Temperature Scaling Law for Quantum Annealing Optimizers

Tameem Albash,[1,2] Victor Martin-Mayor,[3,4] and Itay Hen[1,2,*]
[1]Information Sciences Institute, University of Southern California, Marina del Rey, California 90292, USA
[2]Department of Physics and Astronomy and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA
[3]Departamento de Física Teórica I, Universidad Complutense, 28040 Madrid, Spain
[4]Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Zaragoza 50018, Spain
(Received 27 March 2017; published 14 September 2017)

Physical implementations of quantum annealing unavoidably operate at finite temperatures. We point to a fundamental limitation of fixed finite temperature quantum annealers that prevents them from functioning as competitive scalable optimizers and show that to serve as optimizers annealer temperatures must be appropriately scaled down with problem size. We derive a temperature scaling law dictating that temperature must drop at the very least in a logarithmic manner but also possibly as a power law with problem size. We corroborate our results by experiment and simulations and discuss the implications of these to practical annealers.

DOI: 10.1103/PhysRevLett.119.110502

Introduction.—Quantum computing devices are becoming sufficiently large to undertake computational tasks that are infeasible using classical computing [1–7]. The theoretical underpinning for whether such tasks exist with physically realizable quantum annealers remains lacking, despite the excitement brought on by recent technological breakthroughs that have made programmable quantum annealing (QA) [8–12] optimizers consisting of thousands of quantum bits commercially available. Thus far, no examples of practical relevance have been found to indicate a superiority of QA optimization, i.e., to find bit assignments that minimize the energy, or cost, of discrete combinatorial optimization problems, faster than possible classically [13–20]. Major ongoing efforts continue to build larger, more densely connected QA devices, in the hope that the capability to embed larger optimization problems would eventually reveal the coveted quantum speedup [21–25].

Understanding the robustness of QA optimization to errors that reduce the final ground state probability is critical. In this work, we consider perhaps the most optimistic setting where the only source of error is due to nonzero temperature. We analyze the theoretical scaling performance of ideal fixed-temperature quantum annealers for optimization. We show that even in the case where annealers are assumed to thermalize instantly (rather than only in the infinite runtime limit), the energies, or costs, of their output configurations would be computationally trivial to achieve (in a sense that we explain). We further derive a scaling law for QA optimizers and provide corroboration of our analytical findings by experimental results obtained from the commercial D-Wave 2X QA processor [26–30] as well as numerical simulations (our results equally apply to ideal thermal annealing devices). We discuss the implications of our results for both past benchmarking studies and for the engineering requirements of future QA devices.

Fixed-temperature quantum annealers.—In the adiabatic limit, closed-system quantum annealers are guaranteed to find a ground state of the target cost function, or final Hamiltonian H, they are to solve. The adiabatic theorem of quantum mechanics ensures that the overlap of the final state of the system with the ground state manifold of H approaches unity as the duration of the process increases [31,32]. For physical quantum annealers that operate at positive temperatures ($T > 0$), there is no equivalent guarantee of reaching the ground state with high probability. For long runtimes, an ideal finite-temperature quantum annealer is expected to sample the Boltzmann distribution of the final Hamiltonian at the annealer temperature [33].

In what follows, we argue that even instantly thermalizing quantum annealers [34] are severely limited as optimizers due to their finite temperature. For concreteness, we restrict to annealers for which (i) the number of couplers scales linearly with the number of qubits N [45], (ii) the coupling strengths are discretized and are bounded independently of problem size, and (iii) the scaling of the free energy with problem size is not pathological, i.e., that our system is not tuned to a critical point. Other than the above standard assumptions, our treatment is general (we discuss the performance of quantum annealers when some of these conditions are lifted later on). For clarity, we consider optimization problems written in terms of a Hamiltonian of the Ising-type

$$H = \sum_{\langle ij \rangle} J_{ij} s_i s_j + \sum_i h_i s_i, \tag{1}$$

where $\{s_i = \pm 1\}$ are binary Ising spin variables that are to be optimized over, $\{J_{ij}, h_i\}$ are the coupling strengths

between connected spins and external biases, respectively, and $\langle ij \rangle$ denotes the underlying connectivity graph of the model. The discussion that follows, however, is not restricted to any particular model.

Under the above assumptions, the ground state energies, denoted E_0, of any given problem class, scale linearly with increasing problem size (i.e., the energy is an extensive property as is generically expected from physical systems) while the classical minimal gap $\Delta = E_1 - E_0$ remains fixed. It follows then [46] that the thermal expectation values of the intensive energy

$$\langle e \rangle_\beta = \langle H \rangle_\beta / N, \tag{2}$$

and specific heat

$$c_\beta = \partial \langle e \rangle_\beta / \partial \beta = -N[\langle e^2 \rangle_\beta - \langle e \rangle_\beta^2], \tag{3}$$

remain finite as $N \to \infty$ for any fixed inverse-temperature $\beta = 1/T$. The intensive energy is discretized in steps of Δ/N, yet its statistical dispersion $\sigma_\beta(e) = \sqrt{-c_\beta/N}$ is much larger. Treating e as a stochastic variable, for large enough values of N it can be treated as a continuous variable as the ratio of discretization versus dispersion is $\sqrt{-\Delta^2/(c_\beta N)}$ decaying to zero for large N. From the Boltzmann distribution it follows that the probability density of e goes as $p_\beta(e) = Z_\beta^{-1} e^{N[s(e)-\beta e]}$, where $Z_\beta = \sum_n g_n e^{-\beta E_n}$ is the partition function, g_n is the degeneracy of the nth level, i.e., the number of microstates with $H(\{s_i\}) = E_n$, satisfying $2^N = \sum_{n \geq 0} g_n$, and $s(e)$ is the entropy density [48]. The linear combination $\Psi_\beta(e) = s(e) - \beta e$ plays the role of a large-deviation functional for e. The most probable value of e, which we denote by e^*, is given by the maximum of Ψ_β. Solving $\Psi'_\beta(e^*) = 0$, we find [49]

$$\beta = \left. \frac{\partial s}{\partial e} \right|_{e=e^*}. \tag{4}$$

Close to e^*, Ψ_β can be Taylor expanded as $\Psi_\beta(e) \approx \Psi_\beta(e^*) - [|\Psi''_\beta(e^*)|/2](e - e^*)^2$, from which it follows that

$$p_\beta(e) \approx \frac{e^{N\Psi_\beta(e^*)}}{Z_\beta} \exp\left(-\frac{N|\Psi''_\beta(e^*)|}{2} (e - e^*)^2 \right). \tag{5}$$

The probability density is thus approximately Gaussian in the vicinity of e^*, although deviations from the Gaussian behavior are crucial [50]. Moreover, in the limit of large N, we find

$$\langle e \rangle_\beta = e^* \quad \text{and} \quad c_\beta = \frac{-1}{|\Psi''_\beta(e^*)|}. \tag{6}$$

Therefore, the probability of finding by Boltzmann sampling any energy $e < e^*$ (equivalently, $E < e^* N$) is exponentially suppressed in N, scaling in fact as $\exp\{-N[\Psi_\beta(e^*) - \Psi_\beta(e)]\}$. We thus arrive at the conclusion that even ideal fixed temperature quantum annealers that thermalize instantaneously to the Gibbs state of the classical Hamiltonian are exponentially unlikely to find the ground state since $e^* > e_0 \equiv E_0/N$.

We now corroborate the above derivation by runs on the commercial DW2X quantum annealer [26–29]. To do so, we first generate random instances of differently sized subgraphs of the DW2X Chimera connectivity graph [51,52] and run them multiple times on the annealer, recording the obtained energies [53]. Figure 1 depicts typical resultant residual energy $(E - E_0)$ distributions. As is evident, increasing the problem size N "pushes" the energy distribution farther away from E_0, as well as broadening the distribution and making it more Gaussian-like. In the inset, we measure the departure of $\langle H \rangle_\beta$ from E_0 and the spread of the energies $\sigma_\beta(H)$ over 100 "planted-solution" [18] instances per subgraph size as a function of problem size N [54]. For sufficiently large problem sizes, we find that the scaling of $\langle H - E_0 \rangle_\beta$ is close to linear while $\sigma_\beta(H)$ scales slightly faster than $\sqrt{N}$. While the slight deviations from our analytical predictions suggest that the DW2X configurations have not fully reached asymptotic behavior [55], they exhibit a trend that closely matches our assumptions with the agreement getting better with growing problem sizes.

Given the scaling of the mean and standard deviation, we conclude that fixed-temperature quantum annealers will generate energies e with a fixed distance from e_0, or, in terms of extensive energies, configurations obtained from fixed-temperature annealers will have energies concentrated around $E = (1 - \epsilon)E_0$ for some $\epsilon > 0$ and $E_0 < 0$.

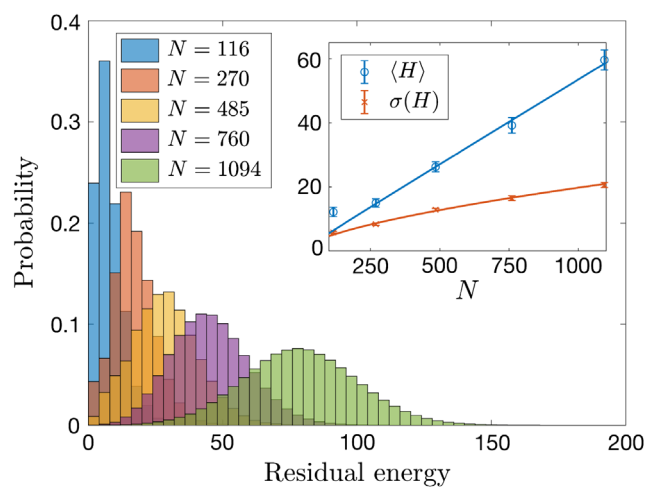

FIG. 1. Distributions of residual energy, $E - E_0$, from DW2X runs. As problem sizes grow, the distributions become more Gaussian-like. Inset: Gaussians' mean (blue) and standard deviation (red) as a function of problem size, averaged over 100 instances per size. The solid lines correspond to power-law fits of the average mean with power 0.98 ± 0.14 and average standard deviation scaling with power 0.63 ± 0.09, taking into account all sizes but the smallest (1.01 ± 0.62 and 0.57 ± 0.37, respectively, if the two smallest sizes are omitted).

One could now ask what the difficulty is for classical algorithms to generate energy values in the above range. This question has been recently answered by the discovery of a polynomial time approximation scheme (PTAS) for spin-glasses defined on a Chimera graph [57] (and which can be easily generalized to any locally connected model), where reaching such energies can be done efficiently [58]. While the scaling of the PTAS with ϵ is not favorable, scaling as $c^{1/\epsilon}$ for some constant c, in practice there exist algorithms (e.g., parallel tempering that we discuss later on) that are known to scale more favorably than PTAS.

Scaling law for quantum annealing temperatures.—In light of the above, it may seem that quantum annealers are doomed to fail as optimizers as problem sizes increase. We now argue that success may be regained if the temperature of the QA device is appropriately scaled with problem size. Specifically, we address the question of how the inverse-temperature β should scale with N such that there is a probability of at least q of finding the ground state.

An estimate for the required scaling can be given as follows. From the above analysis, it should be clear that the probability of finding a ground state at inverse temperature β will not decay exponentially with system size only if the ground state falls within the variation of the mean energy, specifically, if

$$\sigma_\beta(H) = N\sigma_\beta(e) = \sqrt{-Nc_\beta} \tag{7}$$

is comparable to

$$\langle H\rangle_\beta - E_0 = -N\int_\beta^\infty d\beta c_\beta. \tag{8}$$

The third law of thermodynamics dictates that the specific heat $c_T \equiv d\langle e\rangle/dT$ goes to zero when $T \to 0$. Assuming a scaling of the form $c_T \sim T^\alpha$, or, equivalently, $-c_\beta \sim \beta^{-\alpha-2}$, gives

$$\sigma_\beta(H) \sim \sqrt{\frac{N}{\beta^{\alpha+2}}} \quad \text{and} \quad \langle H\rangle_\beta - E_0 = \frac{N}{\beta^{\alpha+1}}. \tag{9}$$

For a power-law specific heat, it thus follows that the sought scaling is $\beta \sim N^{1/\alpha}$. If on the other hand c_β vanishes exponentially in β, the inverse-temperature scaling will be milder, of the form $\beta \sim \log N$.

To illustrate the above, we next present an analysis of simulations of randomly generated instances on Chimera lattices (we study several problem classes and architectures; see the Supplemental Material [37]). To study the energy distribution generated by a thermal sampler on these instances, we use parallel tempering (PT) [59,60], a Monte Carlo method whereby multiple copies of the system at different temperatures are simulated [61]. In Fig. 2, we show an example of how the energy distribution of a planted-solution instance changes with β. The qualitative behavior is similar to what we observe with increasing problem size, whereby decreasing β (increasing the temperature) pushes the energy distribution to larger energies and makes it more Gaussian-like.

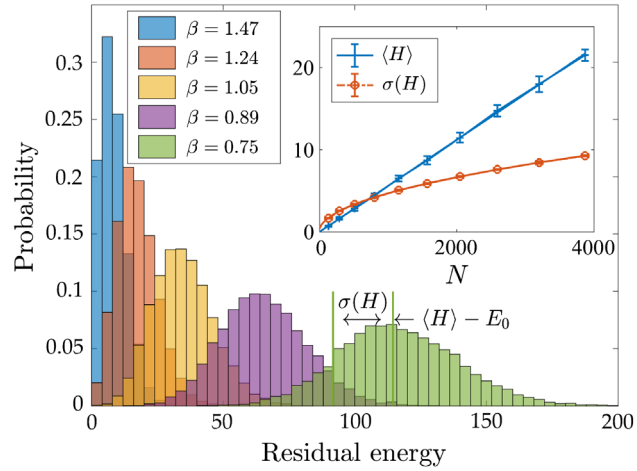

FIG. 2. Distributions of residual energy, $E - E_0$, from PT simulations. For a planted-solution instance defined on an $L=12$ Chimera graph, the distributions become more Gaussian-like as β decreases. For the case of $\beta = 0.75$, the mean residual energy and standard deviation are indicated. Inset: Scaling with problem size of the median mean energy and median standard deviation of the energy for $\beta = 1.47$ over 100 instances.

The behavior of the specific heat c_β as the inverse-temperature β becomes large is shown in Fig. 3. At large sizes, the scaling becomes $c_\beta \propto \exp(-\Delta\beta)$ as expected (here, $\Delta = 4$ is the gap). Based on our predictions above, this should mean that if for a fixed q, the minimum β^* such that $p_{\beta^*}(E_0) \geq q$ falls in this exponential regime, then we

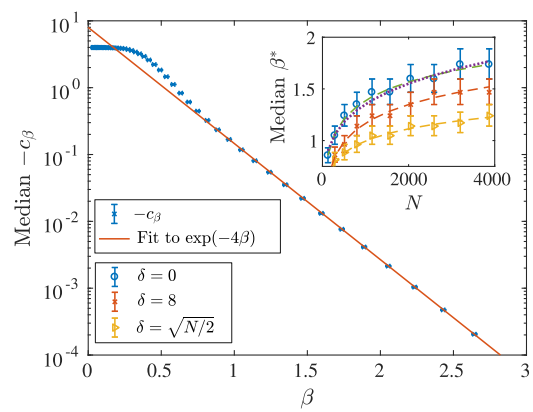

FIG. 3. Typical specific heat with inverse temperature. Behavior of the median specific heat (over 100 instances) for planted-solution instances with inverse-temperature β for $N = 3872$. The behavior transitions from a polynomial scaling with β to an exponential scaling. Inset: Typical minimum inverse temperature required for instances of size N such that the probability of the target energy $E_T = E_0 + \delta(N)$ is at least $q = 10^{-1}$. Also shown are fits to $\log N$ for all three cases and a power-law fit to cN^α that finds $\alpha = 0.19 \pm 0.05$ for the $\delta = 0$ case, which is almost indistinguishable from the logarithmic fit.

should observe a scaling $\beta^* \propto \log N$. Indeed, the inset of Fig. 3, which shows simulation results of β^* versus N, exhibits the expected $\log N$ behavior [62].

While for problem classes with a fixed minimum gap Δ, one may naively expect c_β to vanish exponentially in general, implying that a logarithmic scaling of β will generally be sufficient as our simulations indeed indicate, it is important to note that two-dimensional spin glasses are known to exhibit a crossover between an exponential behavior to a power law [63–66]. This crossover is characterized by a constant $\theta \approx 1/2$, whereby the discreteness of the gap Δ is evident only for sizes $N^{\theta/2} \ll \beta$. Beyond $N^{\theta/2} \sim \beta$, the $2d$ system behaves as if the coupling distribution is continuous [64,65] at which point the system can be treated as if with continuous couplings, for which the specific heat c_T scales as T^α with $\alpha_c = 2\nu$ [63], where $\nu = 3.53(7)$ [66]. Therefore, for an ideal quantum annealer operating beyond the crossover, a scaling of $\beta \sim N^{1/(2\nu)\approx 0.14}$ is required. We may thus expect the same crossover to appear for instances defined on the Chimera lattice, which is $2d$ like. Interestingly, for the temperature scaling shown in the inset of Fig. 3, a power-law fit $\beta \sim N^\alpha$ with $\alpha = 0.19 \pm 0.05$ is almost indistinguishable from the logarithmic one, with a power that is consistent with the $2d$ prediction.

Suboptimal metrics for optimization problems.—For many classically intractable optimization problems, when formulated as Ising models, it is crucial that solvers find a true minimizing bit assignment rather than low-lying excited states. This is especially true for NP-complete or -hard problems [67] where suboptimal costs generally correspond to violated constraints that *must* be satisfied (otherwise the resultant configuration is nonsensical despite its low energy). Nonetheless, it is plausible to assume the existence of problems for which slightly suboptimal configurations would still be of value [68]. We thus also study the necessary temperature scaling for cases where the target energies obey $E_T \leq E_0 + \delta(N)$ with $\delta(N)$ scaling sublinearly with problem size. In the inset of Fig. 3, we plot the required scaling of β for $\delta(N) = \text{const}$ and $\delta(N) \propto \sqrt{N}$. In both cases we find that a logarithmic scaling is still essential, albeit with smaller prefactors.

Conclusions and discussion—We have shown that fixed temperature quantum annealers can only sample "easily reachable" energies in the large problem size limit, thereby posing fundamental limitation on their performance. We derived a temperature scaling law to ensure that quantum annealing optimizers find nontrivial energy values with subexponential probabilities. The scaling of the specific heat with temperature controls this scaling: if β lies in the regime where the specific heat scales exponentially with β, then the inverse temperature of the annealer must scale as $\log N$. However, further considerations are needed because of a possible crossover behavior in the specific heat with temperature and problem size. For Chimera graphs, because of their essentially two-dimensional structure, this may lead to a crossover to power law scaling. Little is known about this crossover in three dimensions or for different architectures, so this concern may not be mitigated by a more complex connectivity graph.

Our results shed important light on benchmarking studies that have found no quantum speedups [17,18, 69–71], identifying temperature as a relevant culprit for their unfavorable performance. Our analysis is particularly relevant for both the utility as well as the design of future QA devices that have been argued to sample from thermal or close-to-thermal distributions [72], calling their role as optimization devices into question.

One approach to scaling down the temperature with problem size is the (theoretically) equivalent scaling up of the overall energy scale of the Hamiltonian. However, the rescaling of the total Hamiltonian is also known to be challenging and may not represent a convenient approach for a scalable architecture. An alternative approach is to develop quantum error correction techniques to effectively increase the energy scale of the Hamiltonian by coupling multiple qubits to form a single logical qubit [73–78] in conjunction with classical postprocessing [79–82] or to effectively decouple the system from the environment [83–86].

Our results reiterate the need for fault-tolerant error correction for scalable quantum annealing; however, they do not preclude the utility of quantum annealing optimizers for large finite size problems, where engineering challenges may be overcome to allow the device to operate effectively at a sufficiently low temperature such that problems of interest of a finite size may be solved even in the absence of fault tolerance. Our results only indicate that this "window of opportunity" cannot be expected to continue as devices are scaled without further improvements in the device temperature or energy scale.

While our arguments above indicate that fixed-temperature quantum annealers may not be scalable as optimizers, the current study does not pertain to the usage of quantum annealers as *samplers* [72,87,88], where the objective is to sample from the Boltzmann distribution. The latter objective is known to be a very difficult task (it is #*P* hard [89–91]) and little is known about when or if quantum annealers can provide an advantage in this regard [92].

T. A. and I. H. thank Daniel Lidar for useful comments on the manuscript. The computing resources were provided by the USC Center for High Performance Computing and Communications. T. A. was supported under ARO MURI Grant No. W911NF-11-1-0268, ARO MURI Grant No. W911NF-15-1-0582, and NSF Grant No. INSPIRE- 1551064. V. M.-M. was partially supported by MINECO (Spain) through Grant No. FIS2015-65078-C2-1-P (this contract partially funded by FEDER).

*itayhen@isi.edu

[1] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Achieving quantum supremacy with sparse and noisy commuting quantum computations, Quantum Topol. **1**, 8 (2017).
[2] X. Gao, S.-T. Wang, and L.-M. Duan, Quantum Supremacy for Simulating a Translation-Invariant Ising Spin Model, Phys. Rev. Lett. **118**, 040502 (2017).
[3] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-Case Complexity Versus Approximate Simulation of Commuting Quantum Computations, Phys. Rev. Lett. **117**, 080501 (2016).
[4] T. Morimae, K. Fujii, and J. F. Fitzsimons, Hardness of Classically Simulating the One-Clean-Qubit Model, Phys. Rev. Lett. **112**, 130502 (2014).
[5] M. A. Broome, A. Fedrizzi, S. Rahimi-Keshari, J. Dove, S. Aaronson, T. C. Ralph, and A. G. White, Photonic boson sampling in a tunable circuit, Science **339**, 794 (2013).
[6] J. B. Spring, B. J. Metcalf, P. C. Humphreys, W. Steven Kolthammer, X.-Min Jin, M. Barbieri, A. Datta, N. Thomas-Peter, N. K. Langford, D. Kundys, J. C. Gates, B. J. Smith, P. G. R. Smith, and I. A. Walmsley, Boson sampling on a photonic chip, Science **339**, 798 (2013).
[7] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, arXiv:1608.00263.
[8] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Quantum annealing: A new method for minimizing multidimensional functions, Chem. Phys. Lett. **219**, 343 (1994).
[9] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli, Quantum annealing of a disordered magnet, Science **284**, 779 (1999).
[10] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse Ising model, Phys. Rev. E **58**, 5355 (1998).
[11] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, Quantum computation by adiabatic evolution, arXiv:quant-ph/0001106.
[12] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Theory of quantum annealing of an Ising spin glass, Science **295**, 2427 (2002).
[13] A. P. Young, S. Knysh, and V. N. Smelyanskiy, Size Dependence of the Minimum Excitation Gap in the Quantum Adiabatic Algorithm, Phys. Rev. Lett. **101**, 170503 (2008).
[14] A. P. Young, S. Knysh, and V. N. Smelyanskiy, First-Order Phase Transition in the Quantum Adiabatic Algorithm, Phys. Rev. Lett. **104**, 020502 (2010).
[15] Itay Hen and A. P. Young, Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems, Phys. Rev. E **84**, 061152 (2011).
[16] E. Farhi, D. Gosset, I. Hen, A. W. Sandvik, P. Shor, A. P. Young, and F. Zamponi, Performance of the quantum adiabatic algorithm on random instances of two optimization problems on regular hypergraphs, Phys. Rev. A **86**, 052334 (2012).
[17] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Defining and detecting quantum speedup, Science **345**, 420 (2014).
[18] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, Probing for quantum speedup in spin-glass problems with planted solutions, Phys. Rev. A **92**, 042325 (2015).
[19] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar, Experimental signature of programmable quantum annealing, Nat. Commun. **4**, 2067 (2013).
[20] T. Albash, W. Vinci, A. Mishra, P. A. Warburton, and D. A. Lidar, Consistency tests of classical and quantum models for a quantum annealer, Phys. Rev. A **91**, 042314 (2015).
[21] Quantum enhanced optimization (qeo), https://www.iarpa.gov/index.php/research-programs/qeo.
[22] S. K. Tolpygo, V. Bolkhovsky, T. J. Weir, L. M. Johnson, M. A. Gouker, and W. D. Oliver, Fabrication process and properties of fully planarized deep-submicron nb/al-alox/nb Josephson junctions for vlsi circuits, IEEE Trans. Appl. Supercond. **25**, 1 (2015).
[23] S. K. Tolpygo, V. Bolkhovsky, T. J. Weir, C. J. Galbraith, L. M. Johnson, M. A. Gouker, and V. K. Semenov, Inductance of circuit structures for mit ll superconductor electronics fabrication process with 8 niobium layers, IEEE Trans. Appl. Supercond. **25**, 1 (2015).
[24] X. Y. Jin, A. Kamal, A. P. Sears, T. Gudmundsen, D. Hover, J. Miloshi, R. Slattery, F. Yan, J. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Thermal and Residual Excited-State Population in a 3d Transmon Qubit, Phys. Rev. Lett. **114**, 240501 (2015).
[25] J. King, S. Yarkoni, J. Raymond, I. Ozfidan, A. D. King, M. M. Nevisi, J. P. Hilton, C. C. McGeoch, Quantum Annealing amid Local Ruggedness and Global Frustration, arXiv:1502.02098.
[26] 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, A scalable control system for a superconducting adiabatic quantum optimization processor, Supercond. Sci. Technol. **23**, 065004 (2010).
[27] 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, A scalable readout system for a superconducting adiabatic quantum optimization system, Supercond. Sci. Technol. **23**, 105014 (2010).
[28] R. Harris *et al.*, Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor, Phys. Rev. B **82**, 024511 (2010).
[29] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz, and J. Whittaker, Architectural considerations in the design of a superconducting quantum annealing processor, IEEE Trans. Appl. Supercond. **24**, 1 (2014).
[30] J. King, S. Yarkoni, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, Benchmarking a quantum annealing processor with the time-to-target metric, arXiv:1508.05087.
[31] T. Kato, On the adiabatic theorem of quantum mechanics, J. Phys. Soc. Jpn. **5**, 435 (1950).
[32] S. Jansen, M.-Beth Ruskai, and R. Seiler, Bounds for the adiabatic approximation with applications to quantum computation, J. Math. Phys. (N.Y.) **48**, 102111 (2007).
[33] L. C. Venuti, T. Albash, D. A. Lidar, and P. Zanardi, Adiabaticity in open quantum systems, Phys. Rev. A **93**, 032118 (2016).

[34] The existence of small minimum gaps prior to the end of the anneal suggests that it is extremely unlikely that the Gibbs state before crossing these gaps would have a larger overlap with the ground state manifold of the final Hamiltonian than after. Therefore, measuring the system midway through a quantum annealing process will generically yield lower success probabilities than measurements taking place at the end of it [16,35,36]. We give two analytical examples in the Supplemental Material [37].

[35] B. Altshuler, H. Krovi, and J. Roland, Anderson localization makes adiabatic quantum optimization fail, Proc. Natl. Acad. Sci. U.S.A. **107**, 12446 (2010).

[36] Sergey Knysh, Zero-temperature quantum annealing bottlenecks in the spin-glass phase, Nat. Commun. **7**, 12370 (2016).

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.119.110502 for additional details about the derivations and simulations, which includes Refs. [38–44].

[38] M. Falcioni, E. Marinari, M. L. Paciello, G. Parisi, and B. Taglienti, Complex zeros in the partition function of the four-dimensional su(2) lattice gauge model, Phys. Lett. **108B**, 331 (1982).

[39] A. M. Ferrenberg and R. H. Swendsen, Optimized Monte Carlo Data Analysis, Phys. Rev. Lett. **63**, 1195 (1989).

[40] A. D. King, T. Lanting, and R. Harris, Performance of a quantum annealer on range-limited constraint satisfaction problems, arXiv:1502.02098.

[41] F. Hamze and N. de Freitas, From fields to trees, in *UAI*, edited by D. M. Chickering and J. Y. Halpern (AUAI Press, Arlington, Virginia, 2004), p. 243.

[42] A. Selby, Efficient subgraph-based sampling of ising-type models with frustration, arXiv:1409.3934.

[43] L. K. Grover, Quantum Mechanics Helps in Searching for a Needle in a Haystack, Phys. Rev. Lett. **79**, 325 (1997).

[44] J. Roland and N. J. Cerf, Quantum search by local adiabatic evolution, Phys. Rev. A **65**, 042308 (2002).

[45] This is equivalent to having a bounded degree connectivity graph.

[46] The analysis is based on the equivalence between the canonical and the microcanonical ensembles of statistical mechanics. This equivalence is reviewed in many places, see, e.g., Ref. [47].

[47] V. Martín-Mayor, Microcanonical Approach to the Simulation of First-Order Phase Transitions, Phys. Rev. Lett. **98**, 137207 (2007).

[48] Equivalently, $Ns(e)$ is the logarithm of the number of microstates with intensive energy e.

[49] A relation best known as the second law of thermodynamics $\mathrm{d}e = T\mathrm{d}s$.

[50] An energy probability density that is precisely Gaussian implies that the energy density is a linear function of the inverse temperature β and hence the specific heat is a constant. We elaborate on this point in the Supplemental Material [37].

[51] V. Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem, Quantum Inf. Process. **7**, 193 (2008).

[52] V. Choi, Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design, Quantum Inf. Process. **10**, 343 (2011).

[53] The reader is referred to the Supplemental Material [37] for further details.

[54] Details of these instances as well as similar results obtained for other problem classes are given in the Supplemental Material [37].

[55] The D-Wave processors are known to suffer from additional sources of error such as problem specification errors [70,71] and freeze-out before the end of the anneal [56] that prevent thermalization to the programmed problem Hamiltonian.

[56] M. H. Amin, Searching for quantum speedup in quasistatic quantum annealers, Phys. Rev. A **92**, 052323 (2015).

[57] R. Saket, A PTAS for the Classical Ising Spin Glass Problem on the Chimera Graph Structure, arXiv:1306.6943.

[58] By no means, however, is it meant that PTAS is able to thermally sample from a Boltzmann distribution of the input problem. In fact, it should be clear that PTAS does not.

[59] C. J. Geyer, Parallel tempering, in *Computing Science and Statistics Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidas (American Statistical Association, New York, 1991), p. 156.

[60] K. Hukushima and K. Nemoto, Exchange monte carlo method and application to spin glass simulations, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[61] Details of our PT implementation can be found in the Supplemental Material [37].

[62] Similar scaling behavior for other classes of Hamiltonians, specifically 3-regular 3-XORSAT instances and random ± 1 instances, is also observed, and we give the results for these instances in the Supplemental Material [37].

[63] T. Jörg, J. Lukic, E. Marinari, and O. C. Martin, Strong Universality and Algebraic Scaling in Two-Dimensional Ising Spin Glasses, Phys. Rev. Lett. **96**, 237205 (2006).

[64] C. K. Thomas, D. A. Huse, and A. A. Middleton, Zero and Low Temperature Behavior of the Two-Dimensional $\pm j$ Ising Spin Glass, Phys. Rev. Lett. **107**, 047203 (2011).

[65] F. P. Toldin, A. Pelissetto, and E. Vicari, Finite-size scaling in two-dimensional ising spin-glass models, Phys. Rev. E **84**, 051116 (2011).

[66] L. A. Fernandez, E. Marinari, V. Martin-Mayor, G. Parisi, and J. J. Ruiz-Lorenzo, Universal critical behavior of the two-dimensional ising spin glass, Phys. Rev. B **94**, 024402 (2016).

[67] A. Lucas, Ising formulations of many NP problems, Front. Phys. **2**, 5 (2014).

[68] W. Vinci and D. A. Lidar, Optimally Stopped Optimization, Phys. Rev. Applied **6**, 054016 (2016).

[69] S. Boixo, T. F. Ronnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, Evidence for quantum annealing with more than one hundred qubits, Nat. Phys. **10**, 218 (2014).

[70] V. Martin-Mayor and I. Hen, Unraveling quantum annealers using classical hardness, Sci. Rep. **5**, 15324 (2015).

[71] Z. Zhu, A. J. Ochoa, S. Schnabel, F. Hamze, and H. G. Katzgraber, Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities, Phys. Rev. A **93**, 012317 (2016).

[72] M. H. Amin, Searching for quantum speedup in quasistatic quantum annealers, Phys. Rev. A **92**, 052323 (2015).

[73] K. L Pudenz, T. Albash, and D. A. Lidar, Error-corrected quantum annealing with hundreds of qubits, Nat. Commun. **5**, 3243 (2014).

[74] K. L. Pudenz, T. Albash, and D. A. Lidar, Quantum annealing correction for random Ising problems, Phys. Rev. A **91**, 042302 (2015).

[75] W. Vinci, T. Albash, G. Paz-Silva, I. Hen, and D. A. Lidar, Quantum annealing correction with minor embedding, Phys. Rev. A **92**, 042310 (2015).

[76] S. Matsuura, H. Nishimori, T. Albash, and D. A. Lidar, Mean field Analysis of Quantum Annealing Correction, Phys. Rev. Lett. **116**, 220501 (2016).

[77] W. Vinci, T. Albash, and D. A Lidar, Nested quantum annealing correction, npj Quantum Inf. **2**, 16017 (2016).

[78] S. Matsuura, H. Nishimori, W. Vinci, T. Albash, and D. A. Lidar, Quantum-annealing correction at finite temperature: Ferromagnetic p-spin models, Phys. Rev. A **95**, 022308 (2017).

[79] N. Chancellor, Modernizing quantum annealing using local searches, New J. Phys. **19**, 023024 (2017).

[80] N. Chancellor, Modernizing quantum annealing II: Genetic algorithms and inference, arXiv:1609.05875.

[81] H. Karimi and G. Rosenberg, Boosting quantum annealer performance via sample persistence, Quantum Inf. Process. **16**, 166 (2017).

[82] H. Karimi, G. Rosenberg, and H. G. Katzgraber, Effective optimization using sample persistence: A case study on quantum annealers and various Monte Carlo optimization methods, arXiv:1706.07826.

[83] S. P. Jordan, E. Farhi, and P. W. Shor, Error-correcting codes for adiabatic quantum computation, Phys. Rev. A **74**, 052322 (2006).

[84] A. D. Bookatz, E. Farhi, and L. Zhou, Error suppression in Hamiltonian-based quantum computation using energy penalties, Phys. Rev. A **92**, 022317 (2015).

[85] Z. Jiang and E. G. Rieffel, Non-commuting two-local Hamiltonians for quantum error suppression, Quantum Inf. Process. **16**, 89 (2017).

[86] M. Marvian and D. A. Lidar, Error Suppression for Hamiltonian-Based Quantum Computation using Subsystem Codes, Phys. Rev. Lett. **118**, 030504 (2017).

[87] S. H. Adachi and M. P. Henderson, Application of quantum annealing to training of deep neural networks, arXiv:1510.06356.

[88] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, Quantum-assisted learning of graphical models with arbitrary pairwise connectivity [Phys. Rev. X (to be published)].

[89] C. H. Papadimitriou, *Computational Complexity* (Addison Wesley Longman, Reading, MA, 1995).

[90] L. G. Valiant, The complexity of enumeration and reliability problems, SIAM J. Comput. **8**, 410 (1979).

[91] V. Dahllöf, P. Jonsson, and M. Wahlström, Counting models for 2sat and 3sat formulae, Theor. Comput. Sci. **332**, 265 (2005).

[92] L. Campos Venuti, T. Albash, M. Marvian, D. Lidar, and P. Zanardi, Relaxation vs adiabatic quantum steady state preparation: which wins?, Phys. Rev. A **95**, 042302 (2017).