OPEN ACCESS
IOP Publishing
Journal of Physics G: Nuclear and Particle Physics
J. Phys. G: Nucl. Part. Phys. **48** (2021) 125003 (12pp)
https://doi.org/10.1088/1361-6471/ac1391

Application of quantum machine learning using the quantum variational classifier method to high energy physics analysis at the LHC on IBM quantum computer simulator and hardware with 10 qubits

Sau Lan Wu[1,*], Jay Chan[1], Wen Guan[1], Shaojun Sun[1], Alex Wang[1], Chen Zhou[1], Miron Livny[2], Federico Carminati[3], Alberto Di Meglio[3], Andy C Y Li[4], Joseph Lykken[4], Panagiotis Spentzouris[4], Samuel Yen-Chi Chen[5], Shinjae Yoo[5] and Tzu-Chieh Wei[6]

[1] Department of Physics, University of Wisconsin, Madison WI, United States of America
[2] Department of Computer Sciences, University of Wisconsin, Madison WI, United States of America
[3] CERN Quantum Technology Initiative, IT Department, CERN, Geneva, Switzerland
[4] Quantum Institute, Fermi National Accelerator Laboratory, Batavia, IL, United States of America
[5] Computational Science Initiative, Brookhaven National Laboratory, Upton, NY, United States of America
[6] C. N. Yang Institute for Theoretical Physics, State University of New York, Stony Brook, NY, United States of America

E-mail: Sau.Lan.Wu@cern.ch

Received 28 April 2021, revised 17 June 2021
Accepted for publication 12 July 2021
Published 26 October 2021

Abstract

One of the major objectives of the experimental programs at the Large Hadron Collider (LHC) is the discovery of new physics. This requires the identification of rare signals in immense backgrounds. Using machine learning algorithms greatly enhances our ability to achieve this objective. With the progress of quantum technologies, quantum machine learning could become a powerful tool

*Author to whom any correspondence should be addressed.

for data analysis in high energy physics. In this study, using IBM gate-model quantum computing systems, we employ the quantum variational classifier method in two recent LHC flagship physics analyses: $t\bar{t}H$ (Higgs boson production in association with a top quark pair, probing the Higgs boson couplings to the top quark) and $H \to \mu^+\mu^-$ (Higgs boson decays to two muons, probing the Higgs boson couplings to second-generation fermions). We have obtained early results with 10 qubits on the IBM quantum simulator and the IBM quantum hardware. With small training samples of 100 events on the quantum simulator, the quantum variational classifier method performs similarly to classical algorithms such as SVM (support vector machine) and BDT (boosted decision tree), which are often employed in LHC physics analyses. On the quantum hardware, the quantum variational classifier method has shown promising discrimination power, comparable to that on the quantum simulator. This study demonstrates that quantum machine learning has the ability to differentiate between signal and background in realistic physics datasets. We foresee the usage of quantum machine learning in future high-luminosity LHC physics analyses, including measurements of the Higgs boson self-couplings and searches for dark matter.

Keywords: quantum machine learning, high energy physics, quantum computer, LHC

(Some figures may appear in colour only in the online journal)

The discovery of the Higgs boson by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) in 2012 [1, 2] was a major milestone for high energy physics. Since then, LHC experiments have been using the Higgs boson as a tool to pursue the discovery of new physics. The discovery of new physics requires the identification of rare signals against immense backgrounds. Using machine learning greatly enhances our ability to achieve this objective.

The intersection between machine learning and quantum computing has been referred to as quantum machine learning, and can possibly offer a valuable alternative to classical machine learning by providing more efficient solutions [3]. In 2018, a quantum variational classifier method was experimentally implemented with a quantum circuit of 2 qubits on a superconducting processor and successfully tested on synthetic datasets [4]. This method provides 'tools for exploring the applications of noisy intermediate-scale quantum computers to machine learning' [4]. With the progress of quantum technologies, quantum machine learning could possibly become a powerful tool for data analysis on real-world datasets such as those seen in high energy physics.

In this study, we employ the quantum variational classifier method in a $t\bar{t}H$ ($H \to 2$ photons) physics analysis and an $H \to \mu^+\mu^-$ physics analysis, two recent flagship physics analyses at the LHC, using IBM gate-model quantum computers. Our goal is to explore and to demonstrate, in a proof of principle experiment, the potential of quantum computers can be a new computational paradigm for big data analysis in high energy physics. An earlier study in a ggH ($H \to 2$ photons) physics analysis using D-wave quantum annealers was performed by Mott *et al* [5].

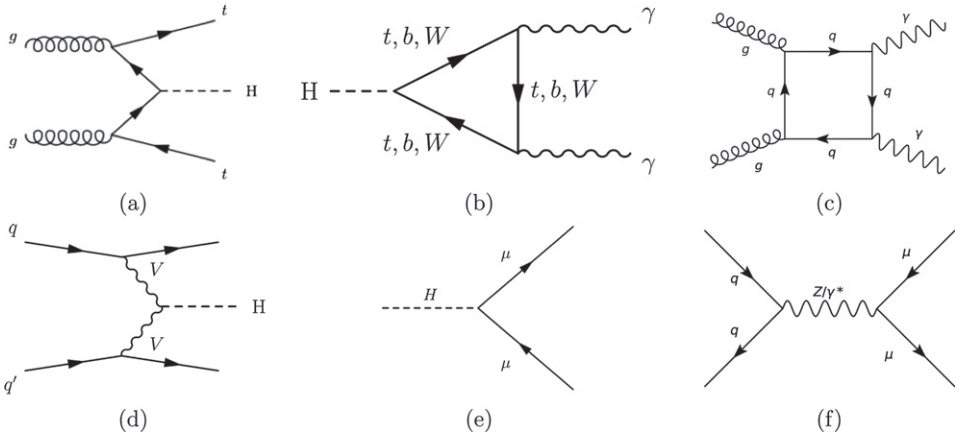


Figure 1. Representative Feynman diagrams for (a) $t\bar{t}H$ production, (b) $H \to \gamma\gamma$ decay, (c) non-resonant two-photon production, (d) VBF Higgs production, (e) $H \to \mu^+\mu^-$ decay, and (f) $Z/\gamma^* \to \mu\mu$ production. In these diagrams, H denotes a Higgs boson, g denotes a gluon, q denotes a quark, t denotes a top quark, b denotes a bottom quark, μ denotes a muon, W denotes a W boson, Z denotes a Z boson, V denotes a W boson or Z boson, and γ denotes a photon.

1. Two recent LHC flagship physics analyses

The observation of $t\bar{t}H$ production (Higgs boson production in association with a top quark pair) in 2018 by the ATLAS and CMS experiments [6, 7] was a significant milestone for the understanding of fundamental particles and interactions. It confirmed the interactions between the Higgs boson and the top quark, which is the heaviest known fundamental particle. The measurement of the Higgs-top coupling strength could refine our understanding of the Higgs mechanism and provide important handles to new physics. As $t\bar{t}H$ only accounts for about 1% of the total Higgs boson production at the LHC, its observation was extremely challenging. Here we address a channel where the Higgs boson decays into two photons ($H \to \gamma\gamma$) and the two top quarks decay into jets. To ensure the results are as realistic as possible, we closely follow an analysis strategy similar to that employed by ATLAS [6]. Starting from reconstructed events with two photons and at least three jets, we train classifiers to separate the $t\bar{t}H$ ($H \to \gamma\gamma$) signal from the dominant background of this analysis, the non-resonant two-photon production. See figure 1 for representative Feynman diagrams for $t\bar{t}H$ production (a), $H \to \gamma\gamma$ decay (b), and non-resonant two-photon production (c). The training is using 23 kinematic variables similar to those in [6]: the transverse momentum p_{T}, pseudo-rapidity η and b-tagging status of up to 6 leading jets, the magnitude of the missing transverse momentum, as well as the $p_{\mathrm{T}}/m_{\gamma\gamma}$ ($m_{\gamma\gamma}$ denotes invariant mass of the photon pair) and η of the two photons.

The searches for $H \to \mu^+\mu^-$ decay (Higgs boson decay into two muons) at the ATLAS and CMS experiments [8, 9] have become one of the most important topics in the LHC physics program. Although the coupling between the Higgs boson and third-generation fermions (e.g. top quark) has been observed, currently there exist only first indications of the coupling between the Higgs boson and second-generation fermions. $H \to \mu^+\mu^-$ decay is the most promising process by which to observe such a coupling at the LHC. The strength of the Higgs-muon coupling could be significantly modified by new physics. With more data in the future, the LHC experiments could establish the Higgs couplings to muons, or exclude the Higgs-muon coupling:

both will be an exciting discovery. In searches for $H \to \mu^+\mu^-$ decay, the challenge is mainly due to the small $H \to \mu^+\mu^-$ decay branching ratio of about 0.02%. Again following an analysis strategy similar to that used by ATLAS [8], we divide reconstructed two-muon events into several n_j (jet multiplicity) channels, and focus on the $n_j \geqslant 2$ channel to target vector boson fusion (VBF) Higgs production, whose signature is two forward jets. We train classifiers to distinguish between the $H \to \mu^+\mu^-$ signal and the dominant background of this analysis, production of a pair of muons through the exchange of a Z boson or a virtual photon ($Z/\gamma^* \to \mu\mu$). See figure 1 for representative Feynman diagrams for VBF Higgs production (d), $H \to \mu^+\mu^-$ decay (e), and $Z/\gamma^* \to \mu\mu$ production (f). The training is based on 13 kinematic variables similar to those in [8]: the p_T and rapidity Y of the two-muon system, the absolute value of the cosine of the lepton decay angle $\cos\theta^*$ in the Collins–Soper frame, the p_T and η of the two leading jets, the relative azimuthal angle of each jet with respect to the di-muon system, the p_T, Y and invariant mass of the two-jet system, as well as the relative azimuthal angle between the two-jet system and the two-muon system.

In both the $t\bar{t}H$ and $H \to \mu^+\mu^-$ cases, we generate the signal and background events using Madgraph5_aMC@NLO [10] plus Pythia6 [11]. The center-of-mass energy of the proton-proton collisions of the generated events is set to 13 TeV (same as the ATLAS publications). For each generated event, we simulate the detector response using Delphes [12]. A principal component analysis (PCA) method [13, 14] is employed for data compression, converting the kinematic variables to a smaller number of PCA variables so that the number of encoded variables matches the number of available qubits (which is 10 in this study). After PCA, the data is transformed using MinMaxScaler in the scikit-learn package [15] so that it ranges from $-\pi$ to π. The events are then passed to the machine learning algorithms, whether classical or quantum.

2. Quantum variational classifier algorithm and workflow

Following [4], we use the quantum variational classifier algorithm to classify physics events of interest from background events. This quantum approach exploits the mapping of classical input data to an exponentially large quantum feature space, which is based on quantum circuits that are hard to simulate classically. It can be summarized in four main steps:

(a) Apply a feature map circuit $U_{\Phi(\vec{x})}$ to encode the input data $\vec{x}$ (containing 10 PCA variables) into a quantum state $\Phi(\vec{x})$, as shown in figure 2(a). In our study the feature map encodes N classical variables to the quantum state space of an N-qubit system.

(b) Apply a quantum variational circuit $W(\vec{\theta})$ parameterized by gate angles $\vec{\theta}$, as shown in figure 2(b). Here the variational circuit takes the form

$$W(\vec{\theta}) = U_{\text{rot}}(\theta)U_{\text{ent}} \ldots U_{\text{rot}}(\theta)U_{\text{ent}}, \tag{1}$$

where $U_{\text{rot}}(\theta)$ refers to a variational circuit consisting of rotations on different qubits and U_{ent} refers to entanglement unitary operations [4].

(c) Measure the qubit state in the computational basis. The qubit measurement error is one of the largest error sources on the quantum hardware and results in imprecision on the classification result. To reduce this imprecision, we entangle every two qubits and then measure half of the N qubits, as shown in figure 2(c). Additionally, a measurement error mitigation method implemented by the Qiskit framework [16] is applied when measuring qubits. This error mitigation method derives a relation matrix between the ideal results and the noisy results, which is later used to correct the noisy results.

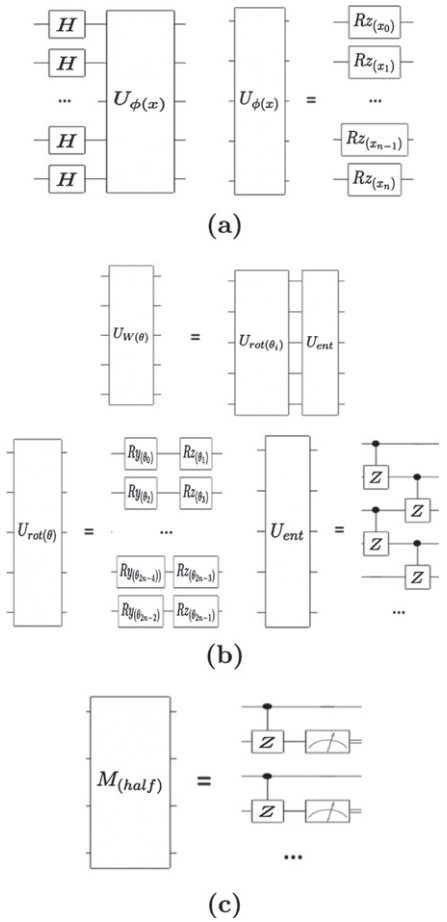


Figure 2. Quantum circuits used in our quantum variational classifier studies. (a) The quantum feature map. First, Hadamard gates H initialize every qubit to an equal superposition of all basis states. Then the feature map circuit $U_{\Phi(\vec{x})}$ encodes classical variables to quantum states by applying a phase rotation RZ of an angle x_i. The parameterization of x_i with the corresponding PCA variable is slightly different between the $t\bar{t}H$ and the $H \to \mu^+\mu^-$ cases. The feature map circuits can be duplicated multiple times with a depth parameter. (b) The quantum variational circuits $W(\vec{\theta})$. The variational rotation circuit $U_{\rm rot}(\theta)$ is parameterized by θ and is followed by the entanglement circuit $U_{\rm ent}$. $U_{\rm rot}(\theta)$ consists of two rotations, RY and RZ, on every qubit with parameterized θ. For the entanglement step $U_{\rm ent}$, we use the controlled phase gate CZ to entangle adjacent qubits. In order to parallelize the qubit operations, we optimized the $U_{\rm ent}$ circuit to reduce execution dependency by at first only entangling each even qubit to its following odd qubit and then only entangling each odd qubit to its following even qubit. Multiple copies of the variational and entanglement circuits with another depth parameter can be applied to increase the number of degrees of freedom in a machine learning model [4]. (c) The measurement circuit $M_{\rm (half)}$. To measure half of the qubits, a controlled phase gate operation CZ is applied on every two qubits and only one of the two entangled qubits is measured.

(d) Classify the state through the action of a diagonal operator f in computational basis with the eigenvalue being either $+1$ or -1. A discriminant is evaluated for the input data $\vec{x}$

according to

$$\langle \Phi(\vec{x})|W^{\dagger}(\vec{\theta})2^{-1}(1+f)W(\vec{\theta})|\Phi(\vec{x})\rangle \tag{2}$$

and used to assign an output label $y \in \{1, 0\}$ denoting either a signal or background process [4].

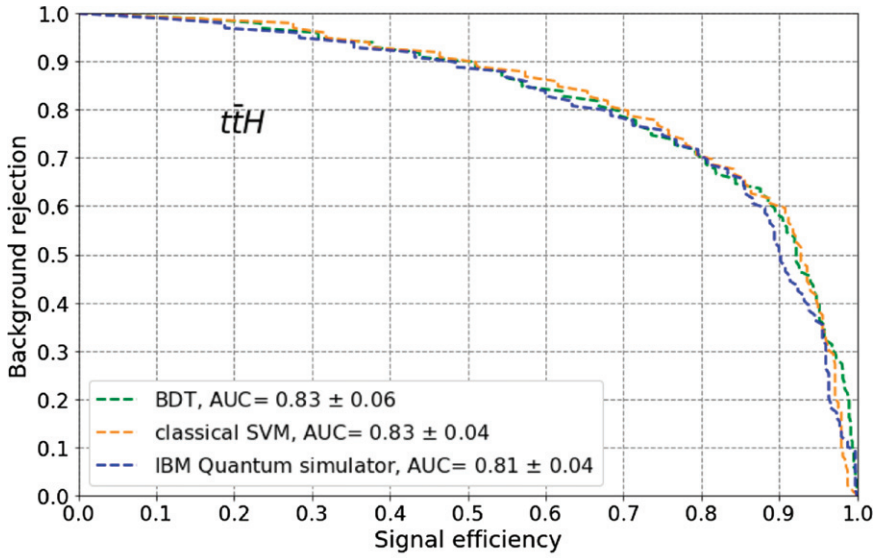
During the training phase, a set of input data $\vec{x}$ and corresponding outputs y are used to train the circuit $W(\vec{\theta})$ to reproduce the correct classification. The set of optimized parameters $\vec{\theta}$ is then kept fixed for all future classifications of the physical data.

3. Result from the IBM quantum computer simulator with 10 qubits

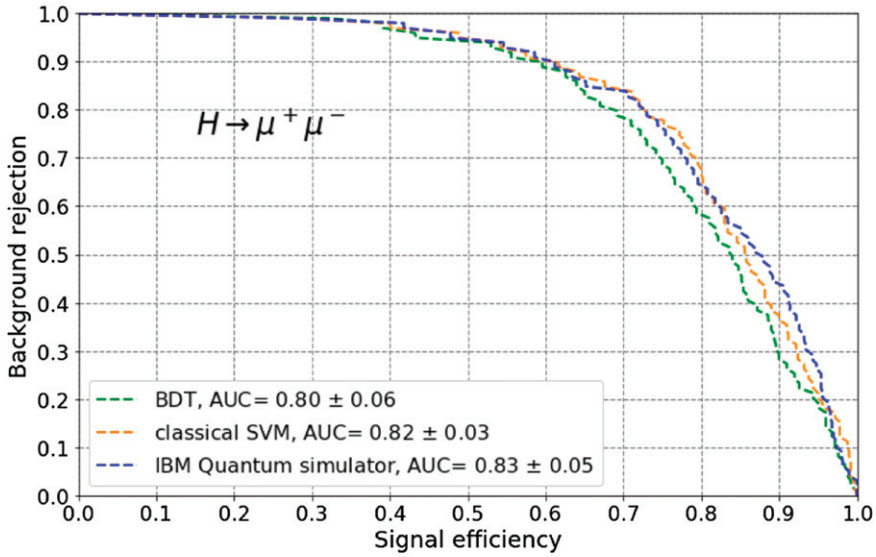
We employ quantum machine learning with 10 qubits on the *ibmq QasmSimulator* [16] to classify signal and background processes for the $t\bar{t}H$ analysis and the $H \to \mu^+\mu^-$ analysis. The *ibmq QasmSimulator* simulates executions and measurements on quantum circuits of the IBM quantum computer hardware. The simulation incorporates a noise model generated from the properties of real hardware device. In each analysis, we apply the quantum variational classifier algorithm to ten independent datasets, each consisting of 100 events for training and 100 events for testing. The quantum circuits are optimized to best fit the constraints imposed by the hardware (e.g. qubit connectivity, gate set availability, and hardware noise), as well as the nature of the data. In the optimized configuration, the feature map depth is 1 and the variational circuit depth is 1. With the present status of our accessed hardware, the limited circuit depths are adopted to overcome the hardware noise when utilizing 10 qubits. The circuit implementation uses linear qubit connectivity. The Spall's simultaneous perturbation stochastic approximation algorithm [17, 18] is used as the optimizer for the variational circuit parameters $\vec{\theta}$ in the training process. With the same ten datasets and the same 10 variables processed with the PCA method, we also train a classical SVM [19] classifier using the scikit-learn package [15] and a BDT [20, 21] classifier using the XGBoost package [22]. The classical SVM and the BDT serve as benchmarks for classical machine learning algorithms. Hyper-parameter tuning was performed on these classical algorithms.

To study the discrimination power of each algorithm for both $t\bar{t}H$ and $H \to \mu^+\mu^-$, the testing events of the ten datasets are combined to make receiver operating characteristic (ROC) curves as a benchmark in the plane of background rejection versus signal efficiency, as shown in figure 3. We observe that in both the $t\bar{t}H$ analysis and the $H \to \mu^+\mu^-$ analysis, the quantum variational classifier method on the *ibmq QasmSimulator* (blue) performs similarly to the classical SVM (yellow) and the BDT (green). We quantify the discrimination power of each classifier by the AUC (area under the ROC curve). In the $t\bar{t}H$ analysis, the AUC for the quantum variational classifier method reaches 0.81 ± 0.04 on the *ibmq QasmSimulator*, compared to 0.83 ± 0.04 for the classical SVM and 0.83 ± 0.06 for the BDT. Similarly, in the $H \to \mu^+\mu^-$ analysis, the AUC for the quantum variational classifier method reaches 0.83 ± 0.05 on the *ibmq QasmSimulator*, compared to 0.82 ± 0.03 for the classical SVM and 0.80 ± 0.06 for the BDT. The quoted errors are the standard deviations for the AUC values of the ten datasets. This demonstrates the quantum algorithm can accurately distinguish signal from background on realistic physics datasets; the performance is comparable to (within the margin of error) state-of-the-art classical methods.

Ultimately, we can select events based on the classifier discriminant to maximize the quantity $S/\sqrt{(B)}$, where S is the number of signal events and B is the number of background events remaining after the selection. $S/\sqrt{(B)}$ is an approximation of the signal significance and is typically correlated with the classifier AUC, as both indicate the degree of separation between



(a)



(b)

Figure 3. The ROC curves (as a benchmark in the plane of background rejection versus signal efficiency) of the quantum variational classifier method on the *ibmq QasmSimulator* (blue), the classical SVM (yellow), and the BDT (green) for (a) the $t\bar{t}H$ analysis and (b) the $H \to \mu^+\mu^-$ analysis. In each analysis, the classifiers are constructed using ten independent datasets, each consisting of 100 events for training and 100 events for testing. All classifiers are trained with the same 10 variables processed with the PCA method. In this study, 10 qubits are employed on the quantum computer simulator. To visualize the discrimination power of each algorithm, the testing events of the ten datasets are combined to make the ROC curves. We observe that the quantum variational classifier method on the *ibmq QasmSimulator* performs similarly to the classical SVM and the BDT for both the $t\bar{t}H$ analysis and the $H \to \mu^+\mu^-$ analysis.

signal and background. In the $t\bar{t}H$ analysis, a selection on the variational quantum classifier discriminant with a signal acceptance of 0.70 is associated with a background rejection of 0.78 (see the ROC curve in figure 3(a)), and hence improves $S/\sqrt{(B)}$ by approximately $0.70/\sqrt{(0.22)} - 1 = 50\%$ with respect to no selection. Similarly, in the $H \to \mu^+\mu^-$ analysis, a selection on the variational quantum classifier discriminant with a signal acceptance of 0.70 is associated with a background rejection of 0.84 (see the ROC curve in figure 3(b)), hence improving $S/\sqrt{(B)}$ by approximately $0.70/\sqrt{(0.16)} - 1 = 75\%$.

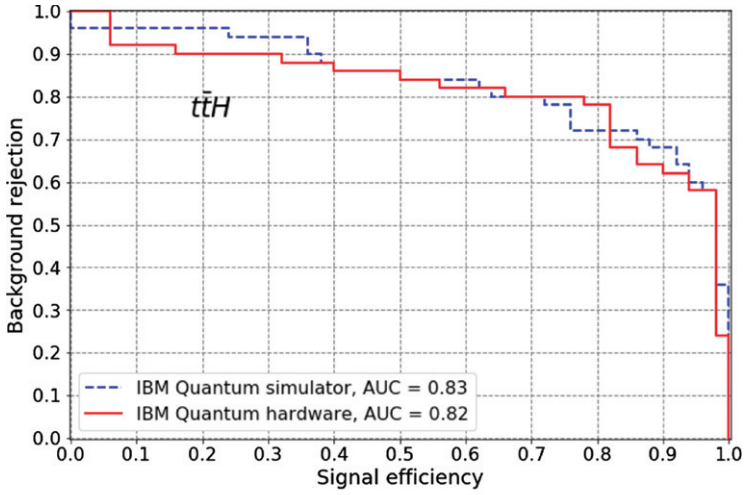
4. Result from the IBM quantum computer hardware with 10 qubits

At this point, it is interesting to assess the potential of quantum hardware calculations on the classification of the data presented in the previous section, and quantify the effect of the device noise. For the $t\bar{t}H$ analysis and the $H \to \mu^+\mu^-$ analysis, we employ the quantum variational classifier algorithm with 10 qubits on the '*ibmq_boeblingen*' and '*ibmq_paris*' quantum computer hardware. '*ibmq_boeblingen*' is a 20-qubit quantum processor and '*ibmq_paris*' is a 27-qubit quantum processor. Both are based on superconducting electronic circuits. Due to current limitation of the access time to the quantum processors, the quantum variational classifier algorithm is only applied to one of the ten datasets for each physics analysis. We pick the dataset whose simulator AUC is closest to the average simulator AUC of the ten datasets. The circuit, optimizer, and error mitigation configuration on the hardware is kept the same as for the simulator jobs.

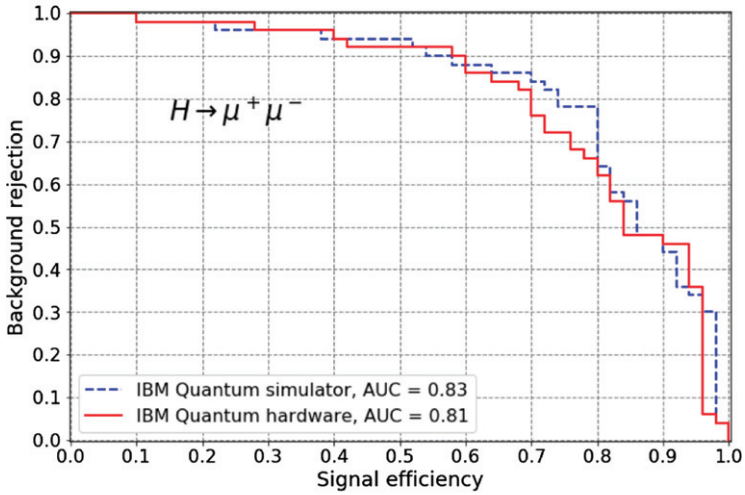
The ROC curves of the quantum variational classifier algorithm on the '*ibmq_boeblingen*' quantum hardware (for $t\bar{t}H$) and '*ibmq_paris*' quantum hardware (for $H \to \mu^+\mu^-$) are shown in red in figure 4. The ROC curves for the *ibmq QasmSimulator* with the same datasets are overlaid in blue. We observe that for the quantum variational classifier method, the quantum simulator and quantum hardware results appear to be in good agreement. In the $t\bar{t}H$ analysis, the quantum hardware AUC is 0.82, while the quantum simulator AUC is 0.83. Similarly, in the $H \to \mu^+\mu^-$ analysis, the quantum hardware AUC is 0.81, while the quantum simulator AUC is 0.83. In each analysis, the difference between the hardware AUC and the simulator AUC is found to be compatible with the test sample statistical error evaluated using a bootstrapping re-sampling method. With the circuit configuration optimized for 10 qubits, the gate-model quantum computers have achieved reasonable performance in exploiting a quantum state space with 2^{10} dimensions to distinguish signal from background at the LHC.

Figure 5 shows the evolution of the loss function versus the number of iterations during the training process of the quantum variational classifier on the hardware (red) and simulator (blue) for the $t\bar{t}H$ analysis and the $H \to \mu^+\mu^-$ analysis. The number of iterations indicates the number of times the variational circuit parameters are updated in the training process. The empirical loss function is defined by the error probability of incorrect assignment compared to the exact solutions available for the training set. During the training process, the loss function is minimized to penalize misassignment and to optimize classifier parameters. The loss function improves and converges as the number of iterations increases, indicating that the quantum algorithm on hardware is indeed learning the difference between signal and background in a realistic high energy physics analysis at the LHC.

In our study, 200 hours are required to run 500 training iterations on 100 events of the $t\bar{t}H$ or $H \to \mu^+\mu^-$ analyses on quantum hardware. This is longer than that of the classical algorithms because today's quantum hardware is not yet fully mature. However, with the present rapid developments in quantum hardware, we expect in the future to see speed ups in quantum machine learning applications to high energy physics.



(a)



(b)

Figure 4. The ROC curves of the quantum variational classifier method with the '*ibmq_boeblingen*' and '*ibmq_paris*' quantum computer hardware (red) and with the *ibmq QasmSimulator* (blue) for (a) the $t\bar{t}H$ analysis (using '*ibmq_boeblingen*') and (b) the $H \to \mu^+\mu^-$ analysis (using '*ibmq_paris*'). For each physics analysis, one dataset consisting of 100 events for training and 100 events for testing is utilized to construct the classifiers. This dataset is one of the ten datasets used in figure 3. All classifiers are trained with the same 10 variables processed with the PCA method. In this study, 10 qubits are employed on the quantum computer hardware and the quantum computer simulator. To visualize the discrimination power of both the quantum simulator and quantum hardware, the testing events of the dataset are used to make the ROC curves. We observe that, for the quantum variational classifier method, the quantum simulator and quantum hardware results appear to be in good agreement.

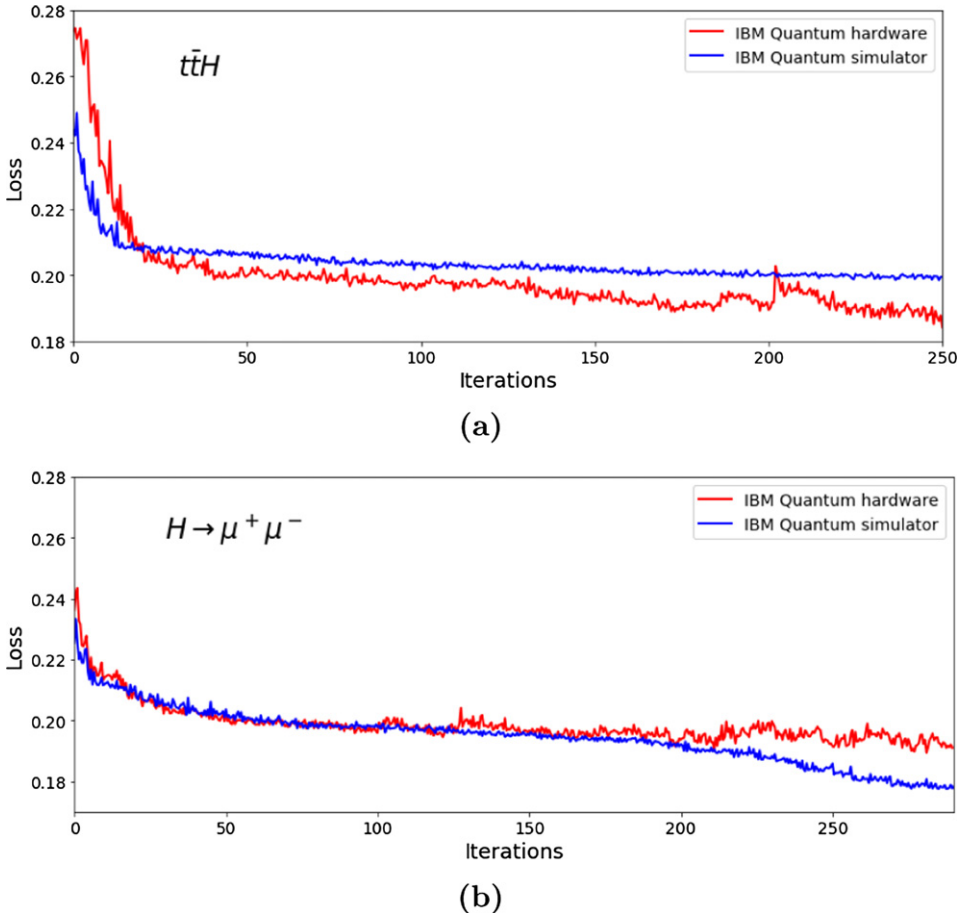


Figure 5. The evolution of the loss function versus the number of iterations during the training process of the quantum variational classifier on the quantum computer hardware (red) and quantum computer simulator (blue) for (a) the $t\bar{t}H$ analysis and (b) the $H \to \mu^+\mu^-$ analysis. The number of iterations indicates the number of times the variational circuit parameters are updated in the training process. The empirical loss function is defined by the error probability of incorrect assignment compared to the exact solutions available for the training set. The loss function improves and converges as the number of iterations increases, indicating that the quantum algorithm on hardware is indeed learning the difference between signal and background.

5. Conclusion

In this study, we have obtained early results in the application of quantum machine learning with 10 qubits on the *imbq QasmSimulator* and the '*ibmq_boeblingen*' and '*ibmq_paris*' quantum hardware to two recent LHC flagship physics analyses: $t\bar{t}H$ and $H \to \mu^+\mu^-$. $t\bar{t}H$, Higgs boson production in association with a top quark pair, probes the Higgs boson couplings to the top quark, while $H \to \mu^+\mu^-$, Higgs boson decays to two muons, probes the Higgs boson couplings to second-generation fermions. In this study we do not attempt to do a complete analysis of $t\bar{t}H$ and $H \to \mu^+\mu^-$. Rather our goal is to perform proof of principle in using quantum machine learning compared with popular classical machine learning

methods, BDT for example. With small training samples of 100 events, the quantum variational classifier method on the *ibmq QasmSimulator* performs similarly to the classical SVM algorithm and the BDT algorithm. The quantum variational classifier method on the quantum hardware has shown promising discrimination power comparable to that on the quantum simulation.

To study the discrimination power of quantum machine learning classifiers, we make use of ROC curves in the plane of background rejection versus signal efficiency as a standard metric in machine learning application for the high energy physics. We further quantify the discrimination power of the classifiers by the AUC (area under the ROC curve). The use of ROC curves and AUCs to be the metric of discrimination power compared with classical machine learning methods is inspired by reference [5]. A difference is that reference [5] uses quantum annealers while our work uses gate-based quantum computers. In the $t\bar{t}H$ analysis, the quantum hardware AUC is 0.82, while the quantum simulator AUC is 0.83. Similarly, in the $H \to \mu^+\mu^-$ analysis, the quantum hardware AUC is 0.81, while the quantum simulator AUC is 0.83. These results demonstrate that quantum machine learning on the hardware of the gate-model quantum computers has the ability to differentiate between signal and background in a realistic high energy physics analysis at the LHC. Furthermore, although we have demonstrated that the quantum and classical machine learning algorithms perform similarly, with the rapid advance of the quantum computing technology, the use of quantum machine learning may offer a 'speed up' advantage [3], which can be critical for the future of the high energy physics community.

In the future, by exploiting the high dimensional feature space defined by a larger number of qubits and by mitigating the impact of quantum hardware noise, quantum machine learning classifiers could possibly outperform classical classifiers. We plan to explore quantum algorithms to extend our analysis to more qubits and larger sample sizes. Moreover, we plan to apply the quantum kernel classifier method proposed in [4] to our LHC flagship physics analyses. We foresee the usage of quantum machine learning in future high-luminosity LHC physics analyses, including measurements of the Higgs boson self-couplings and searches for dark matter.

Acknowledgments

This project is supported in part by the United States Department of Energy, Office of Science, HEP-QIS Research Program, under Award Number DE-SC0020416. This project is supported in part by the United States Department of Energy, Office of Science, Office of High Energy Physics program under Award Number DE-SC-0012704 and the Brookhaven National Laboratory LDRD #20-024. This research used resources of the Oak Ridge Leadership Computing Facility, which is a United States Department of Energy Office of Science User Facility supported under Contract DE-AC05-00OR22725. The Wisconsin group would like to thank the ATLAS Collaboration for the inspiration of the two LHC flagship analyses used in this publication.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Sau Lan Wu https://orcid.org/0000-0001-5866-1504
Samuel Yen-Chi Chen https://orcid.org/0000-0003-0114-4826

References

[1] Aad G *et al* (ATLAS Collaboration) 2012 Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC *Phys. Lett.* B **716** 1–29
[2] Chatrchyan S *et al* (CMS Collaboration) 2012 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC *Phys. Lett.* B **716** 30–61
[3] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 Quantum machine learning *Nature* **549** 195–202
[4] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 Supervised learning with quantum-enhanced feature spaces *Nature* **567** 209–12
[5] Mott A, Job J, Vlimant J-R, Lidar D and Spiropulu M 2017 Solving a Higgs optimization problem with quantum annealing for machine learning *Nature* **550** 375–9
[6] Aaboud M *et al* (ATLAS Collaboration) 2018 Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector *Phys. Lett.* B **784** 173–91
[7] Sirunyan A M *et al* (CMS Collaboration) 2018 Observation of $t\bar{t}H$ production *Phys. Rev. Lett.* **120** 231801
[8] Aad G *et al* (ATLAS Collaboration) 2021 A search for the dimuon decay of the standard model Higgs boson with the ATLAS detector *Phys. Lett.* B **812** 135980
[9] Sirunyan A M *et al* (CMS Collaboration) 2021 Evidence for Higgs boson decay to a pair of muons *J. High Energy Phys.* JHEP01(2021)148
[10] Alwall J *et al* 2014 The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations *J. High Energy Phys.* JHEP07(2014)079
[11] Sjöstrand T, Mrenna S and Skands P 2006 PYTHIA 6.4 physics and manual *J. High Energy Phys.* JHEP05(2006)026
[12] de Favereau J, Delaere C, Demin P, Giammanco A, Lemaître V, Mertens A and Selvaggi M 2014 DELPHES 3, a modular framework for fast simulation of a generic collider experiment *J. High Energy Phys.* JHEP02(2014)057
[13] Pearson K 1901 LIII. On lines and planes of closest fit to systems of points in space *London, Edinburgh Dublin Phil. Mag. J. Sci.* **2** 559–72
[14] Jolliffe I T 2002 *Principal Component Analysis* 2nd edn (Berlin: Springer)
[15] Pedregosa F *et al* 2011 Scikit-learn: machine learning in python *J. Mach. Learn. Res.* **12** 2825–30
[16] Aleksandrowicz G *et al* 2019 Qiskit: an open-source framework for quantum computing https://qiskit.org
[17] Spall J C 1997 A one-measurement form of simultaneous perturbation stochastic approximation *Automatica* **33** 109–12
[18] Spall J C 2000 Adaptive stochastic approximation by the simultaneous perturbation method *IEEE Trans. Autom. Control* **45** 1839–53
[19] Boser B E, Guyon I M and Vapnik V N 1992 A training algorithm for optimal margin classifiers *Proc. 5th Annual ACM Workshop on Computational Learning Theory* pp 144–52
[20] Friedman J H 2002 Stochastic gradient boosting *Comput. Stat. Data Anal.* **38** 367–78
[21] Hastie T, Tibshirani R and Friedman J 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn (Berlin: Springer)
[22] Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* pp 785–94