

## Quantum adiabatic machine learning by zooming into a region of the energy surface

Alexander Zlokapa <sup>1</sup>, Alex Mott <sup>2</sup>, Joshua Job,<sup>3</sup> Jean-Roch Vlimant <sup>1</sup>, Daniel Lidar,<sup>4</sup> and Maria Spiropulu<sup>1</sup>

<sup>1</sup>*Division of Physics, Mathematics & Astronomy, Alliance for Quantum Technologies, California Institute of Technology, Pasadena, California 91125, USA*

<sup>2</sup>*DeepMind Technologies, London, United Kingdom*

<sup>3</sup>*Lockheed Martin Advanced Technology Center, Sunnyvale, California 94089, USA*

<sup>4</sup>*Departments of Electrical and Computer Engineering, Chemistry, and Physics & Astronomy, and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*



(Received 6 July 2020; accepted 9 November 2020; published 4 December 2020)

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, an 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 receiver operating characteristic 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 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.

DOI: [10.1103/PhysRevA.102.062405](https://doi.org/10.1103/PhysRevA.102.062405)

### 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 such algorithms. Similarly, quantum random access memory is typically required to store data, but engineering challenges persist in developing a sufficiently large memory [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], train deep generative models [25], and implement support vector machines [26]. Quantum annealing is the only current quantum computing paradigm that has resulted in architectures with a large enough number of—albeit relatively noisy—qubits [27–29] to address both real-world and fundamental science prob-

lems, e.g., in air traffic control [30], computational biology [31–33], and high-energy physics [34–36]. Under the adiabatic theorem of quantum mechanics, quantum annealing evolves from an initial transverse field Hamiltonian to the target problem 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 [37–39]. The ground state of the problem Hamiltonian is then the solution (as in adiabatic quantum computing [40,41]), although thermal excitations may move the system out of the ground state [42–48], which can be beneficial [49–52]. It is crucial to observe that evidence of a quantum speedup in quantum annealing remains uncertain [53–55], although quantum phenomena have been observed in D-Wave quantum annealers [56–58]. While this remains a speculative topic, quantum annealers may exhibit advantages other than a speedup, such as sampling from nonequilibrium distributions prepared during the anneal [59–61].

Here we propose a quantum algorithm inspired by the previous state-of-the-art quantum annealing for machine learning (QAML) algorithm [34], 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 [62] and data [63].

## 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 input 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 and 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  $R$ :

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

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

Removing the spin-independent term  $\|y\|^2$  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  $\lambda$  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 and (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.

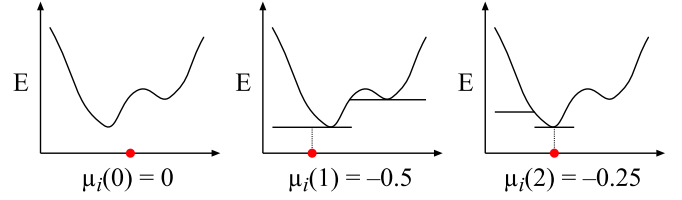


FIG. 1. Zooming extension. While QAML only performs one anneal, QAML-Z iteratively updates the weight  $\mu_i$  (indicated by the dot) of the  $i$ th weak classifier in the strong classifier ensemble by performing a binary search over the energy surface using spin-up and spin-down outcomes.

### 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 (Fig. 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.

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  $t = 0, 1, \dots, T-1$  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  $s_i c_i(\mathbf{x}_\tau) \rightarrow \sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau)$ . The full expression is

$$H(t) = \sum_{i=1}^N \sum_{j>i}^N \sum_{\tau=1}^S (\sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau)) \times (\sigma(t) s_j c_j(\mathbf{x}_\tau) + \mu_j(t) c_j(\mathbf{x}_\tau)) - \sum_{i=1}^N \sum_{\tau=1}^S (\sigma(t) s_i c_i(\mathbf{x}_\tau) + \mu_i(t) c_i(\mathbf{x}_\tau)) y_\tau \quad (7a)$$

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

where Eq. (7b) is derived after dropping constants from the Hamiltonian and applying the same  $C_i$  and  $C_{ij}$  notation as in QAML. This new Hamiltonian may be iteratively optimized for  $t = 0, 1, \dots, T-1$  to update  $\mu_i(t+1) = \mu_i(t) + s_i \sigma(t+1)$ , resulting in the 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 it-

eration, for each qubit  $s_i$  that the energy worsens by the update (i.e.,  $E[\mu_1(t+1), \dots, \mu_i(t+1), \dots, \mu_N(t+1)] > E[\mu_1(t), \dots, \mu_i(t), \dots, \mu_N(t)]$ ), we apply the flip  $s_i \rightarrow -s_i$  with monotonically decreasing probability  $p_f(t)$ , similar 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  $\pm 1$ :

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

where  $N$  is the number of classifiers,  $l \in \mathbb{Z} : -A \leq l \leq A$  is the offset, and  $\delta$  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

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

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

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

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

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 eight 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 (back-

ground) are used to identify simulated Higgs bosons [34]. The simulation of  $H \rightarrow \gamma\gamma$  (Higgs to diphoton decay) is limited to the main process of gluon fusion, produced with PYTHIA 6.4 [64], while subleading contributions are not included. The simulation of the background closely follows the one from Ref. [65], produced with SHERPA 1.1 [66], and is limited to the prompt diphoton production, that includes quark-antiquark annihilation as well as gluon fusion at the two-loops level. The simulated events (provided publicly [63]) are restricted to realistic detector acceptance and with trigger requirements that lie directly under the Higgs peak, that is, restricting to events having photons with  $|\eta| < 2.5$  ( $\eta = -\log[\tan(\theta/2)]$  is the pseudorapidity, where  $\theta$  is the angle with the beam axis), with one photon having  $p_T > 32$  GeV and the other having  $p_T > 25$  GeV ( $p_T$  is the transverse momentum, i.e., the momentum perpendicular to the axis defined by the colliding protons). The total diphoton invariant mass is restricted to be between 122.5 and 127.5 GeV. 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 [29]<sup>1</sup>. 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 were 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 [67] and also allows a minor embedding operation [45–48] 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 [68]. 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 (gauge averag-

<sup>1</sup>This device has since been replaced by the newer D-Wave Advantage quantum annealer.

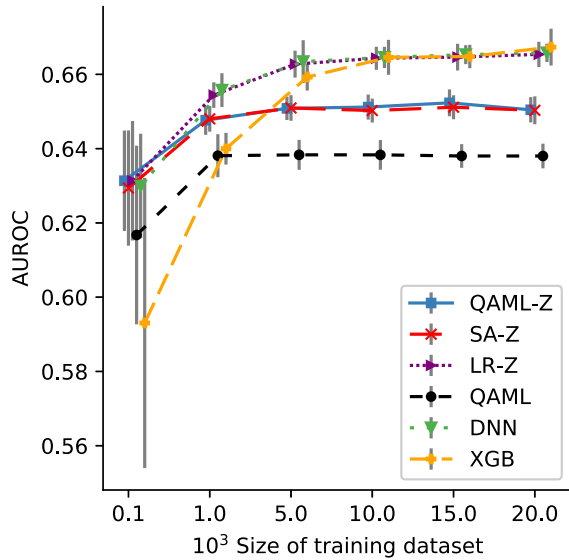


FIG. 2. Area under the ROC curve for the QAML-Z extension, simulated annealing (SA-Z), a logistic regression (LR-Z), the original QAML, a deep neural network (DNN), and XGBoost (XGB) [69] as a function of training set size. While QAML-Z matches DNN performance at small training set sizes, it decreases the margin between QAML and DNN by 47% for the largest training sets. Error bars indicate  $1\sigma$  error, including both variation over training sets and statistical error estimated by reweighting samples from a Poisson distribution.

ing). Annealing is performed with a  $5 \mu\text{s}$  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 deep neural network (DNN) hyperparameter tuning], selecting the best-performing excited states for different false-positive rates.

## B. Results

Compared to the QAML algorithm, the area under the receiver operating characteristic (AUROC) curve is significantly improved by QAML-Z on all training set sizes (Fig. 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

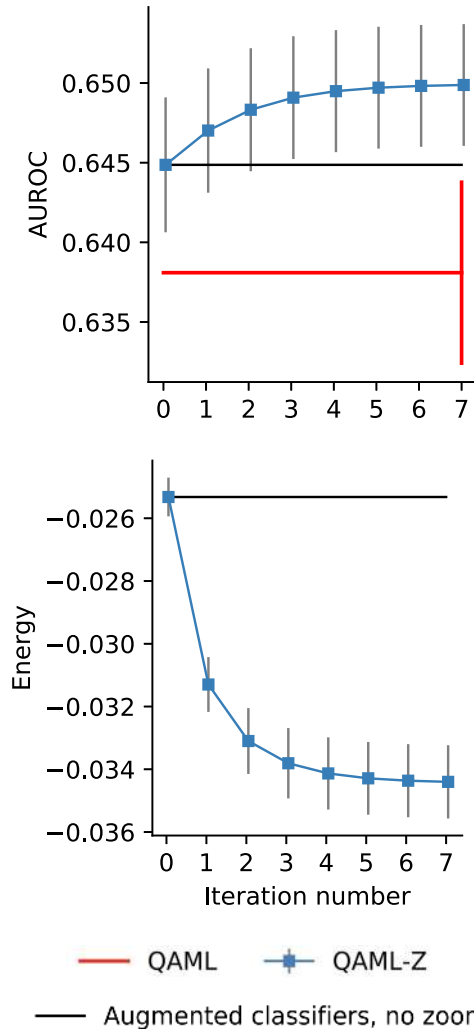


FIG. 3. QAML-Z performance on the test set vs zooming iteration number (training set size of 1000). Top: Significant improvements by QAML-Z can be separately seen for classifier augmentation and zooming over the original QAML algorithm. Bottom: Using model energy on the test set improves monotonically, indicating negligible overfitting. Error bars indicate  $1\sigma$  error. Here and in all subsequent figures, the energy is given in units corresponding to the Hamiltonian in Eq. (9), which for our purposes may be considered to be dimensionless.

the classical algorithms to further improve their performance from Ref. [34]. 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.

We observe the effectiveness of both the zooming and augmentation aspects of QAML-Z (Fig. 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 analog 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. Moreover, we expect a quantum annealer that evolves adiabatically to sample from the quantum Gibbs distribution. Hence sampling from excited states reflects the inclusion of information that would not be easily classically available, and can reasonably be expected to generate a performance difference in comparison with classical methods such as simulated annealing.

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  $\vec{s}$  [70]. 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 vector 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  $\beta$  is stepped with a linear inverse temperature schedule from  $\beta_i = 0.1$  to  $\beta_f = 5$  over  $W = 1000$  sweeps, incrementing the inverse 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. Inverse temperature schedules reaching  $\beta$  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.

QAML-Z performs better than SA-Z on all training sets, with a statistically significant advantage at larger training set sizes (Fig. 4). This suggests that both simulated and quantum annealing (QA) 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 (Fig. 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

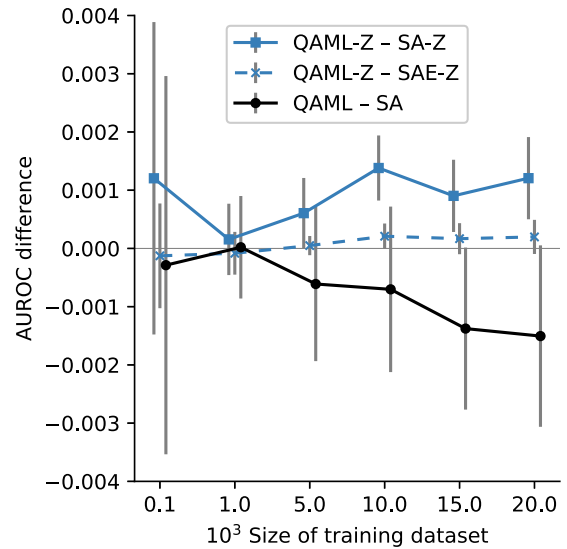


FIG. 4. Comparison of quantum and simulated annealing for the new and original algorithms, measured by area under ROC (AUROC) curve. Although QAML-Z outperforms QAML and SA-Z, the inclusion of excited states in the SAE-Z variant reproduces QAML-Z performance to one standard deviation. Error bars indicate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

SA matches or bests QA with regards to minimum observed energy when they are each supplied identical quadratic unconstrained binary optimization problems generated during the zooming algorithm (Fig. 5).

**D. Other classical benchmarks**

We provide three additional classical benchmarks to compare QAML-Z performance: a deep neural network (DNN), XGBoost algorithm (XGB), and logistic regression (LR-Z).

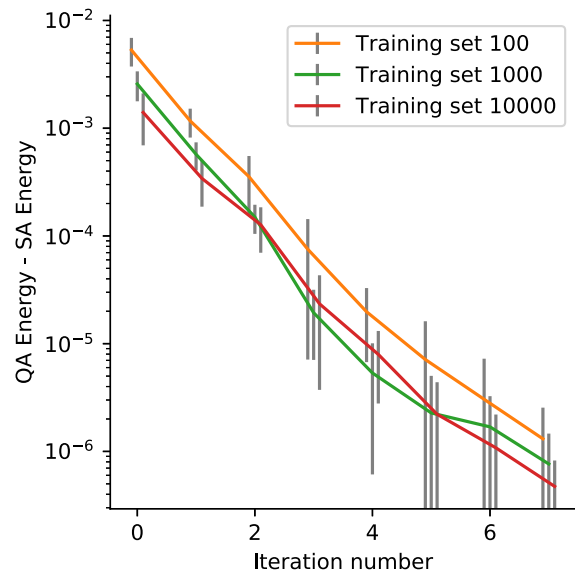


FIG. 5. Difference between the lowest energy of quantum annealing (QA) and simulated annealing (SA). SA finds a lower minimum energy than QA given identical initial Hamiltonians. Error bars show  $1\sigma$  error.

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 [34] 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), data set 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 a state-of-the-art quantum machine learning algorithm for quantum annealers, which was previously benchmarked in a study of Higgs decay classification [34]. Although QAML-Z remains at a disadvantage to a deep neural network (DNN) for sufficiently large data sets, the performance gap between QAML and DNN has been reduced by a factor of 2 by applying QAML-Z. Moreover, the successful performance at small training set sizes and short 5  $\mu$ s anneal times associated with QAML-Z suggests promising applications for online learning on problems that change rapidly, either using a quantum annealer or a field-programmable gate array 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-squares 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 effective-

ness 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 various metrics affecting the performance of quantum annealing technology continue to improve [71–77], 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.

#### ACKNOWLEDGMENTS

Part of this work was conducted at “iBanks,” the AI GPU cluster at Caltech. We acknowledge NVIDIA, SuperMicro, and the Kavli Foundation for their support of “iBanks.” This work is partially supported by a DOE/HEP QuantISED program grant, Quantum Machine Learning and Quantum Computation Frameworks (QMLQCF) for HEP, Grant No. DE-SC0019227. The work is also supported in part by the AT&T Foundry Innovation Centers through INQNET, a program for accelerating quantum technologies. The research is based upon work (partially) supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) and the Defense Advanced Research Projects Agency (DARPA), via the US Army Research Office Contract No. W911NF-17-C-0050. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the US Government. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

- 
- [1] S. P. Leelananda and S. Lindert, Computational methods in drug discovery, *Beilstein J. Org. Chem.* **12**, 2694 (2016).
  - [2] A. L. Tarca, V. J. Carey, X.-W. Chen, R. Romero, and S. Drăghici, Machine learning and its applications to biology, *PLoS Comput. Biol.* **3**, e116 (2007).
  - [3] W. W. Hsieh, *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*, 1st ed. (Cambridge University Press, New York, 2009).
  - [4] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).
  - [5] P. Baldi, P. Sadowski, and D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, *Nat. Commun.* **5**, 4308 (2014).
  - [6] 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).
  - [7] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
  - [8] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum Support Vector Machine for Big Data Classification, *Phys. Rev. Lett.* **113**, 130503 (2014).
  - [9] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).

- [10] N. Wiebe, D. Braun, and S. Lloyd, Quantum Algorithm for Data Fitting, *Phys. Rev. Lett.* **109**, 050505 (2012).
- [11] M. Schuld, I. Sinayskiy, and F. Petruccione, Prediction by linear regression on a quantum computer, *Phys. Rev. A* **94**, 022342 (2016).
- [12] S. Lloyd, S. Garnerone, and P. Zanardi, Quantum algorithms for topological and geometric analysis of data, *Nat. Commun.* **7**, 10138 (2016).
- [13] P. Reberntrost, M. Schuld, L. Wossnig, F. Petruccione, and S. Lloyd, Quantum gradient descent and Newton's method for constrained polynomial optimization, *New J. Phys.* **21**, 073023 (2019).
- [14] S. Aaronson, Read the fine print, *Nat. Phys.* **11**, 291 (2015).
- [15] E. Tang, A quantum-inspired classical algorithm for recommendation systems, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC 2019)* (Association for Computing Machinery, New York, 2019), pp. 217–228.
- [16] S. Arunachalam, V. Gheorghiu, T. Jochym-O'Connor, M. Mosca, and P. V. Srinivasan, On the robustness of bucket brigade quantum RAM, *New J. Phys.* **17**, 123010 (2015).
- [17] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse Ising model, *Phys. Rev. E* **58**, 5355 (1998).
- [18] J. Job and D. Lidar, Test-driving 1000 qubits, *Quantum Sci. Technol.* **3**, 030501 (2018).
- [19] P. Hauke, H. G. Katzgraber, W. Lechner, H. Nishimori, and W. D. Oliver, Perspectives of quantum annealing: Methods and implementations, *Rep. Prog. Phys.* **83**, 054401 (2020).
- [20] J. Job, The theory and practice of benchmarking quantum annealers, Ph.D. thesis, University of Southern California, 2018.
- [21] H. Neven, V. S. Denchev, G. Rose, and W. G. Macready, Training a binary classifier with the quantum adiabatic algorithm, [arXiv:0811.0416](https://arxiv.org/abs/0811.0416).
- [22] K. L. Pudenz and D. A. Lidar, Quantum adiabatic machine learning, *Quantum Inf. Process.* **12**, 2027 (2013).
- [23] B. O'Gorman, R. Babbush, A. Perdomo-Ortiz, A. Aspuru-Guzik, and V. Smelyanskiy, Bayesian network structure learning using quantum annealing, *Eur. Phys. J.: Spec. Top.* **224**, 163 (2015).
- [24] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytsky, and R. Melko, Quantum Boltzmann Machine, *Phys. Rev. X* **8**, 021050 (2018).
- [25] W. Winci, L. Buffoni, H. Sadeghi, A. Khoshaman, E. Andriyash, and M. H. Amin, A path towards quantum advantage in training deep generative models with quantum annealers, *Mach. Learn.: Sci. Technol.* **1**, 045028 (2020).
- [26] D. Willsch, M. Willsch, H. De Raedt, and K. Michielsen, Support vector machines on the D-Wave quantum annealer, *Comput. Phys. Commun.* **248**, 107006 (2020).
- [27] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose, Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor, *Phys. Rev. B* **82**, 024511 (2010).
- [28] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, Quantum annealing with manufactured spins, *Nature* **473**, 194 (2011).
- [29] 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 Trans. Appl. Supercond.* **24**, 1 (2014).
- [30] T. Stollenwerk, B. O. Gorman, D. Venturelli, S. Mandrà, O. Rodionova, H. Ng, B. Sridhar, E. G. Rieffel, and R. Biswas, Quantum annealing applied to de-conflicting optimal trajectories for air traffic management, *IEEE Trans. Intell. Transp. Syst.* **21**, 285 (2020).
- [31] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, Finding low-energy conformations of lattice protein models by quantum annealing, *Sci. Rep.* **2**, 571 (2012).
- [32] R. Y. Li, R. Di Felice, R. Rohs, and D. A. Lidar, Quantum annealing versus classical machine learning applied to a simplified computational biology problem, *npj Quantum Inf.* **4**, 14 (2018).
- [33] R. Y. Li, S. Gujja, S. R. Bajaj, O. E. Gamel, N. Cilfone, J. R. Gulcher, D. A. Lidar, and T. W. Chittenden, Unconventional machine learning of genome-wide human cancer data, [arXiv:1909.06206](https://arxiv.org/abs/1909.06206).
- [34] 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).
- [35] A. Zlokapa, A. Anand, J.-R. Vlimant, J. M. Duarte, J. Job, D. Lidar, and M. Spiropulu, Charged particle tracking with quantum annealing-inspired optimization, [arXiv:1908.04475](https://arxiv.org/abs/1908.04475).
- [36] F. Bapst, W. Bhimji, P. Calafiura, H. Gray, W. Lavrijsen, L. Linder, and A. Smith, A pattern recognition algorithm for quantum annealers, *Comput. Software Big Sci.* **41** (2019).
- [37] T. Kato, On the adiabatic theorem of quantum mechanics, *J. Phys. Soc. Jpn.* **5**, 435 (1950).
- [38] S. Jansen, M.-B. Ruskai, and R. Seiler, Bounds for the adiabatic approximation with applications to quantum computation, *J. Math. Phys.* **48**, 102111 (2007).
- [39] D. A. Lidar, A. T. Rezakhani, and A. Hama, Adiabatic approximation with exponential accuracy for many-body systems and quantum computation, *J. Math. Phys.* **50**, 102106 (2009).
- [40] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, Quantum computation by adiabatic evolution, [arXiv:quant-ph/0001106](https://arxiv.org/abs/quant-ph/0001106).
- [41] T. Albash and D. A. Lidar, Adiabatic quantum computation, *Rev. Mod. Phys.* **90**, 015002 (2018).
- [42] A. M. Childs, E. Farhi, and J. Preskill, Robustness of adiabatic quantum computation, *Phys. Rev. A* **65**, 012322 (2001).
- [43] M. H. S. Amin, D. V. Averin, and J. A. Nesteroff, Decoherence in adiabatic quantum computation, *Phys. Rev. A* **79**, 022107 (2009).
- [44] T. Albash and D. A. Lidar, Decoherence in adiabatic quantum computation, *Phys. Rev. A* **91**, 062320 (2015).
- [45] V. Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem, *Quantum Inf. Process.* **7**, 193 (2008).
- [46] V. Choi, Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design, *Quantum Inf. Process.* **10**, 343 (2011).

- [47] C. Klymko, B. D. Sullivan, and T. S. Humble, Adiabatic quantum programming: Minor embedding with hard faults, *Quantum Inf. Process.* **13**, 709 (2014).
- [48] J. Cai, W. G. Macready, and A. Roy, A practical heuristic for finding graph minors, [arXiv:1406.2741](https://arxiv.org/abs/1406.2741).
- [49] M. H. S. Amin, P. J. Love, and C. J. S. Truncik, Thermally Assisted Adiabatic Quantum Computation, *Phys. Rev. Lett.* **100**, 060503 (2008).
- [50] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez *et al.*, Thermally assisted quantum annealing of a 16-qubit problem, *Nat. Commun.* **4**, 1 (2013).
- [51] L. C. Venuti, T. Albash, M. Marvian, D. Lidar, and P. Zanardi, Relaxation versus adiabatic quantum steady-state preparation, *Phys. Rev. A* **95**, 042302 (2017).
- [52] A. Mishra, T. Albash, and D. A. Lidar, Finite temperature quantum annealing solving exponentially small gap problem with non-monotonic success probability, *Nat. Commun.* **9**, 2917 (2018).
- [53] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Defining and detecting quantum speedup, *Science* **345**, 420 (2014).
- [54] T. Albash and D. A. Lidar, Demonstration of a Scaling Advantage for a Quantum Annealer over Simulated Annealing, *Phys. Rev. X* **8**, 031016 (2018).
- [55] S. Mandrà and H. G. Katzgraber, A deceptive step towards quantum speedup detection, *Quantum Sci. Technol.* **3**, 04LT01 (2018).
- [56] T. Lanting, A. J. Przybysz, A. Yu. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, N. Dickson, C. Enderud, J. P. Hilton, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, R. Neufeld, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, S. Uchaikin, A. B. Wilson, and G. Rose, Entanglement in a Quantum Annealing Processor, *Phys. Rev. X* **4**, 021041 (2014).
- [57] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. H. Amin, A. Yu. Smirnov, M. Mohseni, and H. Neven, Computational multiqubit tunneling in programmable quantum annealers, *Nat. Commun.* **7**, 10327 (2016).
- [58] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, What Is the Computational Value of Finite-Range Tunneling? *Phys. Rev. X* **6**, 031015 (2016).
- [59] M. H. Amin, Searching for quantum speedup in quasistatic quantum annealers, *Phys. Rev. A* **92**, 052323 (2015).
- [60] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare, K. Boothby, P. I. Bunyk, C. Enderud, A. Fréchette, E. Hoskinson, N. Ladizinsky, T. Oh, G. Poulin-Lamarre, C. Rich, Y. Sato, A. Yu. Smirnov, L. J. Swenson, M. H. Volkmann, J. Whittaker, J. Yao, E. Ladizinsky, M. W. Johnson, J. Hilton, and M. H. Amin, Observation of topological phenomena in a programmable lattice of 1,800 qubits, *Nature* **560**, 456 (2018).
- [61] 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).
- [62] Source code for the QAML-Z algorithm is provided at <https://github.com/quantummind/qaml-z/>.
- [63] A. Mott, J. Job, J.-R. Vlimant, D. Lidar, and M. Spiropulu, Kinematics variables of Higgs and background data (2020), doi: [10.22002/D1.1335](https://doi.org/10.22002/D1.1335).
- [64] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 physics and manual, *J. High Energy Phys.* **05** (2006) 026.
- [65] V. Khachatryan *et al.* (CMS Collaboration), Observation of the diphoton decay of the Higgs boson and measurement of its properties, *Eur. Phys. J. C* **74**, 3076 (2014).
- [66] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, Event generation with SHERPA 1.1, *J. High Energy Phys.* **02** (2009) 007.
- [67] A. Pearson, A. Mishra, I. Hen, and D. A. Lidar, Analog errors in quantum annealing: Doom and hope, *npj Quantum Inf.* **5**, 107 (2019).
- [68] D. Venturelli, S. Mandrà, S. Knysh, B. O’Gorman, R. Biswas, and V. Smelyanskiy, Quantum Optimization of Fully Connected Spin Glasses, *Phys. Rev. X* **5**, 031040 (2015).
- [69] 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 (KDD ’16)* (Association for Computing Machinery, New York, 2016), pp. 785–794.
- [70] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science* **220**, 671 (1983).
- [71] F. Yan, S. Gustavsson, A. Kamal, J. Birenbaum, A. P. Sears, D. Hover, T. J. Gudmundsen, D. Rosenberg, G. Samach, S. Weber, J. L. Yoder, T. P. Orlando, J. Clarke, A. J. Kerman, and W. D. Oliver, The flux qubit revisited to enhance coherence and reproducibility, *Nat. Commun.* **7**, 12964EP (2016).
- [72] S. J. Weber, G. O. Samach, D. Hover, S. Gustavsson, D. K. Kim, A. Melville, D. Rosenberg, A. P. Sears, F. Yan, J. L. Yoder, W. D. Oliver, and A. J. Kerman, Coherent Coupled Qubits for Quantum Annealing, *Phys. Rev. Appl.* **8**, 014004 (2017).
- [73] S. Novikov, R. Hinkey, S. Disseler, J. I. Basham, T. Albash, A. Risinger, D. Ferguson, D. A. Lidar, and K. M. Zick, Exploring more-coherent quantum annealing, in *2018 IEEE International Conference on Rebooting Computing (ICRC)* (IEEE, New York, 2018), pp. 1–7.
- [74] M. Khezri, J. A. Grover, J. I. Basham, S. M. Disseler, H. Chen, S. Novikov, K. M. Zick, and D. A. Lidar, Anneal-path correction in flux qubits, [arXiv:2002.11217](https://arxiv.org/abs/2002.11217).
- [75] J. A. Grover, J. I. Basham, A. Marakov, S. M. Disseler, R. T. Hinkey, M. Khalil, Z. A. Stegen, T. Chamberlin, W. DeGottardi, D. J. Clarke, J. R. Medford, J. D. Strand, M. J. A. Stoutimore, S. Novikov, D. G. Ferguson, D. Lidar, K. M. Zick, and A. J. Przybysz, Fast, lifetime-preserving readout for high-coherence quantum annealers, *Phys. Rev. Appl.* **1**, 020314 (2020).
- [76] K. Boothby, P. Bunyk, J. Raymond, and A. Roy, Next-generation topology of D-Wave quantum processors, [arXiv:2003.00133](https://arxiv.org/abs/2003.00133).
- [77] A. D. King and W. Bernoudy, Performance benefits of increased qubit connectivity in quantum annealing 3-dimensional spin glasses, [arXiv:2009.12479](https://arxiv.org/abs/2009.12479).