

This is the accepted manuscript made available via CHORUS. The article has been published as:

Scalable effective-temperature reduction for quantum annealers via nested quantum annealing correction

Walter Vinci and Daniel A. Lidar

Phys. Rev. A **97**, 022308 — Published 6 February 2018

DOI: [10.1103/PhysRevA.97.022308](https://doi.org/10.1103/PhysRevA.97.022308)

Scalable effective temperature reduction for quantum annealers via nested quantum annealing correction

Walter Vinci^{1,2,3} and Daniel A. Lidar^{1,2,3,4}

¹*Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089, USA*

²*Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA*

³*Center for Quantum Information Science & Technology,*

University of Southern California, Los Angeles, California 90089, USA

⁴*Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA*

Nested quantum annealing correction (NQAC) is an error correcting scheme for quantum annealing that allows for the encoding of a logical qubit into an arbitrarily large number of physical qubits. The encoding replaces each logical qubit by a complete graph of degree C . The nesting level C represents the distance of the error-correcting code and controls the amount of protection against thermal and control errors. Theoretical mean-field analyses and empirical data obtained with a D-Wave Two quantum annealer (supporting up to 512 qubits) showed that NQAC has the potential to achieve a scalable effective temperature reduction, $T_{\text{eff}} \sim C^{-\eta}$, with $\eta \leq 2$. We confirm that this scaling is preserved when NQAC is tested on a D-Wave 2000Q device (supporting up to 2048 qubits). In addition, we show that NQAC can be also used in sampling problems to lower the effective temperature of a quantum annealer. Such effective temperature reduction is relevant for machine-learning applications. Since we demonstrate that NQAC achieves error correction via an effective reduction of the temperature of the quantum annealing device, our results address the problem of the “temperature scaling law for quantum annealers”, which requires the temperature of quantum annealers to be reduced as problems of larger sizes are attempted to be solved.

I. INTRODUCTION

Quantum annealing (QA) [1–5] is a predecessor of the quantum adiabatic algorithm (QAA) [6, 7] and adiabatic quantum optimization (AQO) [8, 9], originally conceived as an algorithm that exploits simulated quantum (rather than thermal) fluctuations and tunneling to providing a quantum-inspired version of simulated annealing (SA) [10] for the solution of combinatorial optimization problems. Nowadays it is considered a special case of adiabatic quantum computation (AQC) [11], a paradigm that is universal for quantum computation [12–19]. For reviews see [20, 21].

In AQC, the computation proceeds from an initial Hamiltonian whose ground state is easy to prepare, to a final Hamiltonian whose ground state encodes the solution to the computational problem. In the closed-system setting, the adiabatic theorem [22, 23] guarantees that the system will track the instantaneous ground state provided the Hamiltonian varies sufficiently slowly, and relates the required running time to the inverse of the minimum gap encountered during the computation. The only errors present in this setting are control errors and non-adiabatic transitions. The latter can be made arbitrarily small by a suitable choice of the path from the initial to the final Hamiltonian [24–27].

In the open-system setting, interactions with the environment produce additional errors in the form of dephasing between computational basis states and thermal excitations. Thermal excitation errors are suppressed due to the presence of finite gaps during the computation and corrected due to a certain amount of relaxation back to the ground state [28–32]. This picture is generally appropriate in the weak-coupling regime and when the temperature of the bath is small compared to the minimum gap [33, 34].

In practical implementations of AQC and QA, however, the system may be strongly coupled to the environment and

the temperature of the thermal bath may be much larger than the small gaps expected when implementing computationally hard problems. It is not a surprise that in this scenario, AQC and QA require quantum error correction just like any other form of quantum information processing [35]. Moreover, it has been shown that for Hamiltonians that are sums of commuting two-body interactions it is not possible to even store quantum information reliably in the ground subspace (let alone compute), even at zero temperature [36]. Unfortunately, despite the existence of various error suppression and correction techniques [37–53], it is not yet known how to achieve fault-tolerance in AQC and QA. Effective error suppression codes for AQC and QA typically require highly non-local Hamiltonians and very strong energy penalties [40, 41, 51–54] which are difficult to implement physically. Moreover, fault-tolerance assumes that it is possible to implement codes of arbitrary size (as measured in terms of the number of physical and logical qubits), in order to allow for scalable error correction of arbitrarily large computations. This is achieved in the circuit model, e.g., via concatenated codes [55]. While a comparable approach is still missing in AQC, nested quantum annealing correction (NQAC) [56] represents the first attempt to introduce scalable error correction for quantum annealing. NQAC is a scalable generalization of quantum annealing correction (QAC) [45–49], an error correcting techniques tailored to commercially available quantum annealing devices [57–61].

Such devices operate in a noisy regime in which both thermal and analog control errors play a crucial role in determining their performance [62–69]. Indeed, both theoretical and empirical studies suggest current quantum annealing devices work in a *quasi-static regime* [70–74]. In this regime, due to the strong interaction with the thermal bath, there is an initial phase of quasi-static evolution in which thermalization times are much shorter than the annealing time. Towards the end of

the anneal, thermalization times grow and eventually become much longer than the annealing time. The system thus enters a regime in which dynamics is frozen. The final outcomes of a quasi-static quantum annealing process thus provides a snapshot of the thermal (Gibbs) state of the system at the freezing point. The faster the freezing process, the closer to a thermal distribution at the freezing point the system will be at the end of the annealing process. A rough measure of performance for a quasi-static quantum annealer is thus given by an effective freezing temperature T_{eff} , i.e., the temperature of the approximate thermal distribution at the freezing point.

Performing quantum annealing in a quasi-static regime may hide signs of potential quantum speedup, as discussed in Ref. [70]. On the other hand, this provides an opportunity for using quantum annealers as thermal samplers that can be effectively used in machine-learning applications, e.g., in the training of Boltzmann machines [75–78]. On general grounds we expect the effective freezing temperature T_{eff} to grow with the problem size (earlier freezing or longer thermalization times for larger systems). The effects of intrinsic control errors typically also grows with problem size. Thus, no matter what the application, error correction remains crucial in order to ensure a scalable use of quantum annealers.

A successful error correction scheme in quasi-static quantum annealing should effectively reduce the freezing temperature T_{eff} :

$$T_{\text{eff}} \rightarrow T_{\text{eff}}/\mu, \quad \mu > 1. \quad (1)$$

The temperature reduction can be equivalently interpreted as achieving an effective energy boost

$$(h_i, J_{ij}) \rightarrow \mu(h_i, J_{ij}), \quad (2)$$

where h_i and J_{ij} are the couplings defining the given problem Hamiltonian [see Eq. (5)]. A scalable error correction scheme should allow for μ to scale with the code size.

In Ref. [56] it was shown that NQAC indeed achieves error correction by introducing an energy boost μ and thus effectively reducing the effective temperature at which a D-Wave Two (DW2) quantum annealer operated. This immediately serves to address pessimistic “temperature scaling law” conclusions of the type presented in Ref. [74], which do not account for error correction. Moreover, Ref. [56] showed that the energy boost obeys a power-law scaling:

$$\mu \sim \left(\frac{N_{\text{phys}}}{N} \right)^\eta, \quad (3)$$

where N_{phys} is the number of physical qubits used to encode N logical qubits. These empirical results were theoretically interpreted in terms of a mean field analysis in Refs. [49, 56, 79].

This paper is organized as follows. Section II provides a summary and preview of our main results. We briefly review NQAC in Sec. III. In Sec. IV we show that the power law scaling found on a DW2 device continues to hold on a DW2KQ device, albeit with a reduced η value. The DW2KQ device enabled us to use up to 728 physical qubits for NQAC encoding, whereas NQAC on the DW2 was limited to 288 physical

qubits. We also show that NQAC outperforms classical repetition coding when NQAC is implemented in a regime in which the strength of the energy penalties is not a limiting factor. In Sec. V we show that NQAC can be implemented in conjunction with sampling applications. In particular, we show that NQAC can be used to reduce the sampling temperature of a quantum annealing device without significantly reducing the quality of the sampling. We conclude with a discussion of the implications of our results in Sec. VI.

II. PREVIEW OF MAIN RESULTS

Our work extends the results and analysis of Ref. [56], which introduced NQAC, in two significant ways. First, we confirm that NQAC provides a performance improvement that can be interpreted as an effective temperature reduction when the method is implemented on a D-Wave 2000Q (DW2KQ) quantum annealer (the DW2KQ device has four times as many physical qubits as the DW2 device used in Ref. [56]). This allows us to consider NQAC encodings with larger nesting levels (or code distance) C ; e.g., from $C = 8$ on the DW2 device to $C = 13$ in the present case, for K_4 problem instances. Moreover, we are able to study the encoding of larger problems (up to $C = 2$ for K_{24} problem instances). We confirm the existence of a scaling law $T_{\text{eff}} \sim C^{-\eta}$ that is valid in the scaling regime, i.e., when the energy scale factor α is small enough that the NQAC implementation is not limited by the strength of the energy penalties. In this scaling regime NQAC outperforms classical repetition, a benchmark for any bona fide error correcting scheme.

Second, for the first time we study the use of NQAC in sampling applications, specifically estimating the gradient required for training Boltzmann machines. We argue that there are two important quantities that characterize the performance of quantum annealers as samplers: the sampling temperature and the quality of sampling. Improving the performance of a quantum annealing device for sampling via error correction should reduce the sampling temperature of the device without diminishing the quality of the sampling. We show that NQAC can indeed achieve a monotonic reduction of the sampling temperature. As in the case of optimization applications, this holds in the scaling regime, i.e., when the encoding is not limited by the strength of the energy penalties used in the encoding.

III. NESTED QUANTUM ANNEALING CORRECTION

We start by briefly reviewing the nested quantum annealing correction (NQAC) construction [56]. A standard quantum annealing protocol is defined by the following transverse-field quantum annealing Hamiltonian:

$$H(t) = A(t)H_X + B(t)H_P, \quad t \in [0, t_f], \quad (4)$$

in which the annealing schedules $A(t)$ and $B(t)$ (monotonically decreasing and increasing respectively) control the annealing schedule. The driver $H_X = -\sum_i \sigma_i^x$ serves to set

up an initial uniform superposition (its ground state) and controls the tunneling rate, while the solution to an optimization problem of interest is encoded in the ground state of the Ising problem Hamiltonian H_P :

$$H_P = \sum_{i \in \mathcal{V}} h_i \sigma_i^z + \sum_{(i,j) \in \mathcal{E}} J_{ij} \sigma_i^z \sigma_j^z. \quad (5)$$

The sums above are performed over the vertices \mathcal{V} and edges \mathcal{E} of a graph $G = (\mathcal{V}, \mathcal{E})$ that corresponds to the hardware (or connectivity) graph of a given quantum annealing device. The local fields $\{h_i\}$, couplings $\{J_{ij}\}$, and annealing time t_f are programmed in order to represent the correct computational problem [57–59].

In quantum annealing correction, instead of trying to solve a computational problem by directly implementing the quantum annealing Hamiltonian $H(t)$, one builds an “encoded Hamiltonian” $\bar{H}(t)$ [45–49]:

$$\bar{H}(t) = A(t)H_X + B(t)\tilde{H}_P, \quad t \in [0, t_f], \quad (6)$$

defined over a set of physical qubits N_{phys} larger than the number of logical qubits N . We call \tilde{H}_P the “encoded problem Hamiltonian”. The logical states of the logical Hamiltonian H_P are then recovered through an appropriate decoding procedure of the states of \tilde{H}_P . Note that, due to practical limitations of current QA devices that prevent the programmability of the driver term H_X , only H_P is encoded in QAC.

Nested quantum annealing correction was introduced in Ref. [56]¹ with the goals of obtaining a QAC scheme that (1) can be implemented on arbitrary quadratic unconstrained optimization problems (QUBO), (2) is potentially scalable by allowing a variable code-size and (3) could be implemented on a generic quantum annealing device. These three goals were achieved by means of a “nesting” procedure that we now outline.

The most general QUBO has arbitrary pairwise interactions between a set of N logical variables. We thus consider the encoding of a problem Hamiltonian H_P defined on a complete graph K_N . NQAC involves two maps between three types of qubits: (i) each logical qubit (representing a logical variable) is mapped to a set of code qubits, whereupon $H_P \mapsto \tilde{H}_P$, (ii) each code qubit is mapped to a set of physical qubits, whereupon $\tilde{H}_P \mapsto \bar{H}_P$, the encoded problem Hamiltonian in Eq. (6). The role of the code qubits is primarily to provide protection against thermal and control errors. The code qubits typically comprise a repetition code, though more general stabilizer codes are certainly possible in principle [42]. The map from code qubits to physical qubits is required for embedding purposes, since the code qubits of a given logical qubit are fully connected in NQAC, and the hardware graph G is typically not fully connected. The repetition code can be decoded at the end of each annealing run, which means that excited states can be used profitably to recover the sought-after

ground state of H_P . We now describe each of these components of NQAC in turn, in more detail

A. From logical qubits to code qubits

The first step of the NQAC scheme consists in transforming the logical problem H_P into a “nested” Hamiltonian \tilde{H}_P that is defined on a larger $K_{C \times N}$. The nesting level C controls the amount of hardware resources (qubits, couplers, etc.) used in the transformation and enables the scalability of the error correction method. Each logical qubit i ($i = 1, \dots, N$) in \tilde{H}_P is represented by a C -tuple of code qubits (i, c) , with $c = 1, \dots, C$. This C -tuple occupies the vertices of a K_C with equally weighted ferromagnetic edges, which also serves as a distance- C repetition code. Code qubits c and c' belonging to different logical qubits i and j are coupled with strength J_{ij} . Thus, the “nested” couplers $\tilde{J}_{(i,c),(j,c')}$ and local fields $\tilde{h}_{(i,c)}$ are defined as follows:

$$\tilde{J}_{(i,c),(j,c')} = J_{ij}, \quad \forall c, c', i \neq j, \quad (7a)$$

$$\tilde{h}_{(i,c)} = Ch_i, \quad \forall c, i, \quad (7b)$$

$$\tilde{J}_{(i,c),(i,c')} = -\gamma, \quad \forall c \neq c'. \quad (7c)$$

Note that each logical coupling J_{ij} has C^2 copies $\tilde{J}_{(i,c),(j,c')}$ while the local fields h_i have C copies $\tilde{h}_{(i,c)}$. For each logical qubit i , there are $C(C-1)/2$ ferromagnetic couplings $\tilde{J}_{(i,c),(i,c')}$ of strength $\gamma > 0$ representing energy penalties that facilitate the alignment of the C code qubits that represent one logical qubit.

B. From code qubits to physical qubits

The nested Hamiltonian \tilde{H}_P constructed over the code qubits in the previous step must be implemented in given QA hardware. The required transformation from code qubits to physical qubits, i.e., from \tilde{H}_P to \bar{H}_P can be accomplished using, e.g., minor embedding (ME) [80–84], or the Lechner-Hauke-Zoller (LHZ) scheme [85–87]. We focus here on the ME scheme primarily since it is the relevant one for D-Wave quantum annealers.

The ME step replaces each code qubit in \tilde{H}_P by a ferromagnetically coupled chain of physical qubits, such that all couplings in \tilde{H}_P are represented by inter-chain couplings. The intra-chain coupling represents additional energy penalties that force the chain of physical qubits to behave as a single code qubit. The minor embedding of a $K_{C \times N}$ graph requires that each code qubit (i, c) is represented by a physical chain of length $L = \lceil CN/4 \rceil + 1$ on the Chimera graph [81]. The number of physical qubits necessary to implement the C -th level of nested encoding (i.e., a distance- C code) of a problem with N logical variables is thus

$$N_C^{\text{phys}} = CNL \sim C^2 N^2 / 4, \quad (8)$$

i.e., it scales quadratically with both C and N .

¹ A 9-bit instance of the scheme first appeared in Ref. [42] in the context of correcting uncorrelated random control errors on the problem Hamiltonian. This work showed that such control errors can be quadratically suppressed in the code distance using an NQAC-like scheme.

C. Decoding

Finally, a decoding procedure must be employed to recover the logical state from the readout of the physical qubits. This is a two-step process since we must first decode the length- L chain of physical qubits representing each code qubit (i, c) , and then decode the C code qubits to yield the state of the i -th logical qubit. For simplicity we only consider majority vote decoding for both steps, although other decoding strategies are possible and have been explored [47].

D. Boosting effect

As mentioned above, error correction for quantum annealing in the quasi-static scenario should achieve a scalable effective temperature reduction. In the NQAC construction, each logical coupling is represented by C^2 physical copies and could ideally provide a maximal energy boost that grows quadratically with the nesting level C : $\mu \sim C^\eta$ with $\eta = 2$. A mean-field analysis has confirmed this by showing that the free energy \mathcal{F} after nesting is indeed equal to the free energy without encoding provided the logical couplings are boosted by a factor of C^2 [49, 56, 79]:

$$\mathcal{F}_C(\beta, J, \lambda, \Gamma) = \mathcal{F}_1(\beta, C^2 J, C^2 \lambda, C\Gamma). \quad (9)$$

Note, however, that the transverse field is only sub-optimally boosted by a factor of C : $\Gamma \mapsto C\Gamma$. This is due to the aforementioned limitation of not being able to encode the driver Hamiltonian. Despite this limitation, NQAC was shown to provide a scalable energy boost in experiments performed with a DW2 quantum annealer, in which small logical problems were nested until the physically encoded Hamiltonian used up to 288 physical qubits. The empirical findings confirmed a power law scaling of the energy boost with nesting level, but with a suboptimal scaling factor $\eta < 2$. This suboptimal scaling is a consequence of the combined detrimental effects of minor embedding, control errors, and the role played in the dynamics by the unencoded transverse field [56].

IV. NQAC FOR OPTIMIZATION

In this and the next section we discuss NQAC in experiments with up to 728 physical qubits performed on a DW2KQ device, up from 288 reported in Ref. [56], which used a DW2 device. The results of this section confirm and extend those obtained in Ref. [56], while the results of the next section address NQAC in the new context of sampling. Experimental background details are given in Appendix A.

A. Scalable energy boost

We closely follow the methodology of Ref. [56] to show that NQAC provides an energy boost that scales as a power law with the number of physical qubits used in the encoding.

We start by introducing an overall energy scale α [65] for the problem Hamiltonian: $H_P \mapsto \alpha H_P$, with $0 < \alpha \leq 1$. The dimensionless parameter α controls the level of thermal and dynamically induced errors, as well as the relevance of analog control errors, and thus controls the hardness of an optimization problem when implemented on a quantum annealing device. The introduction of the scale α also allows us to extract the energy boost μ_C provided by an NQAC encoding with C levels of nesting. The energy boost μ_C is defined as the rescaling factor required so that

$$P_C(\alpha) \approx P_1(\mu_C \alpha), \quad (10)$$

where P_C is the success probability of QA when a given optimization problem is implemented with C levels of nesting, and $C = 1$ corresponds to the unencoded case. Equation (10) states that the performance enhancement obtained at nesting level C can be interpreted as increasing the energy scale α by a factor of μ_C relative to the unencoded case.

We are interested in the scaling of μ_C for large C . We thus consider the implementation of a small logical problem that can be nested several times. On the DW2KQ device, a completely antiferromagnetic ($h_i = 0, \forall i$) Ising problem over K_4 ($J_{ij} = 1 \forall i, j$) can be nested up to $C = 13$ with balanced minor embeddings.² The results are shown in Fig. 1. Figure 1(a) shows the success probability $P_C(\alpha)$ as a function of the nesting level C and of the energy scale α . The strength of the energy penalty γ was optimized to obtain the best performance for the NQAC encoding among the 9 values $\gamma \in \{0.1, 0.2, \dots, 0.9, 1\}$. For all experiments we used 25 different, randomly generated balanced embeddings using the algorithm introduced in Ref. [84]. For each embedding we performed 1000 annealing runs. As expected, $P_C(\alpha)$ drops from $P_C(1) = 1$ (solution always found) to $P_C(0) = 6/16$ (random sampling of 6 ground states, where 4 out of the 6 couplings are satisfied, out of a total of 16 states).

In Fig. 1(a) we see two regions in which NQAC has a different behavior: small and large α . In the small α region the success probability grows monotonically with α and C . In the large α region the success probability starts dropping for large C . This drop is due to the fact that the largest implementable value for the energy penalty γ is suboptimal, i.e., too small. For this reason, we refer to the large α region as “penalty-limited”. The monotonic improvement of the success probability in the small α region is consistent with a scalable increase of μ_C . We thus refer to the small α region as the “scaling” region.

Figure 1(b) shows that the data from the left panel can be collapsed using the scaling ansatz given in Eq. (10), and hence that the effects of nesting can be interpreted, at least in the scaling region, as providing an energy boost μ_C . Figure 1(c) shows the energy boost μ_C as a function of C , as determined

² In a balanced embedding all chains have same length. We limited our experiments to balanced embedding since we found that the use of more general, unbalanced, embeddings reduces the performance enhancement provided by NQAC.

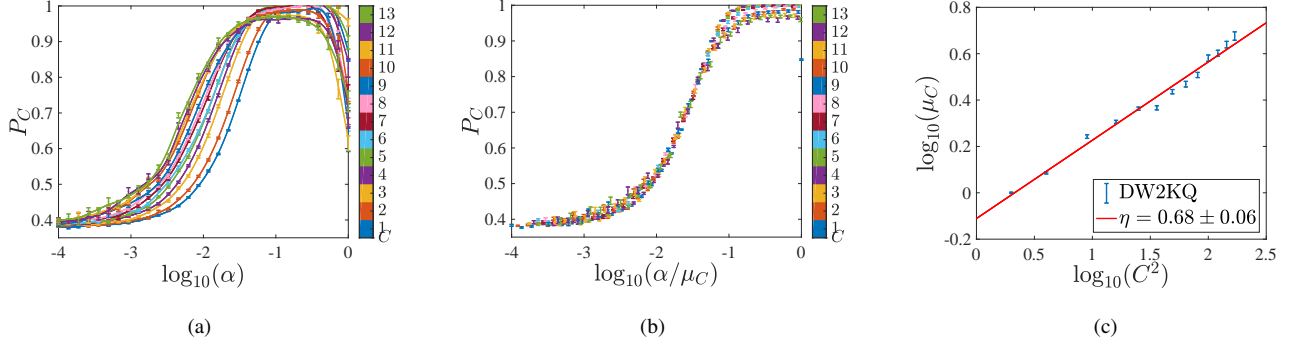


FIG. 1. Empirical results for the antiferromagnetic K_4 , after encoding, followed by ME and decoding. (a) DW2KQ success probabilities $P_C(\alpha)$ for 13 nesting levels C . Increasing C generally increases $P_C(\alpha)$ at fixed α . (b) Rescaled $P_C(\alpha) \mapsto P_C(\alpha/\mu_C)$ data, exhibiting data-collapse. Here μ_C is chosen for each C such that the quality of the data-collapse is maximized. Using the values of μ_C found from data-collapse, (c) shows the scaling of the energy boost μ_C vs the amount of physical resources used ($N_{\text{phys}} \sim C^2$). Empirical data are consistent with a power-law scaling $\mu_C \sim C^\eta$, where $\eta = 0.68 \pm 0.06$. In all panels $N_{\text{phys}} \in [8, 728]$. The error bars in the left and center figures correspond to the median, 25-th and 75-th percentiles of $P_C(\alpha)$ computed for each embedding. See Appendix B for more details on the determination of μ_C and the associated error bars.

via the data collapse shown in the middle panel. We see that the energy boost μ_C grows monotonically with C following the power law scaling $\mu_C \sim C^\eta$ with η determined empirically to be $\eta \sim 0.68$.

We have previously observed that a nested graph $K_{C \times N}$ contains C^2 equivalent copies of the same logical coupling J_{ij} , thus intuitively providing a maximal energy boost in which $\mu_C \sim C^2$. We empirically find $\eta \approx 0.68 < 2$. This suboptimal scaling was explained in Ref. [56] as being attributable mainly to the overhead cost of minor embedding and the detrimental effect of control errors. The scaling parameter η was found to be $\eta \approx 1.1 < 2$ on a DW2 device. We expect to observe different scaling due to the various differences between the DW2 and DW2KQ devices: smaller operating temperature for the DW2KQ, smaller analog control error strength for the DW2KQ, different embedding used for NQAC due to uncalibrated qubits (fewer in DW2KQ but more uniformly distributed over the hardware graph).

B. Classical repetition

The existence of two separate regions (scaling and penalty-limited) becomes more evident when we consider the encoding of larger logical problems. In Fig. 2(a) we show the empirical results for an example instance (with couplings randomly generated from the set $J_{ij} \in \{\pm 0.1, \pm 0.2, \dots, \pm 0.9, \pm 1\}$) defined on a K_{16} with up to $C = 3$ nesting levels. In the scaling region the success probability is monotonically increasing. In the penalty-limited region the success probability peaks at a given value where the optimal penalty strength is close to the largest allowed value and then decreases again at larger values of α . Moreover, the success probability is typically not a monotonically increasing function of the nesting level C .

The presence of a penalty-limited region prevents NQAC from providing a scalable energy boost for a wider range of

energy scales α . This effects is particularly important when comparing NQAC to classical repetition. For NQAC to be considered as a practically useful scheme for error correction in quantum annealing, it must be more effective than a classical repetition scheme that uses the same amount of physical resources as the NQAC encoding. In particular, let us denote by C_{max} the largest nesting level C that can be implemented on the QA device. Then $M_C = \lfloor N_{C_{\text{max}}}^{\text{phys}} / N_C^{\text{phys}} \rfloor$ is the number of copies that can be implemented in parallel for a given nesting level C . The “corrected” success probability of an NQAC encoded optimization problem is thus the probability of succeeding at least once after a run with M_C statistically independent, parallel copies:

$$P_C^{\text{corr}}(\alpha) = 1 - [1 - P_C(\alpha)]^{M_C}. \quad (11)$$

Figure 2(b) shows the parallelized success probability for the example instance of Fig. 2(a). We see that nesting does not help in the penalty-limited region (the unencoded $C = 1$ case gives the best results). However, nesting does help in the scaling region, where the strength of the energy penalties is not a bottleneck. In the case of the example shown we find the existence of an optimal nesting level $C = 2$ in the scaling region.

We now show that the empirical results discussed regarding the example instance shown in Figs. 2(a) and 2(b) are representative of the general behavior of NQAC. We have confirmed this by studying four ensembles of 100 randomly generated instances on 16 and 24 logical variables (K_{16} and K_{24}). Couplings are randomly generated from the set $J_{ij} \in \{\pm 0.1, \pm 0.2, \dots, \pm 0.9, \pm 1\}$. As in the previous section we used 25 randomly generated balanced embeddings and for each embedding we performed 1000 annealing runs. To keep the running time on the DW2KQ manageable, we did not optimize the energy penalty γ , which we fixed to its maximum value $\gamma = 1$. Figures 2(c)-2(f) show the experimental results.

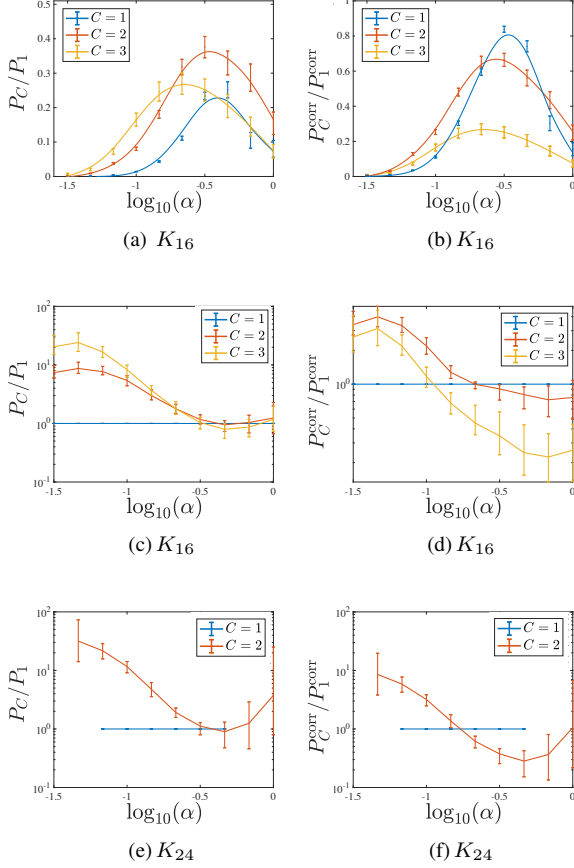


FIG. 2. (a) Success probabilities for a random K_{16} instance with up to $C = 3$ nesting levels. Note the monotonic and non-monotonic improvements of P_C in the scaling (small α) and penalty-limited (large α) regions. (b) Same as (a), but with success probabilities accounting for classical repetition via Eq. (11). Classical repetition outperforms NQAC in the penalty-limited region while there is an optimal nesting level ($C = 2$) for NQAC in the scaling region. Panels (c) and (d) display the normalized success probabilities without (c) and with (d) classical repetition for the K_{16} ensemble. Panels (e) and (f) display the same for the K_{24} ensemble. The behavior of the example instance in (a) and (b) is representative for the class of random instances it belongs to.

The figures show the median (error bars represent the 25th and 75th percentiles) ratios (P_C/P_1) and ($P_C^{\text{corr}}/P_1^{\text{corr}}$). Nesting is preferable when the $C > 1$ curves are above the horizontal breakeven line. Figures 2(c) and 2(e) show that nesting generically provides a success probability improvement that is monotonically increasing with C in the scaling region (small α), while it does not necessarily improve performance in the penalty-limited region. Figures 2(d) and 2(f) show that nesting typically outperforms the unencoded ($C = 1$) case in the scaling regime, while nesting does not help in the penalty-limited region. After correcting for classical repetition, the success probability improvement is not necessarily monotonic in the scaling region. For example Fig. 2(d) shows that $C = 2$ is optimal for K_{16} , in the sense that $P_2^{\text{corr}} > P_1^{\text{corr}}, P_3^{\text{corr}}$.

V. NQAC FOR SAMPLING

In this section we explore NQAC in the context of quantum annealing used for sampling applications. The limited hardware connectivity of quantum annealing devices and the finite strength and precision of their couplings are major limitations not only for optimization (previous section), but also for sampling applications. We are particularly interested in machine-learning applications in which quantum annealers are used as samplers for training Boltzmann machines [75–78].

A. Training Boltzmann machines

Boltzmann machines are generative probabilistic models that can be used for both supervised and non-supervised machine-learning applications. Boltzmann machines can be used as a building block for deep networks thus playing a role in the booming fields of artificial intelligence and deep-learning. A Boltzmann machine associates a given data point $z \equiv \{z_i\}$ (here represented as a string $z_i = \pm 1, i = 1, \dots, N$) to an Ising “energy” function $E(z)$

$$E(z) = \sum_{i \in \mathcal{V}} b_i z_i + \sum_{(i,j) \in \mathcal{E}} w_{ij} z_i z_j, \quad (12)$$

and a corresponding probability distribution $P(z)$:

$$P(z) = e^{-E(z)} / Z, \quad Z = \sum_z e^{-E(z)}. \quad (13)$$

This is a Boltzmann distribution with unit temperature [the connection to the physical temperature is discussed below — see Eq. (16)]. The energy function in Eq. (12) corresponds to an Ising model defined on a graph $G = (\mathcal{V}, \mathcal{E})$. A Boltzmann machine is thus also a “graphical model”. Training a Boltzmann machine consists of finding the values of the weights b_i and w_{ij} such that the probability distribution $P(z)$ generated by the model will extract and generalize as many “features” as possible from the data set. The training of a Boltzmann machine is achieved by minimizing the negative log-likelihood of the data set with respect to the probability distribution $P(z)$. This results in iteratively adjusting the weights of the model according to the following update rules:

$$\begin{aligned} \delta b_i &\sim \langle z_i \rangle_{\text{data}} - \langle z_i \rangle_T \\ \delta w_{ij} &\sim \langle z_i z_j \rangle_{\text{data}} - \langle z_i z_j \rangle_T, \end{aligned} \quad (14)$$

where the first terms are averages over the data set and are usually called “positive phases” while the second terms are thermal averages [i.e., $\langle X \rangle_T \equiv \sum_i x_i P(x_i)$ where x_i are the values taken by the random variable X], and are usually called “negative phases”. Computing the negative phases is known to be computationally hard with classical algorithms, essentially since it requires sampling from the Gibbs state of a spin glass, and is usually replaced with the contrastive divergence approximation [88, 89]. Instead, one may try to compute the negative phases using physical quantum annealing devices as Boltzmann samplers [75–78]. The hope is that this can be

done faster than a classical computation of the associated partition function, which would be an example of a quantum advantage in quantum machine learning [90].

B. Relating the Boltzmann machine parameters to the quantum annealer parameters

As we discussed previously, performing quantum annealing in a quasi-static regime allows one to approximately sample from the instantaneous thermal state at the freezing point [70]. We shall assume here that the freezing point happens late enough in the anneal such that the system freezes when the problem Hamiltonian H_P dominates. Performing quantum annealing with a device that realizes Eqs. (4) and (5) allows to thermally sample from the Boltzmann distribution of the model given by Eqs. (12) and (13).

Let β_{phys} denote the inverse physical temperature of the quantum annealing device, $B(t)$ the profile function controlling the energy scale of the problem Hamiltonian H_P [as in Eq. (4)], and $t^* < t_f$ the freezing time. Because of freezing, the quantum annealer (approximately) samples from a Gibbs distribution at an effective temperature $T_{\text{eff}} = 1/\beta_{\text{eff}}$ given by:

$$\beta_{\text{eff}} = \beta_{\text{phys}} B(t^*)/B(t_f). \quad (15)$$

The existence of an effective sampling temperature allows us to establish the connection between the physical parameters of a quantum annealing device as defined in Eq. (5) and the weights of a Boltzmann machine as defined in Eq. (12):

$$b_i = \beta_{\text{eff}} h_i, \quad w_{ij} = \beta_{\text{eff}} J_{ij}. \quad (16)$$

C. Methodology

Our goal in this section is to study whether NQAC can be used to improve the performance of quantum annealers in the training of fully connected (as opposed to restricted [89]) Boltzmann machines, i.e., when the graphical model underlying the model of Eq. (12) is defined on a complete graph K_N . Due to the physical implementation of quantum annealing on the Chimera graph, this means that the Boltzmann machine is necessarily defined on the logical problem Hamiltonian H_P [Eq. (4)]. In order for NQAC to be successful for sampling it should both allow for reliably sampling from logical thermal distributions (defined over the logical problem Hamiltonian), and result in a reduction of the effective temperature $T_{\text{eff}} = 1/\beta_{\text{eff}}$ [or, equivalently, an effective boost of the annealing couplers (h_i, J_{ij})]. Of course, this is accomplished by sampling from the corresponding minor embedded implementations defined via the encoded Hamiltonian \tilde{H}_P [Eq. (6)].

1. The effective temperature associated with nesting level C

We associate an effective inverse temperature with each nesting level C , which we denote by $\beta_{C,\text{eff}}$. To find this quan-

tity, consider the “energy histogram”:

$$p(E_a) = \sum_{z|E(z)=E_a} P(z). \quad (17)$$

This is the probability $p(E_a)$ to obtain a state with given energy E_a .

For each logical instance defined on a K_N , characterized by a set of logical local fields and couplings $\{h_i, J_{ij}\}$, we use Eq. (16) to define b_i and w_{ij} , and numerically compute via Eq. (17) the thermal energy histogram $p_T(E_a, \beta)$ [shorthand for $p_T(E_a, b_i = \beta h_i, w_{ij} = \beta J_{ij})$]. To do so we first compute $E(z)$ [Eq. (12)] and the Boltzmann distribution $P(z)$ [Eq. (13)]. Similarly, for the same logical instances we evaluate the corresponding empirical DW2KQ energy histogram $p_{\text{DW}}(E_a, C)$ [shorthand for $p_{\text{DW}}(E_a, C, h_i, J_{ij})$], after the logical problem is encoded with C nesting levels and the physical qubit readouts have been decoded via majority vote.

The effective inverse temperature $\beta_{C,\text{eff}}$ is now obtained by minimizing the total variation distance:

$$\beta_{C,\text{eff}} = \arg \min_{\beta} D(p_{\text{DW}}, p_T) \quad (18a)$$

$$= \arg \min_{\beta} \left(\frac{1}{2} \sum_a |p_{\text{DW}}(E_a, C) - p_T(E_a, \beta)| \right). \quad (18b)$$

In other words, the effective sampling temperature $1/\beta_{C,\text{eff}}$ for the quantum annealer is defined as the temperature that minimizes the distance between the empirical and theoretical energy histograms $p_{\text{DW}}(E, C)$ and $p_T(E, \beta)$, respectively.³ Note that the theoretical energy histograms involve computing the partition function [Eq. (13)] and so become numerically very demanding as N grows. Indeed, this is precisely the reason that we are interested in the quantum annealing alternative.

2. Assessing the quality of the sampling distributions obtained with NQAC

The effective sampling temperature is an indirect measure of the quality of the sampling distribution obtained by the annealer. We thus consider an additional quantity that is more directly connected to the training of Boltzmann machines. Recall that the quantum annealer is used to evaluate the negative phases [Eq. (14)] via sampling, i.e.,

$$\langle z_i z_j \rangle_T \equiv \vec{\nabla}^T \mapsto \langle z_i z_j \rangle_{\text{DW}} \equiv \vec{\nabla}^{\text{DW}}, \quad (19)$$

where we have not included the quantities $\langle z_i \rangle_{\text{DW}}$ and $\langle z_i \rangle_T$ since for all the instances we have considered we have vanishing local fields, or biases, h_i . We regard the negative phases

³ The total variation distance $D(p, q) \equiv \frac{1}{2} \sum_a |p_a - q_a|$ between the probability distributions p and q provides a simple way to estimate the effective sampling temperature which is good enough for our purposes. For more sophisticated approaches, see for example Ref. [91].

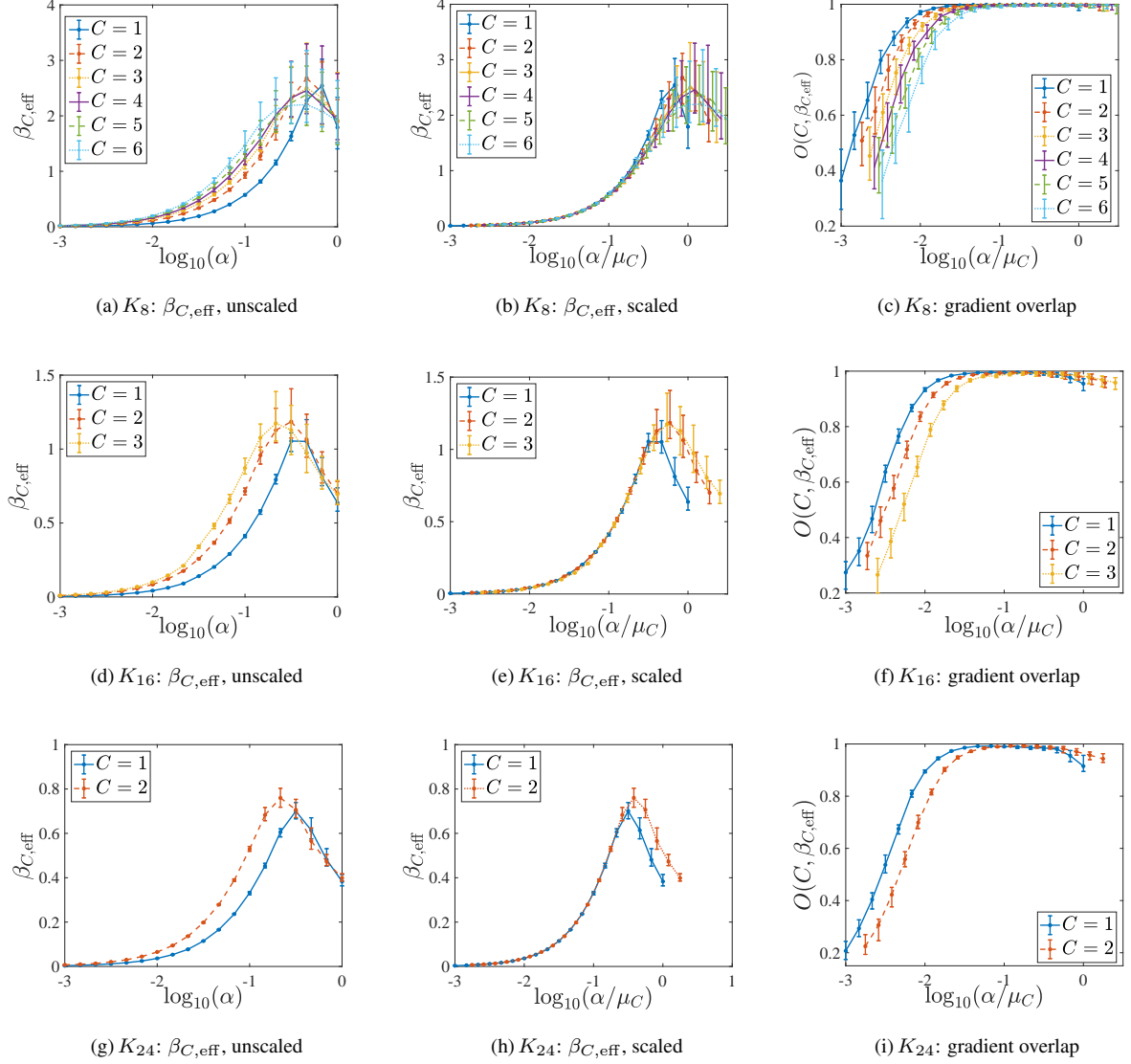


FIG. 3. Results for the K_8 ensemble (top row), K_{16} ensemble (center row) and the K_{24} ensemble (bottom row), of 100 instances each. Leftmost panels show the effective inverse temperature $\beta_{C,\text{eff}}$. Center panels show the collapsed inverse temperature $\beta_{C,\text{eff}}$. Rightmost panels show the gradient overlap $O(p_{\text{DW}}, p_T)$. In all plots we display the median values with the error bars (25th and 75th percentiles) showing the variation over the ensembles.

as vectors in the space of weights. This makes sense since the negative phases enter the definition of a “gradient vector” as in Eq. (14). We thus consider the overlap between the thermal and experimental gradients:

$$O(C, \beta_{C,\text{eff}}) \equiv \hat{\nabla}^{\text{DW}}(C) \cdot \hat{\nabla}^T(\beta_{C,\text{eff}}) \quad (20)$$

where we have included a reminder that O compares the empirical phases estimated with an NQAC encoding at nesting level C and the negative phases numerically computed at temperature $\beta = \beta_{C,\text{eff}}$.

D. Empirical results

1. The effective temperature associated with nesting level C

In Fig. 3 we show the empirical results for the ensemble of instances K_8 [3(a)-3(c)], K_{16} [3(d)-3(f)] and K_{24} [3(g)-3(i)]. Figures in the leftmost panels show the effective temperature $\beta_{C,\text{eff}}$ computed via Eq. (18b). We observe a behavior similar to that of P_C [see, e.g., Fig. 1(a)]. In the scaling region, $\beta_{C,\text{eff}}$ grows monotonically with respect to both α and C . This shows that NQAC provides a systematic reduction of the effective sampling temperature. This interpretation is further confirmed in the center-left panels, in which we show a data

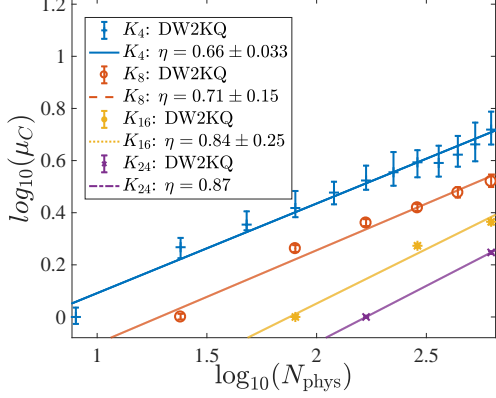


FIG. 4. Scaling of the inverse effective temperature reduction coefficient μ_C as a function of C for the four ensembles of instances studied. μ_C was computed as the rescaling coefficient that gives the data collapse in the center panels of Fig. 3. Empirical data are consistent with a power-law scaling $\mu_C \sim C^\eta$, where $\eta(K_4) = 0.66 \pm 0.033$, $\eta(K_8) = 0.71 \pm 0.15$, $\eta(K_{16}) = 0.84 \pm 0.25$, $\eta(K_{24}) = 0.87 \pm 0.1$. See Appendix B for more details on the determination of μ_C and the associated error bars.

collapse similar to that of Fig. 1(b). In the scaling region we can find an energy boost μ_C such that

$$\beta_{C,\text{eff}}(\alpha) \approx \beta_{1,\text{eff}}(\mu_C \alpha), \quad (21)$$

in analogy to Eq. (10) for P_C . Thus, in the scaling region the effective sampling temperature of the annealer can be decreased at a given energy scale α of the problem Hamiltonian by increasing the nesting level C [the scaling of μ_C is shown in Fig. 4].

2. Assessing the quality of the sampling distributions obtained with NQAC

The rightmost panels of Fig. 3 show the gradient overlap $O(p_{\text{DW}}, p_T)$. We see that the numerically and empirically computed gradients $\vec{\nabla}$ are almost completely aligned for sufficiently large values of α , in the scaling region. At small values of α the gradient overlap shows a decay in sampling quality. Thus, there is a sweet-spot for using NQAC in sampling applications. For the examples shown here, this is in the range $-1.5 \lesssim \log_{10}(\alpha) \lesssim -0.5$. In this range the empirically computed gradients are indistinguishable from the numerically computed one [as measured by $O(C, \beta_{C,\text{eff}})$], and do not decay significantly with increased nesting level C ; moreover, NQAC provides the added benefit of a monotonic reduction of the sampling temperature.

Note that at small α (outside the “sweet-spot region”) we find that $O(C, \beta_{C,\text{eff}})$ is smaller for large C , so that higher nesting appears to be less helpful. However, recall that we did not optimize the energy penalty γ (we set $\gamma = 1$), and in the

small α region we expect that we can improve the sampling quality for larger C by optimizing γ . Contrariwise, even without optimizing the energy penalty strength nesting improves matters for α larger than the sweet-spot region (this effect is only visible for the K_{16} and K_{24} cases, since for K_8 the overlap is pegged at 1 for large α).

3. Scaling of the energy boost

In Fig. 4 we show the scaling of μ_C as a function of the nesting level C given by the data-collapse of the center panels of Fig. 3. For the K_4 case, Fig. 4 shows a power-law scaling consistent with that we observed for the fully antiferromagnetic K_4 , where we found $\eta = 0.68 \pm 0.06$ [Fig. 1(c)]. Interestingly, the result progressively improves for the larger graphs, with K_{24} having the largest value of η , though this is admittedly tenuous as it is based on only two data points due to the size limitations imposed by the DW2KQ. With this caveat in mind, we conclude that NQAC provides a lower effective sampling temperature scaling that improves with problem size.

VI. DISCUSSION

Very recently it was pointed out that fixed finite temperature quantum annealers satisfy an adversarial “temperature scaling law” [74], that prevents them from functioning as competitive scalable optimizers unless annealer temperatures are appropriately scaled down with problem size. This result places on a more rigorous footing the folklore wisdom that, in a realistic open system setting, AQC and QA require the temperature to be reduced as problem sizes grow and (generically) gaps shrink, since at a fixed temperature thermal transitions out of the ground state would become inevitable. Because the third law of thermodynamics (Nernst’s “unattainability principle” formulation, though see Ref. [92]) forces the cooling rate to vanish as the temperature approaches absolute zero, a scalable temperature reduction is considered impractical, known colloquially as there being “no scaling law for refrigerators”.

However, these perspectives ignore the possibility of error correction, which acts as an effective entropy sink, and can be used to attain a scalable temperature reduction at the expense of using up physical qubits to create colder code qubits [37, 53]. In this work we addressed and provided a potential solution for the “temperature scaling law” problem, by demonstrating that NQAC allows for a scalable effective temperature reduction, by using codes of increasing code distance C to achieve an effective temperature reduction scaling as a power law in C . This potentially addresses even the most adversarial scenario considered in Ref. [74], wherein the temperature must drop as a power law with problem size.

We do not expect the NQAC method studied here to provide an indefinite effective temperature reduction at arbitrary problem sizes; such a result belongs in the realm of the open problem of fault-tolerant AQC and QA. However, we do expect that the effective temperature reduction we have demon-

strated here for both optimization and sampling applications to fuel the continued study of practical and implementable error suppression schemes for QA, and perhaps even carry the next generation of quantum annealers to the point of demonstrating a quantum advantage.

ACKNOWLEDGMENTS

We thank D-Wave Systems Inc. for access to their DW2000Q device in Burnaby. This work was (partially) supported under ARO grant number W911NF-12-1-0523, ARO MURI Grant Nos. W911NF-11-1-0268 and W911NF-15-1-0582, and NSF grant number INSPIRE-1551064. The research is based upon work partially supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the U.S. Army Research Office contract 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 U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Appendix A: Experimental details

The annealing schedules of the D-Wave 2000Q (DW2KQ) processor housed in Burnaby is shown in Fig. 5(a). The Chimera hardware graphs of the DW2KQ processor we used in this work is shown in Fig. 5(b). For additional experimental details see, e.g., Refs. [59, 61, 93].

Appendix B: Determination of the energy boost μ_C and error bars

Figures 1(c) and 4 show the experimental scaling of μ_C as extracted by using, respectively, Eq. (10) (via success probability) and Eq. (21) (via effective temperature). To determine the values of μ_C and estimate error bars, we proceeded as follows. Let us define $M_C(\alpha) \equiv P_C(\alpha)$ or $M_C(\alpha) \equiv \beta_{C,\text{eff}}(\alpha)$ depending on whether we are using $P_C(\alpha)$ [Fig. 1(c)] or $\beta_{C,\text{eff}}(\alpha)$ (Fig. 4).

First, we used smoothing splines to determine a continuous interpolation $M_C^{\text{mid}}(\alpha)$ of the median data points of $M_C(\alpha)$. In the same way we also determined the higher and lower interpolating curves $M_C^{\text{high}}(\alpha)$ and $M_C^{\text{low}}(\alpha)$ for the data points of the 75-th and 25-th percentiles of $M_C(\alpha)$ respectively. Then, a reference value α_C^{mid} was determined such that $M_C^{\text{mid}}(\alpha_C^{\text{mid}}) = M_0$, where we used the smooth interpolation of the experimental data. The energy boost was then determined as $\mu_C = \alpha_1^{\text{mid}}/\alpha_C^{\text{mid}}$. M_0 is an arbitrarily chosen reference value where the different $M_C(\alpha)$ curves are made to overlap. This reference serves as a base point for computing μ_C . As shown in the center panels of Fig. 1 and 3, the overlap

of the P_C and $\beta_{C,\text{eff}}(\alpha)$ data over the entire α range means that the specific choice of M_0 is arbitrary. We similarly determined $\mu_C^{\text{high}} = \alpha_1^{\text{high}}/\alpha_C^{\text{high}}$ and $\mu_C^{\text{low}} = \alpha_1^{\text{low}}/\alpha_C^{\text{low}}$ using the corresponding interpolating curves. The error bars shown in the figures were then centered at μ_C , with lower and upper error bars being μ_C^{high} and μ_C^{low} , respectively.

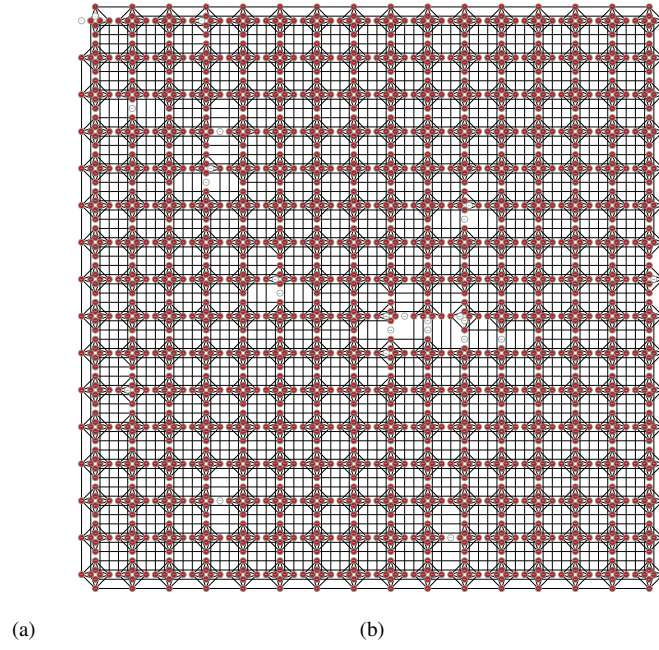


FIG. 5. (a) Annealing schedule for the DW2KQ (with $\hbar = 1$ units). As a reference, we include the operating temperatures of the device, corresponding to 14.1mK for the DW2KQ. (b) "Chimera" hardware graph of the DW2KQ device in Burnaby used in this work. Available qubits are shown in green, and unavailable qubits are shown in red. Programmable couplers are shown as black lines connecting qubits.

-
- [1] B. Apolloni, N. Cesa-Bianchi, and D. de Falco, in *Proceedings of the Ascona/Locarno Conference* (1988) p. 97.
- [2] P. Ray, B. K. Chakrabarti, and A. Chakrabarti, *Phys. Rev. B* **39**, 11828 (1989).
- [3] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, *Chemical Physics Letters* **219**, 343 (1994).
- [4] T. Kadowaki and H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).
- [5] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli, *Science* **284**, 779 (1999).
- [6] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, [arXiv:quant-ph/0001106](https://arxiv.org/abs/quant-ph/0001106) (2000).
- [7] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, *Science* **292**, 472 (2001).
- [8] V. Smelyanskiy, U. V. Toussaint, and D. Timucin, [arXiv:quant-ph/0112143](https://arxiv.org/abs/quant-ph/0112143) (2001).
- [9] B. W. Reichardt, in *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04 (ACM, New York, NY, USA, 2004) pp. 502–510.
- [10] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [11] W. van Dam, M. Mosca, and U. Vazirani, *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on, Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, 279 (2001).
- [12] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, *SIAM J. Comput.* **37**, 166 (2007).
- [13] J. Kempe, A. Kitaev, and O. Regev, *SIAM Journal on Computing* **35**, 1070 (2006).
- [14] A. Mizel, D. A. Lidar, and M. Mitchell, *Phys. Rev. Lett.* **99**, 070502 (2007).
- [15] R. Oliveira and B. M. Terhal, *Quantum Inf. Comput.* **8**, 900 (2008).
- [16] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe, *Communications in Mathematical Physics* **287**, 41 (2009).
- [17] N. P. Breuckmann and B. M. Terhal, *Journal of Physics A: Mathematical and Theoretical* **47**, 195304 (2014).
- [18] D. Gosset, B. M. Terhal, and A. Vershynina, *Phys. Rev. Lett.* **114**, 140501 (2015).
- [19] S. Lloyd and B. M. Terhal, *New Journal of Physics* **18**, 023042 (2016).
- [20] A. Das and B. K. Chakrabarti, *Rev. Mod. Phys.* **80**, 1061 (2008).
- [21] T. Albash and D. A. Lidar, [arXiv:1611.04471](https://arxiv.org/abs/1611.04471) (2016).
- [22] T. Kato, *J. Phys. Soc. Jap.* **5**, 435 (1950).
- [23] S. Jansen, M.-B. Ruskai, and R. Seiler, *J. Math. Phys.* **48**, (2007).
- [24] D. A. Lidar, A. T. Rezakhani, and A. Hamma, *J. Math. Phys.* **50**, (2009).
- [25] A. T. Rezakhani, A. K. Pimachev, and D. A. Lidar, *Phys. Rev. A* **82**, 052305 (2010).
- [26] N. Wiebe and N. S. Babcock, *New J. Phys.* **14**, 013024 (2012).
- [27] Y. Ge, A. Molnár, and J. I. Cirac, [arXiv:1508.00570](https://arxiv.org/abs/1508.00570) (2015).
- [28] A. M. Childs, E. Farhi, and J. Preskill, *Phys. Rev. A* **65**, 012322 (2001).
- [29] M. S. Sarandy and D. A. Lidar, *Phys. Rev. Lett.* **95**, 250503 (2005).
- [30] M. H. S. Amin, P. J. Love, and C. J. S. Truncik, *Phys. Rev. Lett.* **100**, 060503 (2008).
- [31] S. Lloyd, [arXiv:0805.2757](https://arxiv.org/abs/0805.2757) (2008).
- [32] M. H. S. Amin, D. V. Averin, and J. A. Nesteroff, *Phys. Rev. A* **79**, 022107 (2009).
- [33] Q. Deng, D. V. Averin, M. H. Amin, and P. Smith, *Sci. Rep.* **3**, 1479 (2013).
- [34] T. Albash and D. A. Lidar, *Phys. Rev. A* **91**, 062320 (2015).
- [35] D. Lidar and T. Brun, eds., *Quantum Error Correction* (Cambridge University Press, Cambridge, UK, 2013).
- [36] I. Marvian and D. A. Lidar, *Phys. Rev. Lett.* **113**, 260504 (2014).
- [37] S. P. Jordan, E. Farhi, and P. W. Shor, *Phys. Rev. A* **74**, 052322 (2006).
- [38] D. A. Lidar, *Phys. Rev. Lett.* **100**, 160506 (2008).
- [39] G. Quiroz and D. A. Lidar, *Phys. Rev. A* **86**, 042333 (2012).
- [40] K. C. Young, M. Sarovar, and R. Blume-Kohout, *Phys. Rev. X* **3**, 041013 (2013).
- [41] M. Sarovar and K. C. Young, *New J. of Phys.* **15**, 125032 (2013).
- [42] K. C. Young, R. Blume-Kohout, and D. A. Lidar, *Phys. Rev. A* **88**, 062314 (2013).
- [43] A. Ganti, U. Onunkwo, and K. Young, *Phys. Rev. A* **89**, 042313 (2014).
- [44] A. Mizel, [arXiv:1403.7694](https://arxiv.org/abs/1403.7694) (2014).
- [45] K. L. Pudenz, T. Albash, and D. A. Lidar, *Nat. Commun.* **5**, 3243 (2014).
- [46] K. L. Pudenz, T. Albash, and D. A. Lidar, *Phys. Rev. A* **91**, 042302 (2015).
- [47] W. Vinci, T. Albash, G. Paz-Silva, I. Hen, and D. A. Lidar, *Phys. Rev. A* **92**, 042310 (2015).
- [48] A. Mishra, T. Albash, and D. A. Lidar, *Quant. Inf. Proc.* **15**, 609 (2015).
- [49] S. Matsuura, H. Nishimori, T. Albash, and D. A. Lidar, *Physical Review Letters* **116**, 220501 (2016).
- [50] A. D. Bookatz, E. Farhi, and L. Zhou, *Physical Review A* **92**, 022317 (2015).
- [51] I. Marvian, [arXiv:1602.03251](https://arxiv.org/abs/1602.03251) (2016).
- [52] M. Marvian and D. A. Lidar, *Physical Review Letters* **118**, 030504 (2017).
- [53] M. Marvian and D. A. Lidar, *Physical Review A* **95**, 032302 (2017).
- [54] S. Lloyd and B. Terhal, [arXiv:1509.01278](https://arxiv.org/abs/1509.01278) (2015).
- [55] P. Aliferis, D. Gottesman, and J. Preskill, *Quantum Inf. Comput.* **6**, 97 (2006).
- [56] W. Vinci, T. Albash, and D. A. Lidar, *Nature Quantum Information* **2**, 16017 (2016).
- [57] M. W. Johnson, P. Bunyk, F. Maibaum, E. Tolkacheva, A. J. Berkley, E. M. Chapple, R. Harris, J. Johansson, T. Lanting, I. Perminov, E. Ladizinsky, T. Oh, and G. Rose, *Superconductor Science and Technology* **23**, 065004 (2010).
- [58] A. J. Berkley, M. W. Johnson, P. Bunyk, R. Harris, J. Johansson, T. Lanting, E. Ladizinsky, E. Tolkacheva, M. H. S. Amin, and G. Rose, *Superconductor Science and Technology* **23**, 105014 (2010).
- [59] 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, *Phys. Rev. B* **82**, 024511 (2010).
- [60] 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, *Nature* **473**, 194 (2011).

- [61] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. Przybysz, and J. Whittaker, *IEEE Transactions on Applied Superconductivity* **24**, 1 (Aug. 2014).
- [62] S. Boixo, T. F. Ronnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, *Nat. Phys.* **10**, 218 (2014).
- [63] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, [arXiv:1401.7087](#) (2014).
- [64] T. Albash, T. F. Ronnow, M. Troyer, and D. A. Lidar, *Eur. Phys. J. Spec. Top.* **224**, 111 (2015).
- [65] T. Albash, W. Vinci, A. Mishra, P. A. Warburton, and D. A. Lidar, *Phys. Rev. A* **91**, 042314 (2015).
- [66] P. J. D. Crowley, T. Durić, W. Vinci, P. A. Warburton, and A. G. Green, *Phys. Rev. A* **90**, 042317 (2014).
- [67] V. Martin-Mayor and I. Hen, *Scientific Reports* **5**, 15324 EP (2015).
- [68] A. D. King, T. Lanting, and R. Harris, [arXiv:1502.02098](#) (2015).
- [69] W. Vinci, K. Markström, S. Boixo, A. Roy, F. M. Spedalieri, P. A. Warburton, and S. Severini, *Sci. Rep.* **4** (2014).
- [70] M. H. Amin, *Physical Review A* **92**, 052323 (2015).
- [71] L. C. Venuti, T. Albash, D. A. Lidar, and P. Zanardi, *Physical Review A* **93**, 032118 (2016).
- [72] J. Marshall, E. G. Rieffel, and I. Hen, [arXiv:1703.03902](#) (2017).
- [73] N. Chancellor, S. Szoke, W. Vinci, G. Aeppli, and P. A. Warburton, *Scientific Reports* **6**, 22318 EP (2016).
- [74] T. Albash, V. Martin-Mayor, and I. Hen, *Physical Review Letters* **119**, 110502 (2017).
- [75] S. H. Adachi and M. P. Henderson, [arXiv:1510.06356](#) (2015).
- [76] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytsky, and R. Melko, [arXiv:1601.02036](#) (2016).
- [77] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, [arXiv:1609.02542](#) (2016).
- [78] D. Korenkevych, Y. Xue, Z. Bian, F. Chudak, W. G. Macready, J. Rolfe, and E. Andriyash, [arXiv:1611.04528](#) (2016).
- [79] S. Matsuura, H. Nishimori, W. Vinci, T. Albash, and D. A. Lidar, *Physical Review A* **95**, 022308 (2017).
- [80] W. M. Kaminsky, S. Lloyd, and T. P. Orlando, “Quantum computing and quantum bits in mesoscopic systems,” (Springer, New York, 2004) Chap. 25, pp. 229–236.
- [81] V. Choi, *Quant. Inf. Proc.* **10**, 343 (2011).
- [82] C. Klymko, B. D. Sullivan, and T. S. Humble, *Quant. Inf. Proc.* **13**, 709 (2014).
- [83] J. Cai, W. G. Macready, and A. Roy, [arXiv:1406.2741](#) (2014).
- [84] T. Boothby, A. D. King, and A. Roy, [arXiv:1507.04774](#) (2015).
- [85] W. Lechner, P. Hauke, and P. Zoller, *Science Advances* **1** (2015), 10.1126/sciadv.1500838.
- [86] T. Albash, W. Vinci, and D. A. Lidar, *Physical Review A* **94**, 022327 (2016).
- [87] F. Pastawski and J. Preskill, *Physical Review A* **93**, 052325 (2016).
- [88] G. E. Hinton and R. R. Salakhutdinov, *Science* **313**, 504 (2006).
- [89] Y. Bengio, *Foundations and Trends in Machine Learning* **2**, 1 (2009).
- [90] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Nature* **549**, 195 (2017).
- [91] J. Raymond, S. Yarkoni, and E. Andriyash, *Frontiers in ICT* **3**, 23 (2016).
- [92] M. Kolář, D. Gelbwaser-Klimovsky, R. Alicki, and G. Kurizki, *Physical Review Letters* **109**, 090601 (2012).
- [93] T. Albash and D. A. Lidar, [arXiv:1705.07452](#) (2017).