Quantum adiabatic machine learning with zooming

Alexander Zlokapa,[1] Alex Mott,[2] Joshua Job,[3] Jean-Roch Vlimant,[1] Daniel Lidar,[4] and Maria Spiropulu[1]

[1]*Division of Physics, Mathematics & Astronomy, Alliance for Quantum Technologies,*
California Institute of Technology, Pasadena, CA 91125, USA
[2]*DeepMind Technologies, London, UK*
[3]*Lockheed Martin Advanced Technology Center, Sunnyvale, CA 94089, USA*
[4]*Departments of Electrical and Computer Engineering, Chemistry,*
and Physics & Astronomy, and Center for Quantum Information Science & Technology,
University of Southern California, Los Angeles, CA 90089, USA

Recent work has shown that quantum annealing for machine learning, referred to as QAML, can perform comparably to state-of-the-art machine learning methods with a specific application to Higgs boson classification. We propose QAML-Z, a novel algorithm that iteratively zooms in on a region of the energy surface by mapping the problem to a continuous space and sequentially applying quantum annealing to an augmented set of weak classifiers. Results on a programmable quantum annealer show that QAML-Z matches classical deep neural network performance at small training set sizes and reduces the performance margin between QAML and classical deep neural networks by almost 50% at large training set sizes, as measured by area under the ROC curve. The significant improvement of quantum annealing algorithms for machine learning and the use of a discrete quantum algorithm on a continuous optimization problem both opens a new class of problems that can be solved by quantum annealers and suggests the approach in performance of near-term quantum machine learning towards classical benchmarks.

I. INTRODUCTION

Machine learning has gained an increasingly important role in scientific discovery across chemistry, biology, environmental science, and physics [1–5], including in the discovery of the Higgs boson [6]. Various quantum computing algorithms have been proposed for machine learning [7], including support vector machines, principal component analysis, least-squares fitting, topological analysis, and other optimization problems [8–13]. Many of these algorithms include strict data assumptions that provide critical caveats regarding sparsity, state preparation, and rank [14, 15]. Moreover, fault-tolerant quantum computing will be required to implement the large quantum circuits necessary for the proposed algorithms, which has not yet been experimentally established at a scale necessary for the implementation of the proposed algorithms. Similarly, quantum random access memory (qRAM) is typically required to store classical data, but engineering challenges persist in developing a sufficiently large qRAM [16].

One promising near-term avenue for quantum machine learning is quantum annealing [17] (for recent reviews see [18–20]) which can, e.g., perform binary classification [21, 22], learn Bayesian network structure [23], implement quantum Boltzmann machines [24], and train deep generative models [25]. Quantum annealing is the only current quantum computing paradigm that has resulted in architectures with a large enough number of — albeit relatively noisy — qubits [26–28] to address both real-world and fundamental science problems, e.g., in air traffic control [29], computational biology [30–32], and high energy physics [33–35]. Under the adiabatic theorem of quantum mechanics, quantum annealing evolves from an initial transverse field Hamiltonian to the target prob-

lem Hamiltonian, ensuring that the system remains in the ground state if the system is perturbed slowly enough, as given by the energy gap between the ground state and the first excited state [36–38]. The ground state of the problem Hamiltonian is then the solution (as in adiabatic quantum computing [39, 40]), although thermal excitations may move the system out of the ground state [41–47],which can be beneficial [48–51]. It is crucial to observe that evidence of a quantum speedup in quantum annealing remains uncertain [52–54], although quantum phenomena have been observed in D-Wave quantum annealers [55–57]. While this remains a speculative topic, quantum annealers may exhibit advantages other than a speedup, such as sampling from non-equilibrium distributions prepared during the anneal [58–60].

Here we propose a novel quantum algorithm inspired by the previous state-of-the-art quantum annealing for machine learning (QAML) algorithm [33], which constructs a single strong classifier from a linear combination of weak classifiers with binary coefficients of 1 or 0. We propose two modifications to QAML — zooming into the energy surface to optimize real-valued coefficients and artificially augmenting the set of weak classifiers to create a stronger ensemble — and implement the proposed algorithm (QAML-Z) on the D-Wave quantum annealer to benchmark the results on a Higgs classification problem, with available source code [61] and data [62].

II. QAML-Z ALGORITHM

A. Background: QAML Algorithm

In the original QAML algorithm, a training set with S examples of labeled data $\{\mathbf{x}_\tau, y_\tau\}$ (where $\mathbf{x}_\tau$ is an in-

put vector and $y_\tau = \pm 1$ is a binary label for signal and background) is optimized with a set of N weak classifiers c_i, each of which gives $c_i(\mathbf{x}_\tau) = \pm 1/N$ for a signal or background prediction. Given spins $s_i \in \{0, 1\}$ obtained by transforming up/down spins, let $R(\mathbf{x}_\tau)$ be a strong classifier given by

$$R(\mathbf{x}_\tau) = \sum_{i=1}^{N} s_i c_i(\mathbf{x}_\tau), \quad (1)$$

i.e., an ensemble of the weak classifiers where each weak classifier is either turned on or off (weight 1 or 0). To minimize classification error, we simply minimize the distance between y and R:

$$||\mathbf{y} - \mathbf{R}||^2 = \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i c_i(\mathbf{x}_\tau) \right|^2 \quad (2a)$$

$$= ||\mathbf{y}||^2 - 2 \sum_{i=1}^{N} \sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) y_\tau$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) s_j c_j(\mathbf{x}_\tau). \quad (2b)$$

Removing the spin-independent term $||\mathbf{y}||^2$ and the self-spin interactions $c_i^2(\mathbf{x}_\tau)$ to construct a problem suitable for quantum annealing, we rewrite the Hamiltonian as follows (scaling by a factor of 2 for convenience after manipulating indices):

$$H = \sum_{i=1}^{N} \sum_{j>i}^{N} \sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) s_j c_j(\mathbf{x}_\tau) - \sum_{i=1}^{N} \sum_{\tau=1}^{S} s_i c_i(\mathbf{x}_\tau) y_\tau. \quad (3)$$

For convenience, we define the variables:

$$C_{ij} = \sum_{\tau=1}^{S} c_i(\mathbf{x}_\tau) c_j(\mathbf{x}_\tau), \quad (4)$$

$$C_i = \sum_{\tau=1}^{S} c_i(\mathbf{x}_\tau) y_\tau. \quad (5)$$

Hence, in the original QAML algorithm, the following Ising model Hamiltonian is minimized after transforming the range to $s_i \in \{-1, 1\}$, adding an additional λ regularization hyperparameter to penalize nonzero s_i [22]:

$$H = \sum_{i=1}^{N} \left(\lambda - C_i + \frac{1}{2} \sum_{j>i}^{N} C_{ij} \right) s_i + \frac{1}{4} \sum_{i=1}^{N} \sum_{j>i}^{N} C_{ij} s_i s_j, \quad (6)$$

We observe the following limitations in the QAML algorithm: i) arbitrary linear combinations of weak classifiers c_i are forbidden because the strong classifier R is simply formed by turning weak classifiers c_i on or off; ii) the diversity of the ensemble is limited by the selection of weak classifiers. If the set of weak classifiers can be expanded, more nuanced ensembles with more complex decision boundaries can be learned.

B. Zooming Extension

By iteratively performing quantum annealing, the binary weights on the weak classifiers can be made continuous, resulting in a stronger classifier. This is achieved by performing a search on the real numbers, effectively zooming in on a region of the energy surface each iteration (Figure 1). We denote the zooming variant of quantum annealing for machine learning as QAML-Z. Under this reformulation, the weights of the classifiers may be extended from the set $\{0, 1\}$ to the continuous interval $[-1, 1]$, enabling the subtraction of classifiers to reduce cross-correlations between weak classifiers.
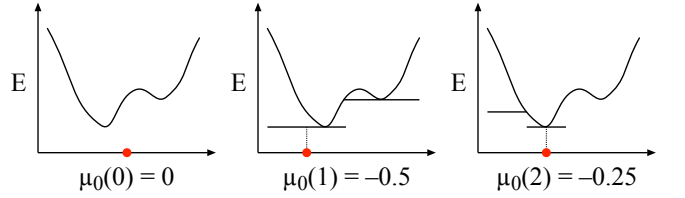

FIG. 1. **Zooming extension.** While QAML only performs one anneal, QAML-Z iteratively updates the weight μ (indicated by the red dot) of a weak classifier (index 0 in the diagram) in the strong classifier ensemble by performing a binary search over the energy surface using spin up/down outcomes.

Let each qubit have a mean $\mu_i(t)$ (starting at $\mu_i(0) = 0$ for all i) and let the search breadth be $\sigma(t) = b^t$, where $t = 0, 1, ..., T-1$ for T iterations and $0 < b < 1$ is a free parameter. Each iteration, the Hamiltonian is centered around the previous mean and the search breadth is narrowed. Receiving spin up or spin down corresponds to shifting the new mean either right or left by a distance given by the search breadth. The weight given to each classifier is thus updated according to the old mean and consequent shift, resulting in a modified Hamiltonian according to the substitution $s_i c_i(\mathbf{x}_\tau) \to \sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau)$. The full expression is:

$$H(t) = \sum_{i=1}^{N} \sum_{j>i}^{N} \sum_{\tau=1}^{S} \left(\sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau) \right)$$

$$\times \left(\sigma(t) s_j c_j(\mathbf{x}_\tau) + \mu_j(t) c_j(\mathbf{x}_\tau) \right)$$

$$- \sum_{i=1}^{N} \sum_{\tau=1}^{S} \left(\sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau) \right) y_\tau \quad (7a)$$

$$= \sum_{i=1}^{N} \left(-C_i + \sum_{j=1}^{N} \mu_j(t) C_{ij} \right) \sigma(t) s_i$$

$$+ \sum_{i=1}^{N} \sum_{j>i}^{N} C_{ij} \sigma^2(t) s_i s_j, \quad (7b)$$

where Eq. (7b) is derived after dropping constants from the Hamiltonian and applying the same C_i and C_{ij} no-

tation as in QAML. This new Hamiltonian may be iteratively optimized for $t = 0, 1, \ldots, T - 1$ to update $\mu_i(t+1) = \mu_i(t) + s_i\sigma(t+1)$, resulting in the new strong classifier $R(\mathbf{x}_\tau) = \sum_{i=1}^{N} \mu_i(T-1)c_i(\mathbf{x}_\tau)$.

Since the zooming algorithm increases the possibility of overfitting, we propose a two-step randomization procedure to regularize the iterative process. After each iteration, for each qubit s_i that the energy worsens by the update (i.e., $E[\mu_1(t+1), \ldots, \mu_i(t+1), \ldots, \mu_N(t+1)] > E[\mu_1(t), \ldots, \mu_i(t), \ldots \mu_N(t)]$), we apply the flip $s_i \to -s_i$ with monotonically decreasing probability $p_f(t)$, similarly to an annealing schedule in classical simulated annealing. Subsequently, all qubits are uniformly randomly flipped from s_i to $-s_i$ with probability $q_f(t)$ where $q_f(t) < p_f(t)$ for all t, akin to a cluster-flip move in variants of simulated annealing. This serves to prevent the strong classifier from overfitting as well as to push it out of local minima. The functions p_f and q_f are specified in the supplementary code.

To take full advantage of these continuous weights, we augment the set of original weak classifiers $h_i(\mathbf{x}_\tau)$ that returns a value in $[-1, 1]$. For each h_i, multiple classifiers are generated by shifting the threshold to round to ± 1:

$$c_{il}(\mathbf{x}_\tau) = \text{sgn}(h_i(\mathbf{x}_\tau) + \delta l)/N, \qquad (8)$$

where N is the number of classifiers, $l \in \mathbb{Z} : -A \leq l \leq A$ is the offset and δ is the step size. With a larger set of weak classifiers to ensemble into a strong classifier, a more complex decision boundary may be formed.

With the augmented set of classifiers, the Hamiltonian is now given by:

$$H(t) = \sum_{l=-A}^{A} \left[\sum_{i=1}^{N} \left(-C_{il} + \sum_{j>i}^{N} \mu_{jl}(t)C_{ijl} \right) \sigma(t)s_{il} \right. $$
$$\left. + \sum_{i=1}^{N} \sum_{j>i}^{N} C_{ijl}\sigma^2(t)s_{il}s_{jl} \right], \qquad (9)$$

where $H(t)$ is iteratively optimized for $t = 0, 1, \ldots$ to update $\mu_{il}(t+1) = \mu_{il}(t) + s_{il}\sigma(t+1)$. Similarly to before, we have defined:

$$C_{ijl} = \sum_{\tau=1}^{S} c_{il}(\mathbf{x}_\tau)c_{jl}(\mathbf{x}_\tau), \qquad (10)$$

$$C_{il} = \sum_{\tau=1}^{S} c_{il}(\mathbf{x}_\tau)y_\tau. \qquad (11)$$

Quantum annealing yields a distribution of excited states, allowing the construction of a stronger classifier than one based purely on ground state results. We take the supremum over the set of excited states' background rejection values for each efficiency in the receiver operating characteristic (ROC) curve according to a validation set of equal size to the training set.

In the experimental demonstration of the algorithm, we set an offset of $A = 3$ and a step size of $\delta = 0.0075$.

Additionally, we set the zoom parameter $b = \frac{1}{2}$ to perform a binary search over the real numbers. Due to the definition of $\sigma(t)$, the marginal impact of each iteration follows an exponential decay. Thus, QAML-Z was trained for only 8 iterations.

III. HIGGS BOSON CLASSIFICATION

As an application of QAML-Z, we revisit the Higgs optimization problem, in which kinematic variables describing diphoton processes corresponding either to a Higgs boson decay (signal) or other Standard Model processes (background) are used to identify simulated Higgs bosons [33]. The simulation of $H \to \gamma\gamma$ is limited to the main process of gluon fusion, while sub-leading contributions are not included. We evaluated the performance of the QAML-Z algorithm on the programmable D-Wave 2X quantum annealer at the University of Southern California's Information Sciences Institute with 1098 physical qubits [28]. The Ising model is generated from weak classifiers developed from kinematic variables such as transverse momentum, pseudorapidity, and the invariant mass of the diphoton system. However, the QAML-Z algorithm augments this set of weak classifiers with regular offsets of the decision boundary, as described above. In our analysis, we seek primarily to use the Higgs classification problem as a context for providing a clear comparison between QAML-Z and state-of-the-art algorithms in both quantum and classical machine learning. A precise measurement of the Higgs mechanism is beyond the scope of this paper.

A. Quantum Annealing on D-Wave

From the 1098 physical qubits of the D-Wave 2X annealer, only 33 fully connected logical qubits are available due to the Chimera graph architecture. Implementing the exact Ising model with all cross-terms on the scale of the Higgs optimization problem would require hundreds of fully connected qubits, therefore we prune the cross-terms in the Ising Hamiltonian, retaining only the largest 5% of weights. This reduces the sensitivity to analog errors associated with small weights [63] and also allows a minor embedding operation [44–47] in combination with the classical polynomial-time `fix_variables` procedure in the D-Wave API to program the problem on the quantum annealer. Each logical qubit is mapped to a chain of physical ferromagnetically coupled qubits on the D-Wave device, where the internal coupling of each chain may be set to prevent thermal excitations and other noise from breaking the chain while still ensuring that the Hamiltonian drives the system dynamics [64]. The coupling within chains is scaled to the largest coupling in the Hamiltonian and decayed with increasing iteration number. Random errors on the local fields and couplers are reduced by randomizing the encoding by sign flips.

Annealing is performed with a 5 μs anneal time, with minimal variation in performance observed for longer anneal times (not shown). The anneal times were selected to attempt to achieve high performance with the shortest anneal times possible using the D-Wave 2X device, suggesting that future quantum annealers may achieve a wall clock time advantage over simulated annealing if the performance is sustained with lower anneal times.

As in QAML, we use an ensemble of excited states to strengthen the classifier. To select the excited states, we place two criteria: a maximum distance d to the lowest-energy state found (i.e., an excited state must have an energy less than $(1 - d)E_{\text{ground}}$ for $E_{\text{ground}} < 0$ or less than $(1 + d)E_{\text{ground}}$ for $E_{\text{ground}} > 0$), and a maximum total number of excited states n_e to be selected. To prevent an exponential increase in the tree of excited states generated by the zooming algorithm, we also decay the values of d and n_e by iteration number. The final classifier is then defined by maximizing the area under the ROC curve on a validation set (equivalent to the validation set used for DNN hyperparameter tuning), selecting the best-performing excited states for different false positive rates.
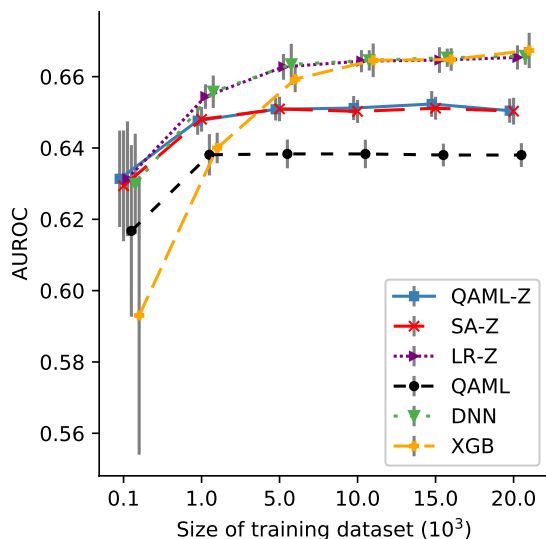
B. Results

FIG. 2. **Area under the ROC curve for the QAML-Z extension, simulated annealing (SA-Z), a logistic regression (LR-Z), the original QAML, a deep neural network (DNN) and XGBoost (XGB) [65] as a function of training set size.** While QAML-Z matches DNN performance at small training set sizes, it decreases the margin between QAML and DNN by 47% for the largest training sets. Error bars indicate 1σ error, including both variation over training sets and statistical error estimated by reweighting samples from a Poisson distribution.

Compared to the QAML algorithm, the area under the receiver operating characteristic curve (AUROC) is significantly improved by QAML-Z on all training set sizes (Figure 2). We select the best-performing classical classifiers (a deep neural network and XGBoost) from the QAML Higgs optimization benchmark, although we optimize additional parameters of the classical algorithms to further improve their performance from Ref. [33]. A logistic regression (LR-Z) directly optimizes the mean-squared error of classification over the set of augmented classifiers that QAML-Z is applied to. When compared to classical simulated annealing (SA-Z), QAML-Z performs slightly better (see Figure 4).
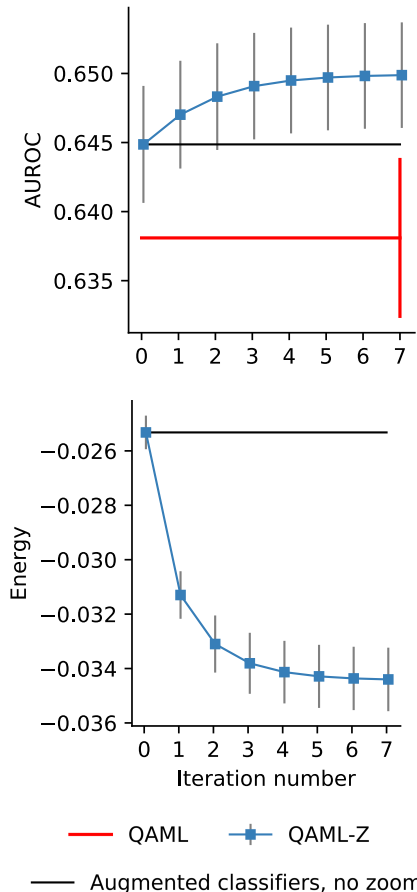

FIG. 3. **QAML-Z performance on the test set vs. zooming iteration number (training set size of 1000).** Top: significant improvements by QAML-Z can be separately seen for classifier augmentation (black) and zooming (blue) over the original QAML algorithm (red). Bottom: Ising model energy on the test set improves monotonically, indicating negligible overfitting. Error bars indicate 1σ error.

We observe the effectiveness of both the zooming and augmentation aspects of QAML-Z (Figure 3). The area under the ROC curve illustrates both the impact of classifier augmentation and the impact of zooming, showing advantages in both the classifier augmentation and

zooming methodologies. Examining the normalized Ising model energy as a function of iteration number, the zooming algorithm is also shown to monotonically decrease the Hamiltonian energy with additional anneals.

C. Simulated Annealing Benchmark

Given the analogue of quantum annealing to simulated annealing [17], we also implement the proposed zooming algorithm in a simulated annealing framework, reporting on simulated annealing with zooming (SA-Z). To attempt to match the improved quantum annealing performance, we also propose simulated annealing with excited states and zooming (SAE-Z), in which the supremum over a set of excited states from simulated annealing is used to improve the area under ROC curve in the same manner as in the quantum algorithm. While a ground state solution minimizes error on the training set, it may overfit to the training data and cause poor generalization on the test set. Hence, the inclusion of excited states — either thermal noise in simulated annealing or sampled from the quantum annealer — can improve performance on the test set.
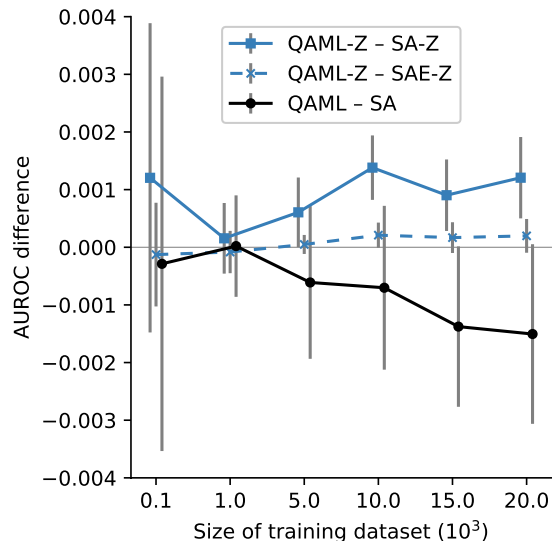

FIG. 4. **Comparison of quantum annealing and simulated annealing for the new and original algorithms, measured by area under ROC curve (AUROC).** Although QAML-Z outperforms QAML and SA-Z, the inclusion of excited states in the SAE-Z variant reproduces QAML-Z performance to one standard deviation. Error bars indicate 1σ error.

We perform simulated annealing using the Metropolis update rule, flipping a random spin to construct a trial spin vector $\vec{s}'$ from the spin vector $\vec{s}$.[66] If the energy $H(\vec{s}') < H(\vec{s})$, then the new vector $\vec{s}'$ is accepted with probability 1. However, if $H(\vec{s}') > H(\vec{s})$, the trial vec-

tor is accepted with probability $\exp[-\beta(H(\vec{s}') - H(\vec{s}))]$. After randomly selecting a spin to flip N times (where $\vec{s}$ has N spins), a sweep has been completed. The inverse temperature β is stepped with a linear inverse temperature schedule from $\beta_i = 0.1$ to $\beta_f = 5$ over $W = 1000$ sweeps, incrementing the temperature by $\frac{\beta_f - \beta_i}{W}$ after each sweep. This process is repeated 1000 times, and the lowest-energy state is selected in the SA-Z algorithm. Temperature schedules reaching β as large as 10 and performing up to 100,000 sweeps per read were found to have no significant impact on the results. To assemble excited states for the SAE-Z benchmark, we perform 5000 sweeps for 5000 reads and select excited states using the same criteria as for quantum annealing.
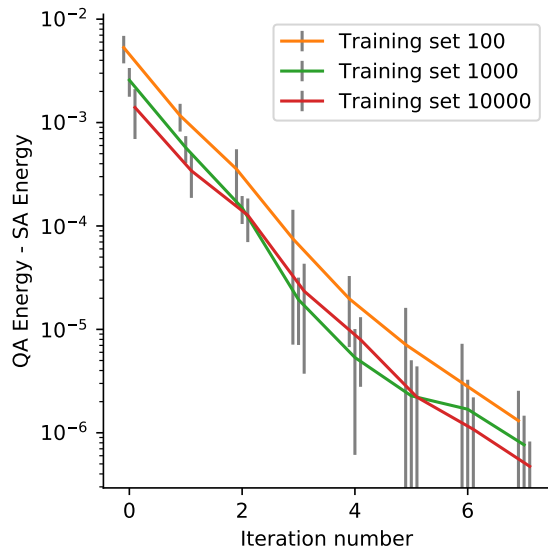

FIG. 5. **Difference between the lowest energy of quantum annealing (QA) and simulated annealing (SA).** SA finds a lower minimum energy than QA given an identical initial Hamiltonian. Error bars indicate 1σ error.

QAML-Z performs better than SA-Z on all training sets, with a statistically significant advantage at larger training set sizes (Figure 4). This suggests that both simulated and quantum annealing methods found similar ground states at the end of the zooming procedure, although they likely took different paths to the final state due to the fact that SA evolves purely under the classical Hamiltonian, whereas QA evolves under the transverse field as well. When including excited states in simulated annealing, SAE-Z achieves statistically equivalent performance to QAML-Z (Figure 4), with excited states selected from a validation set improving the generalization ability by reducing overfitting on the training set. A slight discrepancy remains between the two annealing processes, due to the sampling of excited states from distinct distributions of resulting states from simulated and quantum annealing, as well as the analog errors introduced in the implementation on the D-Wave device

that are absent in the SA case. However, on the training set, we observe that SA matches or bests QA with regards to minimum observed energy when they are each supplied identical QUBOs generated during the zooming algorithm (Figure 5).

D. Other Classical Benchmarks

We provide three additional classical benchmarks to compare QAML-Z performance: an optimized deep neural network (DNN), an optimized XGBoost algorithm (XGB), and a logistic regression (LR-Z). The LR-Z algorithm is a logistic ridge regression on the augmented set of kinematic classifiers used by QAML-Z. The DNN and XGB algorithms are applied to the raw kinematic variables scaled to mean zero and unit standard deviation from the training set and transformed by a principal component analysis to cover 95% of variance. Just as excited states in QAML-Z are selected by a validation set of equal size to the training set, the hyperparameters of the DNN and XGB are optimized over a similar validation set. While the original DNN benchmark for the Higgs optimization problem [33] fixed a given DNN architecture and optimized other hyperparameters, we perform Bayesian optimization over the number of neurons in two hidden layers (2 to 1024 neurons in the first hidden layer and 2 to 4096 neurons in the second hidden layer), L2 regularization (2^{-12} to 0.5) and patience parameter (0.25 to 32) of the DNN. Similarly, we optimize the number of estimators (30 to 10,000), tree depth (1 to 10), learning rate (0.0001 to 0.3), gamma regularization (0 to 15), dataset subsampling (0.2 to 1.0), and feature subsampling (0.3 to 1.0).

IV. CONCLUSION

We find that the QAML-Z extension of quantum annealing over a continuous space of weights on a set of augmented weak classifiers yields strong classifiers that improve the state-of-the-art quantum machine learning algorithm for quantum annealers, which was previously benchmarked in a study of Higgs decay classification [33]. Although QAML-Z remains at a disadvantage to a deep neural network (DNN) for sufficiently large datasets, the performance gap between QAML and DNN has been reduced by a factor of two by applying QAML-Z. Moreover, the successful performance at small training set sizes and short 5-μs anneal times associated with QAML-Z suggests promising applications for online learning on problems that change rapidly, either using a quantum

annealer or an FPGA device.

We observe that a logistic regression with LR-Z performs as well as the DNN, which suggests that the augmented set of weak physics-based classifiers is a highly effective method of feature engineering. Although QAML-Z cannot directly minimize least-squared error due to the lack of self-spin terms in the Ising model, it closely matches the performance of LR-Z and DNN at small training set sizes, validating the effectiveness of the quantum annealing approach. Moreover, QAML-Z significantly outperforms an optimized XGBoost algorithm for small training sets, demonstrating competitiveness with state-of-the-art classical machine learning algorithms.

The extent of improvement of QAML-Z over QAML for Higgs decay classification suggests that noisy intermediate-scale quantum devices may be approaching real-world applicability in machine learning despite their limitations. As the fastest annealing time feasible with quantum technology continues to improve, we anticipate that further work on benchmarking wall-clock times of classical and quantum devices will benefit greatly from practically relevant algorithms such as QAML-Z, where performance equal to classical state-of-the-art machine learning has already been demonstrated in certain regimes. More broadly, the favorable results of zooming in on an Ising model to achieve a solution unreachable by discrete optimization provides future direction for quantum annealing applications, potentially extending to quantum machine learning algorithms beyond QAML.

[1] Sumudu P Leelananda and Steffen Lindert, "Computational methods in drug discovery," Beilstein journal of organic chemistry **12**, 2694–2718 (2016).

[2] Adi L Tarca, Vincent J Carey, Xue-wen Chen, Roberto Romero, and Sorin Drăghici, "Machine learning and its applications to biology," PLoS computational biology **3**, e116 (2007).

[3] William W Hsieh, *Machine learning methods in the environmental sciences: Neural networks and kernels* (Cambridge university press, 2009).

[4] Juan Carrasquilla and Roger G Melko, "Machine learning phases of matter," Nature Physics **13**, 431 (2017).

[5] Pierre Baldi, Peter Sadowski, and Daniel Whiteson, "Searching for exotic particles in high-energy physics with deep learning," Nature communications **5**, 4308 (2014).

[6] Alexander Radovic, Mike Williams, David Rousseau, Michael Kagan, Daniele Bonacorsi, Alexander Himmel, Adam Aurisano, Kazuhiro Terao, and Taritree Wongjirad, "Machine learning at the energy and intensity frontiers of particle physics," Nature **560**, 41 (2018).

[7] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, "Quantum machine learning," Nature **549**, 195 (2017).

[8] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd, "Quantum support vector machine for big data classification," Physical review letters **113**, 130503 (2014).

[9] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, "Quantum principal component analysis," Nature Physics **10**, 631 (2014).

[10] Nathan Wiebe, Daniel Braun, and Seth Lloyd, "Quantum algorithm for data fitting," Physical review letters **109**, 050505 (2012).

[11] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "Prediction by linear regression on a quantum computer," Physical Review A **94**, 022342 (2016).

[12] Seth Lloyd, Silvano Garnerone, and Paolo Zanardi, "Quantum algorithms for topological and geometric analysis of data," Nature communications **7**, 10138 (2016).

[13] Patrick Rebentrost, Maria Schuld, Leonard Wossnig, Francesco Petruccione, and Seth Lloyd, "Quantum gradient descent and newton's method for constrained polynomial optimization," New Journal of Physics (2019).

[14] Scott Aaronson, "Read the fine print," Nature Physics **11**, 291 (2015).

[15] Ewin Tang, "A quantum-inspired classical algorithm for recommendation systems," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019) pp. 217–228.

[16] Srinivasan Arunachalam, Vlad Gheorghiu, Tomas Jochym-O'Connor, Michele Mosca, and Priyaa Varshinee Srinivasan, "On the robustness of bucket brigade quantum ram," New Journal of Physics **17**, 123010 (2015).

[17] Tadashi Kadowaki and Hidetoshi Nishimori, "Quantum annealing in the transverse ising model," Physical Review E **58**, 5355 (1998).

[18] Joshua Job and Daniel Lidar, "Test-driving 1000 qubits," Quantum Science and Technology **3**, 030501 (2018).

[19] Philipp Hauke, Helmut G. Katzgraber, Wolfgang Lechner, Hidetoshi Nishimori, and William D. Oliver, "Perspectives of quantum annealing: Methods and implementations," arXiv:1903.06559 (2019).

[20] Joshua Job, *The Theory and Practice of Benchmarking Quantum Annealers*, Ph.D. thesis, University of Southern California (2018).

[21] Hartmut Neven, Vasil S Denchev, Geordie Rose, and William G Macready, "Training a binary classifier with the quantum adiabatic algorithm," arXiv preprint arXiv:0811.0416 (2008).

[22] Kristen L. Pudenz and Daniel A. Lidar, "Quantum adiabatic machine learning," Quantum Information Processing **12**, 2027–2070 (2013).

[23] B. O'Gorman, R. Babbush, A. Perdomo-Ortiz, A. Aspuru-Guzik, and V. Smelyanskiy, "Bayesian network structure learning using quantum annealing," The European Physical Journal Special Topics **224**, 163–188 (2015).

[24] Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko, "Quantum boltzmann machine," Physical Review X **8**, 021050– (2018).

[25] Walter Vinci, Lorenzo Buffoni, Hossein Sadeghi, Amir Khoshaman, Evgeny Andriyash, and Mohammad H. Amin, "A path towards quantum advantage in training deep generative models with quantum annealers," (2019), arXiv:1912.02119 [quant-ph].

[26] Richard Harris, Mark W Johnson, T Lanting, AJ Berkley, J Johansson, P Bunyk, E Tolkacheva, E Ladizinsky, N Ladizinsky, T Oh, *et al.*, "Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor," Physical Review B **82**, 024511 (2010).

[27] Mark W Johnson, Mohammad HS Amin, Suzanne Gildert, Trevor Lanting, Firas Hamze, Neil Dickson, Richard Harris, Andrew J Berkley, Jan Johansson, Paul Bunyk, *et al.*, "Quantum annealing with manufactured spins," Nature **473**, 194–198 (2011).

[28] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz, and J. Whittaker, "Architectural considerations in the design of a superconducting quantum annealing processor," IEEE Transactions on Applied Superconductivity **24**, 1–10 (2014).

[29] Tobias Stollenwerk, Bryan O'Gorman, Davide Venturelli, Salvatore Mandrà, Olga Rodionova, Hokkwan Ng, Banavar Sridhar, Eleanor Gilbert Rieffel, and Rupak Biswas, "Quantum annealing applied to deconflicting optimal trajectories for air traffic management," IEEE transactions on intelligent transportation systems (2019).

[30] Alejandro Perdomo-Ortiz, Neil Dickson, Marshall Drew-Brook, Geordie Rose, and Alán Aspuru-Guzik, "Finding low-energy conformations of lattice protein models by quantum annealing," Scientific reports **2**, 571 (2012).

[31] Richard Y Li, Rosa Di Felice, Remo Rohs, and Daniel A Lidar, "Quantum annealing versus classical machine learning applied to a simplified computational biology problem," NPJ quantum information **4**, 14 (2018).

[32] Richard Y. Li, Sharvari Gujja, Sweta R. Bajaj, Omar E. Gamel, Nicholas Cilfone, Jeffrey R. Gulcher, Daniel A. Lidar, and Thomas W. Chittenden, "Unconventional machine learning of genome-wide human cancer data," arXiv:1909.06206 (2019).

[33] Alex Mott, Joshua Job, Jean-Roch Vlimant, Daniel Lidar, and Maria Spiropulu, "Solving a higgs optimization problem with quantum annealing for machine learning," Nature **550**, 375 (2017).

[34] Alexander Zlokapa, Abhishek Anand, Jean-Roch Vlimant, Javier M Duarte, Joshua Job, Daniel Lidar, and Maria Spiropulu, "Charged particle tracking with quan-

tum annealing-inspired optimization," arXiv preprint arXiv:1908.04475 (2019).

[35] Frédéric Bapst, Wahid Bhimji, Paolo Calafiura, Heather Gray, Wim Lavrijsen, Lucy Linder, and Alex Smith, "A pattern recognition algorithm for quantum annealers," Computing and Software for Big Science 4 (2019), 10.1007/s41781-019-0032-5.

[36] Tosio Kato, "On the adiabatic theorem of quantum mechanics," Journal of the Physical Society of Japan 5, 435–439 (1950).

[37] Sabine Jansen, Mary-Beth Ruskai, and Ruedi Seiler, "Bounds for the adiabatic approximation with applications to quantum computation," J. Math. Phys. 48, 102111 (2007).

[38] Daniel A. Lidar, Ali T. Rezakhani, and Alioscia Hamma, "Adiabatic approximation with exponential accuracy for many-body systems and quantum computation," J. Math. Phys. 50, 102106 (2009).

[39] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser, "Quantum Computation by Adiabatic Evolution," arXiv:quant-ph/0001106 (2000).

[40] Tameem Albash and Daniel A Lidar, "Adiabatic quantum computation," Reviews of Modern Physics 90, 015002 (2018).

[41] Andrew M Childs, Edward Farhi, and John Preskill, "Robustness of adiabatic quantum computation," Physical Review A 65, 012322 (2001).

[42] Mohammad HS Amin, Dmitri V Averin, and James A Nesteroff, "Decoherence in adiabatic quantum computation," Physical Review A 79, 022107 (2009).

[43] Tameem Albash and Daniel A Lidar, "Decoherence in adiabatic quantum computation," Physical Review A 91, 062320 (2015).

[44] Vicky Choi, "Minor-embedding in adiabatic quantum computation: I. the parameter setting problem," Quantum Information Processing 7, 193–209 (2008).

[45] Vicky Choi, "Minor-embedding in adiabatic quantum computation: Ii. minor-universal graph design," Quantum Information Processing 10, 343–353 (2011).

[46] Christine Klymko, Blair D. Sullivan, and Travis S. Humble, "Adiabatic quantum programming: minor embedding with hard faults," Quant. Inf. Proc. 13, 709–729 (2014).

[47] Jun Cai, William G. Macready, and Aidan Roy, "A practical heuristic for finding graph minors," arXiv:1406.2741 (2014).

[48] M. H. S. Amin, Peter J. Love, and C. J. S. Truncik, "Thermally assisted adiabatic quantum computation," Phys. Rev. Lett. 100, 060503 (2008).

[49] Neil G Dickson, MW Johnson, MH Amin, R Harris, F Altomare, AJ Berkley, P Bunyk, J Cai, EM Chapple, P Chavez, et al., "Thermally assisted quantum annealing of a 16-qubit problem," Nature communications 4, 1–6 (2013).

[50] Lorenzo Campos Venuti, Tameem Albash, Milad Marvian, Daniel Lidar, and Paolo Zanardi, "Relaxation versus adiabatic quantum steady-state preparation," Phys. Rev. A 95, 042302– (2017).

[51] Anurag Mishra, Tameem Albash, and Daniel A. Lidar, "Finite temperature quantum annealing solving exponentially small gap problem with non-monotonic success probability," Nature Communications 9, 2917 (2018).

[52] Troels F. Rønnow, Zhihui Wang, Joshua Job, Sergio Boixo, Sergei V. Isakov, David Wecker, John M. Mar-

tinis, Daniel A. Lidar, and Matthias Troyer, "Defining and detecting quantum speedup," Science 345, 420–424 (2014).

[53] Tameem Albash and Daniel A. Lidar, "Demonstration of a scaling advantage for a quantum annealer over simulated annealing," Physical Review X 8, 031016– (2018).

[54] Salvatore Mandrà and Helmut G Katzgraber, "A deceptive step towards quantum speedup detection," Quantum Sci. Technol. 3, 04LT01 (2018).

[55] Trevor Lanting, Anthony J Przybysz, A Yu Smirnov, Federico M Spedalieri, Mohammad H Amin, Andrew J Berkley, Richard Harris, Fabio Altomare, Sergio Boixo, Paul Bunyk, et al., "Entanglement in a quantum annealing processor," Physical Review X 4, 021041 (2014).

[56] Sergio Boixo, Vadim N. Smelyanskiy, Alireza Shabani, Sergei V. Isakov, Mark Dykman, Vasil S. Denchev, Mohammad H. Amin, Anatoly Yu Smirnov, Masoud Mohseni, and Hartmut Neven, "Computational multiqubit tunnelling in programmable quantum annealers," Nat Commun 7 (2016).

[57] Vasil S. Denchev, Sergio Boixo, Sergei V. Isakov, Nan Ding, Ryan Babbush, Vadim Smelyanskiy, John Martinis, and Hartmut Neven, "What is the computational value of finite-range tunneling?" Phys. Rev. X 6, 031015 (2016).

[58] Mohammad H. Amin, "Searching for quantum speedup in quasistatic quantum annealers," Physical Review A 92, 052323– (2015).

[59] Andrew D. King, Juan Carrasquilla, Jack Raymond, Isil Ozfidan, Evgeny Andriyash, Andrew Berkley, Mauricio Reis, Trevor Lanting, Richard Harris, Fabio Altomare, Kelly Boothby, Paul I. Bunyk, Colin Enderud, Alexandre Fréchette, Emile Hoskinson, Nicolas Ladizinsky, Travis Oh, Gabriel Poulin-Lamarre, Christopher Rich, Yuki Sato, Anatoly Yu. Smirnov, Loren J. Swenson, Mark H. Volkmann, Jed Whittaker, Jason Yao, Eric Ladizinsky, Mark W. Johnson, Jeremy Hilton, and Mohammad H. Amin, "Observation of topological phenomena in a programmable lattice of 1,800 qubits," Nature 560, 456–460 (2018).

[60] R. Harris, Y. Sato, A. J. Berkley, M. Reis, F. Altomare, M. H. Amin, K. Boothby, P. Bunyk, C. Deng, C. Enderud, S. Huang, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, T. Lanting, R. Li, T. Medina, R. Molavi, R. Neufeld, T. Oh, I. Pavlov, I. Perminov, G. Poulin-Lamarre, C. Rich, A. Smirnov, L. Swenson, N. Tsai, M. Volkmann, J. Whittaker, and J. Yao, "Phase transitions in a programmable quantum spin glass simulator," Science 361, 162 (2018).

[61] Source code for the QAML-Z algorithm is provided at https://github.com/quantummind/qaml-z/.

[62] Alex Mott, Joshua Job, Jean-Roch Vlimant, Daniel Lidar, and Maria Spiropulu, "Kinematics variables of higgs and background data," (2020).

[63] Adam Pearson, Anurag Mishra, Itay Hen, and Daniel A. Lidar, "Analog errors in quantum annealing: doom and hope," npj Quantum Information 5, 107 (2019).

[64] Davide Venturelli, Salvatore Mandrà, Sergey Knysh, Bryan O'Gorman, Rupak Biswas, and Vadim Smelyanskiy, "Quantum optimization of fully connected spin glasses," Phys. Rev. X 5, 031040 (2015).

[65] 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, KDD '16 (Association for Computing Machinery, New York, NY, USA, 2016) pp. 785–794.

[66] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).