Quantum Machine Intelligence (2023) 5:4
https://doi.org/10.1007/s42484-022-00092-y

RESEARCH ARTICLE

Boosting the performance of quantum annealers using machine learning

Jure Brence[1,2,3] · Dragan Mihailović[4,5,6] · Viktor V. Kabanov[4] · Ljupčo Todorovski[1,3] · Sašo Džeroski[1,2] · Jaka Vodeb[4,6]

Received: 12 May 2022 / Accepted: 16 December 2022

Abstract
Noisy intermediate-scale quantum (NISQ) devices are spearheading the second quantum revolution. Of these, quantum annealers are the only ones currently offering real world, commercial applications on as many as 5000 qubits. The size of problems that can be solved by quantum annealers is limited mainly by errors caused by environmental noise and intrinsic imperfections of the processor. We address the issue of intrinsic imperfections with a novel error correction approach, based on machine learning methods. Our approach adjusts the input Hamiltonian to maximize the probability of finding the solution. In our experiments, the proposed error correction method improved the performance of annealing by up to three orders of magnitude and enabled the solving of a previously intractable, maximally complex problem.

Keywords Quantum computers · Quantum annealing · Machine learning · Noise · Calibration

1 Introduction

Quantum annealers have proven themselves useful in solving various problems in material science (Harris et al. 2018; King et al. 2018; Bando et al. 2020; Kairys et al. 2020), optimization (Neukart et al. 2017; Orús et al. 2019), and machine learning (Mott et al. 2017; Li et al. 2018; Willsch et al. 2020; Jain et al. 2020), and have shown scaling advantage in problem solving efficiency in some cases (King et al. 2021; Albash and Lidar 2018). However, the quantum computers used in these cases have trouble excluding the impact of device imperfections and the outside environment on the quantum dynamics taking place within the quantum device (Job and Lidar 2018; Bando et al. 2020; Boixo et al. 2016; Gardas et al. 2018; Gardas and Deffner 2018). The resulting errors limit the potential of quantum simulations or quantum speed-up in solving classical optimization problems. Quantum error correction assumes two very different realizations in the two quantum computing models currently used in practice. The quantum gate model (e.g., Google and IBM) already has an established quantum threshold theorem, which states that an arbitrarily long quantum computation circuit can be constructed, provided the qubits involved have a low enough error rate. However, quantum annealing has no such established theorem. Error correction algorithms are therefore scarce and so are examples of research on quantum annealing error correction and topology compensation. The problem of noise in quantum annealing has so far been tackled only in a few different ways that do not involve improvements to hardware. The first is quantum annealing error correction, based on introducing an energy penalty along with encoding and error correction (Pudenz et al. 2014). The second approach involves compensating for differences in chain susceptibility due to the embedding topology (Raymond et al. 2020) and the third evaluates "Hamiltonian noise" (Zaborniak and de Sousa 2021), which

✉ Jure Brence
jure.brence@ijs.si

1 Department for Knowledge Technologies, Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

2 Jozef Stefan International Postgraduate School, Jamova 39, 1000 Ljubljana, Slovenia

3 Department of Mathematics, Faculty for Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

4 Department of Complex Matter, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

5 CENN Nanocenter, Jamova 39, 1000 Ljubljana, Slovenia

6 Department of Physics, Faculty for Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

Published online: 22 January 2023

can then be subtracted in a similar manner as in our work. However, using the terminology considered in this paper, Ref. Zaborniak and de Sousa (2021) considers only the diagonal of the Hamiltonian $Q_{i,i}$, while we consider the full Hamiltonian $Q_{i,j}$.

In this paper, we address the issue of noise in D-Wave's 2000Q quantum annealer. The way a problem is solved on this device is to package the problem at hand into a Hamiltonian, employing the quadratic unconstrained binary optimization (QUBO) formalism

$$H = \sum_{i,j} Q_{i,j} q_i q_j, \tag{1}$$

where specifying the matrix Q corresponds to packaging up the problem as input to the quantum processing unit (QPU). The elements $Q_{i,j}$ are couplers if $i \neq j$ and biases if $i = j$. q_i represents a logical qubit (LQ) with the possible values of 0 and 1, which is used in order to interpret the output solution provided by the QPU. There is an important distinction between LQs and physical qubits (PQs). PQs are non-abstract actual physical qubits, which are manipulated by the couplings and biases applied to them during the quantum annealing process in the form of physical magnetic fields on the chip. Their connectivity is fixed into a so-called Chimera graph (D-Wave Systems Inc 2021), where each PQ is connected to 6 neighbors. However, the connectivity which is required between LQs by the matrix Q is rarely the same as the Chimera graph. In order to realize the desired connectivity, additional PQs and couplings are added, so that each LQ is represented by a chain, composed of several PQs. This mapping between LQs and PQs is called an embedding (D-Wave Systems Inc 2021).

There are many different sources of noise present on the QPU, as specified in the documentation provided by D-Wave: (i) background susceptibility which comes from next-nearest neighbor interactions and leakage of biases, (ii) $1/f$ flux noise is exerted on qubits which manifests itself as a drift of their properties on a larger time-scale, such as between different problem submissions, (iii) the problem Hamiltonian, which is specified as part of the input to the machine has a finite resolution, (iv) the ratio between biases and couplings can vary for different annealing parameters, and (v) qubits cannot be made perfectly identical. In addition, several studies have revealed that the QPU is inherently coupled to its environment, which brings in (vi) thermal effects (Albash et al. 2015; Benedetti et al. 2016; Boixo et al. 2016; Buffoni and Campisi 2020) and (vii) random quantum fluctuations in the Josephson current present in qubits (Bando et al. 2020). We take all of these effects into account by modelling the noise on the QPU as

$$Q_{\text{submitted}} = Q_0 + dQ_{\text{calibration}} + dQ_{\text{noise}}, \tag{2}$$

where Q_0 represents the original Hamiltonian and $Q_{\text{submitted}}$ is the Hamiltonian, corrupted by noise. Each element of dQ_{noise} is a Gaussian random variable with mean 0 and standard deviation σ. Elements of $dQ_{\text{calibration}}$ represent the shifts of the mean of the Gaussian random variables. The idea of this work is to find the appropriate $dQ_{\text{calibration}}$, which we then subtract from $Q_{\text{submitted}}$. We assume that $dQ_{\text{calibration}}$ is constant and represents the combined systematic noise from all the aforementioned noise sources. The degree to which we can reduce the impact of noise is presently unknown and determining it the motivation of this investigation.

The rest of the paper is structured as follows. In Section 2, we introduce the addressed physical problem of electron ordering on a triangular lattice, as well as present our definition of annealing performance and the methodology for calculating it. In Section 3, we report on our investigation of the stability of annealing performance measurements and the effects of different embeddings on annealing performance. In Section 4, we introduce two methods for quantum annealer correction, present the results of their empirical evaluation, and comment on the envisioned applicability of the methods. In Section 5, we present new insights on the nature of systematic noise in quantum annealers, which could inspire new and better approaches to dealing with corresponding issues.

2 Methods

2.1 Electron ordering on a triangular lattice

The physical problem of finding the ground state of a system of electrons on a triangular lattice was already studied in a slightly different version without error correction (Vodeb et al. 2019). The problem is represented by a two-dimensional Hamiltonian

$$Q_0 = \frac{1}{2} \sum_{i,j}^{N} V(i,j)(q_i - \bar{q})(q_j - \bar{q}) - \mu \sum_{i}^{N} q_i \tag{3}$$

where $V(i,j) = V_0 \exp(-r_{i,j}/l_s)/r_{i,j}$, l_s is the screening radius, $r_{i,j} = |r_i - r_j|$, r_i is the ith out of N lattice sites, $q_i \in \{0,1\}$ describes whether an electron occupies the lattice site i or not, the sum runs over all lattice sites, and $\bar{q} = \sum_i^N q_i/N$. $\bar{q}$ is the magnitude of the background and constant positive charge on each lattice site, which needs to be included in order to satisfy electro-neutrality and prevent electrons from escaping to the edges of the system. The number of electrons in the system is varied by the chemical potential μ. In this paper, we considered a 4×4, 5×5 and 6×6 triangular lattice with open boundary conditions with $\mu = -0.2$ and $V_0 = 1$ set to a fixed value for which the

ground state of Q_0 is an electronic lattice with a density of 1 electron per 3 atomic sites.

In order to be able to map the electron ordering problem onto a D-Wave machine, we simplify Q_0 to $Q_0 = \sum_{i,j}^{N} Q_{i,j} q_i q_j$, where $Q_{i,j} = \frac{1}{2} V(i,j) - \frac{1}{2N} \sum_{k}^{N} (V(i,k) + V(k,j)) + \frac{1}{2N^2} \sum_{k,l}^{N} V(k,l) - \mu \delta_{i,j}$. The studied physical problem requires maximal connectivity between logical qubits q_i, which is represented by a fully connected graph or a clique. Figure 1a shows the required connectivity, where i in q_i runs from 1 to 16. A logical qubit is therefore represented by a chain of 5 physical qubits in this case, colored differently for each i. Such a connectivity, when embedded onto D-Wave's Chimera graph of physical qubits, is one of the most complex problems one can attempt to solve on a quantum annealer. Therefore, if we are able to show that our error correction algorithms work in this case, they will also work well with other complex problems.

2.2 Annealing performance

Because the result of a single quantum annealing experiment is stochastic, annealing is typically repeated hundreds or thousands of times, producing a sample of reads. Since the goal is usually to find the global minimum of the energy, generally only the read with the lowest energy is of interest. We consider the performance of the annealer to be better when a larger fraction of the reads finds the ground state. In order to summarize the performance of the annealer in a particular experiment from the sample, we look at three quantities:

- Success rate: the fraction of reads with energy equal to the energy of the ground state.
- Mean energy: the mean across the energies of the entire sample.
- Chain break fraction: the fraction of reads with broken chains where not all qubits in a chain end up in the same state.

In the process of calibrating the quantum annealer, we seek to maximize the success rate and minimize the mean energy and chain break fraction. The three quantities are highly (inversely) correlated and can be used interchangeably in many cases. If the ground state of the system is not known, the success rate can not be calculated. In that case, the simplest solution is seeking to minimize the mean energy of the entire sample. However, if we know the ground state of the system, or at least its energy, we can calculate the relative energy of each read, which makes the interpretation of results easier. If the state vector of a given sample is x and the ground state of a Hamiltonian Q_0 (in QUBO form)

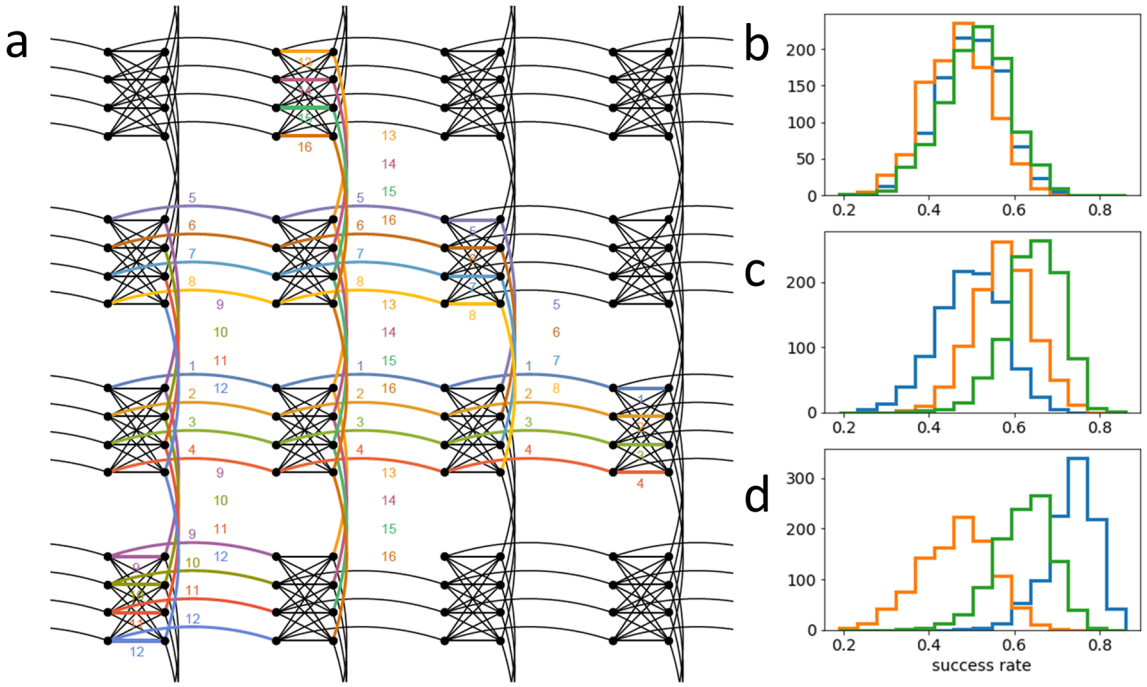

Fig. 1 **a** An example of an embedding. Physical qubits are represented with black dots and the connections between them with black lines. Logical qubits are enumerated and shown with colored connections between physical qubits. Each of the colored connections is forcing the two physical qubits to assume the same value. Therefore, a logical qubit is a chain of physical qubits, all (or most) of which assume the same value. There are post-processing methods in place on the quantum computer, which take care of disagreements between physical qubits in a chain. **b** Distributions of the success rate for three samplings of size 1000, when using the same embedding for the three samplings, **c** when using translated embeddings, and **d** when using distinct embeddings

is labelled x_0, the energy of the state x can be calculated as

$$E(\mathbf{x}) = \mathbf{x}^T Q_0 \mathbf{x}.$$

We can then define the relative energy of a given quantum state as

$$E_r(\mathbf{x}) = E(\mathbf{x}) - E(\mathbf{x}_0)$$

and the success rate can be calculated as the fraction of reads with relative energy zero (or close to it, to account for rounding errors).

3 Embeddings and noise

Before tacking the problem of quantum annealing correction, we first study the reproducibility of experiments on a quantum annealer. Furthermore, we illustrate the effect of different embeddings on the performance of annealing. We performed all experiments in this section on the 4×4 system under identical settings.

3.1 Statistics of annealing performance

Generally, when performing experiments on a quantum annealer, users perform each single experiment with several hundred reads and take the best result. In our work, we are studying the performance of this procedure, quantified through the success rate. To that end, we wish to study the distribution and statistical properties of the success rate, when such an experiment is repeated many times. We repeat the experiment 1000 times, each time reading 500 quantum states. We calculate the success rate of each experiment and thus obtain 1000 samples of the success rate. However, due to variable external conditions affecting noise in the quantum system, the performance of annealing may vary when repeated at different times. To account for this, we repeat the sampling of 1000 experiments two more times, with a delay of around 30 min between sampling. A histogram for each of the three samplings is depicted in Fig. 1b, where different colors correspond to different samplings. The distributions are symmetric, with means close to 0.5 and standard deviations around 0.08. The spread in success rate within one sampling is relatively large — in other words, the performance of annealing can vary significantly between identical experiments. On the other hand, a comparison of the three samplings reveals no significant difference in the distributions. Therefore, repeating samplings at different times to account for external factors is not necessary for the purposes of our study.

3.2 The effect of embeddings on annealing performance

The described experiment used a fixed minor embedding, depicted in Supplementary Fig. 1a. It is well established that the choice of embedding can have a significant effect on the performance of annealing. To study this with our approach, we vary the embedding of the problem in two different ways. First, we merely translate the original embedding within the lattice of the quantum computer, while keeping its shape and topology intact. In a translated embedding, the relative location (i.e., on the edge or in the middle of a chain) of each physical qubit, used for inter-chain coupling, is the same as in the original embedding. The only difference between translated embeddings is which individual physical qubits are used by an embedding. By translating the original embedding twice, we end up with three embeddings with identical topology, but using different physical qubits. The three embeddings are illustrated in Supplementary Fig. 1a–c. Using each of the three embeddings, we perform 1000 experiments, each with 500 reads. The corresponding distributions of success rate are depicted in Fig. 1c. Next, we generate three new minor embeddings that are both translated and have a different topology as compared to the original embedding. They are illustrated in Supplementary Fig. 1d–f. The respective distributions of success rate, obtained by performing 1000 experiments with 500 reads for each of the three embeddings, are depicted in Fig. 1d.

We observe significant differences between the success rate distributions of the translated embeddings, and even greater differences between the distributions of completely different embeddings. The fact that embeddings with an identical topology but different positioning in the D-Wave lattice yield significantly different distributions leads to the conclusion that the topology of the embedding is not the only important factor for performance. In other words, the calibration error is specific to each individual physical qubit. On the other hand, the topology of the embedding has an important effect as well, as evidenced by the larger differences in Fig. 1d for embeddings that are both translated and have a different topology. There is one final important observation we can make from Fig. 1b–d. Because the annealing performance varies significantly between different embeddings, this hints towards a simple but possibly effective method of improving the performance — generating and trying out different embeddings. However, it must be emphasized that due to the high level of random noise present, as well as the stochastic nature of quantum measurements, experiments that seek to evaluate annealing performance must be repeated many times to get reliable performance estimates.

4 Quantum annealing correction

We consider the QUBO Hamiltonian Q to be the input to the quantum annealer. In the process of translation to the physical system, the Hamiltonian is corrupted by both random and systematic noise. Can we learn to correct the input so as to produce the correct output, and in this way counteract systematic noise? We introduce small corrections to the input Hamiltonian:

$$Q = Q_0 + dQ,$$

where Q is the input to the D-Wave sampler, Q_0 is the original Hamiltonian that corresponds to the physical problem we are solving, and dQ is the calibration matrix, with $|dQ_{ij}| \ll |Q_{0_{ij}}|$, or alternatively, $||dQ|| \ll ||Q_0||$, where $||Q_0||$ indicates the matrix norm of Q. We vary the elements of dQ and try to maximize the success rate. The elements of Q correspond to the coupling strength between pairs of logical qubits. As an upper-triangular matrix, Q has $D(D+1)/2$ values for a fully connected system of $D \times D$ logical qubits. In this work, we study the calibration of the sample problem on three system sizes: 4×4, 5×5, and 6×6, with the corresponding dimensionalities of 136, 325, and 666. The calibration of the success rate through varying the matrix elements is a very challenging problem of noisy function optimization in an extremely high-dimensional space. Since we know nothing of the landscape of the systematic noise function, we employ a data-driven approach. We first conduct a large number of experiments by randomly sampling dQ_{ij} values and then employ machine learning methods for regression to analyze the data.

4.1 Sampling the *dQ* space

Before we can sample the dQ matrices for our dataset, we must properly define the space we are interested in. When we invoke the D-Wave sampler for a given Q, the D-Wave system performs a few modifications on the matrix. First the matrix is normalized so that the largest absolute value of its elements is 1. Next, the system transforms the matrix from the logical qubit representation to the physical qubit representation, using the embedding we provide. Finally, the system constructs qubit chains by assigning the chain strength parameter to the couplings between physical qubits in each chain. The normalization of the input after our corrections could compromise our data and confuse modelling. To avoid it, we normalize the Q_0 matrix ourselves and place a constraint on the sampling of dQ so as to ensure that the normalization is not broken by adding dQ. We define the absolutely largest element of Q_0 as

$$Q_{\max} = \max_{ij} |Q_{0_{ij}}|,$$

and derive the allowed range of each matrix element as

$$\eta\left(-1 - \frac{Q_{ij}}{Q_{\max}}\right) \leq dQ_{ij} \leq \eta\left(1 - \frac{Q_{ij}}{Q_{\max}}\right),$$

where $0 \leq \eta \leq 1$ denotes the fraction of the range of each matrix element we wish to consider.

The above formulation, with $\eta = 1$, describes the entire space of possible calibration matrices dQ, constrained only by normalization. We have assumed that the corrections of the input Hamiltonian are small. We control the size of the corrections through the parameter η. Choosing a very small η risks cutting the best dQ out of our search space. On the other hand, a large η vastly increases the size of our search space and consequently the QPU time demands. More importantly, a large η carries the risk that we wander so far away from Q_0 that the problem considered by the quantum computer no longer bears any connection to the physical problem we are studying. For example, the space of all Q also contains a Hamiltonian that represents a degenerate system with all energies equal to the ground state energy of the system we are interested in. Such a Q would achieve seemingly perfect annealing performance, but would be meaningless in practice. In order to ensure that our calibrated system is still solving the correct problem, it is paramount to keep the corrections of the Hamiltonian relatively small. For our experiments, we used the value $\eta = 0.05$ as an educated guess.

Now that we have defined the boundaries of the high-dimensional space we are to explore, we randomly sample it to generate the set of calibration matrices to be evaluated on the quantum annealer. We employ latin hypercube sampling (McKay et al. 2000), a method for random sampling from multidimensional distributions that ensures good coverage of the hyperspace, while keeping the required number of samples low. We study the annealing performance of three samplings, one for each of the three system sizes. In each sampling, we evaluate 10000 calibration matrices, each comprising 500 reads from the quantum annealer.

4.2 Analysis of the data

The three constructed datasets are summarized in Fig. 2. There are significant differences in annealing performance for different system sizes. For the 4×4 system, the mean success rate is close to 50%, for the 5×5 around 14%, and for the largest 6×6 system, the mean success rate is only around 0.5%. In fact, for many calibration matrices for the 6×6 system, not even a single read ended up in the ground state. A comparison between the distributions of the success rate and mean relative energy reveals that while the success rate is a nicer and more intuitive metric, its usefulness drops sharply when the annealing performance is very poor. In such cases, it is better to use the mean relative energy as the

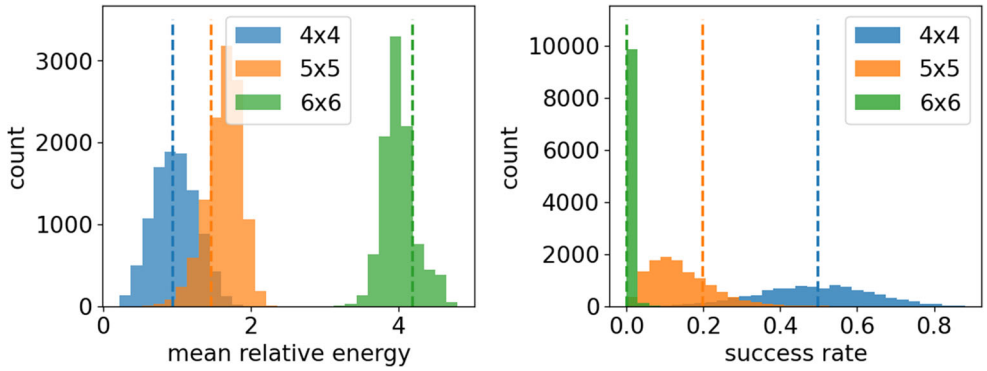

Fig. 2 Distributions of the mean relative energy (left) and the success rate (right), obtained by sampling the space of calibration matrices dQ, for three system sizes. For each system size, 10000 matrices were sampled. Each sample is composed of 500 reads from the quantum annealer, which are summarized with the mean relative energy and success rate statistics. The mean relative energy and success rate for the original Hamiltonian with no calibration ($Q = Q_0$) are depicted with a dashed line of the corresponding color

measure of annealing performance. In Fig. 2, the annealing performance for the uncalibrated Hamiltonian ($Q = Q_0$) is depicted by a dashed line for each system. We can see that, for all system sizes, a good fraction of the evaluated calibration matrices achieves better annealing performance than the original matrix. This observation is surprising, since we expected it to be very difficult to find a calibration that improves performance in the high-dimensional dQ spaces. In fact, taking the calibration matrix that yields the best annealing performance for each system size, we arrive at the results summarized in Table 1. The results are already impressive and demonstrate the effectiveness of error correction by corrections to the input Hamiltonian. Based on these results, we define the first, simple but effective, error correction algorithm — the *argmax* strategy.

4.3 *Argmax* strategy

Algorithm 1 defines the *argmax* strategy for error correction in quantum annealers.

1: compute the ranges for each dQ_{ij} for the problem Q_0 and choose η.
2: sample M calibration matrices dQ using a random sampling method (such as latin hypercube sampling (McKay et al. 2000)).
3: evaluate the annealing performance for each calibration matrix on the quantum annealer,
4: choose the calibration matrix that yields the highest success rate or the lowest mean relative energy.

Algorithm 1 The *argmax* strategy of adaptive quantum annealing correction.

Since the *argmax* strategy is based entirely on latin hypercube sampling, Table 1 already summarizes its performance. In order to assess how the strategy behaves with a lower number of samples, we perform repeated bootstrap subsampling on the gathered data (see SI for details on the methodology) and depict the resulting curves in Fig. 3a. We observe that significant improvements in annealing performance can be achieved even for sample sizes as small as 10 or 100 calibration matrices. Increases in sample size offer diminishing returns, as the curves asymptotically approach the maximum of the full sample.

4.4 *Predictive* strategy

The *argmax* strategy is limited to the calibration matrices we have sampled and evaluated. In the following, we employ machine learning methods to develop a strategy that interpolates or even extrapolates the assembled data in order to find the best calibration matrix. We evaluate the predictive performance of models obtained by different machine learning methods through 5-fold cross-validation. We randomly partition the dataset into 5 fractions. Models are trained on four of the fractions (training set) and attempt to predict the fifth (testing set) fraction. In each of the five iterations, a different fraction is used as the testing set. In the end, the predictions are pooled together and the prediction

Table 1 Annealing performance when using the original Q_0, compared to the best annealing performance measured in the sample of 10000 calibrated matrices $Q = Q_0 + dQ$, which corresponds to the *argmax* strategy

System	4×4	5×5	6×6
Success rate of original Q_0	0.50	0.20	0.0001
Best success rate of dQ sampling	0.88	0.62	0.08
% improvement in success rate	78%	210%	80000%

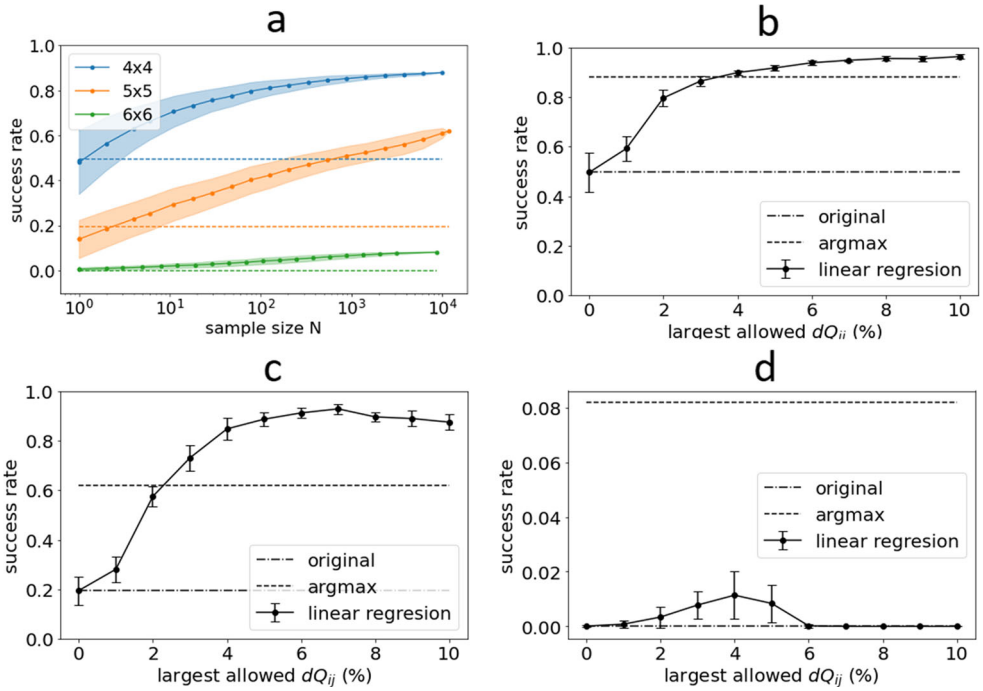

Fig. 3 **a** Dependence of the success rate of the *argmax* calibration strategy on the size of the training dataset (circles and full line). Calculated through bootstrap resampling with 1000 repetitions. The results are compared with the success rate without calibration (dashed line). Standard deviations of success rate are indicated by the shaded areas. The figures below show the dependence of the success rate of the *linear regression* calibration strategy for the **b** 4×4, **c** 5×5, and **d** 6×6 systems on the largest element of the calibration matrix dQ_{ij} allowed (circles and error bars). For each system size, the results are compared with the success rate of the **argmax** strategy (dashed line) and with the success rate without calibration (dotted line)

performance is evaluated. As the metric of comparison, we use the coefficient of determination, defined as

$$R^2 = 1 - \frac{\sum_i (\tilde{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2},$$

where $\tilde{y}_i$ indicates the predicted annealing performance of the ith calibration matrix, $\bar{y}$ is the mean annealing performance for the testing set, and y_i denotes the measured annealing performance of the ith calibration matrix. The coefficient of determination can be interpreted as the portion of variance in the data explained by the model. A perfect model would achieve $R^2 = 1$, whereas $R^2 < 0$ indicates the model is worse than simply taking the average over the test set. In the analysis of the data, we observed high (anti) correlation between mean energy and success rate as measures of annealing performance. Success rate is problematic for the 6×6 system, since many calibration matrices do not end up with any reads in the ground state. For this reason, we have chosen mean relative energy as the measure for annealing performance when training predictive models, even though the success rate offers easier interpretation.

In Table 2, we report the R^2 for ordinary linear regression — LR, linear regression with L1 regularization — Lasso (Tibshirani 1996), linear regression with L2 regularization — Ridge (Hoerl and Kennard 1970), support vector regression with a kernel of radial basis functions — SVR (Drucker et al. 1997), ensembles of decision trees for regression — random forests — RF (Breiman 2001), as well as feedforward neural networks — FNN (Schmidhuber 2015). The details of the employed methods are reported in Supplementary Information. The coefficients of determination are not high, which is consistent with the large variance of the distributions observed in Fig. 1. For all system sizes, all variations of linear regression, as well as SVR and FNN, achieve similar predictive performance. The predictive performance drops as we increase system size. This is expected, due to the curse of dimensionality — the dimensionality of the problem increases quadratically with system size, but our sample size remains constant (Friedman et al. 2001). Since linear regression is the simplest and most robust of the models we have tried and it still achieved competitive predictive performance, we have chosen it for implementing the predictive strategy.

We now have a model that predicts mean relative energy for a given calibration matrix dQ and we can use it as a stand-in for experiments on the quantum

Table 2 The coefficient of determination R^2, achieved by different models for the task of predicting the mean energy, given calibration matrices dQ, for three different system sizes. The best performing method for each system size is highlighted in bold font

	LR	Lasso	Ridge	SVR	RF	FNN
4×4	0.475	**0.476**	0.475	0.454	0.213	0.463
5×5	0.272	0.274	0.273	0.134	0.067	**0.277**
6×6	0.091	**0.113**	0.109	0.076	0.019	0.076

A higher R^2 indicates a better model, with the maximal possible value being $R^2 = 1$. The machine learning methods used are linear regression (LR), linear regression with L1 regularization (Lasso), linear regression with L2 regularization (Ridge), support vector machines with an RBF kernel (SVR), ensembles of decision trees (random forest — RF), and feedforward neural networks (FNN)

annealer. To find the best calibration, we use a method for numerical minimization (in our case, differential evolution as implemented in Python-Scipy (Virtanen et al. 2020)) to find the minimum of the mean relative energy. Minimization requires bounds on the high-dimensional space. Since we know only that the calibration matrix should be small in comparison to Q_0, we once again invoke the dQ_{ij} ranges defined earlier and vary the range fraction η. For each value of η, we minimize the model to obtain a candidate calibration matrix.

1: Compute the ranges for each dQ_{ij} for the problem Q_0 and choose η.
2: Sample N calibration matrices dQ using a random sampling method (such as latin hypercube (McKay et al. 2000) sampling).
3: Evaluate the annealing performance for each calibration matrix on the quantum annealer.
4: Learn predictive models on the data (such as linear regression), employing cross-validation to select the best model.
5: Vary η to define search spaces of increasing size and use an optimization algorithm (such as differential evolution) on the trained model to find the candidate calibration matrix for each η.
6: Test the suggested matrices on the quantum annealer and select the best one.

Algorithm 2 The *predictive* strategy of adaptive quantum annealer correction.

We then evaluate the matrix on the quantum annealer, repeating the measurement 10 times, each time with 500 reads, and compute the mean and standard deviation of the measured success rate. By performing this procedure for each value of η (using the same model) and for each system size (using the respective models), we obtain the curves depicted in Fig. 2b–d.

Despite the relatively low coefficients of determination, the proposed strategy shows large improvements in annealing performance. The *predictive* strategy outperforms the *argmax* strategy for the 4×4 system by 14% and for the 5×5 system by 155%, but falls short of it for the 6×6 system. The observations are easy to understand if we consider the fact that for the 4×4 system, the *argmax* strategy was able to achieve nearly perfect annealing performance, leaving little space for improvement by the *predictive* strategy. For the 6×6 system, the sample size is too low for the model to make meaningful predictions and it is better to simply use the best evaluated calibration matrix. The 5×5 system lies in between the other two, where the predictive model learns enough to dramatically outperform the *argmax* strategy. We expect that with more invested QPU time, similar improvements could be achieved for the 6×6 system as well. See Supplementary Fig. 2 for an estimation of the learning curve of the *predictive* strategy, similar to the performance curve for the *argmax* strategy in Fig. 1a.

4.5 Applicability

The described approaches to quantum annealing correction work, but have certain shortcomings. Firstly, we have demonstrated gains in annealing performance only for the same problem that we performed calibration on. We expect that we could not easily use a discovered calibration matrix on new problems. This is due to the fact that our calibration does not address physical qubits directly, but rather indirectly by tuning the couplings between different chains, which have a different meaning when using a different embedding. However, generalization might be possible for families of physical problems that differ from each other only in the values of nonzero elements of the Hamiltonian and can therefore use the same embedding. Further work is required to investigate this possibility.

The second downside of our approach is that we must perform a number of experiments on a quantum annealer to find a good calibration matrix. On the one hand, we demonstrate that we can achieve significant improvements in annealing performance even sampling only 10–100 calibration matrices. On the other hand, this can quickly become too demanding in terms of processing time for use cases that require the solving of many different problems. Furthermore, since the size of the calibration matrix search space increases exponentially, our quantum annealing correction methods do not scale well to larger problems. However, we developed the approach on a problem featuring fully connected topology, which is the most complex type of problem to address in quantum annealing. We expect that more sparsely connected problems would be less demanding on quantum processor time and would scale better to larger sizes. The described approach to quantum annealing correction is therefore most suitable when we are dealing with a small number of difficult problems and the solution quality is a primary concern.

Finally, note that the first three steps of the *predictive strategy* are identical to the *argmax* strategy, so the dataset constructed by sampling the dQ space can be reused for both strategies. In fact, we recommend the use of both strategies, because it is difficult to predict which of them will be better for a given problem, as evidenced by our experiments on systems of different sizes. It must also be emphasized that the model training and inference, involved in the *predictive* strategy, is performed on a classical computer and is not computationally demanding. In our

experiments, training even the most demanding neural network model took only minutes on a modern CPU.

5 Insights on noise in quantum annealers

When analyzing the dataset constructed by randomly sampling the dQ space, a few observations stand out. Firstly, for all system sizes, nearly half of the sampled calibration matrices result in better annealing performance than using no calibration at all. Secondly, linear regression is able to model the dependence of annealing performance on dQ equally well as more complex, nonlinear methods. The latter can be partially explained by the large amount of random noise present in the measurements, which can increase the risk of overfitting in models and necessitate the use of simple, robust models. However, in such a case, one would expect model regularization to improve the predictive performance, which we observed only to a small degree. Therefore, in combination with the former observation, we are lead to the conclusion that the systematic error landscape is much simpler than we had anticipated. When considering the success rate, dependent upon the elements of the dQ calibration matrix, a single global peak in the high-dimensional space, with few or no local maxima, would explain our observations.

To further explore the noise landscape, we perform a new experiment on the 4×4 system where we perform walks, each starting at Q_0 and following a line in randomly chosen direction. The results are depicted in Fig. 4. We observe two distinct patterns in the measured walks: the annealing performance either monotonically falls towards zero, or it reaches a single maximum before falling off. Furthermore, out of the 14 walks, 4 achieve at some point

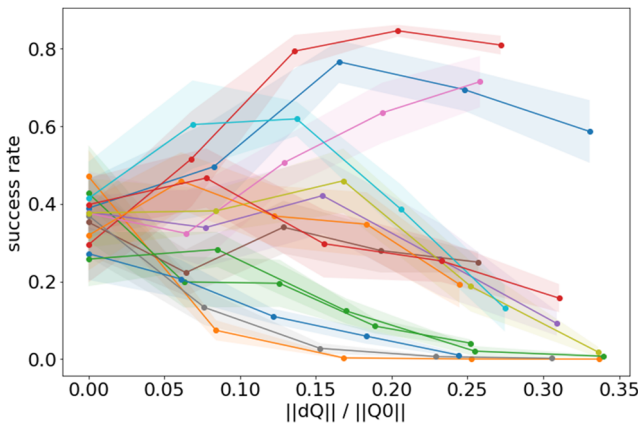

Fig. 4 Annealing performance for walks along straight lines in random directions in the dQ space, starting at Q_0, for the 4×4 system. Each color denotes a distinct random direction. Each measurement is repeated 10 times, with the dots and full line depicting the mean and the shaded areas depicting the standard deviation of each experiment

an annealing performance significantly better than without calibration. Were the noise landscape rugged, we would expect the walks to be rugged as well — experiencing many local extrema before falling towards zero. Furthermore, we would expect walks that improve the starting annealing performance to be extremely rare in a 136-dimensional space. Therefore, the presented observations are consistent with the conclusions we reached when analyzing the dQ dataset. The landscape of systematic error is relatively smooth and likely features only a single maximum for $||dQ|| \ll ||Q_0||$.

This is consistent with a linear approximation of the systematic noise already presented in the D-Wave documentation, where each individual coupling between physical qubits, as well as the external field on each individual qubit, is perturbed by a small constant. Using our calibration methods, we were able to successfully find the appropriate compensation for the specific systematic error present in a specific problem deployed on the D-Wave quantum annealer.

6 Conclusions

We have demonstrated a novel approach to quantum annealing error correction, based on compensating for systematic noise by correcting the input to the quantum annealer. Both the *argmax* and the *predictive* strategy achieve significant improvements of up to 3 orders of magnitude in annealing performance and can be put to good use by practitioners who struggle with the effects of noise in their quantum annealing experiments and applications. However, our method is not well-suited to all use cases. For example, when a large amount of solutions is required under a time constraint or when the problem at hand is native to the topology of the quantum computer, the method will be unlikely to achieve improvements in annealing performance. However, when solution quality is important and the problem is complex or has a random topology, our method can exhibit significant advantage and may be customized to any problem.

Moreover, using various machine learning strategies, we developed important insight into the nature of noise present on the quantum computer, which can be used in order to direct future calibration strategies that are more efficient in the use of QPU time. One avenue for further investigation are gradient-based optimization methods, which are very efficient, but can easily get stuck in local minima. Since the noise landscape appears to be devoid of local minima, gradient-based methods might prove a good choice. However, since the gradient is not known analytically, it must be estimated at each step of the optimization process by performing measurements on the

quantum annealer. In a high-dimensional space, this might prove expensive.

On the other hand, our observations indicate that the interactions between the elements of dQ do not influence annealing performance. This property can be exploited by simplifying the problem from optimizing all the dimensions at the same time, to optimizing each of them independently. A strategy of sequential one-dimensional optimization would scale with system size linearly, instead of exponentially, and allow for scalable calibration of larger systems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42484-022-00092-y.

Author contribution Jure Brence: investigation, formal analysis, writing — original draft; Dragan Mihailovic: supervision, funding acquisition, writing — review and editing; Viktor V. Kabanov: validation, supervision; Ljupčo Todorovski: methodology, validation, writing — review and editing; Sašo Džeroski: supervision, funding acquisition, writing — review and editing; Jaka Vodeb: conceptualization, methodology, writing — original draft.

Funding We acknowledge ARRS core funding P2-0103, P1-0040 and P1-0416, as well as funding from the young researcher grant P08333.

Data availability The data generated and analyzed in the presented study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

Albash T, Lidar DA (2018) Demonstration of a scaling advantage for a quantum annealer over simulated annealing. Phys Rev X, 8

Albash T, Vinci W, Mishra A, Warburton PA, Lidar DA (2015) Consistency tests of classical and quantum models for a quantum annealer. Phys Rev A, 91

Bando Y, Susa Y, Oshiyama H, Shibata N, Ohzeki M, Gómez-Ruiz FJ, Lidar DA, Suzuki S, del Campo A, Nishimori H (2020) Probing the universality of topological defect formation in a quantum annealer Kibble-zurek mechanism and beyond. ArXiv, 2

Benedetti M, Realpe-Gómez J., Biswas R, Perdomo-Ortiz A (2016) Estimation of effective temperatures in quantum annealers for sampling applications: a case study with possible applications in deep learning. Phys Rev A, 94

Boixo S, Smelyanskiy VN, Shabani A, Isakov SV, Dykman M, Denchev VS, Amin MH, Smirnov AY, Mohseni M, Neven H (2016) Computational multiqubit tunnelling in programmable quantum annealers. Nat Commun, 7

Breiman L (2001) Random forests. Mach Learn 45:5–32

Buffoni L, Campisi M (2020) Thermodynamics of a quantum annealer Quantum Science and Technology

D-Wave Systems Inc (2021) Technical description of the D-Wave quantum processing unit technical description of the D-wave quantum processing unit

Drucker H, Burges C, Kaufman L, Smola A, Vapnik V (1997) Support vector regression machines. Adv Neural Inf Process Syst 9:155–161

Friedman J, Hastie T, Tibshirani R et al (2001) The elements of statistical learning, vol 1. Springer, New York

Gardas B, Deffner S (2018) Quantum fluctuation theorem for error diagnostics in quantum annealers. Sci Rep, 8

Gardas B, Dziarmaga J, Zurek WH, Zwolak M (2018) Defects in quantum computers. Sci Rep 8:2–11

Harris R, Sato Y, Berkley AJ, Reis M, Altomare F, Amin MH, Boothby K, Bunyk P, Deng C et al (2018) Phase transitions in a programmable quantum spin glass simulator. Science 361:162–165

Hoerl A, Kennard R (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12:55–67

Jain S, Ziauddin J, Leonchyk P, Yenkanchi S, Geraci J (2020) Quantum and classical machine learning for the classification of non-small-cell lung cancer patients. SN Appl Sci, 2

Job J, Lidar D (2018) Test-driving 1000 qubits. Quantum Sci Technol, 3

Kairys P, King AD, Ozfidan I, Boothby K, Raymond J, Banerjee A, Humble TS (2020) Simulating the Shastry-Sutherland Ising model using quantum annealing. PRX Quantum 1:1

King AD, Carrasquilla J, Raymond J, Ozfidan I, Andriyash E, Berkley A, Reis M, Lanting T, Harris R, Altomare F et al (2018) Observation of topological phenomena in a programmable lattice of 1,800 qubits. Nature 560:456–460

King AD, Raymond J, Lanting T, Isakov SV, Mohseni M, Poulin-Lamarre G, Ejtemaee S, Bernoudy W, Ozfidan I, Smirnov AY et al (2021) Scaling advantage over path-integral monte carlo in quantum simulation of geometrically frustrated magnets. Nat Commun, 12

Li RY, Di Felice R, Rohs R, Lidar DA (2018) Quantum annealing versus classical machine learning applied to a simplified computational biology problem. NPJ Quantum Inf, 4

McKay M, Beckman R, Conover W (2000) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42:55–61

Mott A, Job J, Vlimant JR, Lidar D, Spiropulu M (2017) Solving a higgs optimization problem with quantum annealing for machine learning. Nature 550:375–379

Neukart F, Compostella G, Seidel C, von Dollen D, Yarkoni S, Parney B (2017) Traffic flow optimization using a quantum annealer. Front ICT 4

Orús R, Mugel S, Lizaso E (2019) Forecasting financial crashes with quantum computing. Phys Rev A, 99

Pudenz KL, Albash T, Lidar DA (2014) Error-corrected quantum annealing with hundreds of qubits. Nat Commun, 5

Raymond J, Ndiaye N, Rayaprolu G, King AD (2020) Improving performance of logical qubits by parameter tuning and topology compensation. Proceedings of the IEEE international conference

on quantum computing and engineering (QCE 2020:295–305

Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Ro Stat Soc Ser B (Methodological) 58:267–288

Virtanen P, Gommers R, Oliphant T, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J et al (2020) Scipy 1.0: fundamental algorithms for scientific computing in python. Nat methods 17(3):261–272

Vodeb J, Kabanov V, Gerasimenko Y, Ravnik J, van Midden M, Zupanic E, Sutar P, Mihailovic D et al (2019) Configurational electronic states in layered transition metal dichalcogenides. New J Phys 21(8):083001

Willsch D, Willsch M, De Raedt H, Michielsen K (2020) Support vector machines on the d-wave quantum annealer. Comput Phys Commun, 248

Zaborniak T, de Sousa R (2021) Benchmarking hamiltonian noise in the D-wave quantum annealer. IEEE Trans Quantum Eng 2:1–6