2\epsilon} \tag{3.28}$$

where ϵ represents a infinitesimal parameter and $\hat{e}_i$ is the θ_i unitary direction vector. However, ML is not free of optimization problems where instabilities with respect to numerical differentiation can arise. Therefore analytical gradients are used if available.

Fortunately, analytical gradients for PQC can be calculated exactly using the so-called *parameter shift rule* [66, 67]. Using the chain rule and recalling that $L(\theta)$ is defined as a function of the expectation values $\langle M_k \rangle_\theta$, so, the derivatives of $L(\theta)$ can be rewritten as a function of derivatives of $\langle M_k \rangle_\theta$. Then,

$$\frac{\partial \langle M_k \rangle_\theta}{\partial \theta_i} = \frac{\langle M_k \rangle_{\theta + \frac{\pi}{2}\hat{e}_i} - \langle M_k \rangle_{\theta - \frac{\pi}{2}\hat{e}_i}}{2} \tag{3.29}$$

Therefore, that exact gradients can be computed doing two evaluations of the quantum circuit per parameter, even on quantum hardware.

Chapter 4

Results on b- and c-jet tagging with Quantum Machine Learning

This chapter will be focused to detail all the software implementation and the results from the b- and c- jet classification using different QML models and comparing them with the best classical method doing this classification.

4.1 Training data-set selection

A list of jets is extracted from the data-set, described in section 2.7, with the following requirements:

- Jets with transverse momentum $20 < p_T < 1000$ GeV/c.
- Jets with pseudo-rapidity $2.2 < \eta < 4.2$.
- Jets must be originated from a QCD event.

Initially the amount of available events was around 10^6 jets. After the selection criteria, the data-set was reduced to 4×10^5 jets equally distributed in b- (50%) and c- (50%) jets which have been split in training and test subsets in 70/30 proportion. First, this data should be pre-processed extracting the 13 Secondary Vertex (SV) features among all the available features in order to be used as the input of the QML models. The following SV features have been used for the training:

- The SV mass M.
- The SV corrected mass M_{cor}.
- The transverse flight distance of the two-track SV closest to the PV.
- The fraction of the jet p_T carried by the SV, $p_T(\text{SV})/p_T(\text{jet})$.
- ΔR between the SV flight direction and the jet.
- The number of tracks in the SV.
- The number of SV tracks with $\Delta R < 0.5$ relative to the jet axis.
- The net charge of the tracks that form the SV.
- The flight distance χ^2.
- The sum of all SV track χ^2_{IP}.
- The SV transverse momentum p_T.
- The SV lifetime.
- The SV z-position.

The first row of Figure 4.1 shows the correlation matrix between the 13 SV features either for b- and c- jets. Both correlation matrices looks very similar, as well as with the correlations between features, e.g, the correlation between the SV corrected mass and the SV mass looks the same for the b- and c- jets. Regarding the features distributions, the SV corrected mass is the distribution that presents the most remarkable differences between b- and c- jets, as expected. Besides, each jet carries information about its quark origin (coming from the matched Monte Carlo jet) through the Monte Carlo truth label as described in section 2.4.3, taking a value of 0 if it is b-quark or 1 if it is a c-quark. This variable represents the target feature for the tagging algorithms and is called *Jet Label*.

Additionally, each di-jet ($b\bar{b}$ or $c\bar{c}$) event in the data-set contains two jets of opposite charge associated to the same pp collision, which surely will induce feature correlations. Nevertheless, a jet tagging algorithm must be able to do a prediction on a jet flavour based only on its particle content. Therefore, for training purposes, a shuffled list of jets will be produced to suppress any correlation.

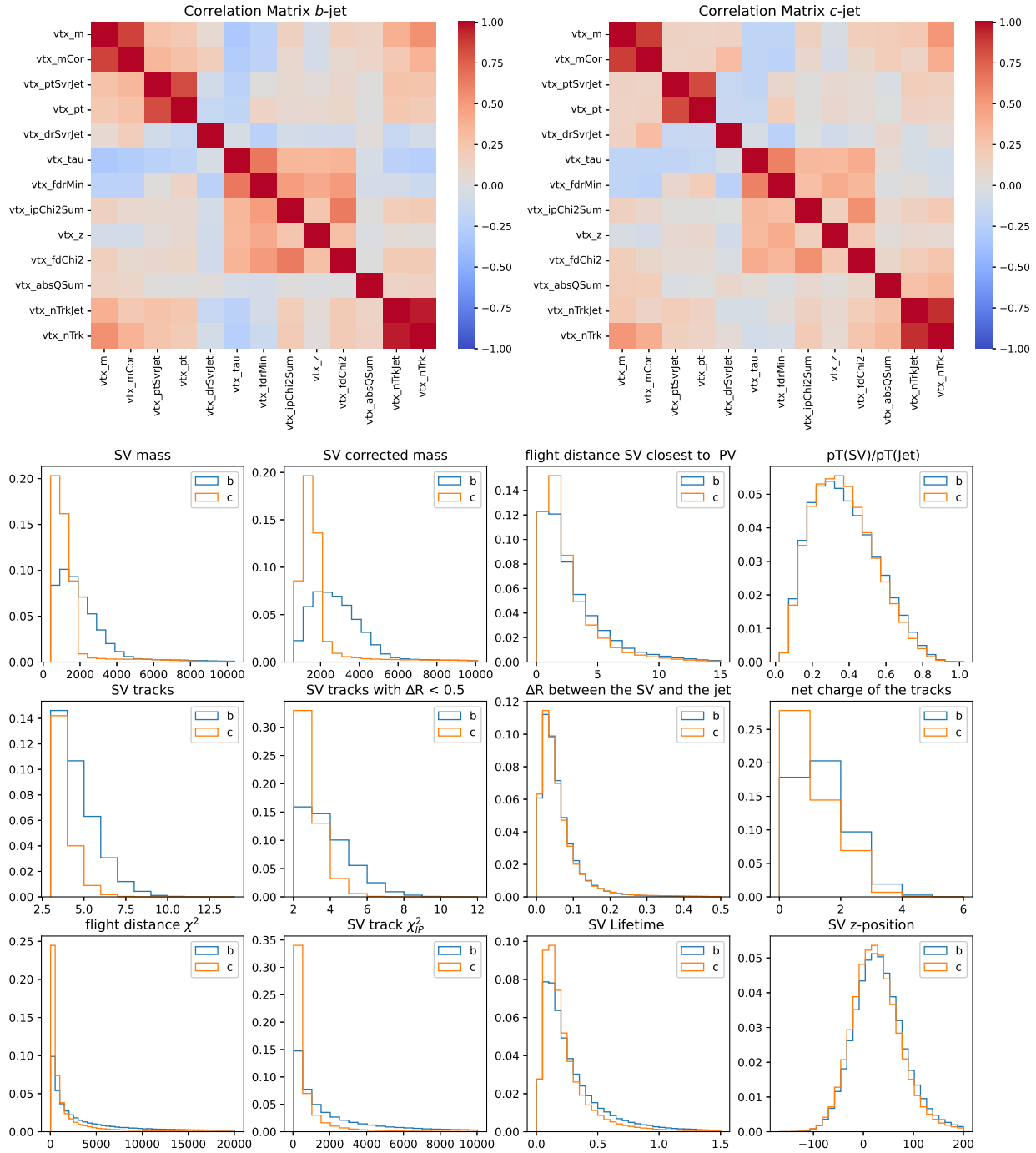


Figure 4.1: Correlation matrices (first row) for the b- and c- jets as well as their feature distributions.

4.2 Quantum Machine Learning models

Different QML models have been tested for the b- and c- jet identification and compared with the classical method [45] used at the LHCb Collaboration, in terms of tagging performance.

4.2.1 Amplitude Encoding classifier

This model consists in a PQC made of an encoding circuit (Amplitude Encoding) followed by a variational circuit, see Figure 4.2. The amplitude encoder circuit is used to embed the 13 features of the data-set into the amplitudes of a 4-qubit circuit. After the encoder, the variational circuit composed of L *strongly entangled layers* are connected. In this stage, each layer consists of parameterized rotational gates $R(\alpha_i, \beta_i, \gamma_i)$ applied to each qubit connected to another qubit through CNOT gates in order to entangle all the qubits. Besides, the complexity of this n-qubit circuit classifier can be tuned adjusting the number of strongly entangles layers L and, consequently, the number of tunable parameters N_{par} equal to

$$N_{\text{par}} = 3 \times n \times L \tag{4.1}$$

Note that even if the dimension of the Hilbert space scales exponentially (2^n) with respect to the number of qubits n, the number of parameters N_{par} just scales linearly. This exponential scaling is exploited by the amplitude encoder which is capable to deal with large number of features using low number of qubits. For this thesis, the number of strongly entangled layers is fixed to $L = 7$, leading to a total number of parameter of $N_{\text{par}} = 84$.

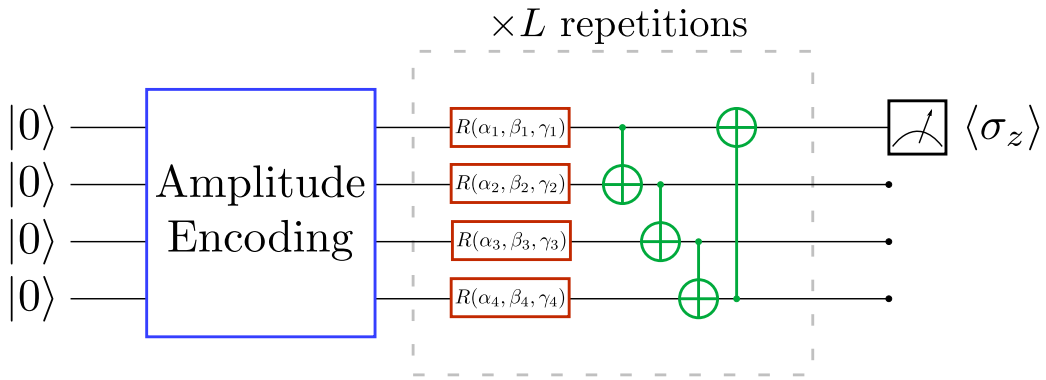


Figure 4.2: Circuit representation of the Amplitude Encoding classifier model. The amplitude encoding stage (blue) is connected to the parametric rotational gates (red) to be optimized during the training phase followed by the entangling CNOT gates (green).

4.2.2 Angle Encoding classifier

This circuit requires a one to one correspondence between the input features and the number of qubits. Therefore, during the first stage of the circuit n features are embedded in n rotational gates $R_x(\theta_i)$ as parameters θ_i, see Figure 4.3. The variational stage of the circuit is identical to the Amplitude Embedding classifier, where L strongly entangled layers can be employed to tune the complexity of the model. Therefore, the number of parameter N_{par} to be optimized are the same as with the Amplitude Embedding classifier, see again Eqn. (4.1) where the N_{par} scales linearly with respect to n. However, this embedding does not exploit the exponential scaling of the Hilbert space. This means that doing predictions with high-dimensional data-set requires a large number of qubits. For this thesis, the number of repetitions of the variational layer L was also fixed to 7, so that the model has a total number of variational parameters $N_{\text{par}} = 273$.

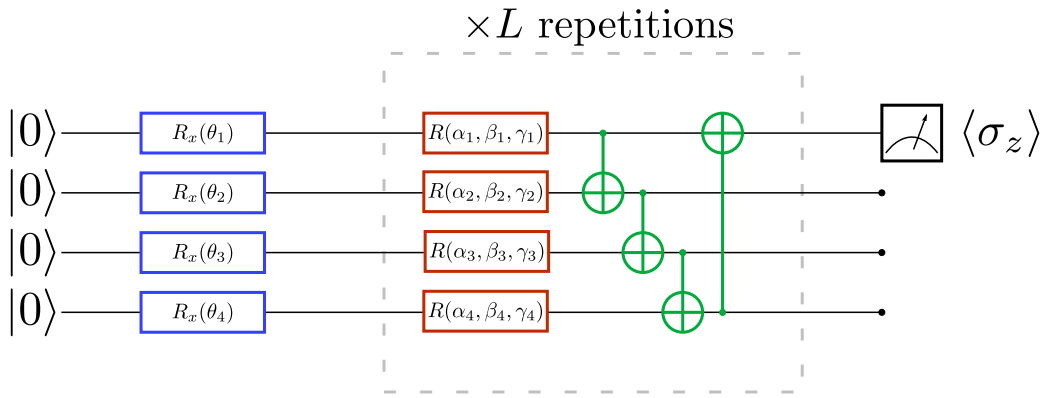


Figure 4.3: Angle Encoding classifier model. The encoding stage is represented by rotational gates (blue) followed by the parametric rotational gates (red) to be optimized during the training phase. Finally couples of qubits are entangled using CNOT gates (green).

4.2.3 QAOA Encoding classifier

This model has a structure inspired from Tree-Tensor Networks (TTNs) [68]. In this circuit, n features are encoded into an n-qubit quantum circuit using the QAOA encoding, see Figure 4.4. This encoding stage contains trainable parameters that presumably enhance the quality of the embedding and must be trained along the other parameters of the circuit. The variational stage of the circuit has a tree structure to process the information coming from the n features, gradually reducing the total number of qubits up to a single qubit which is measured identically as in the previous models. This model contains tunable parameters also in the embedding part where the total number of trainable parameter is $N_{\text{par}} = 2 \times n \times L = 182$, using $L = 7$ variational layers.

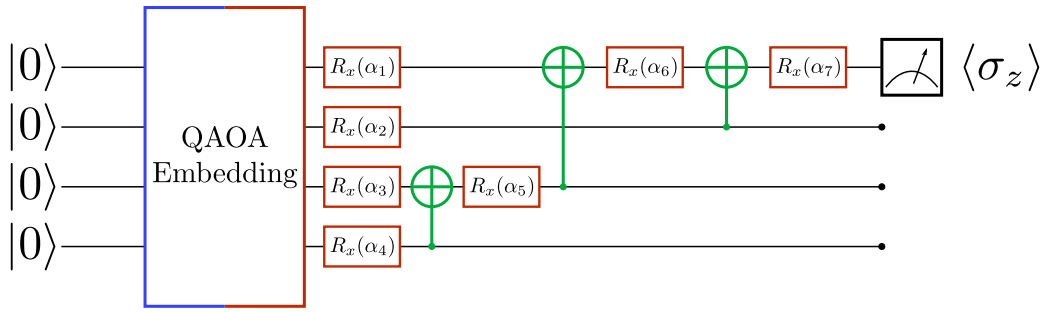


Figure 4.4: Quantum circuit of a QAOA encoding classifier.

4.2.4 Measurement stage

For the measurement stage, the expectation value of the Pauli operator σ_z is measured, with range $\langle \sigma_z \rangle \in [-1, +1]$. Thus, the probability to be a b-jet (P_b) or a c-jet (P_c) are defined as

$$P_b = \frac{\langle \sigma_z \rangle + 1}{2} \tag{4.2}$$

$$P_c = \frac{1 - \langle \sigma_z \rangle}{2} = 1 - P_b \tag{4.3}$$

Using this particular choice, a measurement of $\langle \sigma_z \rangle = -1$ corresponds a 100% probability to be a b-jet, while a measurement of $\langle \sigma_z \rangle = +1$ corresponds to a 100% probability to be a c-jet. Note that $\langle \sigma_z \rangle = 0$ will corresponds to the maximum uncertainty (50% probability). The measurement stage is basically the same for the 3 QML classifiers described before.

4.3 Classical Neural Network model

In addition to the QML models, a classical ML model was used as a reference. This model is the Boosted Decision Tree (BDT) which is commonly used for jet flavour tagging at the LHCb Collaboration [45]. Other classical methods were also used such as Deep Neural Network (DNN) and Convolutional Neural Network (CNN) but the best results were obtained using the BDT model. Therefore only the BDT model will be used to compare its performance with the QML methods.

The BDT is a Gradient boosting classifier which is a machine learning technique that combine many weak learning models (decision trees) together to create a strong predictive model. A decision tree takes a set of input features and splits input feature data at each node recursively based on those features, see Figure 4.5. The terminal nodes is called leaves and represent a class label or probability.

The following parameters can be tuned in a BDT model:

- **Maximum depth**. How tall a tree can grow, sometimes defined by the number of leaves. Set 10 in this work.
- **Maximum features**. How many features can be chosen randomly to build a given tree.
- **Minimum samples per leaf**. How many samples are required to make a new leaf where usually is less that 1% of data and sometimes defined by samples per split.
- **Number of trees.** Set 100 in this work.

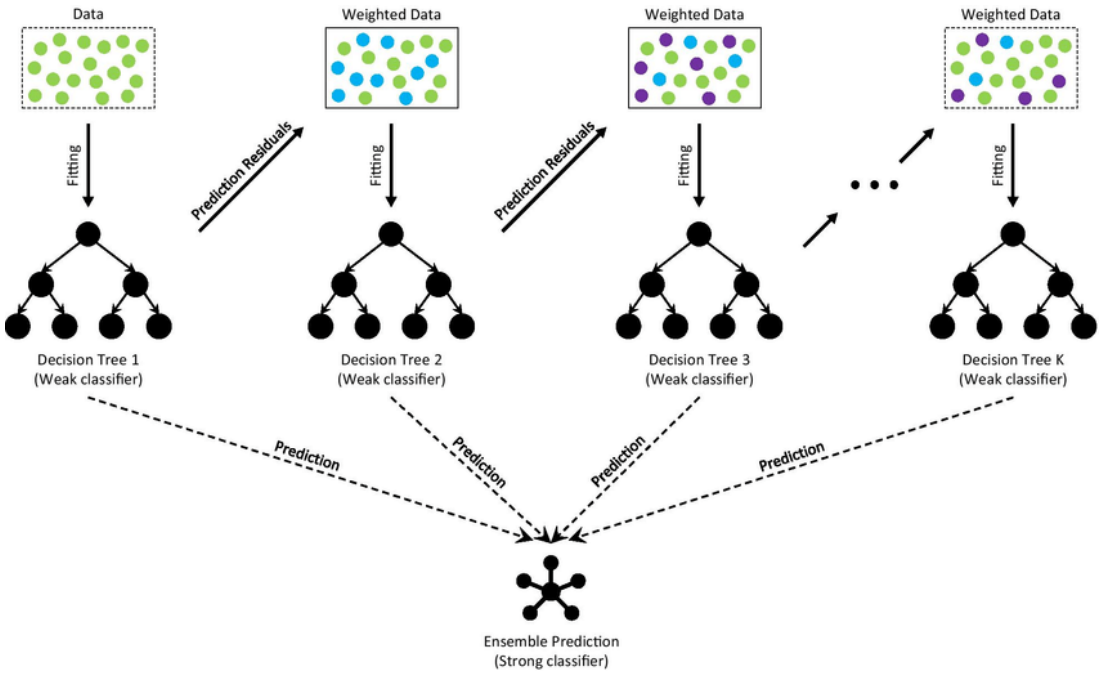


Figure 4.5: Diagrammatic representation of a classical Boosted Decision Tree [69].

A summary of all the analysed models is available in Table 4.1.

Model	$\mathbf{N_{feat.}}$	**n**	L	$\mathbf{N_{par}}$
Amplitude Embedding	13	4	7	84
Angle Embedding	13	13	7	273
QAOA Embedding	13	13	7	182
BDT	13	-	-	1 000

Table 4.1: Summary of all the structure parameters of the QML models used for the jet b- and c-jet tagging.

4.4 Software implementation

Python Programming Language has made a huge effort to incorporate libraries dedicated to the simulation of quantum computers, allowing the creation of QML models. For this thesis, the QML models have been implemented using the `PennyLane` [70] library and for the BDT method the `XGBoost` [71] library was used.

4.4.1 PennyLane

PennyLane is a modern Python library designed specifically for differentiable programming of quantum computers. This allows to simulate the training of a quantum computer in the same way as a neural network, allowing QML applications. Therefore, the core feature of PennyLane is the power to automatically compute gradients of parameterized quantum circuits in a way that is compatible with classical optimization techniques, like back-propagation. This enables the development of hybrid quantum-classical machine learning models built and trained with the same library. Besides, a plugin system supplies access to several quantum simulators and cloud quantum hardware devices.

Figure 4.6 illustrates a basic example of a QML algorithm. The first line loads the PennyLane package followed by the second line that defines a quantum device setting an execution back-end (e.g., `default.qubit`). The default variable for the number of qubit is *wires*. After, a 1-qubit circuit (`wires=1`) is defined using the function `circuit`, built with a R_x rotational gate and the expectation value of the qubit state (using the σ_z). This circuit is executed together with the `qml.qnode(dev)` decorator which maps the circuit to a device, creating a quantum computation node (`QNode`). Then, calling the function `circuit` implies the evaluation of the quantum circuit on the selected device. On the other hand, a `cost` function defines a classical computation node, whose input is a weight array of tunable parameters, which calls the quantum circuit. Finally, the last three rows of the code implement an optimization step of the `cost` function with respect to the weight parameters with the help of a simple Gradient Descent optimizer.

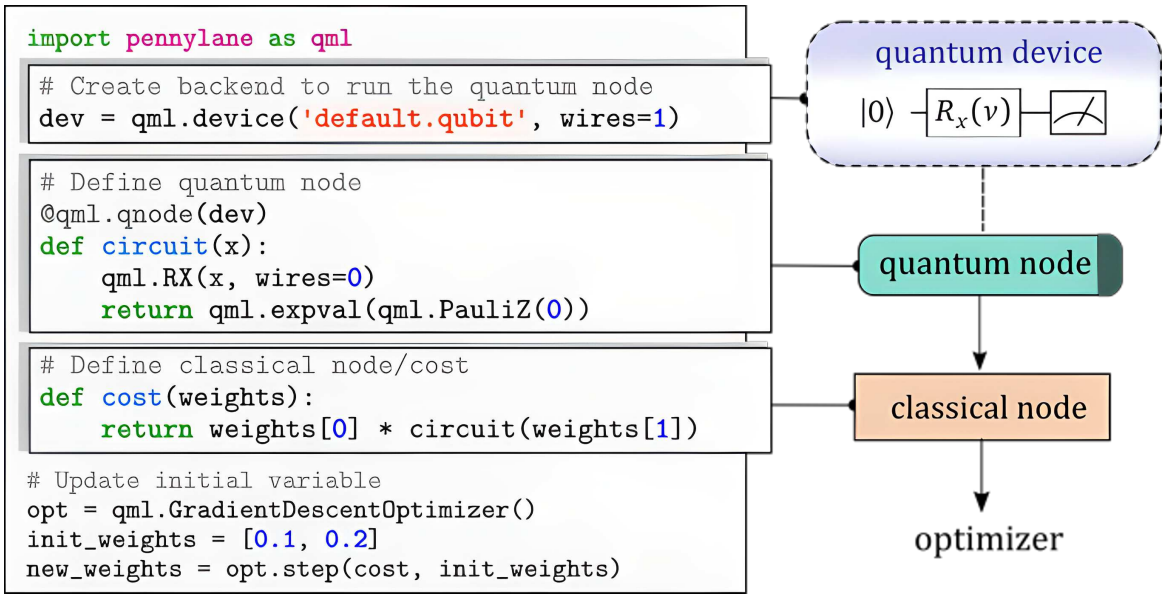


Figure 4.6: Basic example of a QML algorithm built with a quantum node followed by a classical node using the PennyLane package. The output of the classical node is the optimization target [70].

PennyLane supports several type of devices including quantum simulator and hardware devices:

- Simulators:
  - `default.qubit`. It implements needful methods as well as some built-in qubit operations. Therefore, it supplies a very simple pure state simulation of a qubit-based quantum circuit architecture written in Python.

  - `default.qubit.tf`: This module contains a TensorFlow implementation version of the `default.qubit` reference plugin. This allows automatic differentiation through the simulation.
  - `default.qubit.autograd`. This is a state simulator of `default.qubit` but written using the Autograd library for the automatic differentiation.
  - `default.qubit.jax`. This device provides a pure-state qubit simulator written using `JAX` library. `JAX` can automatically differentiate through a large subset of Python's features which compiles and runs on accelerators, like GPUs and TPUs.

- Hardware:
  - **IBM Quantum Experience**. It is an online platform to log into cloud-based quantum computing services provided by IBM.
  - **Amazon Braket**. It is an Amazon Web Services (AWS) cloud-based access containing several quantum simulators and hardware devices.
  - **Rigetti Forest**. It is a cloud-based quantum computing solutions supplied by Rigetti.

PennyLane supports the following automatic differentiation methods

- **Simulation-based differentiation**. It is based on the adjoint [72] or the back-propagation method. This method is not compatible with hardware devices since it needs the manipulation of the quantum state vector.
- **Hardware-compatible differentiation**. It is based on the finite differences method or the parameter-shift rule. This method can be executed on both simulators and hardware devices.
- **Device gradients**. It is provided by some devices (e.g., AWS Braket devices) which evaluates the gradient by themselves.

Other important feature of PennyLane is its support for hybrid quantum and classical models and its compatibility with existing machine learning libraries such as PyTorch [73] and TensorFlow [74], together with its high-level ML library Keras [75]. Therefore, Quantum circuits can be set up to interface with either NumPy, PyTorch, JAX, or TensorFlow, allowing hybrid CPU-GPU-QPU computations. For example, PennyLane can combine `KerasLayer` and `TorchLayer` to convert a `QNode` to a neural network layer that can be used in a Keras/PyTorch model. The main advantages will be:

- The quantum layer can be added to a classical neural network model producing a hybrid model.
- The optimization of the PQC is boosted using the tools provided by Keras and PyTorch.
- `default.qubit.tf` can exploit the GPU computational power for the training within an end-to-end differentiable TensorFlow model.

For this work, the new `default.qubit.jax` device is used because it has shown the best performance overall, specially for its high speed automatic differentiation calculations able to run on accelerators like GPUs.

4.4.2 JAX implementation

The training phase is the most computational demanding phase in a ML model. For the development of a PQC both on quantum hardware and quantum simulators, the gradient descent method needs a large amount of steps where in each step the quantum circuit is evaluated for each data point. The amount of steps to reach the convergence of the model depends on the optimization algorithm and on the differentiation and on number of parameters of the model. GPUs provide the ideal solution to this demanding problem, allowing the circuit evaluation per each data point in parallel making possible a faster evaluation of the whole gradient.

Fortunately, in 2019 Pennylane was able to implement `JAX` transformations to build and optimize quantum circuits. `JAX` was originally designed for classical ML and now many of its transformations are also available for QML, and can be used directly with PennyLane. In this context, the gradient descent technique can be implemented using the `jax.grad` tool, run quantum circuits in parallel using `jax.vmap`, the simulations can be compiled and optimized with the `jax.jit` tool, and control and seed the random nature of quantum computer simulations using the `jax.random`. Therefore, quantum computing programs can be complemented with `JAX` to run in parallel on GPUs as is normally done in classical neural networks nowadays. For that reason, `JAX` was implemented in all the QML models.

4.5 Data-set Pre-processing

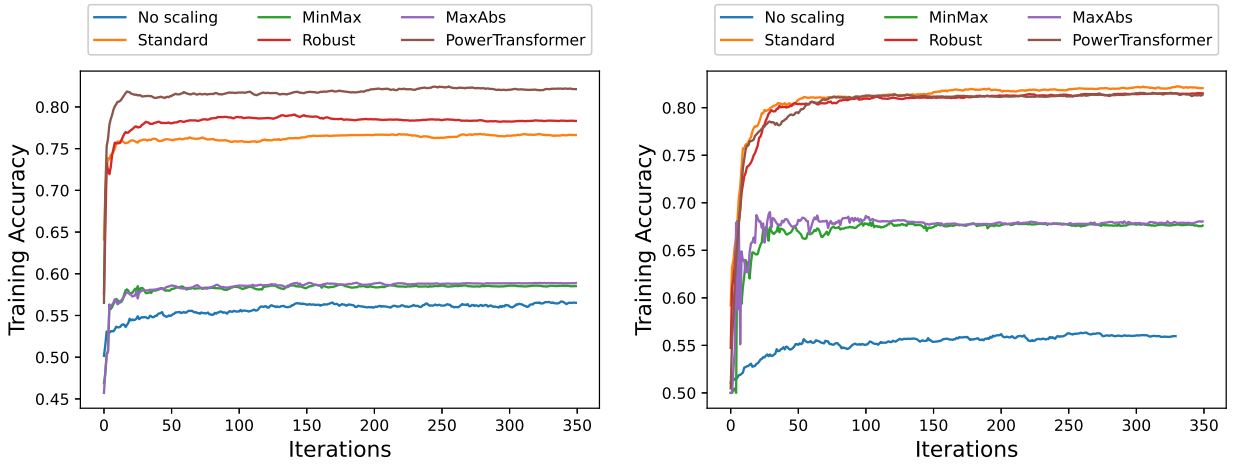
The key for a successful secondary vertex reconstruction is a precise understanding of the primary vertex and the correct selection of the tracks of the decay products. Tracks are reconstructed in each tracking system through different steps. For each track candidate, the track model is fitted to the hits in order to get the best estimates of the track parameters, usually in terms of χ^2. An abnormally large value of the χ^2 suggests a random combination of hits, i.e., a "fake" or "ghost" track [76], which is normally know as an *outlier*. An outlier is defined as a measurement that does not follow the expected behaviour. Outliers can either be removed from the track or down-weighted by employing a robust estimator.

Note that tracking efficiency affects the quality of the Data-set and therefore, the classification accuracy in the last instance. Unfortunately, as 100% tracking efficiency is not possible, some of the events inside the data-set could be outliers. Therefore a pre-processing of the data-set is essential to counteract this effect. For example, possible outlier values in the data could correspond to very large values of the flight distance χ^2 or the sum of all SV tracks χ^2_{IP} ($> 10\,000$) or the fraction of the jet p_T carried by the SV $p_T(\text{SV})/p_T(\text{jet})$ where some values are > 1, see again Figure 4.1.

Then, different processes can be followed in order to deal with the outlier in the data-set:

1. Remove the events with outliers by applying extra cuts.
2. Down-weight the anomalous values by employing a robust scaling.

In this work, the second option was used in order to don't eliminate some possible physical meaning for those anomalous values in the data-set, but for future works is worth exploring the first option. For that reason, several scaler functions were tested to see their impact on the training accuracy. Figure 4.7 shows the training accuracy as a function of the iterations done during the training applying different scaling functions in the data-set, using the Angle Embedding (left) and the QAOA Embedding (right).



(a) Accuracy as function of the number of iterations using the Angle Embedding..

(b) Accuracy as function of the number of iterations using the QAOA Embedding.

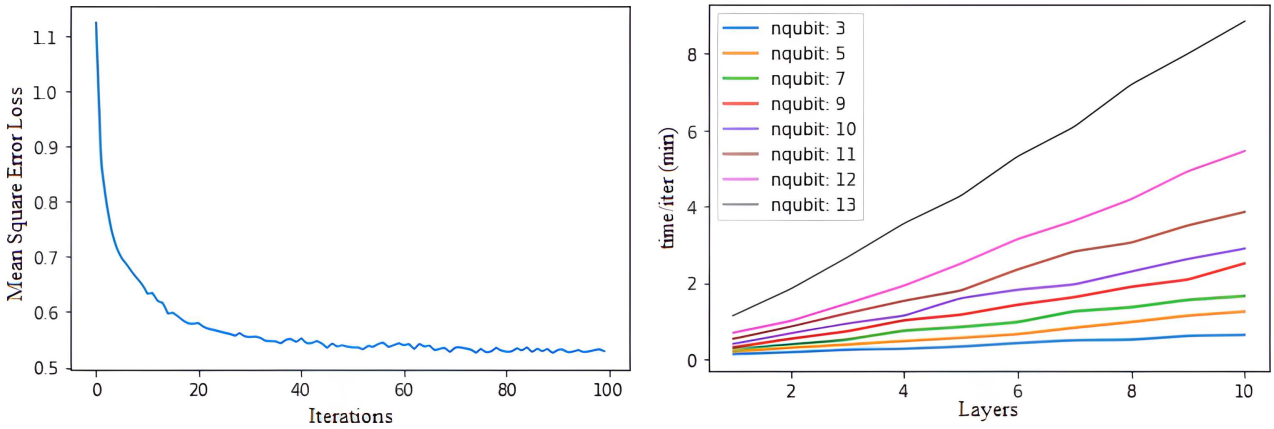
Figure 4.7: Training accuracy dependence on the Scaling function for different QML models.

The most relevant fact that can be observed is a "gap" that appears between different scaling functions in the training accuracy. It is because some scaling functions (e.g., `Standard, Robust, PowerTransformer`) are specialized to deal with outliers in the data-set. Based on the different results observed, `PowerTransformer` scaler is the one which provides the best accuracies. Therefore this scaler function is used to pre-process the data-set in combination with the `MinMax`$(0, \pi)$ scaler to scale the data-set after the first scaling in values inside the $[0, \pi]$ range because the embedding functions work better under angular values (rotational gates). Due to the different nature of the Amplitude Embedding, the best scaling function for this model is the `QuantileTransformer` in combination with the `MinMax`$(0, 1)$ scaler.

4.6 Training Phase

The loss function used for the training phase is the Mean Squared Error, defined in Eqn. 3.20, for all the quantum models as well as for the classical method (BDT). The ADAMAX [63] optimizer is used for all the quantum models fixing the learning rate at 0.01. For each QML model, the parameters matrix is initialized randomly at the beginning of each training stage which provokes a initial bad prediction. However, after the training phase takes place, the prediction distributions of b- and c-quark jets start to separate each other, getting closer to the extreme values of the [0,1] range. A perfect classification stage will place the b-quark jets in 0 and c-quark jets in 1 which means that the final quantum state (before the measurement) is $|0\rangle$ for b-quark jets and $|1\rangle$ for c-quark jets.

Each QML model has been trained for 100 iterations (epochs) where during each epoch, the optimizer processes the training data-set and updates the parameters at the end. Figure 4.8.a shows the behaviour of the loss function as a function of the number of iterations for the QAOA Embedding model as the training proceeds. The training is done using the data-set designated for the training (70%). As expected, the loss function gets progressively smaller, converging after around 80 epochs and a good separation between b-quark and c-quark jets can be appreciated.



(a) Mean Square Error loss as function of the training iterations for the QAOA Embedding method.

(b) Training time required for each iteration as function of the number of layers using the QAOA Embedding method.

Figure 4.8: Loss function and training time per iteration behaviour for the QAOA Embedding method.

Other important parameter to take into account is the training time. As was mentioned before, the `JAX` library will be used in order to speed up the training time using GPUs. For this work, initially 2 GPUs was used for the training phase. Specifically, the evaluations of the simulation on the quantum circuit will be parallelized over the 2 GPUs. Figure 4.8.b shows the required time per iteration using the QAOA embedding running on 2 GPUs. As expected, as the number of qubits increase the training time also increases. The same effect is observed increasing the number of layers. Note that there is a linear dependence between the training time and the number of layers. Thus, a 100-iteration training using the QAOA Embedding model with 13 qubits (all features) and 7 layers will require 10 hours.

4.7 QML optimization

One of the most important steps to build a robust QML is identify the perfect combination of number of qubits and quantum mechanical operations (layers) needed to get the best test accuracy defined, together with its standard uncertainty, as:

$$\text{Accuracy} = \frac{N^{tag}_{b(c)}}{N^{tot}_{b(c)}} \qquad \sigma_{Acc} = \sqrt{\frac{\text{Accuracy}(1+\text{Accuracy})}{N^{tot}_{b(c)}}} \tag{4.4}$$

where $N^{tag}_{b(c)}$ is the number of events correctly classified and $N^{tot}_{b(c)}$ is the total number of events.

Therefore, the test accuracy for different number of qubits and different number of layers was studied for all the QML models. For example, Figure 4.9(left) shows the test accuracy using the QAOA Embedding model for different number of qubits and layers. As expected, increasing the number of qubits (features) the accuracy improves as well as increasing the number of layers. Note that there exits a critical number of layers where the accuracy cannot improve more. On average, the critical number of layers for all the models starts at 7. Therefore, this value is fixed for future analysis.

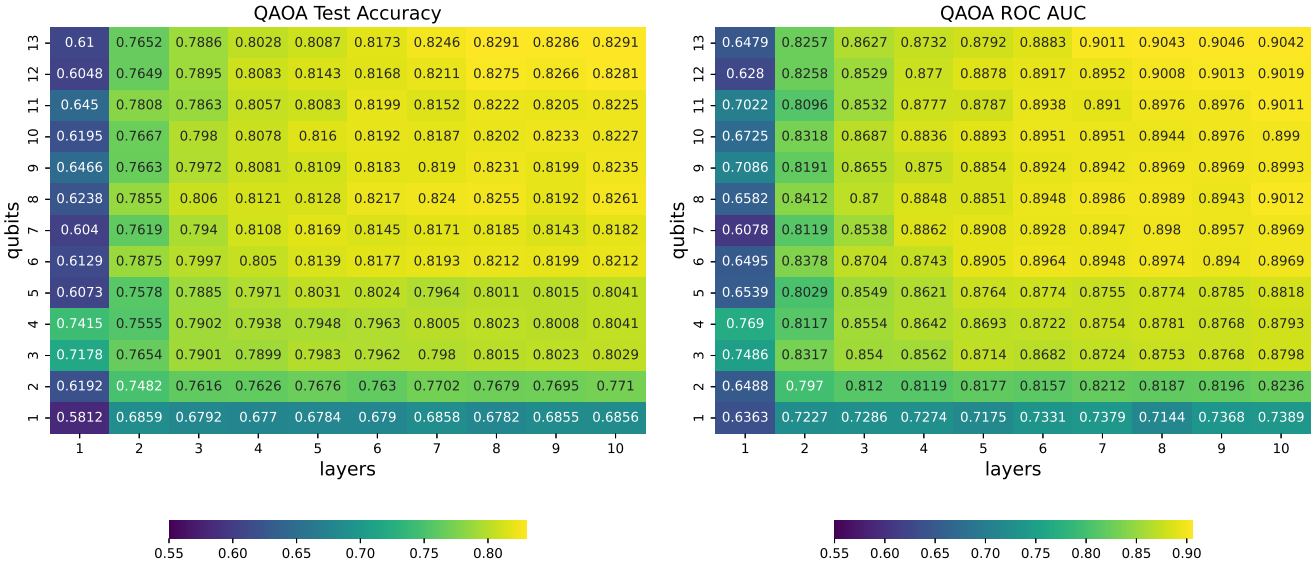


Figure 4.9: Test accuracy (left) and its corresponding ROC area under the curve (right) as a function of number of qubits and number of layers using the QAOA Embedding model.

4.8 QML Performance

There are different scores to measure the performance of ML algorithms. The most used is the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve, where a classifier performs better than other if its AUC is larger. The ROC curve is obtained by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings over the prediction distributions [77] as can be observed in Figure 4.10. The TPR, also known as sensitivity, and the FPR are defined as:

$$TPR = \frac{TP}{TP+FN} \qquad FPR = \frac{FP}{FP+TN} \tag{4.5}$$

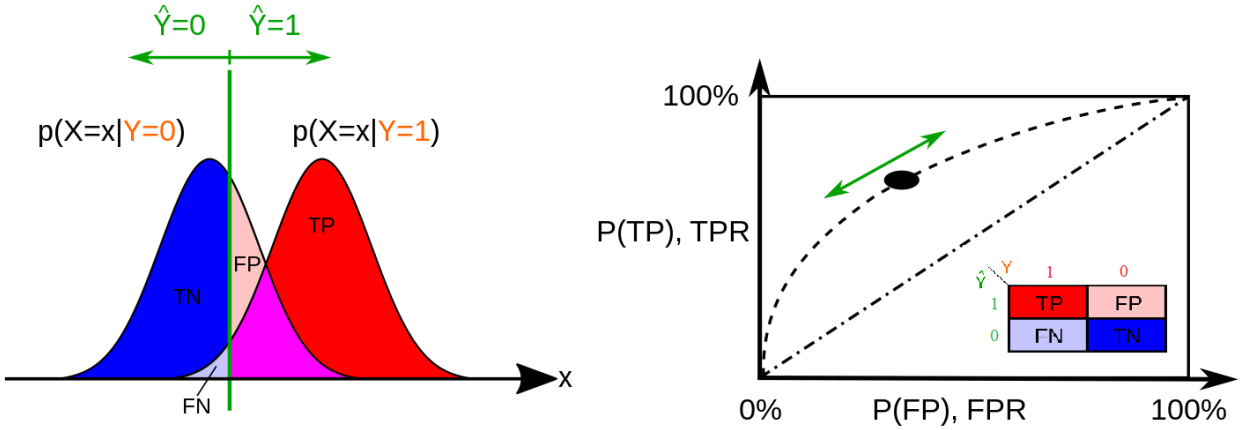
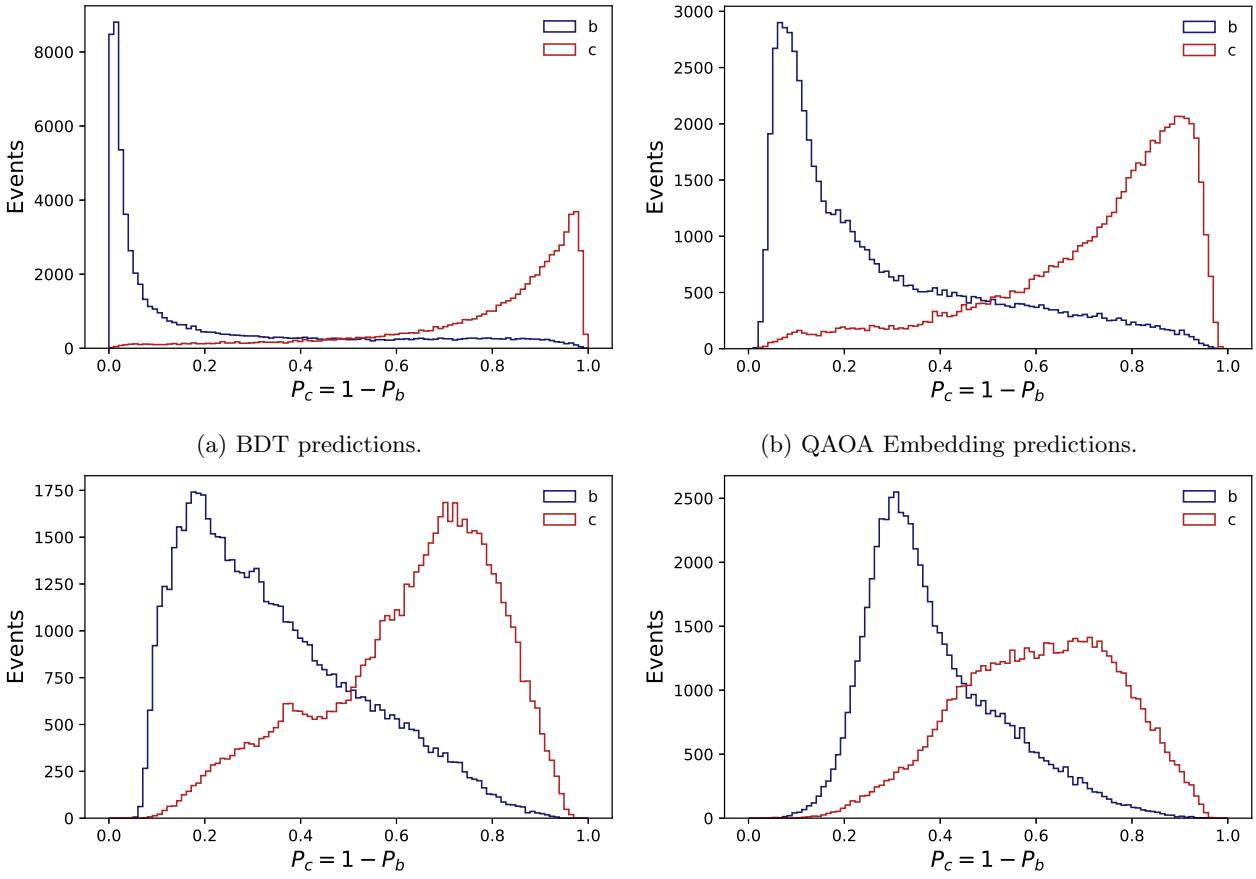


Figure 4.10: ROC curve main parameter.

Figure 4.11 shows the b- and c-quark jet prediction distributions for the different QML models (using 7 layers) and the classical model using all the available features. As can be observed in Figure 4.11.a, the BDT is the model that most separate the prediction distributions of b- from c-quark jet, followed by the distributions produced by the QAOA Embedding model. For the other models, the separation between b- and c-quark jet distributions can be observed but with less effectiveness than the classical method. More deep studies are needed in order to understand where these differences come from.



(a) BDT predictions.

(b) QAOA Embedding predictions.

(c) Angle Embedding predictions.

(d) Amplitude Embedding prediction.

Figure 4.11: Distributions of the output of the different classification algorithms.

Figure 4.12 shows the ROC curves for all the QML models and for the BDT model for comparison. As expected, the BDT model is the one with the largest AUC, followed for the QAOA embedding model. The results of the AUC make sense when looking to the prediction distributions in Figure 4.11.

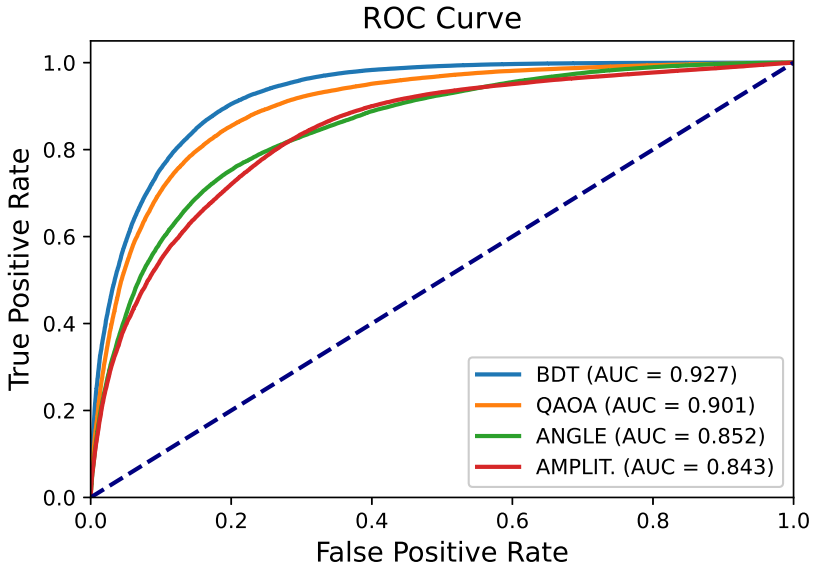
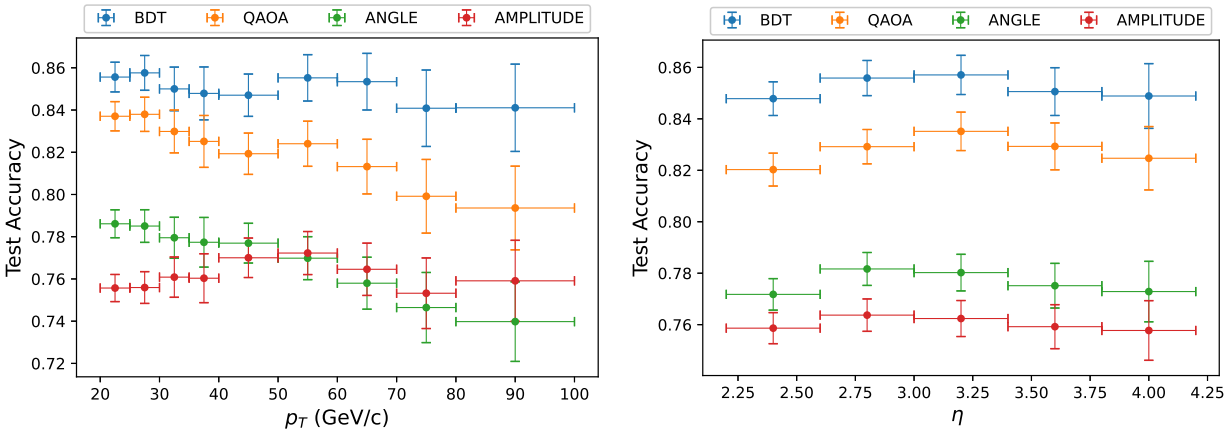


Figure 4.12: ROC curves for different classification models.

Besides, it is important to study the test accuracy of these models under physical quantities such as the jet transverse momentum p_T and the jet pseudorapidity η. In Figure 4.13, the test accuracy is evaluated in several p_T bins in the region of $[20, 100]$ GeV/c and η bins in the $[2.2, 4.2]$ region. As expected, the classical method classifies better than the QML models as function of p_T and η as can be observed in Figure 4.13.a and Figure 4.13.b, respectively. At this point, it is evident that the QAOA Embedding is the QML model with the best results, therefore, this model will be used to calculate the test accuracy for the individual b- and the c-quark jet classification as function of p_T and η.

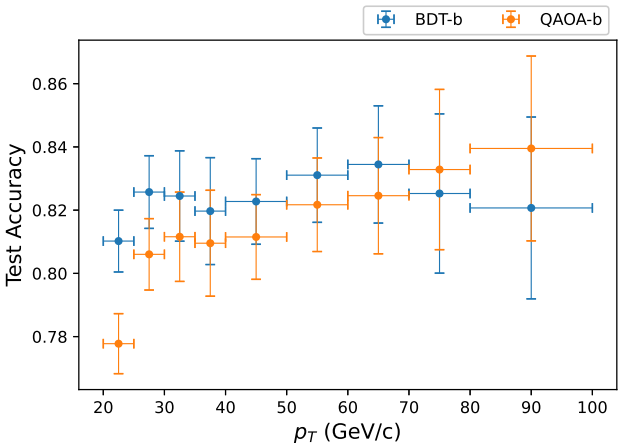


(a) Test accuracy as a function of jet p_T.

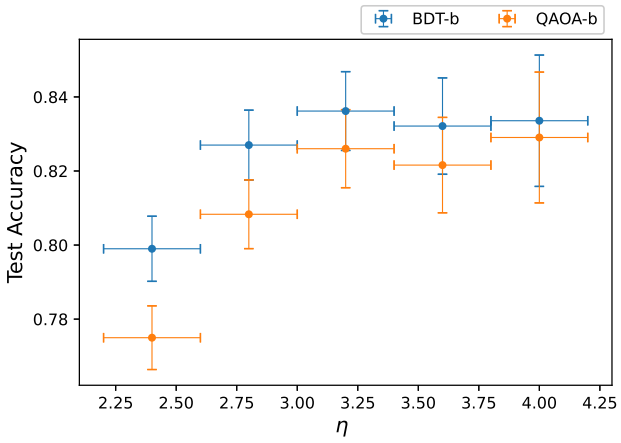
(b) Test accuracy as a function of jet η.

Figure 4.13: Test accuracy as function of jet p_T and jet η.

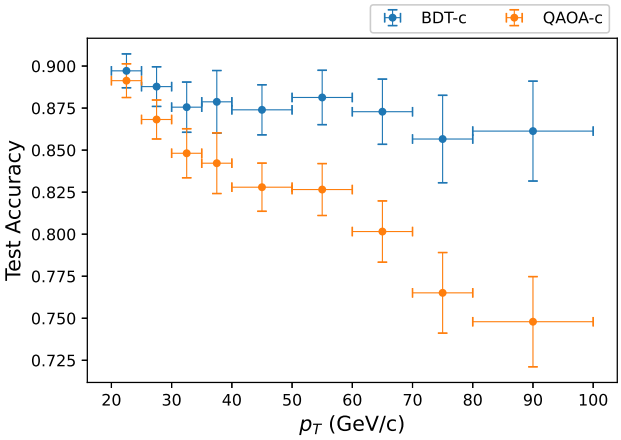
As can be appreciated in Figure 4.14.a and Figure 4.14.b, the b-quark jet test accuracy for the QAOA model and the BDT are very similar between each other, and in some high p_T regions the QAOA accuracy is slightly larger than for the BDT accuracy. Note also that the b-quark jet test accuracy increases as the jet p_T increases. Unfortunately, for the c- jet accuracy the BDT method is considerably better than the QAOA. Figure 4.14.c and Figure 4.14.d show that the BDT test accuracy is larger than the QAOA over all the p_T and η spectrum. This effect is more remarkable specially when increasing the jet p_T, where a huge drop in the accuracy is observed for the QAOA method. Unlike to the b-jet classification, the accuracy for the c-quark jet reduces increasing the jet p_T or the jet η.



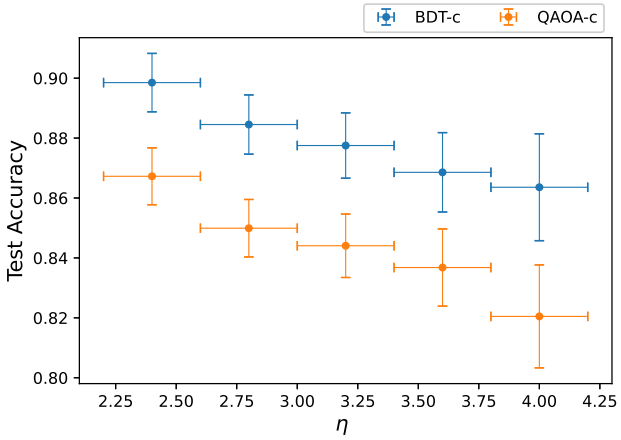
(a) Test accuracy as a function of b-quark jet p_T.



(b) Test accuracy as a function of b-quark jet η.



(c) Test accuracy as a function of c-quark jet p_T.



(d) Test accuracy as a function of c-quark jet η.

Figure 4.14: Test accuracy as function of b- and c-quark jet p_T and η.

4.9 Results

Table 4.2 summarizes the test accuracies and AUCs for the different QML models using all the SV features. Accuracies, ROCs and AUCs suggest that among all the QML algorithms, the QAOA Embedding model is the one with the best classification performance. On the other hand, Angle and Amplitude models performs very similar, but the advantage of the Amplitude model is the low number of qubits necessary to embed the SV features in the quantum circuit, followed by a faster training. However, the classical method (BDT) still performs better than the QML models.

Model	features	qubits	layers	accuracy	AUC
BDT	13	-	-	0.852	0.927
QAOA	13	13	7	0.828	0.901
Angle	13	13	7	0.777	0.852
Amplitude	13	4	7	0.761	0.843

Table 4.2: Result from different ML models.

This work presents for the first time, and represents the starting point, for the jet flavor classification using this new tool. There are still different variables to explore in order to optimize more the QML results. Given the latest results, the BDT model and the QAOA Embedding model will be used for the Higgs search in the next chapter.

Chapter 5

Higgs Search

In this chapter, the best QML classification model (QAOA Embedding) and classical ML model (BDT) for jet flavor tagging are used for the Higgs Search. The Higgs decay channels that will be analyzed are the $H \to b\bar{b}$ and $H \to c\bar{c}$ by means of the Monte Carlo simulated data-sets for the signal and the irreducible background, assuming a negligible light jet contribution as was described in the section 2.7. The main purpose of this analysis is to compare the results for the Higgs search obtained between the classical ML and the QML classifier in terms of significance, tagging efficiency and mis-tag efficiency.

5.1 Event selection

The first task is to calculate the number of expected events N_{exp} for the $H \to b\bar{b}$ and $H \to c\bar{c}$ decay channels and for the $b\bar{b}$ and $c\bar{c}$ irreducible background. Therefore, the following workflow was applied to the available data-sets:

1. Obtain the number of events N_{events} after applying the following criteria on the jet transverse momentum p_T and pseudo-rapidity η:

   - $20 < p_T < 100$ GeV/c
   - $2.2 < \eta < 4.2$

   These requirements ensure that the jets are reconstructed in a phase space region where the reconstruction efficiency is uniform with respect to p_T and η [45] and to remove not desired events like light-quark jets.

2. Calculate the selected events N_{selected} for the $b\bar{b}$ and $c\bar{c}$ samples by re-weighting the events N_{events} because these sample were divided in 4 different intervals:

$$N_{\text{selected}} = \sum_i \omega^i \cdot N^i_{\text{events}} \tag{5.1}$$

   where w^i are the weights which are proportional to the cross sections related to the different intervals of transverse momentum exchanged in the hard interaction $\hat{p}_T$. Therefore, the sum extends over all the files of the data-set. Table 5.1 summarizes these weights obtained using PYTHIA. For the Higgs signals, $N_{\text{events}} = N_{\text{selected}}$.

$\hat{p}_T$ (GeV/c)	$\omega(b\bar{b})$	$\omega(c\bar{c})$
$[10, 15)$	1.0	1.0
$[15, 20)$	0.2497949295	0.2078621513
$[20, 50]$	0.1465910663	0.1203924034
> 50	0.0045055782	0.0034683555

Table 5.1: Production weights as function of the transverse momentum exchanged in the hard interaction.

3. Finally, calculate the N_{exp} as:

$$N_{\text{exp}} = \mathcal{L} \cdot \sigma \cdot \epsilon \tag{5.2}$$

where $\mathcal{L} = 6$ fb^{-1} is the Run-II integrated luminosity at $\sqrt{s} = 13$ TeV, σ is the production cross section of an event within the LHCb geometrical acceptance (computed using PYTHIA8) and ϵ is the efficiency estimated as

$$\epsilon = \frac{N_{\text{selected}}}{N_{\text{generated}}} \tag{5.3}$$

Following the procedure described before and the cross section σ and $N_{\text{generated}}$ values detailed in Table 5.2, the expected events N_{exp} for both Higgs decay channels and the irreducible backgrounds can be calculated.

Channel	**cross section**	$N_{\textbf{generated}}$	**efficiency (%)**	$N_{\textbf{exp}}$
$H \to b\bar{b}$	1.51 pb	2×10^5	12.953	1173
$H \to c\bar{c}$	94 fb	2×10^5	1.531	9
$b\bar{b}$	3.26 $\times10^9$ fb	2×10^6	1.214	237×10^6
$c\bar{c}$	3.26 $\times10^9$ fb	2×10^6	0.135	26 $\times10^6$

Table 5.2: Set of parameters used for the Higgs search.

Note how the expected events for the Higgs decay channels are very low compared with the irreducible background, specially for the $H \to c\bar{c}$ decay channel. That's why the observation of both decays are challenging. Then, the expected events are used in order to calculate the significance for the observation of both Higgs decay channels.

5.2 Higgs significance

A common particle physics experiment involves the search for new phenomena by observing a unique class of events in particle interactions that cannot be described by background hypotheses. This search can be simplified to that of a "counting experiment" where one identifies a class of events using well-defined criteria, counts up the total number of selected events and estimates the events that come from the various background processes. Then, to estimate the p-value (often also expressed in terms of level of statistical significance) of a given observation of signal events, assuming that the probability density for the signal follows a Poisson distribution [78], i.e. the formula for the statistical significance S in terms of standard deviation for the observation is given by:

$$S = \frac{N_{\text{signal}}}{\sqrt{N_{\text{signal}} + N_{\text{background}}}} \tag{5.4}$$

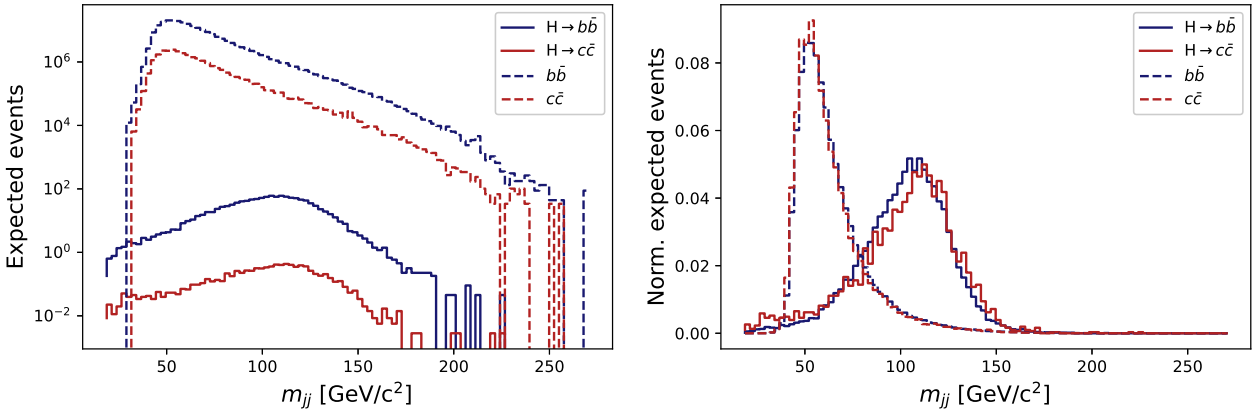
In particle physics, this observation could be represented as a narrow enhancement in a small invariant mass interval of signal events above an expected background events. Assuming that $(E_1, \mathbf{p}_1)$ and $(E_2, \mathbf{p}_2)$ are the four-momenta of the two jets, the invariant mass m_{jj} of di-jet events is calculated as:

$$m_{jj}^2 = (E_1 + E_2)^2 - \|\mathbf{p}_1 + \mathbf{p}_2\|^2 \tag{5.5}$$

Then the significance for the Higgs observation in the $H \to b\bar{b}$ and $H \to c\bar{c}$ decay channels will be done using the calculated expected events N_{exp}, detailed in Table 5.2 and the di-jet invariant mass m_{jj} range produced by the di-jet four-momenta.

5.2.1 $H \to b\bar{b}$ and $H \to c\bar{c}$ search

Figure 5.1 shows the distribution (left) and the normalized (right) distribution of the di-jet invariant mass produced by the $H \to b\bar{b}$ and $H \to c\bar{c}$ decay channels and for the $b\bar{b}$ and $c\bar{c}$ irreducible background. In the normalized distribution is clear the peaks for both Higgs signals are around the 125 GeV/c^2 measured by the CMS [4] and ATLAS experiments [5]. On the other hand, the un-normalized distribution clearly shows why both measurements are very challenging. It can be appreciated that the overwhelming background is several orders of magnitude larger than the Higgs signals, specially for the $H \to c\bar{c}$ decay channel because of its low branching ratio.



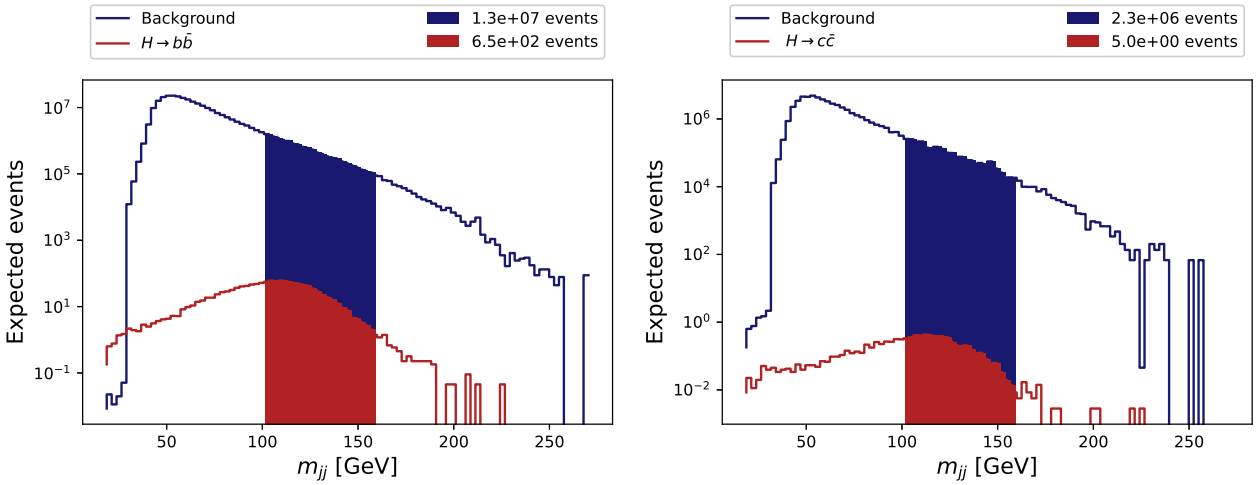
(a) Di-jet invariant mass distributions used for the Higgs search.

(b) Normalized to unit area invariant di-jet mass distribution used for the Higgs search.

Figure 5.1: Di-jet invariant mass distribution used for the Higgs search in the $H \to b\bar{b}$ and $H \to c\bar{c}$ decay channels.

In order to calculate the significance for the observation of both Higgs decays channels, the followed steps were done:

1. Select the invariant mass range where the significance is the maximum for both decay channels. This is done by an optimization process and the results of this calculation correspond to the invariant mass region between [103-160] GeV/c^2, see the shaded region of Figure 5.2.



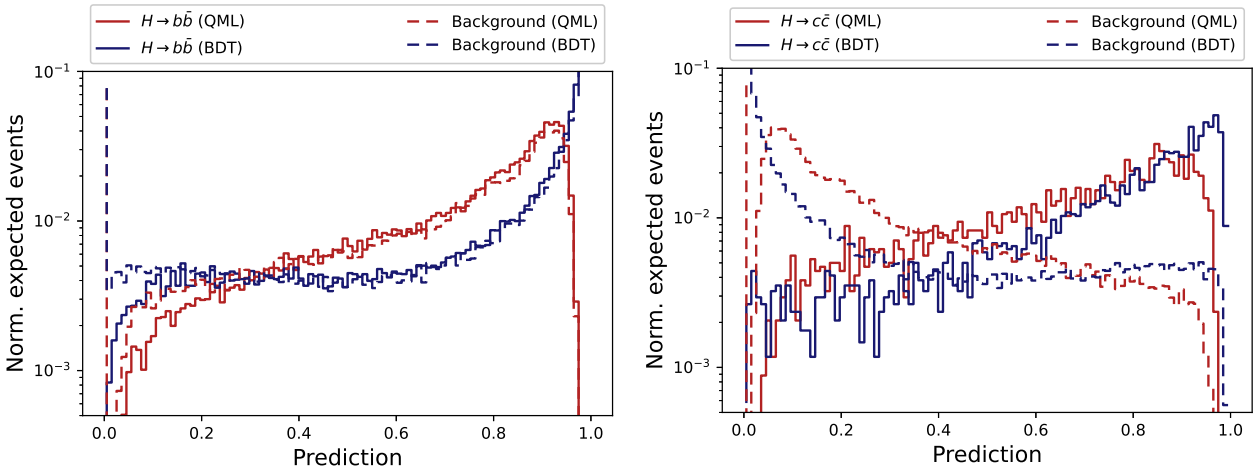
(a) Di-jet invariant mass distribution for the $H \to b\bar{b}$ decay channel and its background.

(b) Di-jet invariant mass distribution for the $H \to b\bar{b}$ decay channel and its background

Figure 5.2: Di-jet invariant mass distribution for the Higgs search in the $H \to b\bar{b}$ and $H \to c\bar{c}$ decays channels and its corresponding background. The shadow regions correspond to the invariant mass region where the significance is the maximum.

For the optimization of the significance of the $H \to b\bar{b}$ observation, the background corresponds to the sum of the expected events of the $b\bar{b}$ and $c\bar{c}$ backgrounds and the H$\to$ $c\bar{c}$ signal. Similarly, for the optimization of the significance of the $H \to c\bar{c}$ observation, the background corresponds to the sum of the expected events of the $b\bar{b}$ and $c\bar{c}$ irreducible background and the $H \to b\bar{b}$ signal. The significance inside the shadow region for the H$\to$ $b\bar{b}$ and H$\to$ $c\bar{c}$ are 0.1845 and 0.0033, respectively. Although, these values could improve by applying ML classifier techniques for jet flavour tagging.

2. At this point, the QML classifier and the classical classifier can be applied to the expected events inside the optimized invariant mass range in order to separate mostly the $c\bar{c}$ background for the $H \to b\bar{b}$ search, and to separate the $b\bar{b}$ background for the $H \to c\bar{c}$ search. Figure 5.3 shows the probability distribution for the signal and the background obtained using the QML and the classical ML models. Note, the probability to be a signal event is 1 and the probability to be a background event is 0.



(a) Normalized to unit area prediction distribution for the $H \to b\bar{b}$ signal and its background.

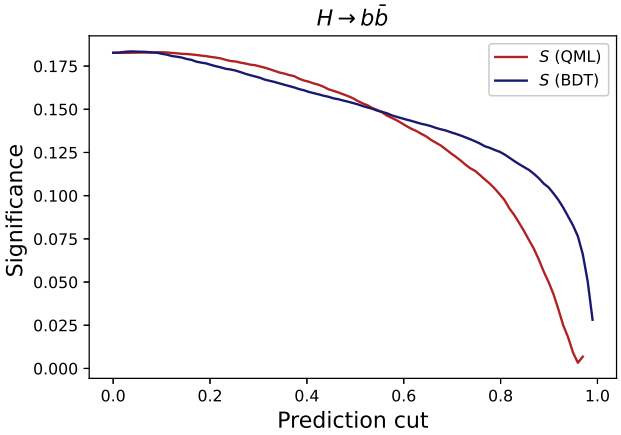
(b) Normalized to unit area prediction distribution for the $H \to c\bar{c}$ signal and its background.

Figure 5.3: Prediction distribution for the $H \to b\bar{b}$ and $H \to c\bar{c}$ signals and their respective backgrounds. The prediction to have a signal event is 1 and the probability to have a background event is 0.

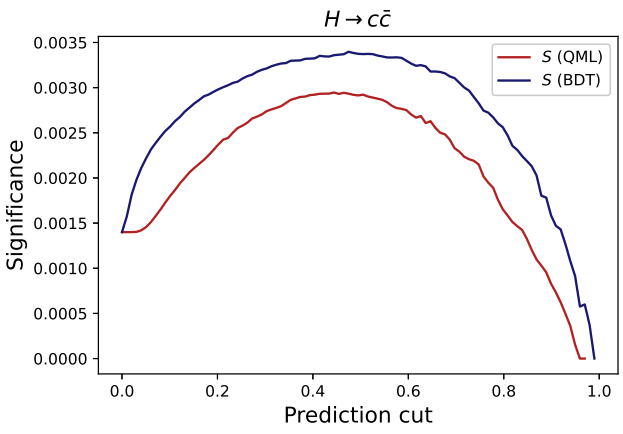
The quality of the separation relies on the classifier performance for the b- and the c-quark jet classification. As can be appreciated in the Figure 5.3.a, the background is not well separated from the $H \to b\bar{b}$ signal for the larger number of $b\bar{b}$ events than the $c\bar{c}$ events. On the other hand, for the previous reasons the $H \to b\bar{b}$ signal is better separated from the background, see Figure 5.3.b. This separation will be useful to cut the signal and background events below a certain probability limit in order to remove background events.

3. Finally, the significance can be improved by applying a probability or prediction cut x over the prediction distribution. This is done by removing all the di-jet events below this cut x either from the signal and the background. Therefore, the significance improvement relies in a good separation between the signal and the background events, where the cut x is mainly applied to suppress the background events because they ideally are located close to 0.

   Figure 5.4 shows the results for the significance as function of the prediction cut x for the $H \to b\bar{b}$ (left) and the $H \to c\bar{c}$ (right) observation. Unfortunately, the significance for the $H \to b\bar{b}$ signal cannot be improved under any prediction cut. On the other hand, the significance for the $H \to c\bar{c}$ signal was improved more than a factor of 2, where the maximum improvement in the significance is located at $x = 0.45$. The drop in the significance after reaching the maximum is because the background becomes huge compared with the signal, even if most of the background events were cut. Then, it is important to quantify the classification efficiency of the QML and the classical ML techniques at the point of maximum significance $x = 0.45$.



(a) Prediction cut scan for the $H \to b\bar{b}$ significance.



(b) Prediction cut scan for the $H \to c\bar{c}$ significance.

Figure 5.4: $H \to b\bar{b}$ and $H \to c\bar{c}$ significance as function of the prediction cut.

5.3 Classifiers efficiency for the $H \to c\bar{c}$ search

Now it is important to quantify the classification efficiency of the QML and the classical ML techniques at the point of maximum significance with respect to some important physical quantifies like the p_T, η and m_{jj}. Due to the significance improvement is just observed for the $H \to c\bar{c}$ signal at $x = 0.45$, this cut will be used to calculate the actual efficiency ϵ and mis-tag ω for the QML and classical ML classifiers, defined as

$$\epsilon(c\text{-jet}) = \frac{N_c(P_c > x)}{N_c} \qquad \omega(b\text{-jet}) = \frac{N_b(P_c > x)}{N_b} \tag{5.6}$$

where N_c is the total number of c- events used for the measurement and $N_c(P_c > x)$ represents total number of c- events above the prediction cut x. Both quantities quantify the discrimination power of these algorithms, ranging from a random classifier ($\omega = 50\%$) to a perfect classifier ($\omega = 0\%$). Also, their corresponding statistical errors are:

$$\sigma_\epsilon = \sqrt{\frac{\epsilon(1+\epsilon)}{N_c}} \qquad \sigma_\omega = \sqrt{\frac{\omega(1+\omega)}{N_b}} \tag{5.7}$$

Figure 5.5 shows the efficiency and the mis-tag as function of the jet transverse momentum p_T (first row), the jet pseudorapidity η (second raw) and the di-jet invariant mass m_{jj} (third raw) using $x = 45$. The efficiency as function of the jet p_T and jet η exhibits similar behaviors compared to the results obtained on the c-jet test accuracy as function of the jet p_T and jet η in Chapter 4, see Figure 4.14.e and Figure 4.14.f, respectively. When the jet p_T increases the $\epsilon(c\text{-jet})$ decreases for both methods, but the efficiency for the BDT classifier is better than the QML classifier. Similarly, as the jet η increases the efficiency slightly decreases, being the classical method again the one with better efficiency. Besides, the QML method gives a $\omega(b\bar{b}\text{-jet})$ larger than the BDT, but it slightly decreases as the p_T or the η increases while for the BDT method it increases.

On the other hand, the di-jet efficiency $\epsilon(c\bar{c}\text{-jet})$ and the di-jet mis-tag $\omega(b\bar{b}\text{-jet})$ have been computed in 4 equidistan bins of the di-jet invariant mass m_{jj} from [34, 214] GeV/c^2, summarised in Table 5.3. As expected from the previous cases, the $\epsilon(c\bar{c}\text{-jet})$ and the $\omega(b\bar{b} - \text{jet})$ are slightly constant along the invariant mass m_{jj} range, but smoothly decreasing as the invariant mass increases. It is a positive outcome since any structures in the efficiencies would indicate a bias introduced by the classifiers in the mass spectrum. Besides, the $\epsilon(c\bar{c}\text{-jet})$ at the Higgs mass (125 GeV/c^2) is 69.77 ± 2.79 for the BDT model while 63.18 ± 3.24 for the QML model, with $\omega(b\bar{b}\text{-jet}) = 3.74 \pm 0.65$ for the BDT model and $\omega(b\bar{b}\text{-jet}) = 3.46 \pm 0.63$ for the QML model. Note, the mis-tag evaluated per pair of jets shows, on average, values around 15% lower than the mis-tag evaluated per single jet, as shown in

Figure 5.5 (bottom left). Finally, these results seem to be consistent between the classical ML method with the QML method due to the statistical uncertainties in most of the cases overlap between each other.

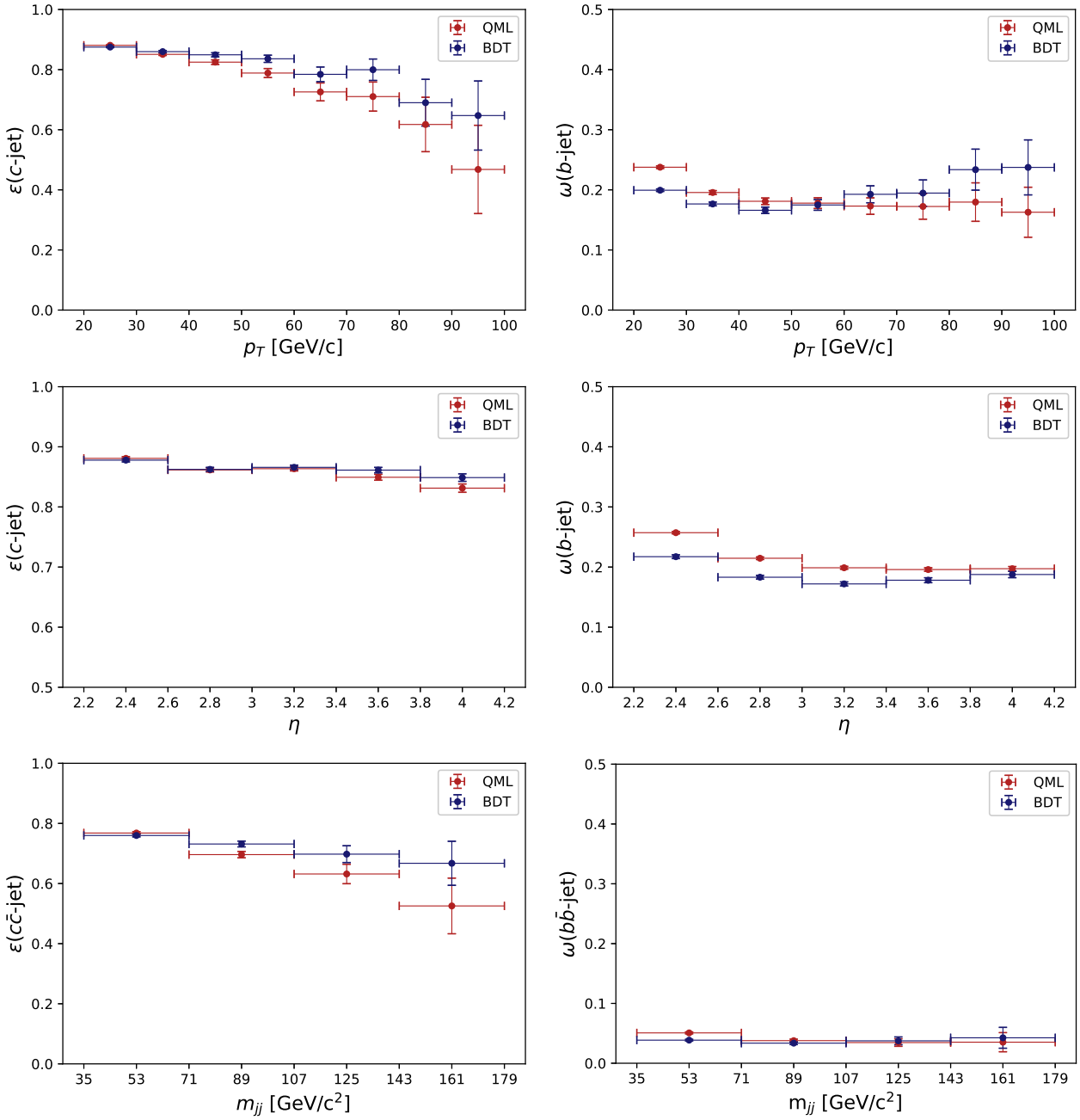


Figure 5.5: Classification efficiency $\epsilon(c\bar{c}\text{-jet})$ and the mis-tag efficiency $\omega(b\bar{b} - \text{jet})$ at $x > 0.45$ using the QML classifier and the ML classifier.

$\mathbf{m}_{jj}$ (GeV/c^2)	$\epsilon(c\bar{c}\text{-jet})$		$\omega(b\bar{b}\text{-jet})$	
	BDT (%)	**QML** (%)	**BDT** (%)	**QML** (%)
$[35, 71]$	75.99 ± 0.47	76.80 ± 0.45	3.85 ± 0.14	5.07 ± 0.16
$[71, 107]$	73.15 ± 0.96	69.62 ± 1.06	3.35 ± 0.24	3.77 ± 0.25
$[107, 143]$	69.77 ± 2.79	63.18 ± 3.24	3.74 ± 0.65	3.46 ± 0.63
$[143, 179]$	66.73 ± 7.32	52.54 ± 9.26	4.25 ± 1.74	3.52 ± 1.59

Table 5.3: Efficiency and mis-tag values in different m_{jj} bins, on the point of maximum efficiency $x = 0.45$ using the classical BDT and the QML classifier.

5.4 Higgs Significance Prediction

An interested question is when the $H \to b\bar{b}$ and $H \to c\bar{c}$ decay channels will have a relevant significance value to proclaim their observation. Taking into account the current significances, $S(H \to b\bar{b}) = 0.1845$ and $S(H \to c\bar{c}) = 0.0033$, is possible to obtain a raw estimation of their next significances for the incoming data taking campaigns. So, using Eqn. (5.2) and Eqn. (5.4) where the significance S_0 scales out with respect to the square root of the number of the expected events N_{exp}, and N_{exp} is proportional to its luminosity $\mathcal{L}_0$, a new significance can be estimated as:

$$S'(\mathcal{L}') \approx S_0 \sqrt{\frac{\mathcal{L}'}{\mathcal{L}_0}} \tag{5.8}$$

assuming $N_{\text{background}} >> N_{\text{signal}}$, a constant cross section σ and efficiency ϵ.

Therefore, using Eqn. (5.8) and the future expected LHCb luminosities, the future significances for both Higgs decays are estimated and summarized in Table 5.4. At the end of the HL-HLC campaign, the significance to have an observation of the $S(H \to b\bar{b})$ decay channel will be still below the value to proclaim this observation. Unfortunately, the expected $S(H \to c\bar{c})$ is still very low. That's why, more sophisticated studies and improvements are necessary. These will be discussed in the next chapter.

Campaign	**Luminosity**	$\sqrt{s}$	$S(H \to b\bar{b})$	$S(H \to c\bar{c})$
Run 2	6 fb^{-1}	13.0 TeV	0.1845	0.0033
Run 3-4	50 fb^{-1}	13.6 TeV	0.5326	0.0095
HL-HLC	300 fb^{-1}	14.0 TeV	1.3046	0.0233

Table 5.4: Expected significances for futures LHC data taking campaigns.

Chapter 6

Conclusions and future developments

6.1 Conclusions

In this thesis, a new proof-of-concept for the application of QML algorithms for the b- and c-quark jet identification on Run2 LHCb simulated data was developed. Several QML models have been tested and compared to the best classical ML algorithm doing this task in terms of classification performance. Then, the best QML model for the b- and c-quark jet classification was used for the $H \to b\bar{b}$ and $H \to c\bar{c}$ search on simulated data to estimate the significance for their observation. The results can be summarized as follows:

- As can be observed in Figure 4.14, at overall the best classical ML algorithm (BDT) has slightly better performance than the best QML algorithm (QAOA Embedding), with AUC = 0.927 for that classical model and AUC = 0.901 for that QML model. Besides, both QML and classical ML algorithm have similar b-quark jet test accuracies, while the classical algorithm gives a better c-quark jet test accuracy.
- Using the best QML model for the Higgs search in the $H \to b\bar{b}$ and $H \to c\bar{c}$ channels, Figure 5.4 shows that the classical ML model gives a better improvement for the $S(H \to c\bar{c})$, while any significance improvement was possible for the $S(H \to b\bar{b})$. At the point of maximum significance improvement for the $S(H \to c\bar{c})$ observation, the efficiency $\epsilon(c\text{-jet})$ for the classical model was slightly better than the QML model and, as expected, the QML has a slightly larger mis-tag efficiency than the classical ML model against different physical quantities like the jet p_T, jet η and di-jet m_{jj}, as can be observed in Figure 5.5. Also, the mis-tag performance gets enhanced when is evaluated on di-jet $b\bar{b}$ pairs.
- In spite of the QML is a field at its early stage, the results obtained in this thesis showed a performance comparable with the classical algorithm for the jet flavour tagging. Therefore, this good result is encouraging to continue with the research in this field looking for a better performance.

6.2 Future Developments

This search has many developments to explore either in the Software and Physics sides.

The QML developments are:

- **Improvement on machine learning techniques**. Despite different QML algorithms were used, future analysis should explore new QML models to exploit the relationship between the quantum circuit ansatz, such as qubits or the trainable parameters, and the structure of the input data-set.
- **Quantum hardware**. Nowadays several quantum hardware are available via quantum computing cloud solutions such as IBM Quantum Experience, Amazon Braket. These hardware devices are not always available, limiting the experience for short periods of user time. However, the availability of quantum resources is expected to grow quickly in the near future [79].
- **Quantum noise**. In this work, only noiseless quantum simulators were used but future studies about it affects the measurements is necessary. It requires a large number of evaluations for the estimation of the gradient and consequently a high computational power. However, software libraries and simulators are quickly improving over time, then the use of noise-aware simulators can be considered soon.

The developments on detector/analysis are:

- **Detector upgrades**. The new data tacking of Run 3 will take advantage of several LHCb detector upgrades such as the Vertex detector upgrades which enables a better SV reconstruction efficiency, a new Calorimeter to measure the jet energy with better resolution allowing to obtain a narrower invariant mass peak, etc.
- **New data**. The future data taking campaigns will run at higher center of mass energies boosting the cross section, the detector acceptance and increasing the luminosity $\mathcal{L}$. These new features obviously will boost the Higgs search.
- **New Higgs analysis techniques**. More sophisticated analysis can be used for the Higgs search, e.g., fitting the QCD background and the signal invariant mass shape instead of counting events in an invariant mass window. This technique (fit the QCD shape) has already been implemented for the $Z \to b\bar{b}$ measurement at the LHCb [80], see Figure 6.1.

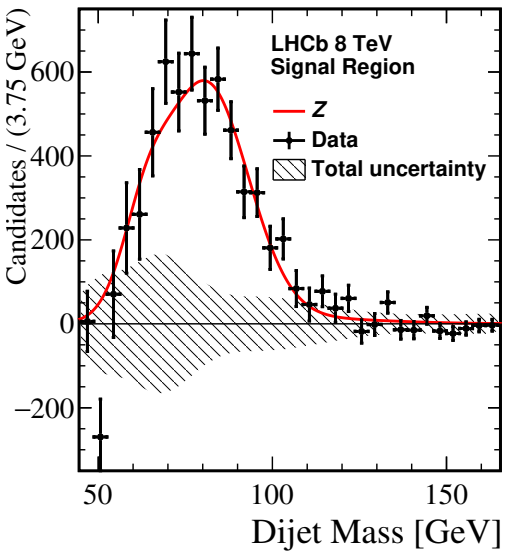


Figure 6.1: Background-subtracted distribution compared with the $Z \to b\bar{b}$ mass model [80].

- **Add other production channels to the analysis**. New Higgs production channels should be included in the analysis like the production channel associated with a W/Z boson.
- **Search for particles correlations inside jets**. The internal structure of jets is the result of non-perturbative QCD mechanisms occurring in the hadronization process. The ability of detecting correlations among particles inside the jet is unique to quantum algorithms and opens new possibilities to build up new methods of jets reconstruction and identification, opening the door for new physics measurements [81].

Bibliography

[1] Andrew Purcell. "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules". In: (Aug. 2012), p. 10. URL: https://cds.cern.ch/record/1473657.

[2] Giovanni Costa and Gianluigi Fogli. *Symmetries and group theory in particle physics: An introduction to space-time and internal symmetries*. Vol. 823. Springer Science & Business Media, 2012.

[3] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf. URL: https://doi.org/10.1093/ptep/ptaa104.

[4] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[5] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[6] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[7] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003.

[8] ATLAS Collaboration. "Evidence for the spin-0 nature of the Higgs boson using ATLAS data". In: *Physics Letters B* 726.1-3 (Oct. 2013), pp. 120–144. DOI: 10.1016/j.physletb.2013.08.026. URL: https://doi.org/10.1016%5C%2Fj.physletb.2013.08.026.

[9] CMS Collaboration. "Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs". In: *Physical Review Letters* 110.8 (Feb. 2013). DOI: 10.1103/physrevlett.110.081803. URL: https://doi.org/10.1103%5C%2Fphysrevlett.110.081803.

[10] ATLAS and CMS collaboration. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Physical Review Letters* 114.19 (May 2015). DOI: 10.1103/physrevlett.114.191803. URL: https://doi.org/10.1103%5C%2Fphysrevlett.114.191803.

[11] ATLAS Collaboration. "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment". In: *Physical Review D* 101.1 (Jan. 2020). DOI: 10.1103/physrevd.101.012002. URL: https://doi.org/10.1103%5C%2Fphysrevd.101.012002.

[12] ATLAS Collaboration. "Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector". In: *Physics Letters B* 784 (Sept. 2018), pp. 345–366. DOI: 10.1016/j.physletb.2018.07.050. URL: https://doi.org/10.1016%5C%2Fj.physletb.2018.07.050.

[13] CMS Collaboration. *Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2706103.

[14] CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Physics Letters B* 805 (June 2020), p. 135425. DOI: 10.1016/j.physletb.2020.135425. URL: https://doi.org/10.1016%5C%2Fj.physletb.2020.135425.

[15] CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Physics Letters B* 805 (June 2020), p. 135425. DOI: 10.1016/j.physletb.2020.135425. URL: https://doi.org/10.1016%5C%2Fj.physletb.2020.135425.

[16] ATLAS Collaboration. "Observation of H→b$\bar{\text{b}}$ decays and VH production with the ATLAS detector". In: *Physics Letters B* 786 (2018), pp. 59–86. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2018.09.013. URL: https://www.sciencedirect.com/science/article/pii/S0370269318307056.

[17] CMS Collaboration. "Observation of Higgs Boson Decay to Bottom Quarks". In: *Physical Review Letters* 121.12 (Sept. 2018). DOI: 10.1103/physrevlett.121.121801. URL: https://doi.org/10.1103%5C%2Fphysrevlett.121.121801.

[18] ATLAS Collaboration. "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physical Review D* 99.7 (Apr. 2019). DOI: 10.1103/physrevd.99.072001. URL: https://doi.org/10.1103%5C%2Fphysrevd.99.072001.

[19] CMS Collaboration. "Observation of the Higgs boson decay to a pair of leptons with the CMS detector". In: *Physics Letters B* 779 (Apr. 2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. URL: https://doi.org/10.1016%5C%2Fj.physletb.2018.02.004.

[20] ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Physics Letters B* 784 (Sept. 2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035. URL: https://doi.org/10.1016%5C%2Fj.physletb.2018.07.035.

[21] CMS Collaboration. "Observation of $t\bar{t}H$ Production". In: *Physical Review Letters* 120.23 (June 2018). DOI: 10.1103/physrevlett.120.231801. URL: https://doi.org/10.1103%5C%2Fphysrevlett.120.231801.

[22] CMS Collaboration. "Evidence for Higgs boson decay to a pair of muons". In: *Journal of High Energy Physics* 2021.1 (Jan. 2021). DOI: 10.1007/jhep01(2021)148. URL: https://doi.org/10.1007%5C%2Fjhep01%5C%282021%5C%29148.

[23] ATLAS Collaboration. "Evidence for Higgs boson decays to a low-mass dilepton system and a photon in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physics Letters B* 819 (Aug. 2021), p. 136412. DOI: 10.1016/j.physletb.2021.136412. URL: https://doi.org/10.1016%5C%2Fj.physletb.2021.136412.

[24] ATLAS Collaboration. *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector*. Tech. rep. Geneva: CERN, June 2021. URL: https://cds.cern.ch/record/2771724.

[25] CMS Collaboration. "A search for the standard model Higgs boson decaying to charm quarks". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). DOI: 10.1007/jhep03(2020)131. URL: https://doi.org/10.1007%5C%2Fjhep03%5C%282020%5C%29131.

[26] E Perez and E Rizvi. "The quark and gluon structure of the proton". In: *Reports on Progress in Physics* 76.4 (Mar. 2013), p. 046201. DOI: 10.1088/0034-4885/76/4/046201. URL: https://doi.org/10.1088%5C%2F0034-4885%5C%2F76%5C%2F4%5C%2F046201.

[27] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[28] O. Zenaiev. "Charm production and QCD analysis at HERA and LHC". In: *The European Physical Journal C* 77.3 (Mar. 2017). DOI: 10.1140/epjc/s10052-017-4620-4. URL: https://doi.org/10.1140%5C%2Fepjc%5C%2Fs10052-017-4620-4.

[29] Kenneth G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445. URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[30] SK Kundu, T Sarkar, and M Maity. "Investigating the parton shower model in PYTHIA8 with pp collision data at $\sqrt{s} = 13$ TeV". In: *arXiv preprint arXiv:2102.13061* (2021).

[31] C. Peterson et al. "Scaling violations in inclusive e^+e^- annihilation spectra". In: *Phys. Rev. D* 27 (1 Jan. 1983), pp. 105–111. DOI: 10.1103/PhysRevD.27.105. URL: https://link.aps.org/doi/10.1103/PhysRevD.27.105.

[32] Alan D. Martin. "Proton structure, Partons, QCD, DGLAP and beyond". In: (2008). DOI: 10.48550/ARXIV.0802.0161. URL: https://arxiv.org/abs/0802.0161.

[33] Eric M. Metodiev. *The Fractal Lives of Jets*. 2020. URL: https://www.ericmetodiev.com/post/jetformation/.

[34] CERN. *The Large Hadron Collider*. 2022. URL: https://home.cern/science/accelerators/large-hadron-collider (visited on 07/01/2022).

[35] Ewa Lopienska. "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022". In: (Feb. 2022). General Photo. URL: http://cds.cern.ch/record/2800984.

[36] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002. URL: https://doi.org/10.1088/1748-0221/3/08/s08002.

[37] LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://cds.cern.ch/record/1129809.

[38] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: http://cds.cern.ch/record/782076.

[39] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088/1748-0221/3/08/s08005.

[40] F Follin and D Jacquet. "Implementation and experience with luminosity levelling with offset beam". In: (Oct. 2014). Comments: 5 pages, contribution to the ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders, CERN, Geneva, Switzerland, 18-22 Mar 2013, 5 p. DOI: 10.5170/CERN-2014-004.183. arXiv: 1410.3667. URL: http://cds.cern.ch/record/1955354.

[41] R Aaij et al. "Performance of the LHCb Vertex Locator". In: *Journal of Instrumentation* 9.09 (Sept. 2014), P09007–P09007. DOI: 10.1088/1748-0221/9/09/p09007. URL: https://doi.org/10.1088/1748-0221/9/09/p09007.

[42] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: https://www.sciencedirect.com/science/article/pii/0168900287908874.

[43] LHCb Collaboration. "LHCb detector performance". In: *International Journal of Modern Physics A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227.

[44] Marcin Kucharczyk, Piotr Morawski, and Mariusz Witek. *Primary Vertex Reconstruction at LHCb*. Tech. rep. Geneva: CERN, Sept. 2014. URL: https://cds.cern.ch/record/1756296.

[45] LHCb collaboration. "Identification of beauty and charm quark jets at LHCb". In: *Journal of Instrumentation* 10.06 (June 2015), P06013–P06013. DOI: 10.1088/1748-0221/10/06/p06013. URL: https://doi.org/10.1088/1748-0221/10/06/p06013.

[46] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[47] Roel Aaij et al. "Measurement of differential $b\bar{b}$- and $c\bar{c}$-dijet cross-sections in the forward region of pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 02 (2021), p. 023. DOI: 10.1007/JHEP02(2021)023. arXiv: 2010.09437 [hep-ex].

[48] Alessio Gianelle et al. "Quantum Machine Learning for b-jet identification". In: *arXiv preprint arXiv:2202.13943* (2022).

[49] A.M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011. URL: http://dx.doi.org/10.1088/1748-0221/13/05/P05011.

[50] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012. URL: http://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[51] ATLAS Collaboration. "Machine Learning Algorithms for b-Jet Tagging at the ATLAS Experiment". In: *J. Phys. Conf. Ser.* 1085.4 (2018), p. 042031. DOI: 10.1088/1742-6596/1085/4/042031. arXiv: 1711.08811 [hep-ex].

[52] Anders Ryd et al. "EvtGen: A Monte Carlo Generator for B-Physics". In: (May 2005).

[53] 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.

[54] Paul Benioff. "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines". In: *Journal of statistical physics* 22.5 (1980), pp. 563–591.

[55] Richard P Feynman. "Simulating physics with computers". In: *Feynman and computation*. CRC Press, 2018, pp. 133–153.

[56] P.W. Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 1994, pp. 124–134. DOI: 10.1109/SFCS.1994.365700.

[57] Lov K. Grover. "A Fast quantum mechanical algorithm for database search". In: (May 1996). arXiv: quant-ph/9605043.

[58] Isaac L Chuang, Neil Gershenfeld, and Mark Kubinec. "Experimental implementation of fast quantum searching". In: *Physical review letters* 80.15 (1998), p. 3408.

[59] Wikimedia Commons. *The Bloch sphere, a geometric representation of a two-level quantum system*. 2009. URL: https://commons.wikimedia.org/wiki/File:Bloch%5C_sphere.svg.

[60] J. S. Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1 (3 Nov. 1964), pp. 195–200. DOI: 10.1103/PhysicsPhysiqueFizika.1.195. URL: https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195.

[61] Michael A Nielsen and Isaac Chuang. *Quantum computation and quantum information*. 2002.

[62] Maria Schuld and Francesco Petruccione. *Supervised learning with quantum computers*. Vol. 17. Springer, 2018.

[63] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[64] Marcello Benedetti et al. "Parameterized quantum circuits as machine learning models". In: *Quantum Science and Technology* 4.4 (Nov. 2019), p. 043001. ISSN: 2058-9565. DOI: 10.1088/2058-9565/ab4eb5. URL: http://dx.doi.org/10.1088/2058-9565/ab4eb5.

[65] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. *A Quantum Approximate Optimization Algorithm*. 2014. arXiv: 1411.4028 [quant-ph].

[66] K. Mitarai et al. "Quantum circuit learning". In: *Physical Review A* 98.3 (Sept. 2018). ISSN: 2469-9934. DOI: 10.1103/physreva.98.032309. URL: http://dx.doi.org/10.1103/PhysRevA.98.032309.

[67] Maria Schuld et al. "Evaluating analytic gradients on quantum hardware". In: *Physical Review A* 99.3 (Mar. 2019). ISSN: 2469-9934. DOI: 10.1103/physreva.99.032331. URL: http://dx.doi.org/10.1103/PhysRevA.99.032331.

[68] Bertrand Michel and Anthony Nouy. *Learning with tree tensor networks: complexity estimates and model selection*. 2021. arXiv: 2007.01165 [math.ST].

[69] Haowen Deng et al. "Ensemble learning for the early prediction of neonatal jaundice with genetic features". In: *BMC Medical Informatics and Decision Making* 21 (Dec. 2021). DOI: 10.1186/s12911-021-01701-9.

[70] Ville Bergholm et al. *PennyLane: Automatic differentiation of hybrid quantum-classical computations*. 2020. arXiv: 1811.04968 [quant-ph].

[71] Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[72] Tyson Jones and Julien Gacon. *Efficient calculation of gradients in classical simulations of variational quantum algorithms*. 2020. arXiv: 2009.02823 [quant-ph].

[73] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: 1912.01703 [cs.LG].

[74] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[75] François Chollet et al. *Keras*. https://keras.io. 2015.

[76] Rudolf Frühwirth and Are Strandlie. *Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors*. Verlag der Österreichischen Akademie der Wissenschaften, 2021.

[77] James P. Egan, Gordon Z. Greenberg, and Arthur I. Schulman. "Operating Characteristics, Signal Detectability, and the Method of Free Response". In: *The Journal of the Acoustical Society of America* 33.8 (1961), pp. 993–1007. DOI: 10.1121/1.1908935. eprint: https://doi.org/10.1121/1.1908935. URL: https://doi.org/10.1121/1.1908935.

[78] Pekka K. Sinervo. *Signal Significance in Particle Physics*. 2002. DOI: 10.48550/ARXIV.HEP-EX/0208005. URL: https://arxiv.org/abs/hep-ex/0208005.

[79] IBM. *IBM's roadmap for building an open quantum software ecosystem*. 2020. URL: https://research.ibm.com/blog/quantum-development-roadmap.

[80] LHCb Collaboration. "First observation of forward $Z \to b\bar{b}$ production in pp collisions at $\sqrt{s} =$ 8 TeV." In: *Physics Letters B* 776 (Jan. 2018), pp. 430–439. DOI: 10.1016/j.physletb.2017.11.066. arXiv: 1709.03458 [hep-ex].

[81] Timo Felser et al. *Quantum-inspired Machine Learning on high-energy physics data*. 2021. arXiv: 2004.13747 [stat.ML].