

Application of Quantum Machine Learning using the Quantum Kernel Algorithm on High Energy Physics Analysis at the LHC

Sau Lan Wu,* Shaojun Sun, Wen Guan, Chen Zhou, Jay Chan, Chi Lung Cheng, Tuan Pham, Yan Qian, Alex Zeng Wang, and Rui Zhang
Department of Physics, University of Wisconsin, Madison WI, USA

Miron Livny
Department of Computer Sciences, University of Wisconsin, Madison WI, USA

Jennifer Glick
IBM Quantum, T.J. Watson Research Center, Yorktown Heights, NY, USA

Panagiotis Kl. Barkoutsos, Stefan Woerner, and Ivano Tavernelli
IBM Quantum, Zurich Research Laboratory, Rüschlikon, Switzerland

Federico Carminati and Alberto Di Meglio
CERN Quantum Technology Initiative, IT Department, CERN, Geneva, Switzerland

Andy C. Y. Li, Joseph Lykken, and Panagiotis Spentzouris
Quantum Institute, Fermi National Accelerator Laboratory, Batavia, IL, USA

Samuel Yen-Chi Chen and Shinjae Yoo
Computational Science Initiative, Brookhaven National Laboratory, Upton, NY, USA

Tzu-Chieh Wei
C.N. Yang Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, NY, USA

Quantum machine learning could possibly become a valuable alternative to classical machine learning for applications in High Energy Physics by offering computational speed-ups. In this study, we employ a support vector machine with a quantum kernel estimator (QSVM-Kernel method) to a recent LHC flagship physics analysis: $t\bar{t}H$ (Higgs boson production in association with a top quark pair). In our quantum simulation study using up to 20 qubits and up to 50000 events, the QSVM-Kernel method performs as well as its classical counterparts in three different platforms from Google Tensorflow Quantum, IBM Quantum and Amazon Braket. Additionally, using 15 qubits and 100 events, the application of the QSVM-Kernel method on the IBM superconducting quantum hardware approaches the performance of a noiseless quantum simulator. Our study confirms that the QSVM-Kernel method can use the large dimensionality of the quantum Hilbert space to replace the classical feature space in realistic physics datasets.

I. INTRODUCTION

In 2012, the ATLAS and CMS experiments discovered the Higgs boson [1, 2] using proton-proton collision data at the Large Hadron Collider (LHC). This discovery completed the fundamental particle spectrum of the Standard Model and was a major achievement in High Energy Physics (HEP). As the LHC experiments enter the post-Higgs discovery era, physicists strive to refine the understanding of the Standard Model and pursue new physics beyond the Standard Model. Machine learning has become one of the most powerful tools for exploring the full physics potential of the huge amount of data collected by the LHC experiments. In HEP, the important usage of machine learning techniques [3–7] includes simulation,

event reconstruction and data analyses. As an example, the ATLAS and CMS experiments utilized supervised machine learning algorithms in the recent data analyses to achieve the observation of the $t\bar{t}H$ processes [8, 9], a very rare Higgs production mode.

Quantum machine learning, where machine learning is performed using quantum algorithms, has the potential to improve the computational complexity of classical machine learning algorithms and obtain computational speed-ups when being executed on quantum computers [10]. Some of the quantum algorithms will benefit from exponential improvements in speed. For classification problems, it may also lead to better separation power than classical machine learning. A recent result [11] demonstrates a significant advantage in prediction accuracy for a quantum algorithm over some classical algorithms on engineered datasets. A key component of quantum machine learning algorithms is exploit-

* sau.lan.wu@cern.ch

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

ing the high dimensional quantum state space through the actions of superposition and entanglement. With the progress of quantum technologies, quantum machine learning could possibly become a valuable alternative to classical machine learning for processing big data (including simulation, reconstruction and analyses) in High Energy Physics [12]. Recent projections from IBM [13], Google [14], and IonQ [15] suggest that quantum computers with thousands of qubits capable of performing practical computational tasks may become available within the next ten years. This coincides with the High Luminosity upgrades of the LHC (HL-LHC) [16], bringing a massive increase in not only the amount of collected physics data, but also the computational power needed to process that data. Present research on quantum machine learning algorithms could be implemented on those future devices, thereby ensuring the timely exploitation of quantum advantages for physics applications, and possibly even contributing to the discovery of new physics.

The challenges we face for the future include working with large numbers of events in the millions, applying large numbers of qubits in the thousands to obtain stellar performance, and achieving excellent computational speeds. It is the purpose of this publication to inch one step closer to conquering those challenges, even under the present limitations of existing quantum computer technology.

Previous studies [17–19] have investigated quantum annealing or quantum classifiers trained with variational circuits. A quantum machine learning algorithm, the support vector machine with a quantum kernel estimator (QSVM-Kernel), was proposed to solve classification problems [20, 21]. This algorithm was experimentally implemented with 2 qubits on a superconducting quantum computer and found to be accurate for artificial datasets [20]. QSVM-Kernel leverages the quantum state space as a direct representation of the feature space, which can give rise to kernel functions that are hard to evaluate classically [20, 22]. However, any potential quantum advantage does not lie solely in the high dimensionality of the quantum state space, as it is well known that classical kernels can map into feature space of arbitrary high dimensions. Rather, a path towards quantum advantage is found in the computational complexity of the quantum circuits used to compute the quantum kernel. Namely, these circuits must be hard to estimate classically. A recent result [23] establishes an exponential quantum speedup for QSVM-Kernel using a fault-tolerant quantum computer to estimate the kernel function for a classically hard classification problem. Although this particular problem was not practically motivated, the result rigorously formalizes the intuition that quantum feature maps can identify patterns classical machines are unable to capture. From this foundation, quantum feature maps can be designed and tested on practical data sets, potentially leading to better classification results than classical feature maps and kernels.

It is therefore interesting to study the QSVM-Kernel

algorithm with a larger number of qubits and evaluate its performance on real-world datasets. In our study, we successfully employ the QSVM-Kernel algorithm in the $t\bar{t}H$ analysis, a recent LHC flagship physics analysis, using up to 20 qubits on quantum computer simulators and up to 15 qubits on quantum computer hardware. Furthermore, we compare the classification result of the QSVM-Kernel algorithm to a few popular classical algorithms that are commonly used by the LHC experiments.

II. $t\bar{t}H$ PHYSICS ANALYSIS AT THE LHC

The observation of $t\bar{t}H$ (Higgs boson production in association with a top quark pair) by the ATLAS and CMS experiments [8, 9] was one of the LHC flagship physics results following the Higgs boson discovery. It established a direct observation of the Higgs boson’s interaction with the top quark, the heaviest known fundamental particle. Study of the Higgs-top interaction may provide crucial test for the Higgs mechanism of the Standard Model and essential clues for new physics beyond the Standard Model. Due to the small $t\bar{t}H$ production rate at the LHC, its observation was highly challenging. To achieve the desired sensitivities to $t\bar{t}H$ production, the ATLAS and CMS collaborations combined results from a number of decay channels. The physics analyses in many of these channels utilize machine learning techniques. For example, classifiers based on machine learning are constructed to analyze kinematic variables of the collision events and distinguish between signal and background.

In our study, we focus on an important $t\bar{t}H$ analysis channel where the Higgs boson decays into two photons ($H \rightarrow \gamma\gamma$) and the two top quarks decay into hadrons. The dominant background in this analysis channel is non-resonant two-photon production. See Figure 1 for representative Feynman diagrams for $t\bar{t}H$ production, $H \rightarrow \gamma\gamma$ decay, and non-resonant two-photon production.

III. METHOD

The support vector machine (SVM) [24, 25] is one of the most commonly used supervised machine learning algorithms for data classification. Here for a data event, \vec{x} denotes the vector of its input features and $y \in \{0, 1\}$ denotes its class label (0 for background and 1 for signal). The SVM algorithm maps \vec{x} into a higher dimensional feature space, where it measures the similarity between any two data events (denoted as “kernel entry”, $k(\vec{x}_i, \vec{x}_j)$). The SVM algorithm then optimizes a hyperplane that separates signal events from background events and classifies a new data event \vec{x}' by

$$y' = \text{sgn}\left(\sum_{i=1}^t \alpha_i y_i k(\vec{x}_i, \vec{x}') + b\right) \quad (1)$$

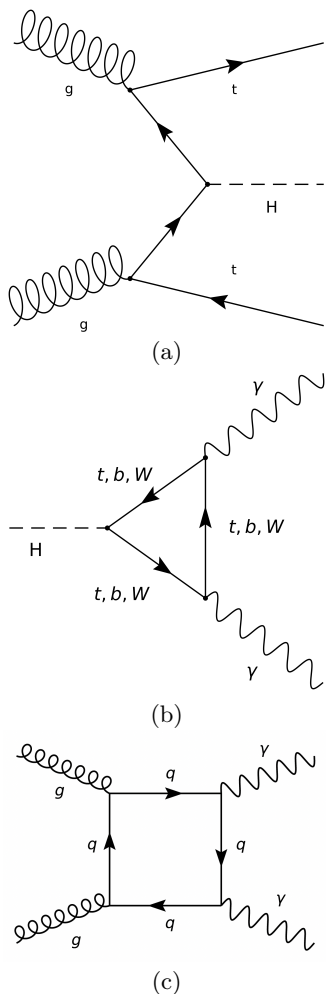


Figure 1. Representative Feynman diagrams for (a) ttH production, (b) $H \rightarrow \gamma\gamma$ decay, and (c) non-resonant two-photon production. In these diagrams, H denotes a Higgs boson, g denotes a gluon, q denotes a quark, t denotes a top quark, b denotes a bottom quark, W denotes a W boson, and γ denotes a photon.

where sgn is the sign function, t is the size of the training dataset with known labels y_i , and (α_i, b) defines the separating hyperplane. Furthermore, a continuous SVM discriminant can be obtained by computing the probabilities of being in the signal class. A main limitation of the classical SVM algorithm is that evaluating kernel entries in a large feature space can be computationally expensive. Three different popular classical kernels are considered to benchmark the performance of the classical SVM method in our study: the linear kernel $\vec{x}_i^T \vec{x}_j$, the polynomial kernel $(\gamma \vec{x}_i^T \vec{x}_j)^d$, and the RBF kernel $\exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2)$ (γ and d are hyperparameters).

A. Quantum Kernel Estimation

A quantum version of the SVM with a quantum kernel estimator (QSVM-Kernel) was introduced in Ref [20, 21], which leverages the quantum state space as a feature

space to efficiently compute kernel entries. This algorithm maps the classical data event \vec{x} non-linearly to a quantum state of N qubits by applying a quantum feature map circuit $\mathcal{U}_{\Phi(\vec{x})}$ to the initial state $|0^{\otimes N}\rangle$:

$$|\Phi(\vec{x})\rangle = \mathcal{U}_{\Phi(\vec{x})}|0^{\otimes N}\rangle \quad (2)$$

It then calculates the kernel entry for data events \vec{x}_i and \vec{x}_j based on the inner product of their quantum states:

$$k(\vec{x}_i, \vec{x}_j) = |\langle \Phi(\vec{x}_i) | \Phi(\vec{x}_j) \rangle|^2 = |\langle 0^{\otimes N} | \mathcal{U}_{\Phi(\vec{x}_i)}^\dagger \mathcal{U}_{\Phi(\vec{x}_j)} | 0^{\otimes N} \rangle|^2 \quad (3)$$

The kernel entry can be evaluated on a quantum computer by measuring the $\mathcal{U}_{\Phi(\vec{x}_i)}^\dagger \mathcal{U}_{\Phi(\vec{x}_j)} | 0^{\otimes N} \rangle$ state in the computational basis with repeated measurement shots and recording the probability of collapsing the output into the $|0^{\otimes N}\rangle$ state. In our study, the general design of the quantum circuit for evaluating the kernel entries is inherited from Ref [20] and shown in Figure 3 (a).

B. Quantum Feature Map

As suggested in Ref [20], the mathematical properties of the quantum feature map should be complex enough to be hard to simulate on a classical computer, and simple enough to be executable on noisy intermediate-scale quantum computers. Furthermore, the classification performance of the QSVM-kernel method is highly dependent on the kernel matrix, which is calculated from the quantum states of the feature map. Therefore, choosing a suitable feature map is essential for optimizing the classification performance of the QSVM-kernel method.

The quantum feature map $\mathcal{U}_{\Phi(\vec{x})}$ gives rise to a 2^N dimensional feature space (N is the number of qubits) that is conjectured to be hard to estimate classically [22]. Following Ref [20], the quantum feature map in our study has two repeated layers:

$$\mathcal{U}_{\Phi(\vec{x})} = U_{\Phi(\vec{x})} H^{\otimes N} U_{\Phi(\vec{x})} H^{\otimes N} \quad (4)$$

where H is a Hadamard gate and $U_{\Phi(\vec{x}_i)}$ is a unitary operator that encodes data from a classical event in its parameters. We analyzed various quantum circuit candidates for the unitary operator $U_{\Phi(\vec{x})}$ and found that the quantum circuit shown in Figure 3 (b) performs the best for our ttH physics analysis. The chosen quantum circuit is constituted by single-qubit rotation gates (A , B and A'), as well as two-qubit CNOT entangling gates. On the k^{th} qubit, given an input feature vector \vec{x} , an A gate rotates the qubit around the z axis of the Bloch sphere by x_k (the k th element of \vec{x}), a B gate rotates the qubit around the y axis by x_k^d , and an A' gate rotates the qubit around the z axis by $(\frac{x_{k-1} + x_k}{2})^d$, where $d = 3$. The entangling operations are arranged in an alternating pattern to yield short-depth circuits for execution on noisy intermediate scale quantum computers. The differences in the quantum feature map between here and Ref [20] are mainly the use of B gates, the different parameterization of rotation angles, and the extension to more qubits.

C. Separating Hyperplane

In the training phase, the kernel entries are evaluated for all data event pairs of the training sample and then used to find a separating hyperplane. In the testing phase, the kernel entries are evaluated between a new data event \vec{x}' and each of the data events from the training sample, which are then used to classify the new data event \vec{x}' according to the separating hyperplane. For both phases, quantum computers are only used to evaluate the kernel entries. Using these kernel entries, the optimization of the separating hyperplane and classification of the new data event are done in classical computers, as for a classical SVM.

D. Analysis dataset

In our study for the $t\bar{t}H$ analysis, the signal and dominant background processes are generated using Madgraph5_aMC@NLO [26] and Pythia6 [27], and simulated using Delphes [28]. To construct classifiers for the physics processes, we utilize a total of 23 object-based kinematic variables based on the ATLAS analysis [8]: the transverse momentum (divided by the photon pair invariant mass) and pseudo-rapidity of the two leading photons, the transverse momentum, pseudo-rapidity and b -tagging status of up to six leading jets, as well as the missing transverse momentum. The signal and background processes differ in the distributions of these variables, providing discriminating power for the machine learning algorithms. Examples of the most powerful variable distributions are shown in Figure 2. To match the N qubits used by the QSVM-Kernel algorithm, the 23 kinematic variables are compressed into N variables, using a Principal Component Analysis (PCA) method [29, 30]. In the case of 15 qubits, for example, each of the 15 variables is formed by combining the 23 original variables. Afterwards, the N variables are rescaled by

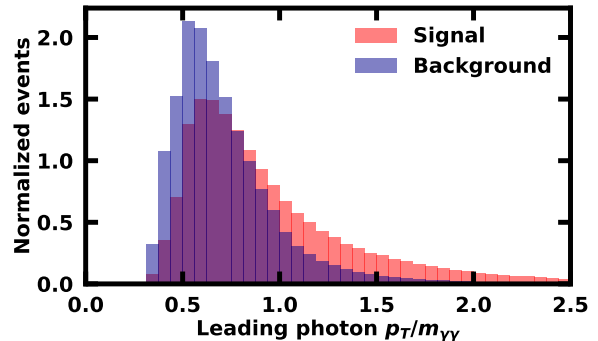
$$x_i \rightarrow -1 + 2 \times \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} \quad (5)$$

where $x_{i,min}$ ($x_{i,max}$) is the minimal (maximal) value of the variable x_i , so that the variable values range from -1 to $+1$. This ensures that the rotation angles of A , B and A' gates are within $[-1, +1]$, which is found to be slightly more optimal for the $t\bar{t}H$ analysis than $[-\pi, +\pi]$ used in Ref [20].

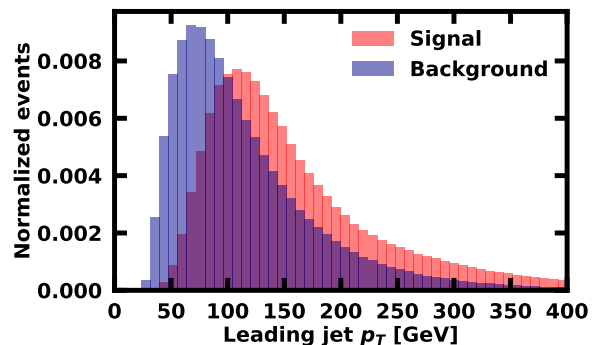
IV. RESULTS

A. Results from Quantum Computer Simulators

To classify the signal and dominant background processes for the $t\bar{t}H$ analysis, we employ the QSVM-Kernel algorithm using up to 20 qubits on the *qsim Simulator* from the Google TensorFlow Quantum framework [31], the *Statevector Simulator* from the IBM Quantum framework [32] and the *Local Simulator* from the Amazon Braket framework [33]. From this point



(a)



(b)

Figure 2. Signal and background distributions for some of the most powerful input variables to the $t\bar{t}H$ analysis: (a), the transverse momentum of the leading photon divided by the photon pair invariant mass, and (b), the transverse momentum of the leading jet.

on, they are referred to as the Google framework, IBM framework and Amazon framework. All three simulators model the noiseless execution of quantum computer hardware and evaluate the resulting quantum state vector. They represent ideal quantum hardware that can perform infinite measurement shots and experience no hardware device noise. With training variables processed by PCA, we perform our analysis for a number of dataset sizes. For a given dataset size, we prepare 60 statistically independent datasets to reduce the impact of statistical fluctuations. Each of the datasets consists of two samples of the same size: a training sample and a test sample. (In this study, a dataset of size n indicates a training sample of n events and a test sample of n events.) We have overcome the challenges of intensive computing resources needed for processing the datasets of up to 50000 events on the quantum computer simulators. Using the training sample, we adopt a cross-validation procedure [34, 35] to tune the SVM regularization hyperparameter that controls the size of the margin between the separating hyperplane and the

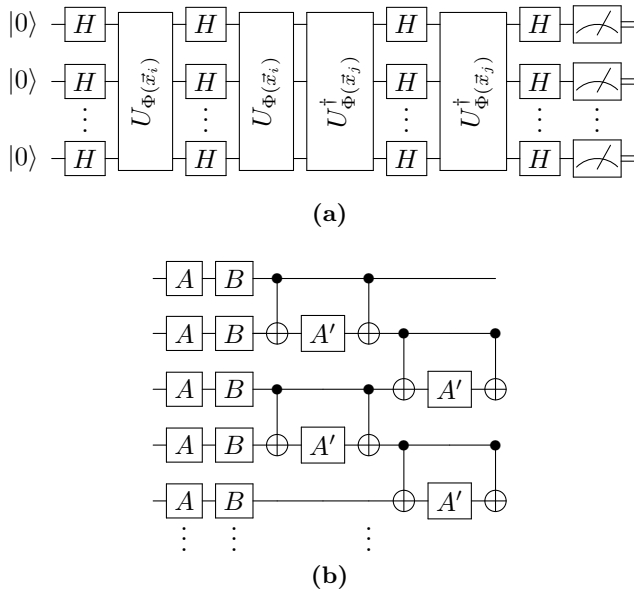


Figure 3. **(a)** Quantum circuit for evaluating the kernel entry for data events \vec{x}_i and \vec{x}_j used in our study. H is a Hadamard gate and $U_{\Phi(\vec{x}_i)}$ is a unitary operator that encodes data from a classical event in its parameters. **(b)** Quantum circuit of the unitary operator $U_{\Phi(\vec{x}_i)}$. It is constituted by single-qubit rotation gates (A , B and A'), as well as two-qubit CNOT entangling gates.

data points in the feature space. With the same datasets and the same training variables, we also construct a classical SVM [24, 25] classifier using the scikit-learn package [36] and a classical BDT [37, 38] classifier using the XGBoost package [39]. The classical SVM and BDT serve as benchmarks for classical machine learning algorithms. We again perform cross-validation on the training sample to tune the hyper-parameters of the two classical algorithms. For the classical SVM, we optimize the choice of the classical kernel, the unique hyperparameters for each kernel, and the SVM regularization hyperparameter. The optimized hyperparameters for the classical BDT include the maximum tree depth and the learning rate. The other BDT hyperparameters were found to be irrelevant to our study.

To study the discrimination power of each classifier, we produce Receiver Operating Characteristic (ROC) curves that plot background rejection versus signal efficiency, as well as areas under the ROC curves (AUCs). ROC curves and AUCs are standard metrics in machine learning applications. The use of ROC curves and AUCs in this study is inspired by Ref. [17]. We first show the ROC curves of various classifiers using the $t\bar{t}H$ analysis datasets of 20000 events and 15 input variables in Figure 4. Each curve represents results averaged over 60 statistically independent datasets. Figure 4 (a) overlays the results of the QSVM-Kernel algorithm (from the Google framework), the classical SVM algorithm and

the classical BDT algorithm. Figure 4 (b) overlays the QSVM-Kernel results from the Google framework, IBM framework and Amazon framework. Here the QSVM-Kernel classifiers employ 15 qubits on the quantum simulators. We observe that, for these $t\bar{t}H$ analysis datasets, the QSVM-Kernel performances are similar to the performances given by the two commonly used classical machine learning algorithms. Furthermore, the three quantum computer simulators, from the Google framework, IBM framework and Amazon framework, provide identical classification performances using the QSVM-Kernel algorithm.

Based on the classifier discriminant, we could perform an event selection in order to maximize $S/\sqrt{(B)}$, where S is the number of selected signal events and B is the number of selected background events. $S/\sqrt{(B)}$ is an approximation of the statistical significance of the signal process, and usually correlated with the AUC of the classifier. For the above-mentioned QSVM-Kernel classifier in the $t\bar{t}H$ analysis, if applying a selection with a signal acceptance of 70%, approximately 92% of background events will be rejected and hence $S/\sqrt{(B)}$ will be improved by around 150% with respect to no selection.

Our observation becomes more clear in Figure 5, where we study the AUC for various classifiers as a function of the $t\bar{t}H$ analysis dataset size (10000 to 50000 events). Figure 5 (a) shows the results of the QSVM-Kernel (from the Google framework), the classical SVM and the classical BDT. Figure 5 (b) further shows the difference between the QSVM-Kernel algorithm and the classical algorithms. Figure 5 (c) shows the QSVM-Kernel results from the Google framework, IBM framework and Amazon framework. Here all the classifiers use the same 15 variables and the QSVM-Kernel classifiers employ 15 qubits on the quantum simulators. The quoted AUCs are averaged over 60 statistically independent datasets and the quoted errors are the standard deviations for the AUCs of the 60 datasets. We find that, the performance of all methods improve with increasing dataset size. For 15 qubits and up to 50000 events, the QSVM-Kernel algorithm performs similarly to the classical SVM and classical BDT algorithms. Furthermore, the QSVM-Kernel performances from the three different quantum computer simulators (Google, IBM and Amazon) are comparable.

We also investigate the AUCs of the QSVM-Kernel algorithm as a function of the number of qubits (10 to 20 qubits), as shown in Figure 6. The number of qubits is equal to the number of input variables using PCA as described in Section 3. The 60 statistically independent $t\bar{t}H$ analysis datasets of 20000 events are used in this study. In Figure 6 (a), we compare the results of the QSVM-Kernel (from the Google framework) with the classical SVM and classical BDT using the same input variables. In Figure 6 (b), we further display the

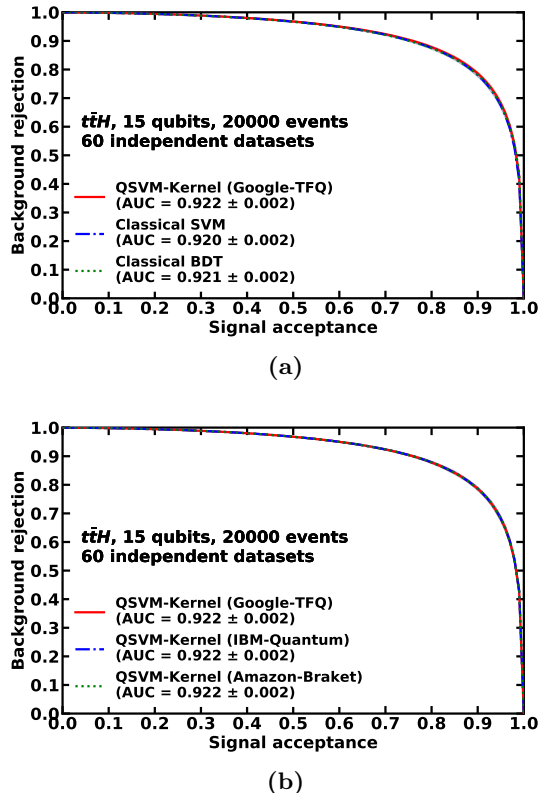


Figure 4. ROC curves of various classifiers using the $t\bar{t}H$ analysis datasets of 20000 events and 15 input variables. Each curve represents results averaged over 60 statistically independent datasets. (a) overlays the results of the QSVM-Kernel algorithm (on the *qsim Simulator* from the Google TensorFlow Quantum framework), the classical SVM algorithm and the classical BDT algorithm. (b) overlays the QSVM-Kernel results on the *qsim Simulator* from the Google TensorFlow Quantum framework, the *Statevector Simulator* from the IBM Quantum framework and the *Local Simulator* from the Amazon Braket framework. Here the QSVM-Kernel classifiers employ 15 qubits on the quantum simulators.

difference between the QSVM-Kernel results and the classical machine learning results. In Figure 6 (c), again, we compare the Google framework, IBM framework and Amazon framework for the QSVM-Kernel results. We find that, the QSVM-Kernel result with 15 qubits is better than 10 qubits and similar to 20 qubits. For 10 to 20 qubits and 20000 events, the performance of the QSVM-Kernel algorithm is similar to that of the classical SVM algorithm. Again, the three quantum computer simulators (Google, IBM and Amazon) yield the same classification power.

B. Results from Quantum Computer Hardware

After the studies using simulation of the ideal quantum computers, it is now of great interest to assess the quantum machine learning performances on today's

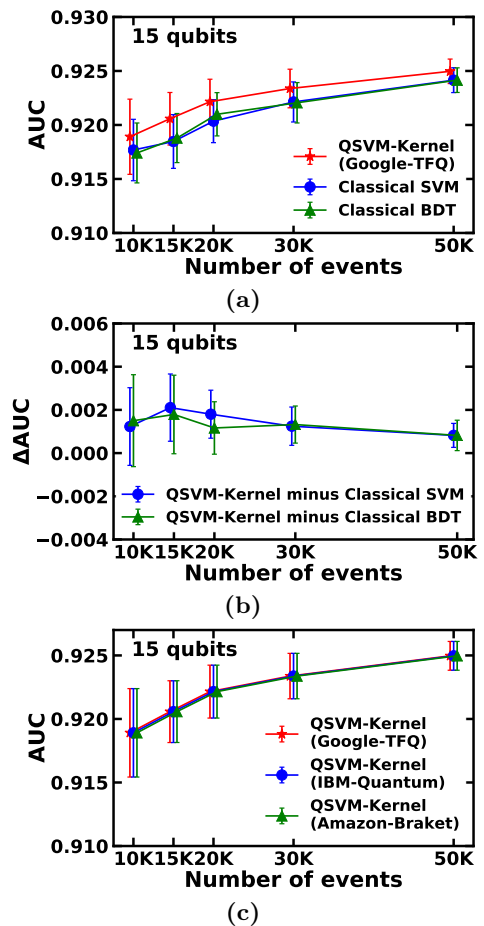


Figure 5. The AUC for various classifiers as a function of the $t\bar{t}H$ analysis dataset size (10000 to 50000 events). (a) shows the results of the QSVM-Kernel (on the *qsim Simulator* from the Google TensorFlow Quantum framework), the classical SVM and the classical BDT. (b) further shows the difference between the QSVM-Kernel algorithm and the classical algorithms. (c) shows the QSVM-Kernel results on the *qsim Simulator* from the Google TensorFlow Quantum framework, the *Statevector Simulator* from the IBM Quantum framework and the *Local Simulator* from the Amazon Braket framework. Here all the classifiers use the same 15 variables and the QSVM-Kernel classifiers employ 15 qubits on the quantum simulators. The quoted AUCs are averaged over 60 statistically independent datasets and the quoted errors are the standard deviations for the AUCs of the 60 datasets.

noisy quantum computer hardware. For the $t\bar{t}H$ physics analysis, we employ the QSVM-Kernel algorithm on the IBM "*ibmq-paris*" quantum computer hardware. "*ibmq-paris*" is a 27-qubit quantum processor based on superconducting electronic circuits. The qubit map of the "*ibmq-paris*" quantum system [40] is shown in Figure 7. Due to limited access time available to us, we performed six runs using 15 qubits on "*ibmq-paris*". Each run processes a statistically independent dataset of 100 events. For these six runs, the average running time on the quantum hardware is approximately 680

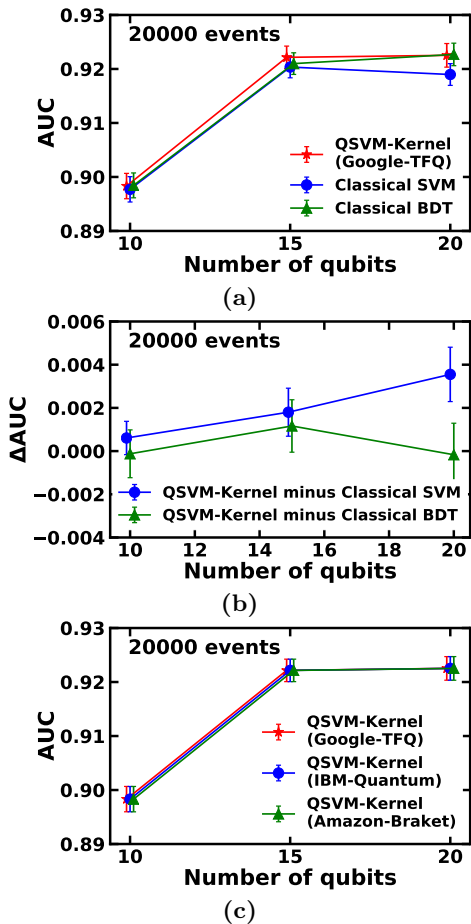


Figure 6. AUCs of the QSVM-Kernel algorithm as a function of the number of qubits (10 to 20 qubits). The number of qubits is equal to the number of input variables. The 60 statistically independent $t\bar{t}H$ analysis datasets of 20000 events are used in this study. In (a), we compare the results of the QSVM-Kernel classifier (on the *qsim Simulator* from the Google TensorFlow Quantum framework) with the results of the classical SVM and classical BDT classifiers using the same input variables. In (b), we further display the difference between the QSVM-Kernel results and the classical machine learning results. In (c), we compare the *qsim Simulator* from the Google TensorFlow Quantum framework, the *StatevectorSimulator* from the IBM Quantum framework and the *Local Simulator* from the Amazon Braket framework for the QSVM-Kernel results.

minutes. With more advanced quantum hardware in the future, the running time is expected to be significantly reduced. The quantum circuit of the hardware runs is kept the same as for the simulator runs, while the SVM regularization hyperparameter is separately optimized for hardware and simulator runs. To reduce statistical uncertainties in evaluating kernel entries on quantum hardware, we use 8192 measurement shots for every kernel entry.

In Figure 8, we present the ROC curve of the QSVM-

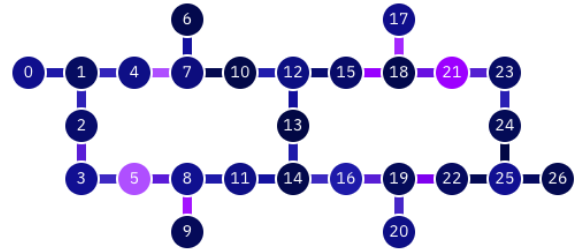


Figure 7. The qubit map of the “*ibmq-paris*” quantum system [40]. The (darker) colors indicate (lower) readout error rates of the qubits and CNOT error rates of the connections. Our study uses qubits 3, 5, 8, 11, 14, 16, 19, 22, 25, 24, 23, 18, 15 and 12.

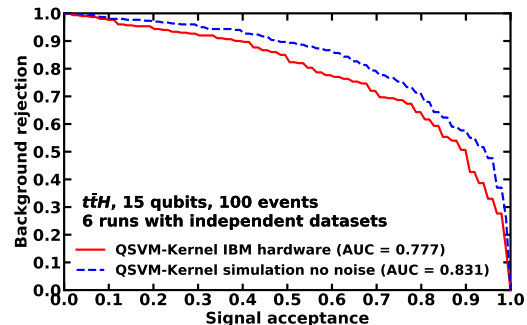


Figure 8. ROC curve of the QSVM-Kernel classifier with the “*ibmq-paris*” quantum computer hardware using the $t\bar{t}H$ analysis datasets of 100 events. For comparison, we overlay the ROC curve with the *StatevectorSimulator* from the IBM Quantum framework using the same datasets. The results are averaged over the six hardware runs. All the QSVM-Kernel classifiers use 15 qubits and the same 15 variables.

Kernel classifier with the “*ibmq-paris*” quantum computer hardware using the $t\bar{t}H$ analysis datasets of 100 events. For comparison, we overlay the ROC curve with the *StatevectorSimulator* from the IBM Quantum framework using the same datasets. The results are averaged over the six hardware runs. All the QSVM-Kernel classifiers use 15 qubits and the same 15 variables. In Figure 9, we compare the ROC curve with the “*ibmq-paris*” quantum computer hardware and the ROC curve with the *StatevectorSimulator* for each of the six hardware runs. With small training samples of 100 events, the performance achieved by the “*ibmq-paris*” quantum computer hardware is promising and approaching the noiseless quantum computer simulator. The difference between the hardware performance and the simulator performance is likely due to the effect of quantum hardware noise and fluctuates among our hardware runs.

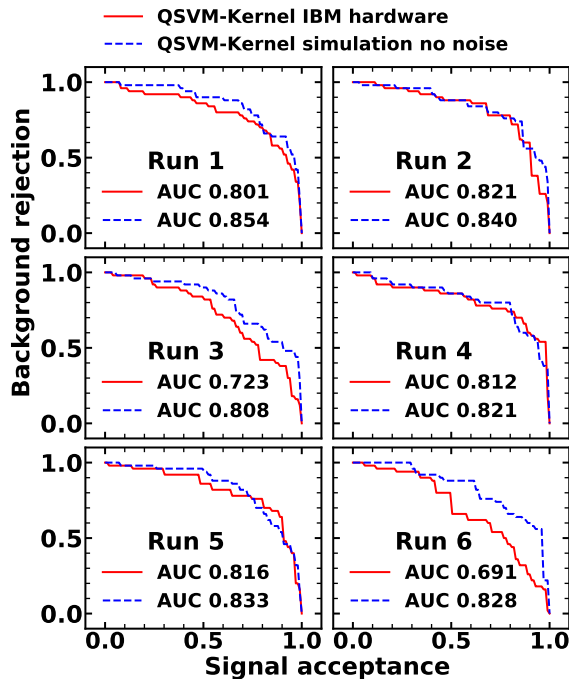


Figure 9. ROC curve with the “*ibmq-paris*” quantum computer hardware and ROC curve with the *StatevectorSimulator* from the IBM Quantum framework for each of the six hardware runs. Each run processes a statistically independent dataset of 100 events. All the QSVm-Kernel classifiers are using 15 qubits and the same 15 variables.

V. CONCLUSION

In this study, we have successfully employed the quantum support vector machine kernel (QSVm-Kernel) method in the $t\bar{t}H$ (Higgs boson production in associa-

tion with a top quark pair) physics analysis, a recent LHC flagship physics analysis, on gate-model quantum computer simulators and hardware. The simulation study has been performed using the Google TensorFlow Quantum framework, IBM Quantum framework and Amazon Braket framework. We have overcome the challenges of intensive computing resources in the cases of up to 20 qubits and up to 50000 events on the quantum computer simulators, in order to perform quantum machine learning studies on physics datasets that closely resemble those used in the official ATLAS publication [8]. The QSVm-Kernel method achieves good classification performance that is similar to the performances of the classical machine learning methods currently used in LHC physics analyses, classical SVM and classical BDT for example. On the “*ibmq-paris*” superconducting quantum computer hardware, we have also employed the QSVm-Kernel algorithm using 100 events and 15 qubits to assess the effect of quantum hardware noise. The performance achieved on the “*ibmq-paris*” quantum hardware is promising and is approaching the performance from the noiseless quantum simulators.

Our quantum simulation result gives an example that quantum machine learning performs as well as its classical counterpart using three different platforms (Google, IBM and Amazon) for realistic high energy physics analysis datasets. Furthermore, our result on noisy quantum hardware provides important validation for the result on noiseless quantum simulators. Our studies confirm that the QSVm-Kernel algorithm can use the large dimensionality of the quantum Hilbert space to replace the classical feature space. In the future, large improvement in computational speed and reduction in device noise on quantum computing hardware will likely be achieved and lead to quantum advantage in quantum machine learning applications. With the large investments in quantum computing and fierce competitions in technology, this expectation is realistic. Therefore, we predict that quantum machine learning will become a powerful tool for data analysis in High Energy Physics.

-
- [1] ATLAS Collaboration G. Aad et al., Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC, *Physics Letters B* **716**, 1 (2012).
 - [2] CMS Collaboration, S. Chatrchyan et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, *Physics Letters B* **716**, 30 (2012).
 - [3] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics, *Nature* **560**, 41 (2018).
 - [4] K. Albertsson *et al.*, Machine Learning in High Energy Physics Community White Paper, *J. Phys. Conf. Ser.* **1085**, 022008 (2018), arXiv:1807.02876 [physics.comp-ph].
 - [5] D. Guest, K. Cranmer, and D. Whiteson, Deep Learning and its Application to LHC Physics, *Ann. Rev. Nucl. Part. Sci.* **68**, 161 (2018), arXiv:1806.11484 [hep-ex].
 - [6] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019), arXiv:1903.10563 [physics.comp-ph].
 - [7] D. Bourilkov, Machine and Deep Learning Applications in Particle Physics, *Int. J. Mod. Phys. A* **34**, 1930019 (2020), arXiv:1912.08245 [physics.data-an].
 - [8] ATLAS Collaboration, M. Aaboud et al., Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector, *Physics Letters B* **784**, 173 (2018).
 - [9] CMS Collaboration, A. M. Sirunyan et al., Observation of $t\bar{t}H$ production, *Physical Review Letters* **120**, 231801 (2018).

- [10] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195202 (2017).
- [11] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Power of data in quantum machine learning, *Nature Communications* **12**, 2631 (2021).
- [12] W. Guan, G. Perdue, A. Pesah, M. Schuld, K. Terashi, S. Vallecorsa, and J.-R. Vlimant, Quantum Machine Learning in High Energy Physics, (2020), arXiv:2005.08582 [quant-ph].
- [13] J. Gambetta, Ibms roadmap for scaling quantum technology (2020), <https://www.ibm.com/blogs/research/2020/09/ibm-quantum-roadmap/>.
- [14] S. Shankland, Quantum computer makers like their odds for big progress (2019), <https://www.cnet.com/tech/computing/quantum-computer-makers-like-their-odds-for-big-progress-soon/>.
- [15] T. Q. Hubs, Ionqs roadmap up to 2025, <https://thequantumhubs.com/ionqs-roadmap-up-to-2025-2/>.
- [16] High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1 **4/2017**, 10.23731/CYRM-2017-004 (2017).
- [17] A. Mott, J. Job, J.-R. Vlimant, D. Lidar, and M. Spiropulu, Solving a Higgs optimization problem with quantum annealing for machine learning, *Nature* **550**, 375 (2017).
- [18] K. Terashi, M. Kaneda, T. Kishimoto, M. Saito, R. Sawada, and J. Tanaka, Event Classification with Quantum Machine Learning in High-Energy Physics, *Comput. Softw. Big Sci.* **5**, 2 (2021).
- [19] S. L. Wu, J. Chan, W. Guan, S. Sun, A. Wang, C. Zhou, M. Livny, F. Carminati, A. Di Meglio, A. C. Y. Li, J. D. Lykken, P. Spentzouris, S. Y.-C. Chen, S. Yoo, and T.-C. Wei, 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, *Journal of Physics G: Nuclear and Particle Physics* (2021), accepted, arXiv:2012.11560 [quant-ph].
- [20] V. Havlek, A. D. Crookes, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature* **567**, 209 (2019).
- [21] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, *Physical Review Letters* **122**, 040504 (2019).
- [22] S. Aaronson and A. Ambainis, Forrelation: A problem that optimally separates quantum from classical computing, *SIAM Journal on Computing* **47**, 982 (2018).
- [23] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, arXiv:2010.02174 (2020).
- [24] B. E. Boser, I. M. Guyon, and V. N. Vapnik, A training algorithm for optimal margin classifiers, in *Proceedings of the fifth annual workshop on Computational learning theory* (1992) pp. 144–152.
- [25] V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer Science & Business Media, 2013).
- [26] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *Journal of High Energy Physics* **07**, 079 (2014).
- [27] T. Sjöstrand, S. Mrenna, and P. Skands, PYTHIA 6.4 physics and manual, *Journal of High Energy Physics* **05**, 026 (2006).
- [28] J. De Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaitre, A. Mertens, M. Selvaggi, D. . Collaboration, *et al.*, DELPHES 3: a modular framework for fast simulation of a generic collider experiment, *Journal of High Energy Physics* **02**, 057 (2014).
- [29] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559 (1901).
- [30] I. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics (Springer, 2002).
- [31] M. Broughton *et al.*, TensorFlow Quantum: A Software Framework for Quantum Machine Learning, arXiv:2003.02989 (2020).
- [32] G. Aleksandrowicz *et al.*, Qiskit: An Open-source Framework for Quantum Computing (2019).
- [33] Amazon Braket, <https://aws.amazon.com/braket/> (2019).
- [34] D. M. Allen, The relationship between variable selection and data augmentation and a method for prediction, *Technometrics* **16**, 125 (1974).
- [35] M. Stone, Cross-validated choice and assessment of statistical predictions, *Journal of the Royal Statistical Society, Series B (Methodological)* **36**, 111 (1974).
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* **12**, 2825 (2011).
- [37] J. H. Friedman, Stochastic gradient boosting, *Computational Statistics & Data Analysis* **38**, 367 (2002).
- [38] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction* (Springer Science & Business Media, 2009).
- [39] T. Chen and C. 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.
- [40] IBM Quantum, <https://quantum-computing.ibm.com/> (2021).

Acknowledgements 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 and by the Vilas foundation at the University of Wisconsin. This project is also 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 LHC flagship analysis used in this

publication.