PROCEEDINGS OF SCIENCE

Quantum Machine Learning for b-jet charge identification

Davide Zuliani[a,b,1,*]

[a] *Università degli Studi di Padova, Padova, Italy*

[b] *INFN Sezione di Padova, Padova, Italy*

E-mail: davide.zuliani@cern.ch

Machine Learning algorithms are playing a fundamental role in solving High Energy Physics tasks. In particular, the classification of hadronic jets at the Large Hadron Collider is suited for such types of algorithms, and despite the great effort that has been put in place to tackle such a classification task, there is room for improvement. In this context, Quantum Machine Learning is a new methodology that takes advantage of the intrinsic properties of quantum computation (e.g. entanglement between qubits) to possibly improve the performance of a classification task. In this contribution, a new study of Quantum Machine Learning applied to jet identification is presented. Namely, a Variational Quantum Classifier is trained and evaluated on fully simulated data of the LHCb experiment, in order to identify jets containing a hadron formed by a b or $\bar{b}$ quark at the moment of production. The jet identification performance of the quantum classifier is compared with a Deep Neural Network using the same input features.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

[1] on behalf of the LHCb Collaboration
[*] Speaker

PoS(ICHEP2022)231

1. Introduction

The physics at the Large Hadron Collider (LHC) heavily relies on the reconstruction and identification of jets: jets are streams of particles that result from the fragmentation and hadronization of quarks, as described by Quantum Chromodynamics (QCD). Several analyses are based on the capability of reconstructing and identifying jets, in order to search for New Physics processes: the search for the Higgs decay into a pair of charm quarks and the $b\bar{b}$ asymmetry [1] are two valid examples where the jet reconstruction and identification performance play a relevant role. The ability to collect vast amounts of data and information has led to new algorithms to reconstruct and identify jets. These algorithms are based on machine learning (ML) tools, which have shown a great improvement with respect to classical non-ML tools [2].
In the last years, an increasing interest has been shown for Quantum Computing (QC) applications in High-Energy Physics (HEP): QC may lead to an advantage with respect to classical tools, both in terms of algorithm performance and algorithm complexity, by exploiting the intrinsic properties of quantum computation, e.g. entanglement between qubits. A promising approach is the use of Quantum Machine Learning (QML) techniques to tackle HEP problems, such as track reconstruction and jet identification [3]. In the following, a QML algorithm for b-jet charge identification at the LHCb experiment is presented [4].

2. The LHCb experiment and b-jet identification

The LHCb experiment [5] is a single arm spectrometer originally designed to study b- and c-hadrons. Despite the lower luminosity and the smaller angular acceptance with respect to ATLAS and CMS, in the last years it has proven itself to perform competitive measurements in the QCD and Electroweak sector.
Given its cleaner environment with low pile-up and excellent track and vertex reconstruction, jets originated from heavy-flavour quarks (namely b and c quarks) are identified with high efficiency. Particularly, it is possible to perform jet charge identification by identifying jets originated from b and $\bar{b}$ quarks. This can be achieved in two ways, as shown in Fig. 1:

- with an *exclusive* approach, which uses a specific physics process inside the jet to infer the quark charge. For the b-jet charge identification, a muon with transverse momentum $p_{\mathrm{T}} > 5$ GeV is found coming from the semileptonic decay of the b-quark, and its charge is used to infer the quark charge; this algorithm is called *muon tagging* and has been used at LHCb during Run 1 [6]. Although being an algorithm with high purity, its efficiency is limited by the b-quark branching ratio $\mathcal{B} \sim 10\%$;

- with an *inclusive* approach, which exploits all the information coming from the jet substructure to infer the quark charge. Being not related to a specific physics process, this approach is not limited by any branching ratio, but the increased amount of information makes it necessary to use ML tools.

The relevant figure of merit for b-jet charge identification is the tagging power $\epsilon_{\text{tag}} = \epsilon(1 - 2\omega)^2$, where ϵ is the tagging efficiency and ω is the *mistag* rate, the latter defined as the number of wrongly

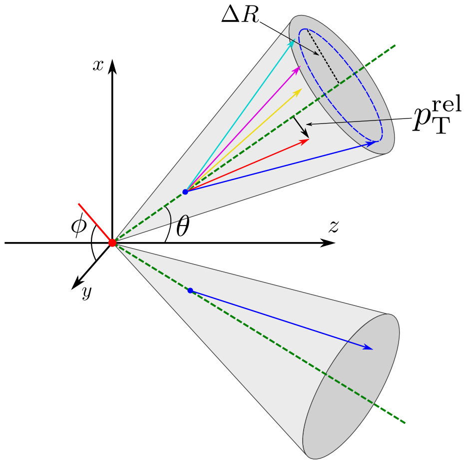

Figure 1: Schematic view of the different approaches for jet charge identification: *exclusive* approach (lower jet) and *inclusive* approach (upper jet) [4].

identified jets over the number of identified jets. The tagging power ϵ_{tag} gives the effective fraction of correctly identified jets, relevant for the physics analysis.

3. Dataset description

In the following, an inclusive approach has been used. The samples of $b\bar{b}$ dijets have been simulated using the official full LHCb simulation framework, from the proton-proton hard scattering up to the trigger lines to select the $b\bar{b}$ dijets signature. Almost 700000 jets have been simulated at a center-of-mass energy $\sqrt{s} = 13$ TeV, simulating Run 2 conditions. For each jet, 16 features of the jet substructure are taken as input: inside each jet, the muon, the kaon, the pion, the electron and the proton with the highest p_{T} are selected. In this way, five charged particles of different types are selected and, for each particle, three quantities are considered: the momentum of the particle relative to the jet axis ($p_{\mathrm{T}}^{\mathrm{rel}}$), the charge of the particle (q), and the distance measured in the space (η,ϕ) space between the particle and the jet axis (ΔR). In this way, the $5 \times 3 = 15$ observables are selected. Finally, the total jet charge Q is considered, defined as the weighted average of the charges of the particles inside the jet, with the particles $p_{\mathrm{T}}^{\mathrm{rel}}$ used as weights:

$$Q = \frac{\sum_i (p_{\mathrm{T}}^{\mathrm{rel}})_i q_i}{\sum_i (p_{\mathrm{T}}^{\mathrm{rel}})_i} \; . \tag{1}$$

If no particle type is found inside the jet, all relative features are set to 0. The dataset is divided into training+validation and testing sub-datasets.

4. QML for b-jet charge identification

The QML algorithm used to tackle this problem is based on a Variational Quantum Classifier (VQC). A VQC is a hybrid quantum-classical algorithm that is based on a Parametrized Quantum Circuit (PQC): classical data are embedded into a quantum state, which is processed by the PQC

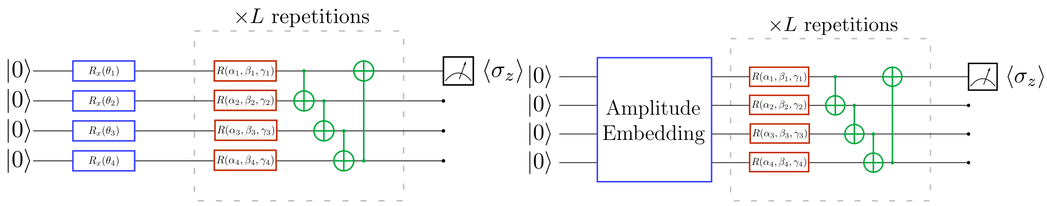

Figure 2: Schematic view of the embeddings and the circuits studied: Angle Embedding (left) and Amplitude Embedding (right) [4].

made by several variational layers, and this leads to a prediction. This structure is included in a classical training procedure that optimizes the PQC parameters to obtain the best classification performance.

Different embeddings have been considered: the Angle Embedding is based on a one-to-one correspondence between feature and qubit, where classical data are loaded in a quantum state using rotational gates; the Amplitude Embedding loads classical data into amplitudes of a given n-qubit state. A scheme of the two embeddings used is shown in Fig. 2. Two sets of features are considered: a *muon* set of features, which uses as input the features relative to the muon and the total jet charge Q, and the *complete* set of features, using all 16 input features. Results from the QML algorithms have been compared with the *muon tagging* approach used at LHCb and to a state-of-the-art Deep Neural Network (DNN), which uses the same input variables of the QML algorithms. Circuits have been simulated without noise contribution using the Pennylane [7] library.

5. Results

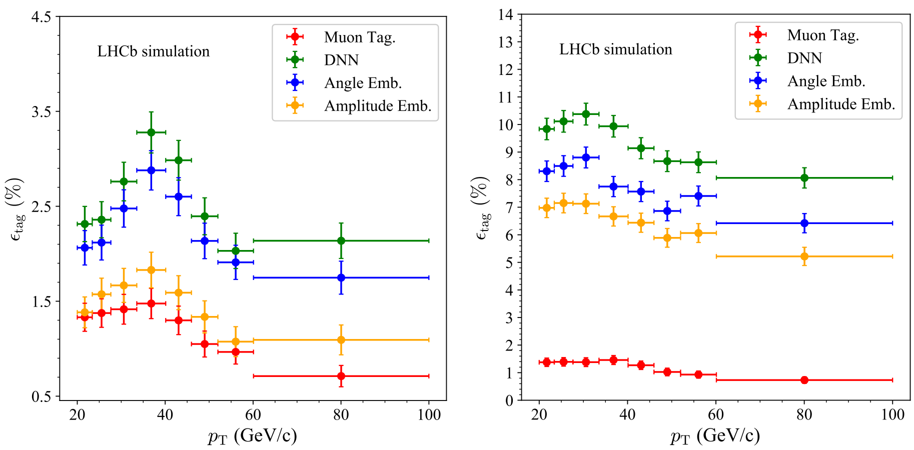

Figure 3: Tagging power ϵ_{tag} as a function of the jet p_{T} for the *muon* set of features (left) and for the *complete* one (right) [4].

Results for the tagging power ϵ_{tag} as a function of the jet p_{T} are shown in Fig. 3, for the *muon* (*complete*) set of features on the left (right): the Angle Embedding circuit shows comparable results with respect to the DNN, particularly for the *muon* set of features; for the *complete* set of features DNN still performs slightly better than Angle Embedding, therefore there is room for

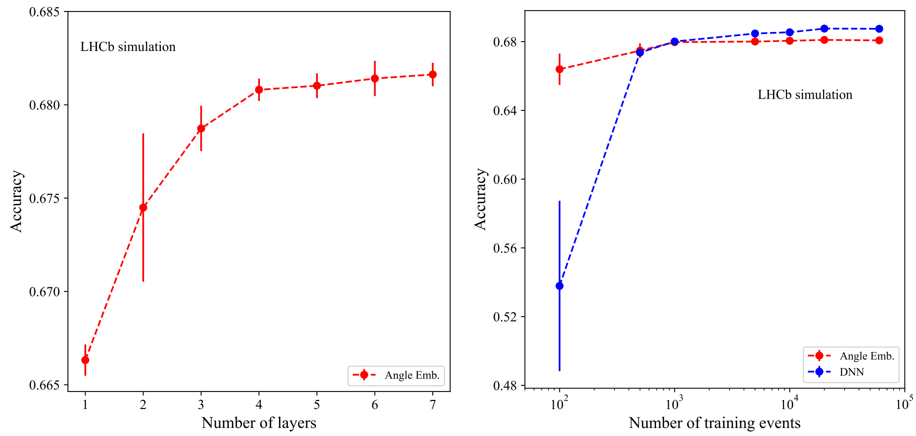

Figure 4: Testing accuracy as a function of the number of layers (left), showing saturation after four variational layers, and on the number of training events (right), showing for low number of training events that the Angle Embedding (red) performs better than the DNN (blue) [4].

improvement. The Angle Embedding performs better than the Amplitude Embedding in both cases, and all algorithms outperform the standard *muon tagging*, which is expected given that the latter is an exclusive algorithm. Further studies have been done to better understand and characterize the algorithm. In particular, performance for the Angle Embedding circuit using the *muon* set of features has been assessed as a function of the number of variational layers and as function of the number of training events. Results are shown in Fig. 4: the testing accuracy seems to saturate after four variational layers, which might be a good indicator to achieve an optimal trade-off between accuracy and circuit complexity; for lower number of training events, the Angle Embedding structure performs better than the classical DNN, while reaching the same performance for higher number of training events.
Finally, the noise contribution has also been taken into account: all the circuits are simulated using noiseless simulation, therefore, it is fundamental to assess performance including noise contributions coming from the interaction between the circuit and the environment. Noise models coming from different IBM backends have been simulated using the *pennylane-qiskit* [8] library. The simpler Angle Embedding structure with four qubits is studied with a smaller number of training events and just three variational layers for computational constraints. The testing accuracy as a function of the training epochs is shown in Fig. 5: all the IBM simulated backends reach the same performance as the noiseless circuit within the statistical error bands, showing that the chosen circuit is rather robust against noise contributions.

6. Conclusions

A recent and first application of a QML technique applied at the LHCb experiment has been presented [4]. In this work, a Variational Quantum Classifier has been used to classify jets originated from b- and $\bar{b}$-quarks, simulated with the official LHCb simulation framework. Results for the tagging power have been compared with a DNN and the *muon tagging* approach used so far at LHCb: the QML algorithms have performance close to the DNN classifier, particularly for the

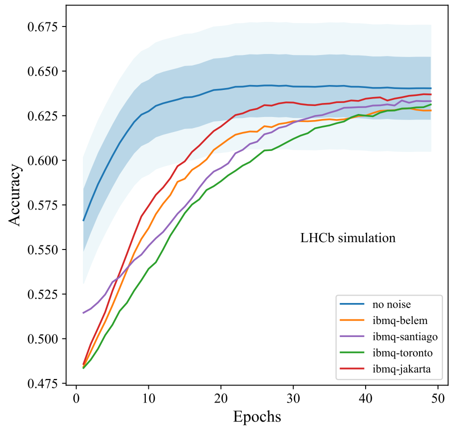

Figure 5: Testing accuracy as a function of the training epochs. Results are shown for the noiseless simulation (blue line with error bands) and for several simulated IBM backends, showing that the noisy simulations have the same performance as the noiseless one [4].

muon set of features. Performance for the QML algorithms as a function on number of variational layers and number of training events has been studied, showing performance saturation after some variational layers and better performance for lower number of training events with respect to DNN. Finally, the noise contribution has been studied, showing that simple circuit structures are rather robust against noise contributions, reaching the same performance as the noiseless simulation. This study might lead to future understanding and improvement in the application of QML to real HEP problems.

References

[1] Gauld, R., Haisch, U. & Pecjak, B.D., Asymmetric heavy-quark hadroproduction at LHCb: predictions and applications. J. High Energ. Phys. 2019, 166 (2019)

[2] D. Guest, K. Cranmer and D. Whiteson, Deep learning and its application to LHC physics, Ann. Rev. Nucl. Part. Sci. 68 (2018) 161

[3] Delgado, A. *et al.*, Quantum Computing for Data Analysis in High-Energy Physics, 2022 Snowmass Summer Study, arXiv::2203.08805

[4] Gianelle, A., Koppenburg, P., Lucchesi, D. *et al.*, Quantum Machine Learning for b-jet charge identification. J. High Energ. Phys. 2022, 14 (2022)

[5] A. A. Alves Jr. *et al.*, *The LHCb Detector at the LHC*, JINST 3 (2008) S08005

[6] LHCb collaboration, First measurement of the charge asymmetry in beauty-quark pair production, Phys. Rev. Lett. 113 (2014) 082003

[7] V. Bergholm *et al.*, Pennylane: automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968

[8] M. Treinish *et al.*, Qiskit/qiskit: Qiskit 0.37.0, Zenodo (2022)