Quantum Science and Technology

ACCEPTED MANUSCRIPT

Analog errors in Ising machines

To cite this article before publication: Tameem Albash *et al* 2019 *Quantum Sci. Technol.* in press https://doi.org/10.1088/2058-9565/ab13ea

Analog Errors in Ising Machines

Tameem Albash,[1,2,3] Victor Martin-Mayor,[4,5] and Itay Hen[1,2,3]

[1] *Information Sciences Institute, University of Southern California, Marina del Rey, California 90292, 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] *Departamento de Física Teórica, Universidad Complutense, 28040 Madrid, Spain*
[5] *Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Zaragoza, Spain*

Recent technological breakthroughs have precipitated the availability of specialized devices that promise to solve NP-Hard problems faster than standard computers. These 'Ising Machines' are however analog in nature and as such inevitably have implementation errors. We find that their success probability decays exponentially with problem size for a fixed error level, and we derive a sufficient scaling law for the error in order to maintain a fixed success probability. We corroborate our results with experiment and numerical simulations and discuss the practical implications of our findings.

Introduction.—Recently there has been a flourishing of experimental quantum information processing devices [1–5]. The scientific community is readying itself for the first demonstration of 'quantum supremacy' [6–13] in the Noisy Intermediate-Scale Quantum (NISQ) era [14], whereby quantum information processing devices will start performing tasks not accessible even for the largest classical supercomputers. At the forefront of this effort has been the development of analog quantum devices performing quantum annealing [15, 16], already realized on various platforms such as arrays of superconducting flux qubits [5, 17–23], to solve optimization problems, that is, to find bit assignments that minimize the cost of discrete combinatorial problems or equivalently the ground states (GSs) of Ising Hamiltonians. With this growing interest in developing special purpose devices for solving such problems, alternatives have also emerged [24–26], touting improved performance over standard computers [27, 28].

These Ising machines are analog in nature: the programmable parameters of the cost functions they aim to solve are controlled by continuous fields, implying that the intended cost function is only implemented to a certain precision. In this Letter, we study the effect of the analog nature of these devices on their ability to find global minima. We argue that even when such devices are assumed ideal, i.e., always find a minimum of the *implemented* cost function, the obtained configurations become exponentially unlikely to be global minima of the *intended* problem. In the absence of fault tolerant error correction, this represents a fundamental limitation to the scalability of Ising machines. We corroborate our analytical findings with numerical simulations as well as experiment carried out on a D-Wave 2000Q processor [29]. We further derive a scaling law for how the magnitude of implementation errors must be reduced with problem size in order to maintain acceptable performance. We find that errors must scale as a power-law with problem size, with a model dependent exponent. Our results imply that fixed-size classical repetition codes are not a feasible approach to help these devices maintain their performance asymptotically with size.

Analog Ising machines.—Analog Ising machines suffer by nature from implementation errors caused by the conversion of the intended problem parameters to analog signals [30]. The Hamiltonian implemented by these devices is generally of the form:

$$H_\sigma(s) = H_0 + \delta H_\sigma = \sum_{\langle ij\rangle} \tilde{J}_{ij} s_i s_j + \sum_i \tilde{h}_i s_i \qquad (1)$$
$$= \sum_{\langle ij\rangle} \left(J_{ij} + \delta J_{ij}\right) s_i s_j + \sum_i \left(h_i + \delta h_i\right) s_i \,,$$

where $\{s_i = \pm 1\}$ are binary variables that are to be optimized over and $\langle ij\rangle$ denotes the underlying connectivity graph of the model. We denote by $\{J_{ij}, h_i\}$ the intended Ising couplings between connected spins and the intended local fields on individual spins respectively of the intended Hamiltonian H_0, and we denote by $\{\tilde{J}_{ij}, \tilde{h}_i\}$ the implemented values, obeying $\tilde{J}_{ij} = J_{ij} + \delta J_{ij}$ and $\tilde{h}_i = h_i + \delta h_i$. The variables δJ and δh represent the noise due to the analog nature of the device, and we take for simplicity $\delta J, \delta h \sim \mathcal{N}(0, \sigma^2)$ [31]. We assume the noise for the different δJ_{ij} and δh_i is statistically independent [32].

To isolate the effect of analog noise on performance [33–36], where success is defined as finding a GS of the intended Hamiltonian, we assume that the machine is otherwise ideal, i.e., that it always finds the GS of the implemented Hamiltonian. This allows us to draw from classical spin glass theory, where it is known that spin glasses are susceptible to a phenomenon referred to as J-chaos, in which perturbations to the intended Hamiltonian change the identity of the GS [37–40].

Analytical treatment.— For concreteness, we consider a spin-glass Hamiltonian of the form of Eq. (1), where the intended couplings take values $J_{ij} = \{1, 0, -1\}$ and $h_i = 0$ [41]. Let us denote a GS of the intended Hamiltonian (in general, there could be exponentially

many of those) by $s^{\rm G}$ and an excited state (ES) of the intended Hamiltonian by $s^{\rm E}$. The energy gap between the two is

$$\Delta E_0 = H_0(s^{\rm E}) - H_0(s^{\rm G}) > 0 \ . \tag{2}$$

The ES has a chance of becoming the GS of the implemented Hamiltonian only if $\Delta E_\sigma = H_\sigma(s^{\rm E}) - H_\sigma(s^{\rm G}) < 0$. Expressed differently,

$$\Delta E_\sigma = \Delta E_0 - \sum_{\langle ij \rangle} \delta J_{ij} s_i^{\rm G} s_j^{\rm G} (1 - q_{ij}^{\rm link}) - \sum_i \delta h_i s_i^{\rm G} (1 - q_i) \,, \tag{3}$$

where we have introduced the spin overlap $q_i = s_i^{\rm G} s_i^{\rm E}$ and the link-overlap $q_{ij}^{\rm link} = s_i^{\rm G} s_j^{\rm G}\, s_i^{\rm E} s_j^{\rm E}$ corresponding to the bond (ij). When $q_{ij}^{\rm link} = 1$, it implies that if the bond (ij) is satisfied (unsatisfied) by the GS of the intended Hamiltonian, it is also satisfied (unsatisfied) by the ES. Similarly $q_{i,j}^{\rm link} = -1$ implies that the corresponding bond is satisfied in one spin assignment and unsatisfied for the other.

We now define W to be the total number of bonds with $q_{ij}^{\rm link} = -1$ and D to be the total number of spins with $q_i = -1$ (or equivalently the Hamming distance between the ES and the GS). Because (i) the δJ_{ij} and δh_i are statistically independent from $s_i^{\rm G}$ and $s_i^{\rm G} s_j^{\rm G}$ and (ii) the δJ_{ij} for different bonds and δh_i for different spins are mutually independent, we can write

$$\Delta E_\sigma = \Delta E_0 + 2\sigma \sqrt{W + D}\, \eta \,, \tag{4}$$

where η is a normal random variable $\mathcal{N}(0,1)$. Defining the parameter $z = \frac{\Delta E_0}{2\sigma\sqrt{W+D}}$, the probability $p(z)$ of having $\Delta E_\sigma < 0$ is given by

$$p(z) \equiv {\rm Prob}[\Delta E_\sigma < 0] = \int_{-\infty}^{-z} \frac{{\rm d}\eta}{\sqrt{2\pi}} {\rm e}^{-\eta^2/2} . \tag{5}$$

Therefore, the likelihood that a particular ES lies below a GS for the implemented Hamiltonian ranges from being infinitesimal [more precisely $p(z) \sim {\rm e}^{-z^2/2}$ for $z \gg 1$], to 50% (for $z \ll 1$). In fact, $p(z)$ become sizable at $z = 1$.

For physical devices, the number of couplers scales linearly with number of spins n, so at most $W \sim n$. Similarly, D can at most scale as $\sim n$. More specifically, we identify two kinds of ESs (see Ref. [34]). On the one hand, there are 'topologically trivial' ESs for which $W \sim 1$. For these, there are no energy barriers separating them from a GS, i.e., gradient descent takes the ES to a GS and therefore, $p(z)$ is extremely small for these ESs up to $\sigma \sim 1$. On the other hand, there exist topologically non-trivial ESs for which $W \sim n$. These low lying ESs produce a sizable $p(z)$ already for $\sigma \sim 1/\sqrt{W+D} \sim 1/\sqrt{n}$. From now on, we consider only the latter kind.

So far we have only considered a pair of states, a single GS and a single ES (both out of possibly exponentially many [42–44]). However, the relative degeneracies of these must be taken into account. To do that, in lieu of relying on a particular model [39], we make a few simplifying assumptions. First, let us restrict to the ESs with a fixed ΔE_0 (in our case $\Delta E_0 \geq 2$) and assume there are $N_{\rm ES}$ of those. We further assume that: i) W does not fluctuate significantly between different ESs. In other words, $p(z)$ defined in Eq. (5) can be roughly regarded as non-fluctuating. ii) In Eq. (4), each ES defines a random variable η. We now assume that the η's for different ESs can be regarded as independent. In other words, $N_{\rm ES}$ becomes the *effective* number of statistically independent η's within the entire population of ESs of the intended Hamiltonian (we later show that the conclusions that follow from these assumptions are consistent with numerical results).

Under these assumptions, the probability of a single GS to remain energetically favorable to *every* ES for the implemented Hamiltonian is

$$p_{1\,{\rm GS}} \equiv [1 - p(z)]^{N_{\rm ES}} \ . \tag{6}$$

Since we expect $N_{\rm ES}$ to scale at least polynomially (if not exponentially) with system size, we conclude that the likelihood for a single GS to remain optimal for the implemented Hamiltonian *decays* exponentially with system size.

In the most general case, there may be an *effective* number of GSs, $N_{\rm GS}$, only one of which is required to remain optimal for an analog Ising machine to succeed. The probability that at least one GS remains more optimal than all the ESs for the implemented Hamiltonian is:

$$p_{\rm S} = 1 - [1 - p_{1\,{\rm GS}}]^{N_{\rm GS}} \ . \tag{7}$$

We expect that the growth of $N_{\rm GS}$ is unlikely to overcome the decay of $p_{1\,{\rm GS}}$ with system size [45], and it would then follow that $p_{\rm S}$ will also become exponentially small with system size, i.e., analog Ising machines are expected to fail for any fixed noise level as we scale up the system size.

At this point, we address how must the noise strength σ scale in order to still have a sub-exponential probability to succeed. In the regime where the asymptotic estimate $p(z) \sim {\rm e}^{-z^2/2} \ll 1$ holds, i.e. a regime where the failure rate is small, rewriting Eq. (7) for a fixed $p_{\rm S}$ requires that z behave as:

$$z^2 \approx 2\left(\log(N_{\rm ES}) - \log(1 - p_{\rm S})/N_{\rm GS}\right) \ . \tag{8}$$

If in addition the effective number of ESs scales as $N_{\rm ES} \sim {\rm e}^{Bn^k}$ (with $k = 0$ denoting sub-exponential scaling) [46], then we conclude the noise level must be scaled down as

$$\sigma_{\rm S} \sim 1/\sqrt{(W+D)n^k} \ . \tag{9}$$

A worst-case estimate would be to take $W, D \sim n$ and $k = 1$, in which case an estimate of the required scaling of the noise to maintain a sub-exponential success

probability is

$$\sigma_{\mathrm{S}}^{(\text{worst case})} \sim 1/n\,. \tag{10}$$

Numerical and experimental results.— To illustrate the validity of the above derivation and accompanying simplifying assumptions, we present results of numerical simulations done on instances with known GSs. Our first problem class is of planted-solution instances [47] defined on a Chimera lattice: an $L \times L$ grid of 8-spin unit cells in a $K_{4,4}$ formation and is the graph of currently available D-Wave quantum annealers [48]. The instances have a degenerate GS corresponding to all spins pointing up or all spins pointing down (due to the Z_2 symmetry of the problem Hamiltonian). The couplings are restricted to have magnitudes $\{1/3, 2/3, 1\}$. (See the SM for further details [49].)

To simulate analog errors, we introduce a zero-mean Gaussian noise with standard deviation σ as in Eq. (1) to the implemented Hamiltonian, and we use the Hamze-Freitas-Selby (HFS) [50, 51] algorithm to find the GSs of the implemented Hamiltonian. While the HFS algorithm is not guaranteed to find the GS with certainty, we can know with certainty if the GS of the implemented Hamiltonian has changed from that of the desired Hamiltonian if (1) the returned state by the algorithm has a lower energy on the implemented Hamiltonian than the two GSs of the intended Hamiltonian, and (2) the returned state is an ES of the intended Hamiltonian. In order to ensure that we only consider topologically non-trivial ESs, we perform 100 sweeps of steepest-gradient updates: we flip each spin, and if the energy is reduced (according to the intended Hamiltonian) we accept the update and reject otherwise. In cases where the returned state fails either condition, we assume that the GS of the intended Hamiltonian has not changed. Since this may in fact be incorrect, our results are an upper bound on the probability that the GS has not changed. For each grid size $L \in [8, 14]$, we generate 10^2 instances representing our intended Hamiltonians, and for each instance we generate 10^3 noisy realizations of the implemented Hamiltonian.

In Fig. 1, we show the probability that the GS of the intended Hamiltonian remains the GS of the implemented Hamiltonian for several sizes and noise strengths. Our results show an exponential dependence as a function of $\sigma^2 n$ for large p_{S} and approaching $\sigma^{7/4} n$ for small p_{S}, consistently with the above analytical derivation. This change in scaling may be expected since as the system size or noise level grows the effective number of ESs grows as well. We find that for these instances W and D scale linearly with n (shown in the inset of Fig. 1), which suggests that the number of effective ESs grows from polynomial to a stretched-exponential with $k = 1/7$ [Eq. (9)]. For these instances then, the noise level must scale as $\sigma_{\mathrm{S}} \sim n^{-4/7}$ in order to maintain a fixed performance level.

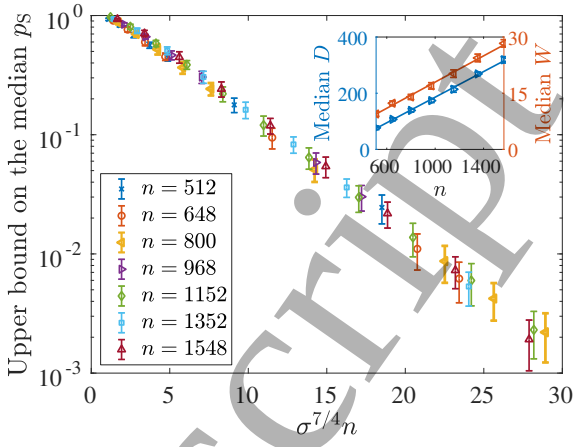

FIG. 1. Upper bound on the median probability p_{S} over 10^2 instances that the GS of the intended Hamiltonian remains the GS of the implemented Hamiltonian for varying planted-solution instances on different problem sizes n and Gaussian noise strengths $\sigma \in [0.01, 0.2]$. Error bars correspond to two standard deviate error bars obtained from bootstrapping over the 10^2 instances. Inset: the scaling of the median D and W from the intended GSs for $\sigma = 0.1$. Solid line corresponds to a linear fit with slope $\alpha \approx 0.23$ and $\alpha \approx 0.02$ for D and W respectively.

While our analysis so far has focused only on whether the intended GS is no longer a GS of the implemented Hamiltonian, it is instructive to study the energy distribution (according to the intended Hamiltonian) of the topologically non-trivial ESs that become the new GSs of the implemented Hamiltonian. This gives us a sense of how far away in energy the implemented GSs are from the intended GS. We show this in Fig. 2, where we see that for a fixed system size and growing noise strength, the distribution becomes more Gaussian-like with the mean moving farther away from the GS energy. Similar behavior occurs for a fixed noise strength and growing system size, which we provide in the SM. This effect is similar to the behavior of a thermal energy distribution for a fixed temperature and increasing system size [52].

We corroborate our analysis with runs on a D-Wave 2000Q quantum annealing processor [5]. The device has an intrinsic noise level [34, 53], however we introduce additional Gaussian noise to simulate the effect of different noise levels. As shown in Fig. 3, even in the absence of noise, the device rarely finds a GS, which we can attribute to both thermal [52] and analog errors on the device. As the noise level is increased, the distribution of states observed moves even farther away from the GS energy, in a manner consistent with Fig. 2.

Conclusions and outlook.— In this work, we have pointed to a fundamental limitation of analog Ising Machines, whereby implementation errors detrimentally affects their scaling performance. We have shown that even under the assumption that these devices are otherwise

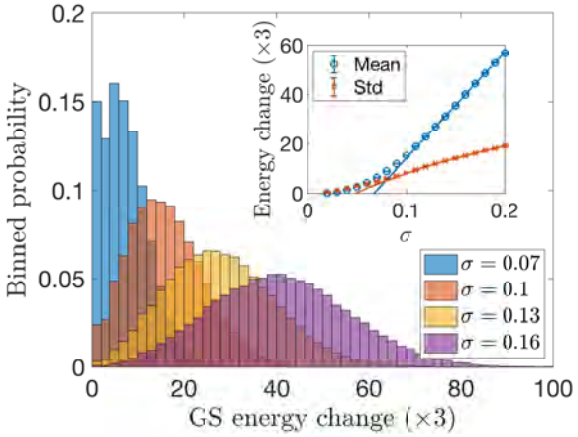

FIG. 2. Histogram of the energy of the GSs of the implemented (noisy) Hamiltonian as measured by the intended Hamiltonian. Here we use 10^3 noise realizations for each of the 10^2 planted-solution Chimera instances defined on a 12×12 grid. Only the topologically non-trivial ESs of the intended Hamiltonian are shown. Inset: the scaling of the mean and standard deviation of the distribution of GS energies. The error bars correspond to two standard deviate error bars generated by a bootstrap over the 10^5 data points. The solid curves are fits to $\mu_E = a + \sigma b$ and $\sigma_E = c + \sqrt{\sigma} d$ for the mean and standard deviation respectively, with $a = -28.82, b = 429.47$ and $c = -18.57, d = 84.28$.

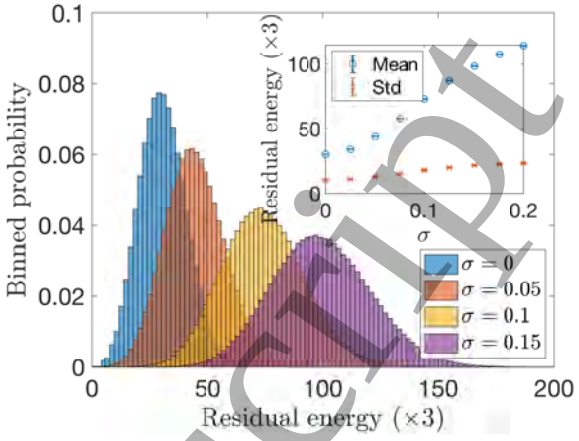

FIG. 3. Residual energy of the states returned by the D-Wave 2000Q quantum annealing processor at NASA's Ames Research Center after post-processing with 100 sweeps of steepest descent for a single instance. The instance here is defined on a 12×12 subgraph of the processor (a total possible of $n = 1142$ spins), and it is generated in an identical way to the Chimera instances studied in Figs. 1 and 2. Gaussian noise with mean 0 and standard deviation σ is added to both the couplings and local fields before the instance is submitted to the processor. For this single instance, 10^4 noise realizations were generated and 10^2 anneals were performed for each noise realization. (Further details are given in the SM.)

ideal, i.e., that they instantaneously find a minimizing configuration of the implemented problem, their success probability decays exponentially with system size for any fixed nonzero noise level. Such errors have important ramifications beyond the context of optimization; in the setting of generating thermal (Boltzmann) samples, we can expect that analog errors may distort the sampled distribution [54]. Even quantum logic gates cannot be implemented perfectly, but the difference is that fault-tolerant error correction can correct for these errors [55].

We emphasize that our results do not mean that 'disorder-chaos' should be expected for every mildly-disordered system. (We provide an example in terms of the 1d chain in the SM that is meant to illustrate this.) The key point is the nature of the low-energy landscape of the problem under consideration. Broadly speaking, 'disorder-chaos' requires the problem Hamiltonian to have a *large* number of very low-energy excitations, each of them wildly differing from the GS (i.e. Hamming distance of order n). Under such circumstances, it is intuitively clear that even tiny errors may cause one of these non-trivial excited states to take over as the true GS of the implemented Hamiltonian. In fact, the hypothesis we made to derive Eqs. (3-7) made quantitative the above-outlined physical picture.

Now, it turns out that disorder chaos is a generic feature of spin-glasses. Because spin-glasses are an archetype of *hard* optimization problems, we should expect to encounter disorder chaos in every optimization problem hard-enough to deserve to be solved with a specialized analog device such as an Ising machine. Indeed, the 3-regular 3-XORSAT instances that we study in the SM exhibit disorder chaos as well. Fortunately, as universality suggests, the scaling laws that we find for 3-XORSAT seem compatible with our findings for instances on the Chimera lattice.

Our results for the Chimera instances differ slightly from the results on J-chaos (sometimes also called bond-chaos) for ground state of the 2d Edward-Anderson model by Krzakala & Bouchaud [39] (a finite temperature analysis was carried out in Ref. [56]), where they found a scaling of $\sigma^2 n$ for the regime of fully developed J-chaos. This scaling suggests a sub-exponential scaling for the number of ESs [$k = 0$ in Eq. (9)]. This discrepancy might be accounted for by the fact, at exactly zero temperature, that the ruling fixed point for the Renormalization Group flow for Gaussian and binary couplings is different [57]. Extending beyond 2d, the example of 3-regular XORSAT, which we present in the SM, exhibits a scaling closer to $\sigma^{4/3} n$, indicating a value of $k = 1/2$ for this class of instances. By deriving appropriate scaling laws, we have found that in order to counteract this reduction in performance as the system size grows, the noise level must be scaled down as a power law (with the worst case being $1/n$).

We can consider how known error correction schemes would effectively reduce the magnitude of this noise. One

way is to use classical repetition codes that implement multiple copies of the intended Hamiltonian [58–62]. This has the feature of effectively rescaling the implemented Hamiltonian norm by a factor K and hence effectively reducing the noise strength by a factor of $1/\sqrt{K}$ [58]. Therefore, a rescaling of $K \sim n^2$ is needed to achieve the necessary noise reduction for the worst case [Eq. (10)] and of $K \sim n^{8/7}$ for the Chimera example studied numerically here (Fig. 1). More generally, for a K that scales as n^α, the encoded Hamiltonian using classical repetition schemes [58, 62] would require a Hamiltonian for which the energy scale grows faster than linear with n. For scalable architectures, this would require the number of physical spins per encoded spin to grow as a power law with n. This highlights the importance of scalable error correction to maintain the performance of an algorithm as the system size scales.

We conclude by emphasizing that our analysis is asymptotic in nature and is valid in the large n limit. If only a specific problem class of a certain finite size are of interest, then it would be possible in principle to engineer a device with a sufficiently low noise for that problem class and size. It remains to be seen whether analog Ising machines will be a viable alternative to digital Ising simulations in light of the fundamental problems illustrated by our work.

Acknowledgements.— Computation for the work described in this paper was supported by the University of Southern California's Center for High-Performance Computing (hpc.usc.edu) and by the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. 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. We acknowledge the support of the Universities Space Research Association (USRA) Quantum AI Lab Research Opportunity Program. V. M.-M. was partially supported by MINECO (Spain) through Grant No. FIS2015-65078- C2-1-P (this contract partially funded by FEDER).

[1] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," Nature **549**, 242 EP – (2017).

[2] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. Schuyler Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, C. A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, R. S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, B. R. Johnson, M. Reagor, M. P. da Silva, and C. Rigetti, "Unsupervised Machine Learning on a Hybrid Quantum Computer," ArXiv e-prints (2017), arXiv:1712.05771 [quant-ph].

[3] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, R. Barends, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, B. Foxen, R. Graff, E. Jeffrey, J. Kelly, E. Lucero, A. Megrant, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, H. Neven, and J. M. Martinis, "A blueprint for demonstrating quantum supremacy with superconducting qubits," ArXiv e-prints (2017), arXiv:1709.06678 [quant-ph].

[4] 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 e-prints (2016), arXiv:1608.00263 [quant-ph].

[5] "D-wave systems previews 2000-qubit quantum system," .

[6] J. Preskill, "Quantum computing and the entanglement frontier," ArXiv e-prints (2012), arXiv:1203.5813 [quant-ph].

[7] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd, "Achieving quantum supremacy with sparse and noisy commuting quantum computations," Quantum **1**, 8 (2017).

[8] Xun Gao, Sheng-Tao Wang, and L.-M. Duan, "Quantum supremacy for simulating a translation-invariant ising spin model," Phys. Rev. Lett. **118**, 040502 (2017).

[9] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd, "Average-case complexity versus approximate simulation of commuting quantum computations," Phys. Rev. Lett. **117**, 080501 (2016).

[10] Tomoyuki Morimae, Keisuke Fujii, and Joseph F. Fitzsimons, "Hardness of classically simulating the one-clean-qubit model," Phys. Rev. Lett. **112**, 130502 (2014).

[11] Matthew A. Broome, Alessandro Fedrizzi, Saleh Rahimi-Keshari, Justin Dove, Scott Aaronson, Timothy C. Ralph, and Andrew G. White, "Photonic boson sampling in a tunable circuit," Science **339**, 794–798 (2013).

[12] Justin B. Spring, Benjamin J. Metcalf, Peter C. Humphreys, W. Steven Kolthammer, Xian-Min Jin, Marco Barbieri, Animesh Datta, Nicholas Thomas-Peter, Nathan K. Langford, Dmytro Kundys, James C. Gates, Brian J. Smith, Peter G. R. Smith, and Ian A. Walmsley, "Boson sampling on a photonic chip," Science **339**, 798–801 (2013).

[13] 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 e-prints (2016), arXiv:1608.00263 [quant-ph].

[14] J. Preskill, "Quantum Computing in the NISQ era and beyond," ArXiv e-prints (2018), arXiv:1801.00862

[quant-ph].
[15] Tadashi Kadowaki and Hidetoshi Nishimori, "Quantum annealing in the transverse Ising model," Phys. Rev. E **58**, 5355 (1998).
[16] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser, "Quantum Computation by Adiabatic Evolution," arXiv:quant-ph/0001106 (2000).
[17] 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 Transactions on Applied Superconductivity **25**, 1–12 (2015).
[18] 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 Transactions on Applied Superconductivity **25**, 1–5 (2015).
[19] 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).
[20] 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," Superconductor Science and Technology **23**, 065004 (2010).
[21] 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," Superconductor Science and Technology **23**, 105014 (2010).
[22] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose, "Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor," Phys. Rev. B **82**, 024511 (2010).
[23] P. I Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, AJ. Berkley, R. Harris, J. P. Hilton, T. Lanting, AJ. Przybysz, and J. Whittaker, "Architectural considerations in the design of a superconducting quantum annealing processor," IEEE Transactions on Applied Superconductivity **24**, 1–10 (Aug. 2014).
[24] Yoshihisa Yamamoto, Kazuyuki Aihara, Timothee Leleu, Ken-ichi Kawarabayashi, Satoshi Kako, Martin Fejer, Kyo Inoue, and Hiroki Takesue, "Coherent ising machines—optical neural networks operating at the quantum limit," npj Quantum Information **3**, 49 (2017).
[25] Peter L. McMahon, Alireza Marandi, Yoshitaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsunomiya, Kazuyuki Aihara, Robert L. Byer, M. M. Fejer, Hideo Mabuchi, and Yoshihisa Yamamoto, "A fully-programmable 100-spin coherent ising machine with all-to-all connections," Science (2016), 10.1126/science.aah5178.
[26] "Quantum neural network: Collective computing at criticality of opo phase transition," .
[27] Yoshitaka Haribara, Hitoshi Ishikawa, Shoko Utsunomiya, Kazuyuki Aihara, and Yoshihisa Yamamoto, "Performance evaluation of coherent ising machines against classical neural networks," Quantum Science and Technology **2**, 044002 (2017).
[28] Takahiro Inagaki, Yoshitaka Haribara, Koji Igarashi, Tomohiro Sonobe, Shuhei Tamate, Toshimori Honjo, Alireza Marandi, Peter L. McMahon, Takeshi Umeki, Koji Enbutsu, Osamu Tadanaga, Hirokazu Takenouchi, Kazuyuki Aihara, Ken-ichi Kawarabayashi, Kyo Inoue, Shoko Utsunomiya, and Hiroki Takesue, "A coherent ising machine for 2000-node optimization problems," Science **354**, 603–606 (2016).
[29] J. King, S. Yarkoni, J. Raymond, I. Ozfidan, A. D. King, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, "Quantum Annealing amid Local Ruggedness and Global Frustration," ArXiv e-prints (2017), arXiv:1701.04579 [quant-ph].
[30] Other sources of errors not discussed here also exist such as $1/f$-noise [**?**].
[31] Andrew D. King, Trevor Lanting, and Richard Harris, "Performance of a quantum annealer on range-limited constraint satisfaction problems," arXiv:1502.02098 (2015).
[32] We expect our results to be robust to noise with a small local correlation but certainly not to noise with long-range statistical correlations.
[33] Davide Venturelli, Salvatore Mandrà, Sergey Knysh, Bryan O'Gorman, Rupak Biswas, and Vadim Smelyanskiy, "Quantum optimization of fully connected spin glasses," Phys. Rev. X **5**, 031040 (2015).
[34] V. Martín-Mayor and I. Hen, "Unraveling quantum annealers using classical hardness," Scientific Reports **5**, 15324 (2015).
[35] Zheng Zhu, Andrew J. Ochoa, Stefan Schnabel, Firas Hamze, and Helmut 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).
[36] Salvatore Mandrà, Zheng Zhu, Wenlong Wang, Alejandro Perdomo-Ortiz, and Helmut G. Katzgraber, "Strengths and weaknesses of weak-strong cluster problems: A detailed overview of state-of-the-art classical heuristics versus quantum approaches," Phys. Rev. A **94**, 022337 (2016).
[37] M. Nifle and H. J. Hilhorst, "New critical-point exponent and new scaling laws for short-range ising spin glasses," Phys. Rev. Lett. **68**, 2992–2995 (1992).
[38] M. Ney-Nifle, "Chaos and universality in a four-dimensional spin glass," Phys. Rev. B **57**, 492 (1998).
[39] F. Krzakala and J. P. Bouchaud, "Disorder chaos in spin glasses," Europhys. Lett. **72**, 472 (2005).
[40] H. G. Katzgraber and F. Krzakala, "Temperature and disorder chaos in three-dimensional ising spin glasses," Phys. Rev. Lett. **98**, 017201 (2007).
[41] Our derivation can be straightforwardly generalized to other models as well.
[42] F. Barahona, T. Maynard, R. Rammal, and J. P. Uhry, "Morphology of ground states of two-dimensional frustration model," J. Phys. A **15**, 673 (1982).
[43] F. Romá, S. Bustingorry, and P. M. Gleiser, "Signature of the ground-state topology in the low-temperature

dynamics of spin glasses," Phys. Rev. Lett. **96**, 167205 (2006).

[44] L. A. Fernandez, V. Martin-Mayor, G. Parisi, and B. Seoane, "Temperature chaos in 3d ising spin glasses is driven by rare events," EPL (Europhysics Letters) **103**, 67003 (2013).

[45] We can always make $p_{1\,\mathrm{GS}}$ smaller than $1/N_{\mathrm{GS}}$ by increasing σ for example. This will increase the typical ΔE_0, which implies that the typical N_{ES} will grow.

[46] While studies on temperature chaos [44] suggest that ESs are exponentially more abundant than GSs [34] which might suggest that $k = 1$, here we are requesting topologically non-trivial and statistically independent ESs.

[47] Itay Hen, Joshua Job, Tameem Albash, Troels F. Rønnow, Matthias Troyer, and Daniel A. Lidar, "Probing for quantum speedup in spin-glass problems with planted solutions," Phys. Rev. A **92**, 042325– (2015).

[48] Vicky Choi, "Minor-embedding in adiabatic quantum computation: I. The parameter setting problem," Quant. Inf. Proc. **7**, 193–209 (2008).

[49] In the SM, we consider two additional problem classes, namely, 3-regular 3-XORSAT instances involving 3-body couplings, which exhibits a similar behavior, and Ising instances defined on a square grid with random ± 1 couplings.

[50] Firas Hamze and Nando de Freitas, "From fields to trees," in *UAI*, edited by David Maxwell Chickering and Joseph Y. Halpern (AUAI Press, Arlington, Virginia, 2004) pp. 243–250.

[51] Alex Selby, "Efficient subgraph-based sampling of ising-type models with frustration," arXiv:1409.3934 (2014).

[52] Tameem Albash, Victor Martin-Mayor, and Itay Hen, "Temperature scaling law for quantum annealing optimizers," Phys. Rev. Lett. **119**, 110502 (2017).

[53] Andrew D. King and Catherine C. McGeoch, "Algorithm engineering for a quantum annealing platform," arXiv:1410.2628 (2014).

[54] Philipp Hauke, Fernando M Cucchietti, Luca Tagliacozzo, Ivan Deutsch, and Maciej Lewenstein, "Can one trust quantum simulators?" Reports on Progress in Physics **75**, 082401 (2012).

[55] Emanuel Knill and Raymond Laflamme, "Theory of quantum error-correcting codes," Phys. Rev. A **55**, 900–911 (1997).

[56] Wenlong Wang, Jonathan Machta, and Helmut G. Katzgraber, "Bond chaos in spin glasses revealed through thermal boundary conditions," Phys. Rev. B **93**, 224414 (2016).

[57] C. Amoruso, E. Marinari, O. C. Martin, and A. Pagnani, "Scalings of domain wall energies in two dimensional ising spin glasses," Phys. Rev. Lett. **91**, 087201 (2003).

[58] Kevin C. Young, Robin Blume-Kohout, and Daniel A. Lidar, "Adiabatic quantum optimization with the wrong hamiltonian," Phys. Rev. A **88**, 062314– (2013).

[59] Kristen L Pudenz, Tameem Albash, and Daniel A Lidar, "Error-corrected quantum annealing with hundreds of qubits," Nat. Commun. **5**, 3243 (2014).

[60] Kristen L. Pudenz, Tameem Albash, and Daniel A. Lidar, "Quantum annealing correction for random Ising problems," Phys. Rev. A **91**, 042302 (2015).

[61] Walter Vinci, Tameem Albash, Gerardo Paz-Silva, Itay Hen, and Daniel A. Lidar, "Quantum annealing correction with minor embedding," Phys. Rev. A **92**, 042310–(2015).

[62] Walter Vinci, Tameem Albash, and Daniel A Lidar, "Nested quantum annealing correction," Npj Quantum Information **2**, 16017 EP – (2016).