Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities

Zheng Zhu,[1] Andrew J. Ochoa,[1] Stefan Schnabel,[2] Firas Hamze,[3] and Helmut G. Katzgraber[1,4,5]

[1]*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA*
[2]*Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ), Universität Leipzig, Postfach 100920, D-04009 Leipzig, Germany*
[3]*D-Wave Systems, Inc., 3033 Beta Avenue, Burnaby, British Columbia, V5G 4M9, Canada*
[4]*Materials Science and Engineering Program, Texas A&M University, College Station, Texas 77843, USA*
[5]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

(Dated: January 19, 2016)

Recent tests performed on the D-Wave Two quantum annealer have revealed no clear evidence of speedup over conventional silicon-based technologies. Here, we present results from classical parallel-tempering Monte Carlo simulations combined with isoenergetic cluster moves of the archetypal benchmark problem—an Ising spin glass—on the native chip topology. Using realistic uncorrelated noise models for the D-Wave Two quantum annealer, we study the best-case resilience, i.e., the probability that the ground-state configuration is not affected by random fields and random-bond fluctuations found on the chip. We thus compute classical upper-bound success probabilities for different types of disorder used in the benchmarks and predict that an increase in the number of qubits will require either error correction schemes or a drastic reduction of the intrinsic noise found in these devices. We restrict this study to the exact ground state, however, the approach can be trivially extended to the inclusion of excited states if the success metric is relaxed. We outline strategies to develop robust, as well as hard benchmarks for quantum annealing devices, as well as any other (black box) computing paradigm affected by noise.

I. INTRODUCTION

Although a useful universal quantum computer [1, 2] is far from reality at the moment, the advent of quantum annealing (QA) machines based on quantum adiabatic optimization techniques [3–14] has sparked a small computing revolution in recent years. Being a novel hardware based on nonsilicon chips used to perform computations exploiting the potential advantages of quantum fluctuations [15], quantum annealing machines might affect the way a multitude of hard optimization problems are solved today.

The first somewhat useful programmable commercial devices that attempt to exploit this unique power are the D-Wave One and Two quantum annealers [16], that are designed to solve quadratic unconstrained binary optimization (QUBO) problems [17], such as finding the ground state of a disordered Ising spin-glass Hamiltonian, a well-known NP-hard problem in this general formulation [18]. Because many problems across disciplines can be mapped onto QUBOs, multiple studies of the D-Wave quantum annealer's performance, compared to some classical optimization approaches, such as simulated annealing (SA) [19], have been performed [20–31]. Tests [21, 22, 24, 27] by different research teams suggest that the D-Wave quantum annealer does benefit from quantum effects. However, it is unclear if this quantum advantage is involved in the optimization of cost functions. Furthermore, to date these studies reveal no clear evidence of limited quantum speedup [26] over classical optimization algorithms on traditional computers.

Recent work by Katzgraber *et al.* [32] suggests that current benchmarking approaches using spin glasses with uniformly-distributed disorder on the Chimera graph [33], such as bimodal or range-k, might not be the best benchmark problems in the quest for quantum speedup. In particular, Ref. [34] proposes an innovative approach based on insights from the study of spin glasses to design hard benchmark problems *within* the constraints of the D-Wave device. To overcome the limitations posed by the D-Wave architecture, Ref. [34] proposes to use instances with a unique ground state, as well as many metastable states. In this work we study the interplay between the generation of hard benchmark instances with the design of problems suitable for the D-Wave device that are robust to noise. Ideally, thus, a two-tier (unfortunately computationally-expensive) data mining approach is needed to produce ideal test instances for any quantum annealing device: First, random benchmark instances are mined for their desired properties (e.g., unique ground state) that make them hard problems to solve. Second, these instances are tested for their robustness to the intrinsic noise present in any hardware device.

The fact that different numerical studies [9, 10, 13, 35] demonstrated that QA might outperform SA in certain problems—especially those with rough energy landscapes—has motivated the authors of Refs. [34] and [36] to design tunable hard benchmarking problems. Reference [34] goes a step further, by being able to carefully tune the barrier thickness between dominant features in the energy landscape, thus putatively allowing for the detection of any quantum advantage that a quantum annealing device might pose over traditional optimization approaches. Despite these efforts, noise due to thermal excitations and control errors on qubits and couplers have a detrimental effect on the performance of the D-Wave quantum annealer [23, 37–40] that likely is masking any potential limited quantum speedup [26]. A simple explanation for these problems is given by the fragility of spin glasses to small perturbations, also known as chaotic effects [41–50] to either couplers (bond chaos), qubits via longitudinal fields (field chaos), or both couplers and qubits (temperature chaos).

Here, small fluctuations can produce large changes in the free energy of the system, thus perturbing the original problem Hamiltonian to be solved.

Although quantum error correction [23, 39, 40] can, in principle, mitigate these errors, it does so at a cost of needing multiple physical qubits to encode one logical qubit, thus reducing the effective system size of problems to be studied. This also means that "error-corrected" benchmark instances, while more robust to noise, will likely be too small to be in the scaling regime of interest for currently available system sizes. As such, designing hard benchmark instances that are robust to noise and require no overhead in the embedding to keep the problem size at a maximum are of utmost importance to detect quantum speedup. In this work we classically study *resilience*, i.e., the probability that the ground-state configuration is not affected by random fields and random-bond fluctuations found on the chip for different benchmark instance classes, by using realistic uncorrelated noise models for the D-Wave Two quantum annealer. Furthermore, we present strategies on how to develop hard benchmark instances that, at the same time, are robust to noise. Note that our methodology is generic, i.e., it can be applied to any architecture or noisy black-box optimization device. Furthermore, the study can be trivially extended to include low-lying excited states if the gold standard of finding the exact ground state is relaxed to include a subset of low-lying excited states.

The paper is structured as follows. In Sec. II we introduce the different benchmark instance classes studied, as well as the noise model. Furthermore, we describe the heuristic used to find the ground-state configurations. Our numerical results on the D-Wave chimera topology are presented in Sec. III, followed by concluding remarks.

II. MODEL, OBSERVABLES, ALGORITHM

Our calculations are for the currently-available D-Wave Two device [51]. However, the ideas can be generalized to any topology.

A. Model

The *native* benchmark for the D-Wave Two quantum annealer is an Ising spin glass [2, 52, 53] defined on the Chimera topology of the system [33]. The Hamiltonian of the problem to be optimized is given by

$$\mathcal{H} = -\sum_{\{i,j\}\in\mathcal{V}} J_{ij} s_i s_j - \sum_{i\in\mathcal{V}} s_i h_i \,. \tag{1}$$

where $s_i \in \{\pm 1\}$ signify Ising spins on the vertices $\mathcal{V}$ of the Chimera lattice. Figure 1 shows a 512 qubit Chimera lattice with 8×8 $\mathrm{K}_{4,4}$ cells. In addition, each spin s_i is coupled to a local random field h_i. The sum is over all edges $\mathcal{E}$ connecting vertices $\{i,j\} \in \mathcal{V}$.. The interactions J_{ij} between the spins are drawn from carefully chosen, discrete disorder distributions within the hardware constraints of the D-Wave Two architecture.

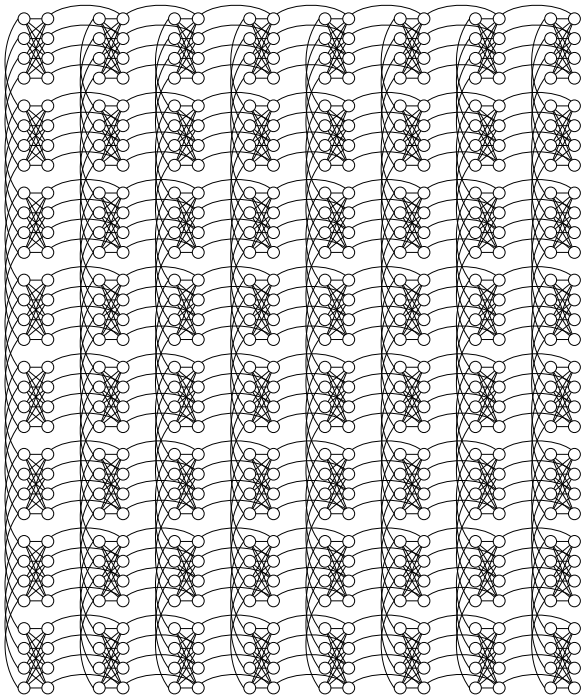

FIG. 1: Adjacency matrix of the D-Wave Two chip with 8×8 $\mathrm{K}_{4,4}$ cells and 512 qubits (circles) connected by couplers (lines).

To emulate the effects of thermal noise in the device, we perturb the discrete values of the couplers J_{ij} by a random amount ΔJ_{ij} drawn from a Gaussian distribution with zero mean and standard deviation ΔJ. For simplicity, we assume the noise is quenched and uncorrelated. This "white noise" represents a realistic (classical) noise model for coupler fluctuations that is typically used to study the effects of noise in electronic devices, as well as telecommunications. Although the qubit noise in the D-Wave Two device is closer to $1/f$ noise with a "pink" power spectrum, for simplicity we couple the individual qubits to uncorrelated quenched random fields drawn from a Gaussian distribution with zero mean and standard deviation h. We do not expect this simplification to qualitatively change our results.

B. Instance classes & observables

Carefully-chosen interactions between the spins determine the hardness and robustness of instance classes [34]. To develop hard instances, multiple requirements have to be fulfilled. First, it is of paramount importance to ensure that the instances have a *unique* ground-state configuration that minimizes the cost function in Eq. (1). Furthermore, it is desirable to have dominant metastable states such that the system is easily trapped – a process that can be accomplished by a post-processing selection and mining of the data based on insights from the study of the dynamics of spin glasses using classical simulation techniques [34, 54]. Ultimately, an ideal

benchmark instance is robust to noise, has a unique ground state and, ideally, many metastable states.

To gauge the fraction of unique ground-state configurations for a particular instance class, we define a quantity we call *yield* ($\mathcal{Y}$), i.e.,

$$\mathcal{Y} = N_{\rm unique}/N_{\rm total}. \quad (2)$$

In Eq. (2) $N_{\rm total}$ is the total number of *randomly-generated* instances for that particular instance class and $N_{\rm unique}$ is the number of instances featuring a unique ground state (no degeneracy).

One simple approach pioneered in Ref. [34] to design instance classes with high yield, is to ensure that as few qubits s_i as possible have zero local fields $\mathcal{F}_i = \sum_{j\neq i} J_{ij}s_j + h_i$. If for a given qubit $\mathcal{F}_i \equiv 0$, then the qubit's value does not change the energy of the system. Therefore, if a system with N qubits has k free qubits with zero local field, the degeneracy of the ground state is increased by a factor 2^k. We have exhaustively computed the probability that a particular combination of two, three, or four integer values [55] in the range $\{\pm 1, \ldots, \pm i_{\rm max}\}$ (with $i_{\rm max} = 28$) [56] for the couplers J_{ij} on the Chimera topology yields the smallest fraction of qubits with zero local fields. Furthermore, we have attempted to "spread out" the integers as much as possible in the range $[-1, 1]$ after a normalization of the coupler values with $i_{\rm max}$. In addition to the previously-studied cases of bimodal disorder, i.e.,

$$\mathrm{U}_1 \in \{\pm 1\},$$

as well as uniform range-k disorder with $k = 4$ [21, 26]

$$\mathrm{U}_4 \in \{\pm 1, \pm 2, \pm 3, \pm 4\},$$

we also study Sidon-type instances [34, 57], namely

$$\mathrm{U}_{5,6,7} \in \{\pm 5, \pm 6, \pm 7\},$$

which are similar to uniform range-7 instances, however only the three largest integers that form a Sidon set are kept. Finally, we study a larger Sidon set

$$\mathrm{S}_{28} \in \{\pm 8, \pm 13, \pm 19, \pm 28\}.$$

The $\mathrm{U}_{5,6,7}$ and S_{28} Sidon instance classes reduce the probability of zero local fields drastically by design, and thus maximize the yield of unique ground states. In fact, while U_1 has an average probability of 23% to have zero local fields, this number is reduced to 6% in the U_4 class. $\mathrm{U}_{5,6,7}$ has only 4.5% zero local fields and S_{28} has 1.5%.

To increase the resilience to noise for a given instance, one has to maximize the change in energy when flipping a spin, i.e., the minimum classical energy gap. Ideally, this change in energy should be considerably larger than the typical noise fluctuations to prevent qubit errors. For Ising spins, this energy gap is given by $\Delta E = 2/i_{\rm max}$, where $i_{\rm max}$ is the largest integer in the unnormalized bond distribution. For example, $\Delta E(\mathrm{U}_1) = 2$, whereas $\Delta E(\mathrm{U}_4) = 1/2$, $\Delta E(\mathrm{U}_{5,6,7}) = 2/7$, and $\Delta E(\mathrm{S}_{28}) = 1/14 \sim 0.07$. For the current D-Wave Two machine with 512 qubits, coupler fluctuations are typically ~ 0.035 if the bonds are normalized to unity ("autoscaling mode"). This means that in this case the S_{28} instance class pushes the limits of the machine because $\Delta E(\mathrm{S}_{28}) \sim 2\Delta J$.

To quantify the robustness of ground-state configurations to noise, we define the resilience R of an instance to be

$$R = N_{\rm same}/N_{\rm trials} \quad (3)$$

where $N_{\rm same}$ is the number of trials with different random noise perturbations (either fields or bonds) that do not change the original ground-state configurations. We perform $N_{\rm trials} = 10$ trials (or gauges) to compute R. The resilience of an instance class is the resilience for each instance R averaged over the bond disorder, i.e., $\mathcal{R} = [R]_{\rm av}$, where $[\cdots]_{\rm av}$ represents an average over multiple random bond configurations. A preference should be given to whole instance classes with high resilience. However, individual instances that are unaffected by the perturbations are also robust instances and can be used for benchmarking purposes. Conversely, to study the effects of noise in quantum annealing machines and how to reduce these, instances with a *small* resilience can also be mined [29].

Finally, we emphasize that a "relaxed" resilience R_k can also be defined, where

$$R_k = N_{\rm same}(E \leq E_k)/N_{\rm trials}. \quad (4)$$

Here $N_{\rm same}(E \leq E_k)$ is the number of times a state with an energy E less or equal than the energy of the k-th excited state is found. This is of importance when the analog machine suffers from high noise levels and where the determination of the exact ground state is difficult or even impossible. For discrete disorder distributions—as commonly used on quantum annealing machines with finite precision—the energy levels are separated by well-defined values, i.e., computing the relaxed resilience of an instance class, $\mathcal{R}_k = [R_k]_{\rm av}$, is well defined.

C. Algorithm details

In order to measure the yield and resilience of a particular instance class, ground states of instances from all instance classes have to be found. We apply a heuristic method that uses the parallel tempering Monte Carlo algorithm [58] combined with isoenergetic cluster moves [59] to speed up the thermalization. Simulation parameters are listed in Table I and thermalization has been determined by a logarithmic binning of the data. Once the last three bins agree within error bars, we deem the system to be in thermal equilibrium. The detailed algorithm to detect ground states was first introduced in Ref. [60]. However, to increase the accuracy of our heuristic, here four instead of two copies of the system with the same disorder are simulated with *independent* Markov chains. We perform $N_{\rm sw}$ updates [61]. For $N_{\rm sw}/8$ updates we keep track of the lowest energy E of each Markov chain at the lowest temperature simulated. If $E^{(1)} = E^{(2)} = E^{(3)} = E^{(4)}$, it is very likely the ground state energy E_0 has been found. For the remaining number of updates we keep statistics of the

TABLE I: Simulation parameters: For each instance class and system size N, we compute $N_{\rm sa}$ instances. $N_{\rm sw} = 2^b$ is the total number of Monte Carlo sweeps for each of the $4N_T$ replicas for a single instance, $T_{\rm min}$ [$T_{\rm max}$] is the lowest [highest] temperature simulated, and N_T is the number of temperatures used in the parallel tempering method. For the lowest $N_{\rm icm}$ temperatures isoenergetic cluster moves are applied.

Class	N	$N_{\rm sa}$	b	$T_{\rm min}$	$T_{\rm max}$	N_T	$N_{\rm icm}$
U_1	512	900	19	0.150	3.050	30	13
U_4	512	900	19	0.150	3.000	30	14
$U_{5,6,7}$	128	900	19	0.150	3.000	30	14
$U_{5,6,7}$	288	900	19	0.150	3.000	30	14
$U_{5,6,7}$	512	900	19	0.150	3.000	30	14
$U_{5,6,7}$	800	900	19	0.150	3.000	30	14
$U_{5,6,7}$	1152	900	19	0.150	3.000	30	14
S_{28}	512	900	19	0.150	3.000	30	14

configurations that minimize the Hamiltonian and thus estimate the degeneracy distribution of the ground state. However, there is no guarantee that any solution obtained by this heuristic method is the true optimum, or that we have found all configurations that minimize the Hamiltonian. Fortunately, for the Sidon-type instance classes the degeneracy is small by construction. Therefore, it is likely that we found all ground-state configurations. Once the ground-state configurations of all instances have been found, the average yields for different instance classes can be computed.

In addition to the effects of the minimum energy gap ΔE on the resilience for each instance class, we also consider the effects of the number of first excited states on the resilience. To estimate the number of first excited states, for the remaining $(7/8)N_{\rm sw}$ sampling updates we also keep track of all configurations that have an energy $E_1 = E_0 + \Delta E$.

III. RESULTS

We focus only on the resilience of the exact ground state in this study for the sake of brevity and to illustrate the developed methodology. Our approach is easily extended to include low excited states. Note that if the resilience R of an instance is large, we also expect the relaxed instance resilience R_k to be large for small enough k.

A. Yield of non-degenerate ground states

For the current D-Wave Two architecture with 512 qubits, the yield of unique ground states is strongly dependent on the instance class (disorder between spins) used. When the disorder is drawn from a bimodal distribution (U_1) the yield in all our experiments was exactly 0%. Surprisingly, uniform range-4 instances (U_4) also have $\mathcal{Y} = 0\%$. However, by increasing the range of the integers and selecting them from a Sidon set while removing the lowest values gives $\mathcal{Y} = 4.5(4)\%$ for the $U_{5,6,7}$ class. Although a small fraction, it is clearly nonzero. Finally, for the large Sidon set S_{28} we obtain a fraction $\mathcal{Y} = 20.0(6)\%$ of unique ground states [62], i.e., optimal for large-scale benchmarking.

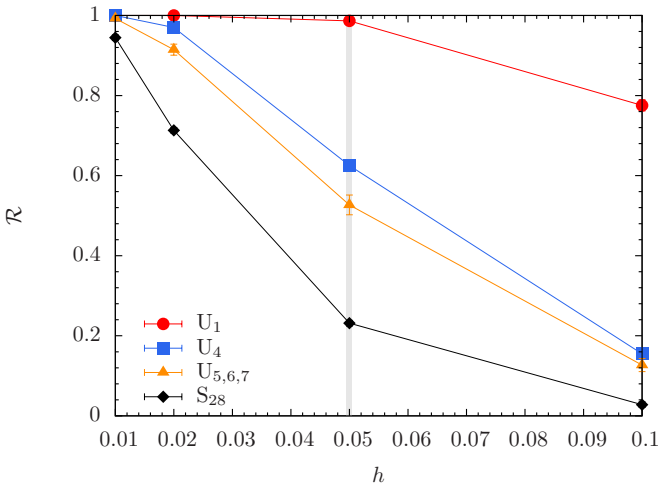

FIG. 2: (Color online) Resilience ($\mathcal{R}$) of different instance classes (see text) for a $N = 512$ qubit system on the Chimera graph as a function of Gaussian random field strength (h). Instance classes are less resilient to noise with increasing field strength and decreasing classical energy gap. The shaded line represents the current field noise strength of approximately 5% in the D-Wave Two system.

B. Resilience to noise

Figure 2 shows the resilience to random-field noise for different instance classes. As the typical field strength h increases, the resilience $\mathcal{R}$ for all instance classes decreases. This is to be expected, because the energy spread due to the splitting of degenerate excited states via the random fields results in more energy levels crossing. Furthermore, for a fixed field strength, instance classes with small energy gaps ΔE tend to have lower resilience. This is to be expected: it is easier for split states to have a lower energy than the original ground state when the gap is small. Note that while instance classes $U_{5,6,7}$ and U_4 have a similar resilience, the yield of unique ground states is considerably higher for $U_{5,6,7}$, i.e., a careful design of the spin-spin interactions is key when attempting to benchmark a quantum annealing device.

Figure 3 shows the resilience of different instance classes as a function of different typical coupler perturbations ΔJ. Again, for all instance classes studied, the resilience decreases as fluctuations increase. In addition, instance classes with small energy gaps have a lower resilience. It is important to note that bond noise has a stronger impact on the resilience than field noise. Considering each qubit has typically ~ 6 neighbors in the Chimera lattice, the impact of bond noise is amplified by multiple connections of qubits. Therefore, reducing the fluctuations of the couplers is more important than dealing with the intrinsic flux noise of each qubit.

Unfortunately, for the D-Wave architecture, to find an instance class that is both hard and robust to noise, compromise has to be made. The U_1 instance class has the highest

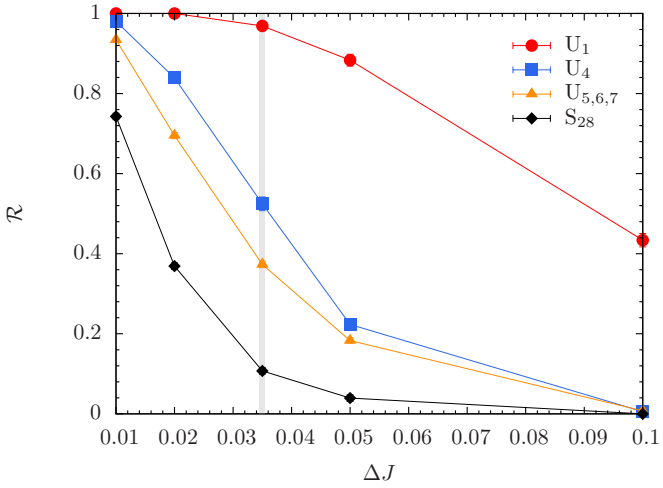

FIG. 3: (Color online) Resilience ($\mathcal{R}$) of different instance classes (see text) for a $N = 512$ qubit system on the Chimera graph as a function Gaussian random bond fluctuation strength (ΔJ). Instance classes are less resilient to noise with increasing bond fluctuation strength and a decreasing classical energy gap. The shaded line represents the current bond noise strength in the D-Wave Two system, i.e., $\sim 3.5\%$. Note that bond noise has a stronger effect than field noise (Fig.2) on the device.

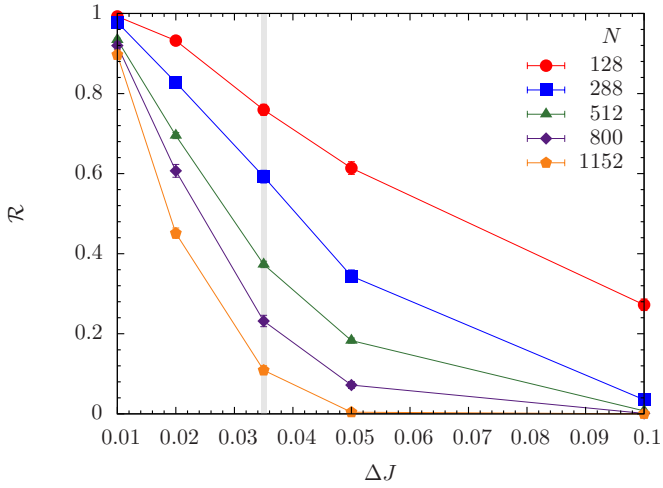

FIG. 4: (Color online) Resilience $\mathcal{R}$ of the $U_{5,6,7}$ instance class as a function of the bond fluctuation strength (ΔJ) for different system sizes N on the Chimera topology. The resilience clearly decreases for increasing noise and system size. The shaded vertical line represents the current bond-noise strength in the D-Wave Two system, approximately 3.5%.

resilience to noise, however, the huge ground-state degeneracy makes it easier for classical algorithms such as SA to find minimum-energy configurations [32, 34]. On the flip side, the Sidon instance class is known to be hard [34] and produces many unique ground states, but its resilience is comparably low due to the small energy gap. A compromising natural choice would therefore be to either use the U_4 or $U_{5,6,7}$ instance classes. However, while the resilience for both U_4 and $U_{5,6,7}$ are comparable, the yield of unique ground states needed to construct hard benchmark problems is much higher for $U_{5,6,7}$. We thus conclude that for the current Chimera topology, the $U_{5,6,7}$ instance class is the optimal compromise to design hard benchmark problems within the D-Wave Two architecture constraints. For the remainder of this paper we thus focus on this particular instance class.

Figure 4 shows the resilience of the $U_{5,6,7}$ instance class for different system sizes N of the Chimera lattice as a function of the random-bond fluctuation strength ΔJ. Clearly, for increasing system size the resilience $\mathcal{R}$ decreases (larger system sizes typically have a higher degeneracy, therefore level crossings are more common than with smaller systems). This means that to scale up the system size of the D-Wave Two—or any other quantum annealing device—in the future, a much more precise control over the device's noise and/or the implementation of error correction schemes [23, 39, 40] are imperative.

We conclude this section by quoting results specifically calculated for the current D-Wave Two (512 theoretical qubits) and the next-generation D-Wave 2X (1152 theoretical qubits) machines when both errors in the couplers and qubits are applied, using the real values provided by D-Wave, Inc. [63]. For the D-Wave Two machine with 512 qubits, $\Delta J = 3.5\%$ and $\Delta h = 5\%$. Applying both coupler and qubit perturbations yields an average resilience of $\mathcal{R} = 0.22(2)$. For the next-generation D-Wave 2X device noise levels have been reduced, i.e., $\Delta J = 2.5\%$ and $\Delta h = 3\%$. This results in $\mathcal{R} = 0.21(3)$. We point out two interesting facts: First, it seems that the resilience for both coupler and qubit noise is approximately the product of the resilience of only noise being considered on the couplers with the resilience of only noise being considered on the qubits. Thus, as a rule of thumb and to obtain an approximate estimate for the combined effects, the individual numbers can be multiplied. Second, despite the lower noise level of the next-generation device, the resilience remains approximately unchanged within error bars. It seems that the increased number of qubits cancels out the additional precision.

C. Effects of the number of first excited states

Figure 5 shows the resilience $\mathcal{R}$ of the $U_{5,6,7}$ instance class as a function of the degeneracy of the first excited state on the Chimera topology with $N = 512$ spins. The higher the degeneracy of the first excited state, the lower the resilience. This can be explained by the increased probability of level crossing. We also color coded each dot in the figure: The heat map represents the number of instances that had a given degeneracy N_1 of the first excited state out of the 900 simulated. In this case, the bulk of the instances have between 4 and 8 degenerate first excited states. This results in a reduction of the resilience, compared to instances that contain only one or two first excited states.

While instances with only one or two first excited states are extremely rare, the effort needed to find these might outweigh the approximately 30% in the resilience reduction by allowing

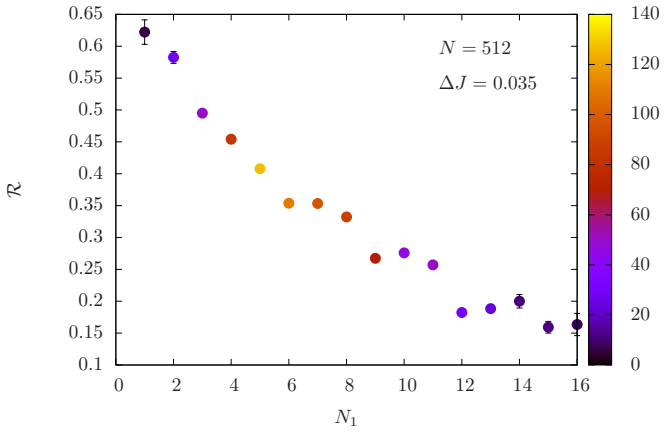

FIG. 5: (Color online) Resilience $\mathcal{R}$ as a function of the number of first excited states N_1 for $N = 512$ spins on the Chimera lattice. The data are for the $U_{5,6,7}$ instance class. The color bar shows approximately how often a given number of first excited states occurs for the 900 instances studied. In this case, between four and eight first excited states are most common.

states with three to four first excited states. We thus recommend to fix the number of first excited states to be less than or equal to four in this case.

We have also computed the Hamming distance between the ground state and all first excited states for a given instance. Our results suggest that when the average Hamming distance is small, the resilience to noise is higher. A simple explanation is that both ground-state and excited configurations are quite similar and therefore the noise affects them comparably, i.e., both the ground state and the first excited states are shifted approximately by the same amount when the Hamming distance is small.

IV. CONCLUSIONS

In order to develop both hard and robust benchmark instances, we have tested different instance classes by computing their yield (fraction of instances with a unique ground-state configuration) and resilience to noise fluctuations. Ideally, hard instances (high yield) with a high resilience are optimal for benchmarking purposes. Both yield and resilience can be tuned by a careful design of the instance classes—within the hardware restrictions of the machine—followed by a mining of the data. Although the numerical effort to do such "designer instances" is nonnegligible, we think this is a key ingredient in designing good benchmarks for quantum annealing devices, as well as any other computing architectures. It seems that both resilience and yield for the Chimera topology are slightly anticorrelated. A good compromise is thus the $U_{5,6,7}$ instance class where $J_{i,j} \in \{\pm 5, \pm 6, \pm 7\}$ that has a good resilience to both field and coupler noise, as well as a nonzero yield of unique ground states, with a small number of first excited states.

We emphasize that our results for the resilience represent a *best-case scenario* for any quantum annealing machine. Any other source of error can only decrease the success probabilities further. However, it could be that the introduction of carefully-crafted correlations between bond and field noise might reduce the errors and increase the resilience. Bond noise is the most limiting issue for the current D-Wave Two quantum annealer and is highly dependent on the connectivity of the graph. While it is desirable to have a high connectivity to be able to embed interesting problems on any putative architecture, one has to also keep in mind that noise levels should be far lower than in the current D-Wave machine.

This classical study of both resilience and yield plays an important role in the design of future adjacency matrices for quantum annealing machines, as well as the study of strategies to reduce noise in quantum annealers. Our results and methods can easily be generalized to other systems and thus should be of general interest when designing hard instance problems that attempt to circumvent the limitations of current hardware. Furthermore, calibration of future generations of the D-Wave device should be improved to allow for the encoding of more complex Sidon sets and thus the design of harder benchmark problems. Similarly, although the main goal of this work is to produce problems that are robust to noise, the methodology presented can be used to design tailored instances that are particularly sensitive to noise. This could play an important role when designing approaches to better calibrate devices, as done in Ref. [64]. Finally, we emphasize that if either noise is large or the instances produced are too difficult to minimize, a relaxed resilience that includes low-lying excited states can be defined.

Acknowledgments

We would like to thank M. H. Amin, P. Bunyk,T. M. Lanting, A. Perdomo-Ortiz and H. Muñoz-Bauza for fruitful discussions. We also thank H. Muñoz-Bauza for assistance. H. G. K. acknowledges support from the National Science Foundation (Grant No. DMR-1151387) and would like to thank the makers of the Dish-Wash Dos for motivation on this project. He also would like to especially thank P. Bunyk and T. M. Lanting for explaining multiple details related to the D-Wave Two device. H. G. K. and S. S. would like to thank the European Commission through the IRSES network DIONICOS under Contract No. PIRSES-GA-2013-612707 (FP7-PEOPLE-2013-IRSES). We would like to thank the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources (Lonestar Linux Cluster) and Texas A&M University for access to their Eos and Ada clusters. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. 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 ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[2] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, New York, 2001).

[3] 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).

[4] T. Kadowaki and H. Nishimori, *Quantum annealing in the transverse Ising model*, Phys. Rev. E **58**, 5355 (1998).

[5] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aepli, *Quantum annealing of a disordered magnet*, Science **284**, 779 (1999).

[6] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, *Quantum Computation by Adiabatic Evolution* (2000), arXiv:quant-ph/0001106.

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

[8] G. Santoro, E. Martoňák, R. Tosatti, and R. Car, *Theory of quantum annealing of an Ising spin glass*, Science **295**, 2427 (2002).

[9] A. Das and B. K. Chakrabarti, *Quantum Annealing and Related Optimization Methods* (Edited by A. Das and B.K. Chakrabarti, Lecture Notes in Physics 679, Berlin: Springer, 2005).

[10] G. E. Santoro and E. Tosatti, *TOPICAL REVIEW: Optimization using quantum mechanics: quantum annealing through adiabatic evolution*, J. Phys. A **39**, R393 (2006).

[11] D. A. Lidar, *Towards Fault Tolerant Adiabatic Quantum Computation*, Phys. Rev. Lett. **100**, 160506 (2008).

[12] A. Das and B. K. Chakrabarti, *Quantum Annealing and Analog Quantum Computation*, Rev. Mod. Phys. **80**, 1061 (2008).

[13] S. Morita and H. Nishimori, *Mathematical Foundation of Quantum Annealing*, J. Math. Phys. **49**, 125210 (2008).

[14] S. Mukherjee and B. K. Chakrabarti, *Multivariable optimization: Quantum annealing and computation*, Eur. Phys. J. Special Topics **224**, 17 (2015).

[15] D. R. Simon, *On the Power of Quantum Computation*, SIAM J. Comp. **26**, 116 (1994).

[16] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, et al., *Quantum annealing with manufactured spins*, Nature **473**, 194 (2011).

[17] A. Lucas, *Ising formulations of many NP problems*, Front. Physics **12**, 5 (2014).

[18] F. Barahona, *On the computational complexity of Ising spin glass models*, J. Phys. A **15**, 3241 (1982).

[19] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Optimization by simulated annealing*, Science **220**, 671 (1983).

[20] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar, *Experimental signature of programmable quantum annealing*, Nat. Commun. **4**, 2067 (2013).

[21] S. Boixo, T. F. Rønnow, 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).

[22] L. Wang, T. F. Rønnow, S. Boixo, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, *Comment on: "Classical signature of quantum annealing"* (2013), (arxiv:quant-phys/1305.5837).

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

[24] W. Vinci, T. Albash, A. Mishra, P. A. Warburton, and D. A. Lidar, *Distinguishing classical and quantum models for the D-Wave device* (2014), (arXiv:1403.4228).

[25] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, *How "Quantum" is the D-Wave Machine?* (2014), (arXiv:1401.7087).

[26] 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).

[27] J. Smolin and G. Smith, *Classical signature of quantum annealing*, Frontiers in Physics **2**, 52 (2014).

[28] G. Smith and J. Smolin, *Putting "Quantumness" to the Test*, Physics **6**, 105 (2013).

[29] D. Venturelli, S. Mandrà, S. Knysh, B. O'Gorman, R. Biswas, and V. Smelyanskiy, *Quantum Optimization of Fully Connected Spin Glasses*, Phys. Rev. X **5**, 031040 (2015).

[30] T. Albash, W. Vinci, A. Mishra, P. A. Warburton, and D. A. Lidar, *Consistency Tests of Classical and Quantum Models for a Quantum Device*, Phys. Rev. A **91**, 042314 (2015).

[31] T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, *Reexamining classical and quantum models for the D-Wave One processor*, Eur. Phys. J. Spec. Top. **224**, 111 (2015).

[32] H. G. Katzgraber, F. Hamze, and R. S. Andrist, *Glassy Chimeras Could Be Blind to Quantum Speedup: Designing Better Benchmarks for Quantum Annealing Machines*, Phys. Rev. X **4**, 021008 (2014).

[33] P. Bunyk, E. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, and J. Whittaker, *Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor*, IEEE Trans. Appl. Supercond. **24**, 1 (2014).

[34] H. G. Katzgraber, F. Hamze, Z. Zhu, A. J. Ochoa, and H. Munoz-Bauza, *Seeking Quantum Speedup Through Spin Glasses: The Good, the Bad, and the Ugly*, Phys. Rev. X **5**, 031026 (2015).

[35] R. D. Somma, D. Nagaj, and M. Kieferová, *Quantum Speedup by Quantum Annealing*, Phys. Rev. Lett. **109**, 050501 (2012).

[36] 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).

[37] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez, et al., *Thermally assisted quantum annealing of a 16-qubit problem*, Nat. Commun. **4**, 1903 (2013).

[38] T. Lanting, A. J. Przybysz, A. Y. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, et al., *Entanglement in a quantum annealing processor*, Phys. Rev. X **4**, 021041 (2014).

[39] K. L. Pudenz, T. Albash, and D. A. Lidar, *Quantum Annealing Correction for Random Ising Problems*, Phys. Rev. A **91**, 042302 (2015).

[40] R. R. Correll, *An Efficient User-Side Nulling Calibration for Quantum Annealing Computers* (2015), (arXiv:1503.00700).

[41] S. R. McKay, A. N. Berker, and S. Kirkpatrick, *Spin-Glass Behavior in Frustrated Ising Models with Chaotic Renormalization-Group Trajectories*, Phys. Rev. Lett. **48**, 767 (1982).

[42] A. J. Bray and M. A. Moore, *Chaotic Nature of the Spin-Glass Phase*, Phys. Rev. Lett. **58**, 57 (1987).

[43] I. Kondor, *On chaos in spin glasses*, J. Phys. A **22**, L163 (1989).

[44] F. Ritort, *Static chaos and scaling behavior in the spin-glass phase*, Phys. Rev. B **50**, 6844 (1994).

[45] M. Ney-Nifle and A. P. Young, *Chaos in a two-dimensional Ising spin glass*, J. Phys. A **30**, 5311 (1997).

[46] M. Ney-Nifle, *Chaos and universality in a four-dimensional spin glass*, Phys. Rev. B **57**, 492 (1998).

[47] A. Billoire and E. Marinari, *Evidence against temperature chaos in mean-field and realistic spin glasses*, J. Phys. A **33**, L265 (2000).

[48] A. Billoire and E. Marinari, *Overlap among states at different temperatures in the SK model*, Europhys. Lett. **60**, 775 (2002).

[49] M. Sasaki, K. Hukushima, H. Yoshino, and H. Takayama, *Temperature Chaos and Bond Chaos in Edwards-Anderson Ising Spin Glasses: Domain-Wall Free-Energy Measurements*, Phys. Rev. Lett. **95**, 267203 (2005).

[50] H. G. Katzgraber and F. Krzakala, *Temperature and Disorder Chaos in Three-Dimensional Ising Spin Glasses*, Phys. Rev. Lett. **98**, 017201 (2007).

[51] See http://www.dwavesys.com.

[52] K. Binder and A. P. Young, *Spin Glasses: Experimental Facts, Theoretical Concepts and Open Questions*, Rev. Mod. Phys. **58**, 801 (1986).

[53] D. L. Stein and C. M. Newman, *Spin Glasses and Complexity*, Primers in Complex Systems (Princeton University Press, 2013).

[54] B. Yucesoy, J. Machta, and H. G. Katzgraber, *Correlations between the dynamics of parallel tempering and the free-energy landscape in spin glasses*, Phys. Rev. E **87**, 012104 (2013).

[55] We restricted the search to four integer values because of the hardware restrictions on the D-Wave Two machine that require the interactions to be well separated in the interval $[0, 1]$.

[56] We chose $i_{\max} = 28$ empirically because it allowed us to encode very hard problems with many unique ground states.

[57] S. Sidon, *Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen*, Mathematische Annalen **106**, 536 (1932).

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

[59] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber, *Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension*, Phys. Rev. Lett. **115**, 077201 (2015).

[60] H. G. Katzgraber and A. P. Young, *Monte Carlo studies of the one-dimensional Ising spin glass with power-law interactions*, Phys. Rev. B **67**, 134410 (2003).

[61] One update corresponds to one Monte Carlo lattice sweep (N attempted spin updates), followed by one isoenergetic cluster move (for $T \lesssim 1$) and a parallel tempering update.

[62] Error bars are computed using a bootstrap analysis.

[63] Numbers quoted from the presentation of Mark Johnson (D-Wave, Inc.), "*Next Generation Quantum Annealing Processor*" held at AQC 2015 – Fourth Conference in Adiabatic Quantum Computing, June 29, 2015, Zurich, Switzerland.

[64] A. Perdomo-Ortiz, B. O'Gorman, J. Fluegemann, R. Biswas, and V. N. Smelyanskiy, *Determination and correction of persistent biases in quantum annealers* (2015), (arXiv:quant-phys/1503.05679).