Quantum annealing with more than one hundred qubits

Sergio Boixo,[1] Troels F. Rønnow,[2] Sergei V. Isakov,[2] Zhihui Wang,[3] David Wecker,[4] Daniel A. Lidar,[5] John M. Martinis,[6] and Matthias Troyer*[2]

[1] *Information Sciences Institute and Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA*
[2] *Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland*
[3] *Department of Chemistry and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
[4] *Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA*
[5] *Departments of Electrical Engineering, Chemistry and Physics, and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
[6] *Department of Physics, University of California, Santa Barbara, CA 93106-9530, USA*

arXiv:1304.4595v1 [quant-ph] 16 Apr 2013

At a time when quantum effects start to pose limits to further miniaturisation of devices and the exponential performance increase due to Moore's law, quantum technology is maturing to the point where quantum devices, such as quantum communication systems [1], quantum random number generators [2] and quantum simulators [3], may be built with powers exceeding the performance of classical computers. A quantum annealer [4–6], in particular, finds solutions to hard optimisation problems by evolving a known initial configuration towards the ground state of a Hamiltonian that encodes an optimisation problem. Here, we present results from experiments on a 108 qubit D-Wave One device based on superconducting flux qubits. The correlations between the device and a simulated quantum annealer demonstrate that the device performs quantum annealing: unlike classical thermal annealing it exhibits a bimodal separation of hard and easy problems, with small-gap avoided level crossings characterizing the hard problems. To assess the computational power of the quantum annealer we compare it to optimised classical algorithms. We discuss how quantum speedup could be detected on devices scaled to a larger number of qubits where the limits of classical algorithms are reached.

Annealing a material by slow cooling is an ancient technique to improve the properties of glasses, metals and steel that has been used for more than four millennia [7]. Mimicking this process in computer simulations is the idea behind *simulated annealing* as an optimisation method [8], which views the cost function of an optimisation problem as the energy of a physical system. Its configurations are sampled in a Monte Carlo simulation using the Metropolis algorithm [9], escaping from local minima by thermal fluctuations to find lower energy configurations. The goal is to find the global energy minimum (or at least a very low energy configuration) by slowly lowering the temperature and thus obtain the best (or a very good) solution to the optimisation problem.

Instead of escaping a local minimum by thermally crossing over energy barriers it can be more efficient to explore the state space *quantum mechanically*. In *simulated quantum annealing* [10, 11], one makes use of this effect by adding quantum fluctuations, which are slowly reduced – ultimately ending up in a low energy configuration of the optimisation problem. Simulated quantum annealing has been observed to be more efficient than thermal annealing for certain spin glass models [11]. Further speedup is expected in physical quantum annealing, either as an experimental technique for annealing a quantum spin glass [12], or – and this is what we will focus on here – as a computational technique in a programmable quantum device. The state of the quantum device evolves according to the laws of quantum mechanics and updates all variables simultaneously, which can be more efficient than local Monte Carlo updates of a path integral configuration performed in a simulated quantum annealer.

In this work we report on computer simulations and experimental tests on a D-Wave One device [13] in order to address central open questions about quantum annealers: is the device actually a quantum annealer, i.e. do the quantum effects observed on eight qubits [13, 14] persist when scaling problems up to more than 100 qubits, or do short coherence times turn the device into a classical, thermal annealer? Which problems are easy and which problems are hard for a quantum annealer, for a simulated classical annealer and for a simulated quantum annealer? How does the effort to find the ground state scale with problem size? Does the device have advantages over classical computers?

We consider the problem of finding the ground state (lowest energy configuration) of an Ising spin glass model with Hamiltonian

$$H_{\text{Ising}} = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i, \qquad (1)$$

with N binary variables $\sigma_i = \pm 1$. This problem is nondeterministic polynomially (NP) hard [15], meaning that it is at least as hard as the hardest problems in NP, a class which includes notoriously difficult problems such as traveling salesman, satisfiablity of logical formulas, and

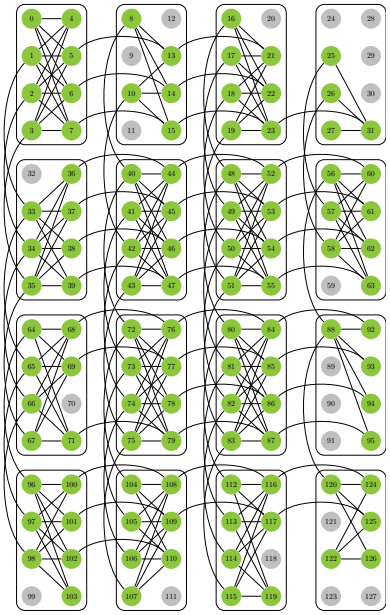

FIG. 1: **Qubits and couplers in the quantum annealer.** The D-Wave One Rainer chip consists of 4×4 unit cells of eight qubits, connected by programmable inductive couplers as shown by lines. Of the 128 qubits in the device, the 108 calibrated qubits shown in green have been used in the experiments. The qubits and couplers can be thought of as the vertices and edges, respectively, of a "Chimera" graph with a bipartite structure. The tree width of this graph is 17 for the ideal (128 qubit) Rainier chip and scales as $\sqrt{N}$ for an N-qubit Chimera graph [16, 17].

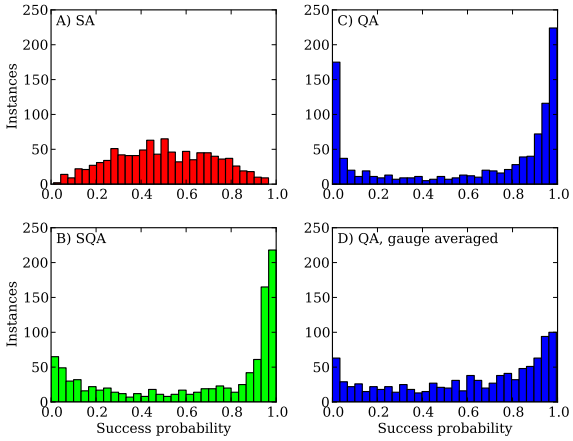

FIG. 2: **Histogram of success probabilities.** Shown are histograms of the probabilities of finding the ground states in 1000 different spin glass instances with $N = 108$ spins, solved by several computational methods. A) A classical simulated annealer shows a unimodal distribution. B) A simulated quantum annealer shows a bimodal distribution with easy and hard instances having high and low success probabilities, respectively. C) The D-Wave device also shows a bimodal distribution. D) Averaging over annealing with sixteen different encodings (gauge choice) retains a bimodal distribution.

factoring. It also implies that no efficient (polynomial time) algorithm to find these ground states is known and the computational effort of all existing algorithms scales exponentially with problem size as $\mathrm{O}(\exp(cN^a))$. While quantum mechanics is not expected to turn the exponential scaling into a polynomial one, the constants c and a can be smaller on quantum devices, potentially giving substantial speedup over classical algorithms.

To perform quantum annealing or adiabatic quantum optimisation (either simulated or in a device), we map each of the Ising variables σ_i to the z-component of a quantum spin-half variable (qubit) and add a transverse magnetic field in the x-direction to induce quantum fluctuations, thus obtaining the time-dependent Hamiltonian

$$H(t) = -A(t)\sum_i \sigma_i^x - B(t)\left[\sum_{i<j} J_{ij}\sigma_i^z\sigma_j^z + \sum_i h_i\sigma_i^z\right]. \quad (2)$$

Quantum annealing at finite temperature T starts in the limit of a strong transverse field and weak couplings $A(0) \gg \max(k_B T, B(0))$, with the system state close to the ground state of the transverse field term where all spins point in the positive x-direction. Decreasing the strength of the transverse field $A(t)$ and increasing the couplings $B(t)$, the system evolves towards the ground state of the optimisation problem when the end of the annealing schedule has been reached, where $B(t_{\mathrm{anneal}}) \gg A(t_{\mathrm{anneal}})$.

To remain adiabatically in the ground state, the evolution must be slow compared to the inverse energy gap between the ground state and the first excited state of the time dependent Hamiltonian $H(t)$. In hard optimisation problems, the smallest gaps of avoided level crossings are expected to close exponentially fast with increasing problem size [18, 19], implying an exponential scaling of the required annealing time with problem size.

We performed our experiments on a D-Wave One chip, a device comprised of superconducting flux qubits with programmable couplings. Of the 128 qubits on the device, the 108 qubits shown in figure 1 had been calibrated and were used in our experiments. Instead of trying to map specific optimisation problem to the connectivity graph of the chip [20, 21], our strategy has been to choose random (NP hard) spin glass problems that can directly be realised. For each coupler J_{ij} in the device we randomly assigned a value of either $+1$ or -1, giving rise to a very rough energy landscape. Local fields $h_i \neq 0$ give a bias to individual spins, tending to make the problem easier to solve for annealers. We thus set all $h_i = 0$ for most data shown here and provide data with local fields in the supplementary material. We performed experiments for problems of various sizes N, using $N = 8$ to $N = 108$ spins. For each problem size, we selected 1000 different *instances* of the spin glass by choosing 1000 sets

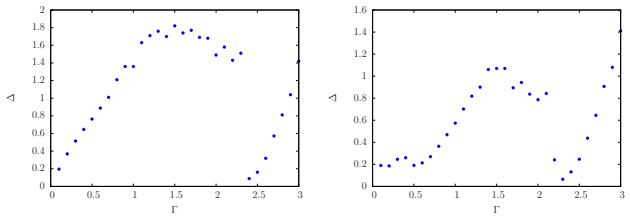

FIG. 3: **Evolution of the lowest spectral gap.** Shown are (upper bounds for) gaps between ground state and the lowest excited state for two typical spin glass instances of $N = 108$ spins as a function of the ratio of transverse field to coupling $\Gamma = A(t)/B(t)$. Note that Γ decreases during the annealing schedule. In the left panel we show the gap (in units of the Ising couplings) for an "easy" instance with success probability 98% and on the right for a "hard" instance with success probability 8%.

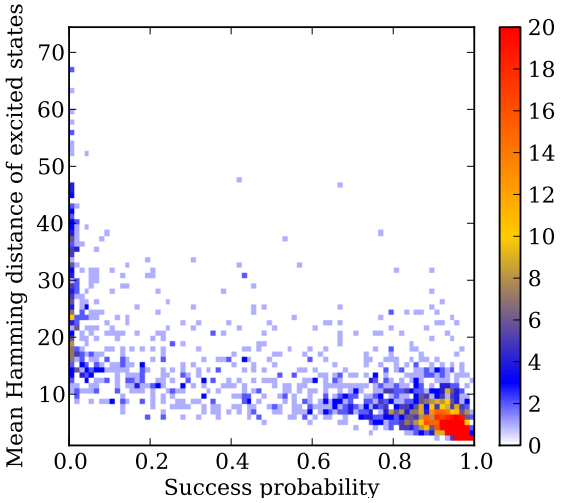

FIG. 4: **Correlation between success probability and Hamming distance of excited states for the D-Wave device.** Shown is a scatter plot of correlation between the success probabilities for $N = 108$ spin instances with local random fields, and the mean Hamming distance of excited states found to the closest ground state. The color scale indicates how many of the instances are found in a pixel of the plot. Easy (hard) instances tend to have a small (large) Hamming distance.

of different random couplings $J_{ij} = \pm 1$ (and for some of the data also random fields $h_i = \pm 1$). For each of these instances, we performed $M = 1000$ annealing *runs*, counted the number of runs M_{GS} in which the ground state was reached, and determined the success probability as $s = M_{GS}/M$.

We then plot a histogram of s for 1000 instances and compare to results of simulated classical and quantum annealing on the same instances (see Methods for details on these algorithms), to obtain the first key results of our study in figure 2. The simulated classical annealer shows a unimodal distribution, with a peak position that moves as one changes the annealing time (see supplementary material). In contrast, both the the D-Wave device and the simulated quantum annealer exhibit a bimodal distribution, with a clear split into easy and hard instances. Increasing the annealing time in the simulated quantum annealer makes the bimodal distribution more pronounced (see supplementary material). From the distinct differences between the success probability histograms of the simulated classical and quantum annealer, and the similarity between the device and the simulated quantum annealer, we tentatively conclude that the device does not perform classical but quantum annealing.

Before we firm up this conclusion by investigating correlations we explain the bimodal distribution, *i.e.*, the origin of "hard" (low success probability) and "easy" (high success probability) instances. We picked five instances that are hard for both the device and the simulated quantum annealer, and five instances that are easy for both, and performed extensive QMC simulations to estimate the spectral gap between the ground state and the first excited state. A representative result of one easy and one hard instance is shown in figure 3, results for the other instances are shown in the supplementary material. For all instances, we found that the gap trivially closes around a ratio $\Gamma = A(t)/B(t) \approx 2.3$ of transverse field to Ising coupling, related to a global Z_2 spin inversion symmetry. The gap also closes towards the end of the schedule as $\Gamma \to 0$, when multiple states are expected to become degenerate ground states. Neither of these small gaps has a detrimental effect on finding the ground state, since even after choosing the wrong branch at these avoided level crossings (either by thermalisation or diabatic transitions) the system still ends up in a ground state at the end of the annealing. The hard instances, however, typically have additional avoided level crossings with small gaps as is seen at $\Gamma \approx 0.5$ in the right panel of figure 3. These additional avoided level crossings cause failures of the annealing due to transitions to higher energy states, thus making the problem "hard". An explanation of the origin of small gap avoided level crossings for the hardest instances is presented in the supplementary material.

Further confirmation for the "hard" instances being due to small gap avoided level crossings comes from investigating the excited states found by the quantum annealer. In figure 4 we show a scatter plot of the mean Hamming distance of excited states versus success probability. The Hamming distance is the number of spins that need to be flipped to reach the closest ground state. We find that for the "easy" instances with high success probability the Hamming distance is typically small. The associated spin flips are often due to thermal errors that can easily be corrected (with linearly scaling effort) by checking if the energy can be reduced by flipping a small number of spins. Such an error correction scheme and its effectiveness is presented in the supplementary material. The "hard" instances on the other hand typically result in excited states with a large Hamming distance. This

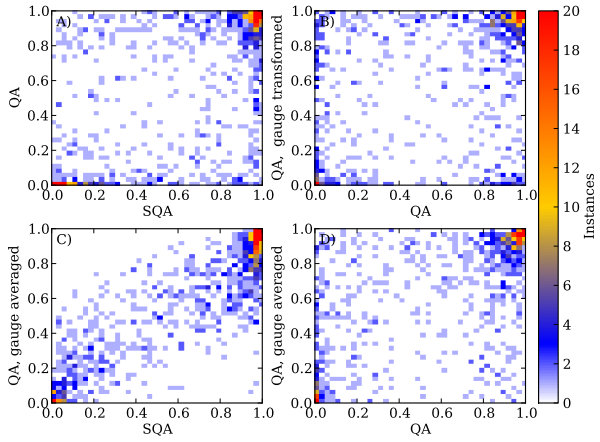

FIG. 5: **Correlations.** Shown are scatter plots of correlations of the success probabilities of 1000 instances of $N = 108$ spins. The color scale indicates how many of the instances are found in a pixel of the plots. A) between a simulated quantum annealer (SQA) and the quantum device (QA) B) the quantum device and a gauge-transformed encoding of the same instance on the device where the sign of all couplings is changed. C) the SQA and an average over 16 random gauges on the device. D) a single gauge choice on the device against the average over 16 random gauges.

means that there the device typically finds local minima far away from ground states. Many spins would need to be flipped to reach a ground state, which results in small tunneling matrix elements between the state found and the true ground state and thus small gap avoided level crossings.

To provide further evidence for quantum annealing in the D-Wave device we investigate correlations between the success probabilities of the device and a simulated quantum annealer. Panel A of figure 5 shows a scatter plot of the hardness of instances in the simulated quantum annealer (SQA) and the hardware quantum annealer (QA). The high density in the lower left corner (hard for both methods) and the upper right corner (easy for both methods) confirms the similarities between the quantum device and a simulated quantum annealer.

However, a small percentage of instances appears in the other corners (hard for one method but easy for the other). We attribute these to calibration errors of the device, as the couplers are not all identical and the programmed couplings vary by about 10% across the device [22]. To test for calibration errors we performed a "gauge transformation" on each instance (see Methods) to realise a different encoding of the same spin glass instance on the device. The correlations between two different encodings (gauge choices), shown in panel B, turn out to be comparable to those between the SQA and the device, demonstrating that the observed deviations can be attributed to calibration errors.

To minimise calibration errors on the device, we performed annealing with multiple encodings of the same spin glass instance related by gauge transformations and averaged the success probabilities. The resulting correlations between the simulated quantum annealer and the gauge-averaged results from the quantum annealer, shown in panel C, are substantially improved compared to a single gauge choice, with a substantial reduction of the number of extreme outliers. These correlations are also much better than those between a single embedding and the gauge averaged results on the device (see panel D), confirming that (within calibration uncertainties) the device behaves consistently with a simulated quantum annealer, but not with a classical thermal annealer, as seen by the bimodal histogram of success probabilities even after gauge averaging in panel D of figure 2.

We finally investigate the scaling of the annealing effort with problem size N and start by showing, in panel A of figure 6, the times needed for three exact classical optimisation algorithms: akmaxsat [23], the biqmac algorithm [24] used in the spin glass server [25], belief propagation [26] and a related optimized divide-and-conquer algorithm. These algorithms take seconds for our problems with $N = 108$ spins, but hours to days (or run out of memory) for $N = 512$ spins. This is orders of magnitude longer than the D-Wave device and the simulated annealers, but in contrast to the annealers these exact algorithms can prove optimality, and were used to verify our SA, SQA, and QA results. Since the tree width of the Chimera graph scales as $\sqrt{N}$ (figure 1) exact solvers making use of the graph structure are expected to scale asymptotically no worse than $\exp(c\sqrt{N})$, as indeed observed in figure 6, panel A. Similar scaling is observed also for SA and SQA (panels D and E).

For the D-Wave device we only take into account the intrinsic annealing time and not overhead from programming the couplers and readout of the results. We calculate the total annealing time Rt_{anneal}, defined as the product of the annealing time t_{anneal} of one annealing run multiplied by the number of repetitions R needed to find the ground state at least once with 99% probability. From the success probability s we calculate the required number of repetitions $R_p = \log(1-p)/\log(1-s)$, with $p = 0.99$.

In panel B of figure 6 we show not only the scaling of the typical (median) instance but various quantiles of hardness from the easiest (1% quantile) to the hardest (99% quantile) problems on the D-Wave device. For the 1% quantile we find constant annealing time: the probability to reach the ground state in a single annealing run is larger than 99%. The rapid increase of the higher quantiles is due to calibration issues that cause problems for a fraction of problems. It can be expected that on the next generation device with improved calibration these quantiles will be substantially reduced.

Focusing on the median we see only a modest increase in total annealing time from 5 μs to 15 μs, corresponding to three repetitions of the annealing. While an extrapolation of the observed experimental scaling is tempt-

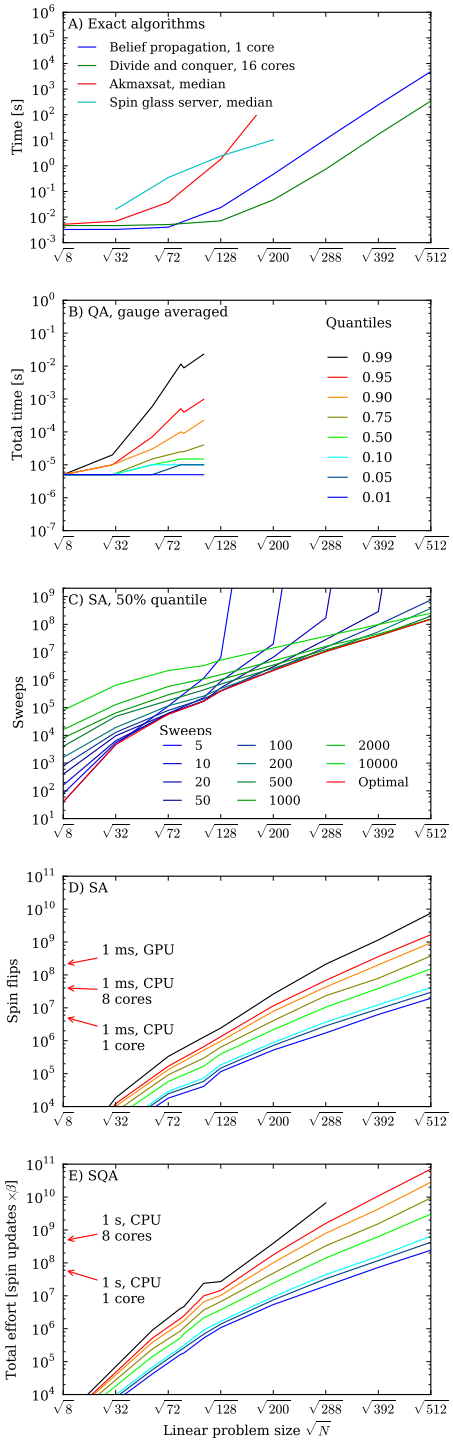

FIG. 6: **Scaling with problem size.** Shown is the scaling of the total effort for A) exact classical solvers, B) the D-Wave device, C) the median total effort of the simulated annealer at constant annealing times measured in sweeps (one sweep is one attempted update per spin), D) the simulated annealer and E) the simulated quantum annealer to find the ground state with 99% probability. The individual lines in panels B, D and E show the scaling of the various quantiles, from the 1% easiest instances (0.01 quantile) to the 1% hardest instances (0.99 quantile). For the simulated annealers the vertical axis shows the total effort in number of spin updates for the simulated classical and quantum annealer. Arrows mark the number of spin updates that can be done in 1ms or 1s on the reference machines.

ing, it can only be used to estimate the performance for slightly larger problem sizes, but will not give the true asymptotic scaling. To see this, consider panel C of figure 6 where we show the scaling of the median time for a simulated annealer for various fixed annealing times (similar behavior is observed for a simulated quantum annealer). The scaling at *fixed* annealing time t_{anneal} increases only modestly for small system sizes, and then suddenly shoots upwards once t_{anneal} becomes too short for the problem size. Since the fastest possible annealing time $t_{\mathrm{anneal}} = 5\mu$s on the D-Wave device is longer than the optimal time for the considered problem sizes (see the supplementary material), our experimental data is in the initial transient regime of modest increase, and thus cannot be used for reliable extrapolation or determination of quantum speedup.

To obtain the true asymptotic scaling one needs to first minimise the total annealing time Rt_{anneal} for each problem size by varying t_{anneal}, and then consider the scaling of this optimal time, as shown for the simulated classical and quantum annealer in panels D and E. The effort here is measured in number of spin updates, defined as RNN_{updates} where N_{updates} is the optimal number of updates per spin to minimise the total effort. Indicated by an arrow is the number of spin updates that can be performed in a millisecond on an 8-core Intel Xeon E5-2670 ("Sandy Bridge") CPU and on an Nvidia K20X ("Kepler") GPU. At $N = 108$ spins the total annealing time of the simulated annealer is 4.3μs on the Sandy Bridge CPU and 0.8μs on the Kepler GPU, slightly faster than the D-Wave device, while the simulated quantum annealer is substantially slower.

Increasing the problem size up to $N = 512$ spins, which will be available on the next generation chip, our simulated quantum annealer shows that the fraction of easy instances drops rapidly (see figure 7). This is perhaps a surprising result: $N = 200$ spins is still a "small" problem, presumably due to the limited connectivity of the device. As a consequence, one finds an increase of the median effort by a factor of ≈ 870 for our simulated classical annealer and ≈ 1400 when increasing the problem size from $N = 108$ to $N = 512$ spins. Unlike the results of Ref. [11], we also observe that the simulated quantum annealer here scales slightly worse than the simulated classical annealer. A similar behavior has previously been reported for k-satisfiablity problems [27]. Investigating for which class of problems simulated quantum annealing is better than simulated classical annealing is an important open question to be addressed in future work.

Conclusions – Our experiments have demonstrated that quantum annealing with more than one hundred qubits takes place in the D-Wave One device, despite limited qubit coherence times. The key evidence is the correlation between the success probabilities on the device and a simulated quantum annealer. In particular we see a bimodal distribution of easy and hard instances, where the hard instances are characterised by avoided level crossings with small gaps. Sensitivity to these small

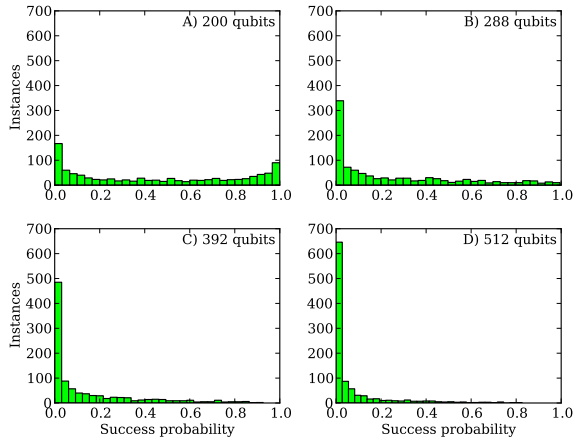

FIG. 7: **Scaling to larger problem sizes.** Histograms of the hardness distribution obtained by simulated quantum annealing with $N_{\text{updates}} = 10000$ Monte Carlo updates per spin. The fraction of easy problems rapidly decreases when increasing the problem size beyond 200 spins.

gaps of the quantum model demonstrates that the device has sufficient ground state quantum coherence to realise a quantum annealing of a transverse field Ising model. Considering the pure annealing time, the performance for typical (median) instances matches that of a highly optimised classical annealing code on a high-end Intel CPU.

While for 108 spins a majority of optimisation problems is still relatively easy, experiments using up to 512 spins on the next generation device will enter a very interesting regime where almost all instances become hard for both simulated annealing and simulated quantum annealing. Simulated annealers require about three orders of magnitude more computational effort for $N = 512$ spins compared to $N = 108$ spins for our problems, and there will be potential to see advantages of a quantum annealer over classical algorithms.

Quantum speedup can then be detected by comparing the scaling results of the simulated classical and quantum annealers to experiments, as we discuss in detail in the supplementary material. Going to even larger problem sizes we soon approach the limits of classical computers. Optimistically extrapolating using the observed scaling, the median time to find the best solution for our test problem will increase from milliseconds to minutes for 2048 variables, and months for 4096 variables, and the scaling might be much worse if fat tailed distributions start to dominate, as we had previously observed for other Monte Carlo algorithms [28, 29]. A quantum annealer showing better scaling than classical algorithms for these problem sizes would be an exciting breakthrough, validating the potential of quantum information processing to outperform its classical counterpart.

Methods

Quantum annealing was performed on the D-Wave One Rainer chip installed at the Information Sciences Institute of the University of Southern California. Details of the device have been presented elsewhere [13]. After programming the couplings, the device was cooled for 2.5 seconds, and then one thousand annealing runs were performed using an annealing time of $t_{\text{anneal}} = 5\mu$s. Details of the annealing schedule and results for longer annealing times are provided in the supplementary material.

Simulated annealing was performed using the Metropolis algorithm with local spin flips with codes optimized for the ± 1 couplings used as test problems. A total of N_{updates} flips per spin were attempted, increasing the inverse temperature $\beta = 1/k_B T$ linearly in time from 0.1 to 3. The *simulated quantum annealing* simulations were performed in both discrete and continuous time path integral quantum Monte Carlo simulations with cluster updates along the imaginary time direction to account for the transverse field, combined with Metropolis rejection sampling for the Ising interactions (see the supplementary material for details).

Gauge averaging was performed on the device by using gauge symmetries to obtain a new model with the same spectrum. This was achieved by picking a gauge factor $a_i = \pm 1$ for each qubit, and transforming the couplings as $J_{ij} \to a_i a_j J_{ij}$ and $h_i \to a_i h_i$. success probabilities s_g obtained from G gauge choices were arithmetically averaged for the correlation plots and as $\overline{s} = \prod_{g=1}^{G} (1 - s_g)^{1/G}$ for the scaling of total effort (see supplementary material for a derivation).

The ground state energies were obtained using exact optimisation algorithms, belief propagation [26] and a related optimized divide-and-conquer algorithm described in the supplementary material.

Acknowledgements We acknowledge useful discussions with M.H. Amin, M.H. Freedman, H.G. Katzgraber, C. Marcus, B. Smith and K. Svore, and help with optimizing the GPU codes from P. Messmer. Simulations were performed on the Brutus cluster at ETH Zurich and on computing resources of Microsoft Research with the help of J. Jernigan. This work was supported by the Swiss National Science Foundation through the NCCR QSIT, by ARO grant number W911NF-12-1-0523, by ARO MURI grant number W911NF-11-1-0268, by the Lockheed Martin Corporation, by DARPA grant number FA8750-13-2-0035, and by NSF grant number CHE-1037992. MT acknowledges hospitality of the Aspen Center for Physics, supported by NSF grant PHY-1066293. The initial planning of the experiments by MT was funded by Microsoft Research.

Author Contributions MT, JM and DL designed the experiments. SB and ZW performed the experiments on D-Wave One. TR, SI and MT wrote the simulated classical and quantum annealing codes and TR, SI, MT and DW performed the simulations. SB and TR wrote the bucket sort code and divide and conquer codes. TR, SI, MT, SB, ZW and DL evaluated the data. All authors contributed to the discussion, the presentation of the results and writing of the manuscript.

Supplementary material for "Quantum annealing with more than one hundred qubits"

I. OVERVIEW

Herel we provide additional details in support of the main text. Section II expands upon the algorithms employed in our study. Section III presents additional success probability histograms for different numbers of qubits and for instances with magnetic fields, explains the origin of easy and hard instances, and explains the error correction scheme. Section IV presents further correlation plots and provide more details on gauge averaging. Section V gives details on how we determined the scaling plots and how quantum speedup can be detected on future devices. Finally, section VI explains how the spectral gaps were calculated by quantum Monte Carlo (QMC) simulations.

II. CLASSICAL ALGORITHMS

A. Simulated annealing

Simulated annealing is performed by using the Metropolis algorithm to sequentially update one spin after the other. One pass through all spins is called one *sweep*. Our highly optimised simulated annealing code, based on a variant of the algorithm in Ref. [30, 31], uses multi-spin coding to simultaneously perform 64 simulations in parallel on a single CPU core: each bit of a 64-bit integer represents the state of a spin in one of the 64 simulations and all 64 spins are updated at once. A similar code for GPUs uses 32-bit integers and additionally performs many independent annealing runs in parallel in multiple threads.

The performance of our codes on the classical reference hardware is shown in Table I. We use high end chips at the time of writing this paper, an 8-core Intel Xeon E5-2670 "Sandy Bridge" CPU and an Nvidia Tesla K20X "Kepler" GPU. To find a ground state of our hardest 108-spin instances with a probability of 99%, this translates to a median annealing 32μs on a single core of the CPU, 4μs on eight cores, and 0.8μs on the GPU, which should be compared to 15μs pure annealing time on the D-Wave device for the same problems.

	spin flips per ns	relative speed
Intel Xeon E5-2670, 1 core	5	1
Intel Xeon E5-2670, 8 cores	40	8
Nvidia Tesla K20X GPU	210	42

TABLE I: Performance of the classical annealer on our reference CPU and GPU.

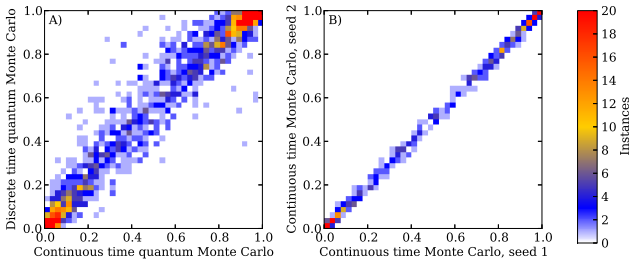

FIG. 1: **Correlation between simulated quantum annealers.** Axes corresponds to success probabilities and pixels are color-coded according to the number of instances. A) correlations between continuous- and discrete time Monte Carlo simulations. The scatter observed here is a measure for the dependence of success probabilities on details of the simulated quantum annealing implementation, for instances with $N = 108$ spins performing 10,000 sweeps. B) Correlations between two independent sets of 1000 simulations with different initial starting points. Schedule II and 10,000 sweeps are used, see figure 2.

B. Simulated quantum annealing

For simulated quantum annealing we use both a continuous time algorithm and a discrete time algorithm.

The continuous time algorithm [32] constructs segments of a world line in the (imaginary) time direction and flips them using the Metropolis algorithm. Specifically, we pick a random site and introduce new cuts in the time direction via a Poisson process. Then we calculate the overlaps with the neighbouring sites and use these overlaps to calculate the Metropolis acceptance ratios $P_{\rm Met}$ and flip a segment with probability $P_{\rm Met}/2$. We cannot grow the cluster along the space directions as in Ref. 32, which connects segments into larger clusters, since such cluster updates are inefficient in frustrated models like our spin glass.

In order to implement the fastest possible simulated quantum annealing code we also implemented a discrete time algorithm similar to that outlined in Ref. 33. However, unlike Ref. 33. we again used cluster updates along the imaginary time direction, typically with 64 time slices.

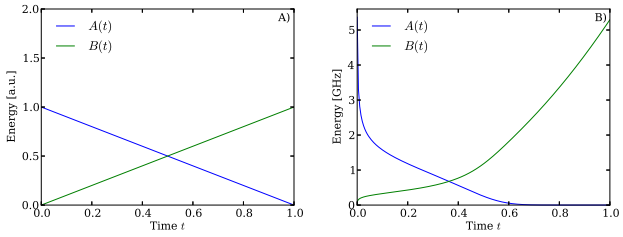

FIG. 2: **Annealing schedules used for Monte Carlo codes** A) schedule I, the linear schedule. B) schedule II, the schedule of the D-Wave device.

To verify that the discrete time code produces results similar to the continuous time algorithm – *i.e.*, that the error in the imaginary time direction is small – we show a correlation plot between the discrete- and continuous time in figure 1A). The scatter observed in this plot is a measure for the dependence of success probabilities on details of the simulated quantum annealing implementation. We also show correlations for a continuous time Monte Carlo using two different sets of random seeds for initial configurations and updates in figure 1B). The scatter of points is within the 3% ($1/\sqrt{1000}$) error expected for the success probabilities when performing 1000 annealing runs per instance.

In Table II, we summarise the performance of these two codes for typical cases using the linear schedule of figure 2A).

We have performed simulated quantum annealing with two different schedules shown in figure 2: A) a linear schedule (referred to as schedule I) where the transverse field is ramped down linearly in time and the Ising couplings are ramped up linearly, and B) the schedule used in the D-Wave device (referred to as schedule II). The performance was similar in both cases. For the scaling plots we used the linear schedule I at an optimised temperature ranging from $T = 0.33$ to $T = 1$ depending on system size, which gives slightly better performance than schedule II.

For the correlation plots we use the continuous time code (CTQ) with schedule II – the average over slightly different schedules for the individual qubits on the device – but at up to ten times lower temperature than the device temperature of 20mK (0.4 GHz). The simulated quantum annealer requires a temperature at least three times lower than the nominal temperature of the device to exhibit a bimodal distribution. This can be explained and motivated as follows. When the transverse field is strong the quantum Monte Carlo updates mimic the quantum tunneling taking place in the device, however when the transverse field becomes smaller these Monte Carlo updates turn into local spin flips of a classical (thermal) annealer. The device in this regime, on the other hand, has high tunneling barriers between the two states of the qubits of the device that suppress thermal tunneling over the barrier. To achieve a similar suppression of thermal excitations we need to lower the temperature in the quantum Monte Carlo simulations by at least a factor two. Lowering by more than a factor of ten does not significantly change histograms or correlations, indicating that at the chosen temperature the simulated quantum annealer is dominated by quantum tunneling and not thermal effects.

	spin updates per μs	relative speed
CTQ, 1 core	1.3	1
CTQ, 8 cores	3.8	8
DTQ, 1 core	5.8	4.5
DTQ, 8 cores	46	35

TABLE II: **Performance of simulated quantum annealers.** We show performance figures of both our continuous time (CTQ) and discrete time (DTQ) implementations on an Intel Xeon E5-2670 CPU using schedule II at inverse temperature $\beta = 10$.

C. Divide and conquer algorithm

In addition to belief propagation using bucket sort [26], we implemented an exact divide-and-conquer algorithm specifically designed for the chimera graph, generalising divide-and-conquer for the square lattice. This algorithm can be considered a specialisation of the more general bucket sort algorithm. We consider a chimera graph of $M \times M$ 8-spin unit cells ($8M^2$ spins). For each possible configuration of the $4M$ spins on the left side of the unit cells in the first row (which couple vertically) we find and store the optimal configuration of the remaining $4M$ spins with effort $4M$, giving a total effort of $4M\,2^{4M}$. Finding the 2^{4M} optimal configurations of the next row of unit cells, one builds up the solution row by row, scaling as $\mathcal{O}\left(M^2\,2^{4M}\right)$. Since $M \propto \sqrt{N}$, where N is the number of spins, this scaling demonstrates explicitly the claim made in the main text that exact solutions scale no worse than $\mathcal{O}(\exp(c\sqrt{N}))$, as for the bucket sort algorithm.

III. SUCCESS PROBABILITY HISTOGRAMS

In this section we provide additional experimental and simulation data, complementing the data shown in the paper.

A. Experimental annealing histograms

In addition to the histogram for 108 qubits shown in the main text, we also show histograms for 58, 84, 87 and 108 qubits without local fields in figure 3 and with local fields in figure 4. By comparing these two figures, one notes that the cases with local fields are in general easier

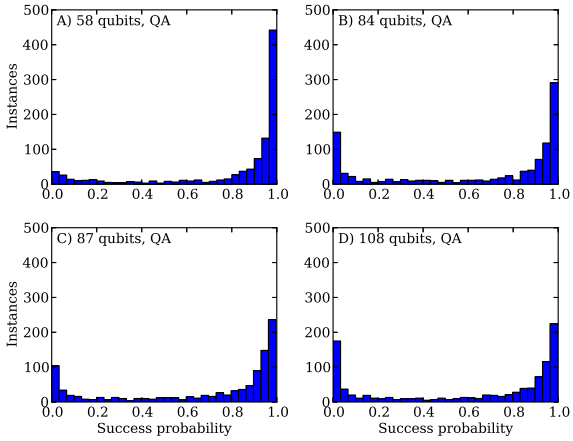

FIG. 3: **Success probability histograms for the D-Wave device for instances without local fields.** A) using 58 qubits, B) 84 qubits, C) 87 qubits and D) 108 qubits. The peak at low success probability grows with the number of qubits, reflecting the increasing hardness of the corresponding problem instances.

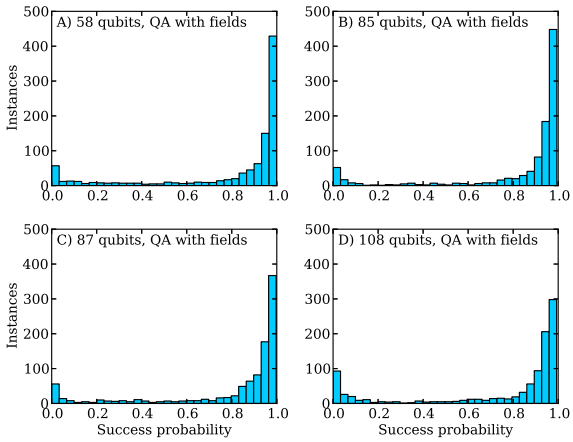

FIG. 4: **Success probability histograms for the D-Wave device for instances with local fields.** A) using 58 qubits, B) 85 qubits, C) 87 qubits and D) 108 qubits.

for the D-Wave device. We also verified that this is the case for the simulated annealers - see figure 5.

To check that the bimodality is not due to faulty couplers or qubits we performed the following analysis for the hardest instances of $N = 108$ spin problems where the D-Wave device did not find the ground state with one gauge choice. For each of these instances we looked at the lowest energy configurations reached and determined which spins differ compared to the closest ground state configuration. Closeness is measured in terms of the Hamming distance, which is the number of spins that have to be flipped to reach a ground state configuration. We did not observe a strongly peaked distribution which would have indicated a singly faulty qubit or coupler.

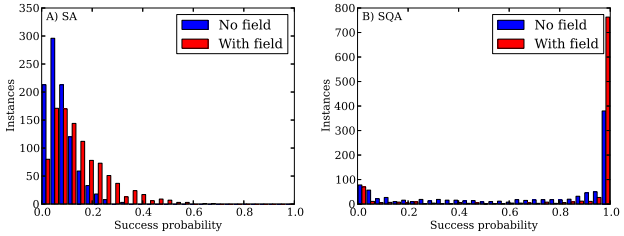

FIG. 5: **Comparison of instances with and without local fields**. A) success probability histograms for the simulated classical annealer with 50 sweeps (updates per spin). B) success probability histograms for the simulated quantum annealer using 10,000 sweeps.

B. Simulated annealing and simulated quantum annealing

Figures 6 and 7 show the success probability histograms for $N = 108$ spin problems for simulated annealing and simulated quantum annealing with different number of annealing sweeps (updates per spin). While in the simulated classical annealer the distribution is always unimodal and shifts towards larger success probabilities upon increasing the annealing time, the simulated quantum annealer becomes more strongly bimodal when increasing annealing times.

Figure 8 shows the success probability histograms for simulated quantum annealing for instances with local fields. The bimodality of easy and hard instances remains for up to about 512 spins for instances with local

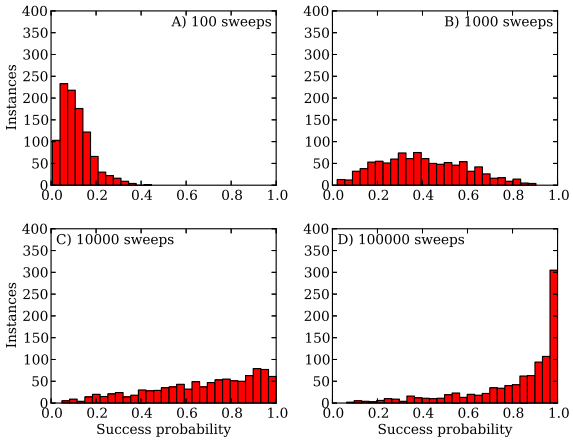

FIG. 6: **Success probability histogram for the simulated thermal annealer.** Annealing times are A) 100 sweeps, B) 1,000 sweeps, D) 10,000 sweeps and D) 100,000 sweeps for instances without local fields.

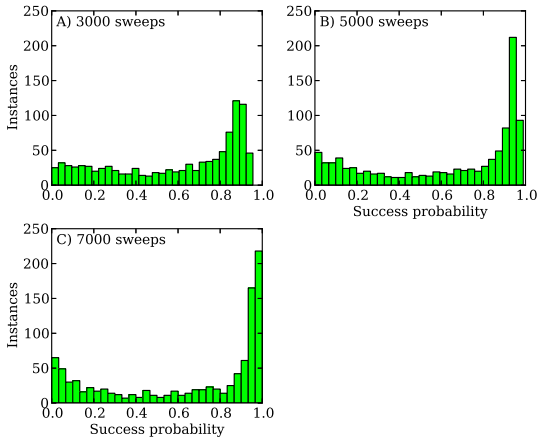

FIG. 7: **Success histograms for simulated quantum annealing.** The bimodal structure gets more pronounced upon increasing the annealing time from A) 3,000 sweeps to B) 5,000 and C) 7,000 sweeps. All three histograms were obtained for instances without local fields using schedule II.

fields, indicating that there are more easy instances here than in the case without fields. This can also be used to test for quantumness in a device scaled up to more qubits. Once the "easy" problems disappear beyond 512 spins alternative analysis methods will be necessary to test for quantumness.

Supplementing figure 8, and figure 7 of the main text, we show the evolution of the success probability histograms for a simulated classical annealer in figure 9. It is seen that at fixed annealing time the unimodal peak gradually shifts to smaller success probabilities.

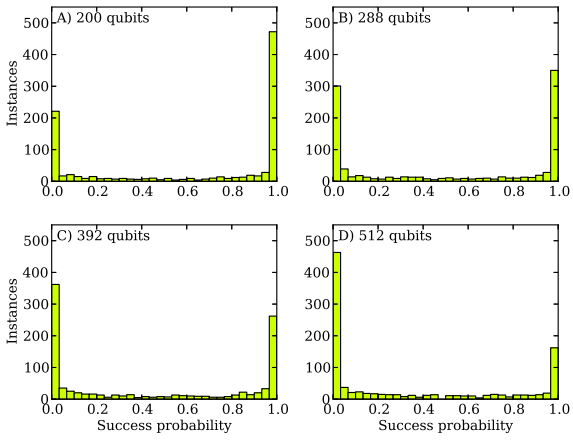

FIG. 8: **Scaling to larger problem sizes (without fields) for the simulated quantum annealer for instances with local fields.** Histograms of the hardness distribution obtained by simulated quantum annealing with $N_{\text{updates}} = 10000$ Monte Carlo updates per spin.

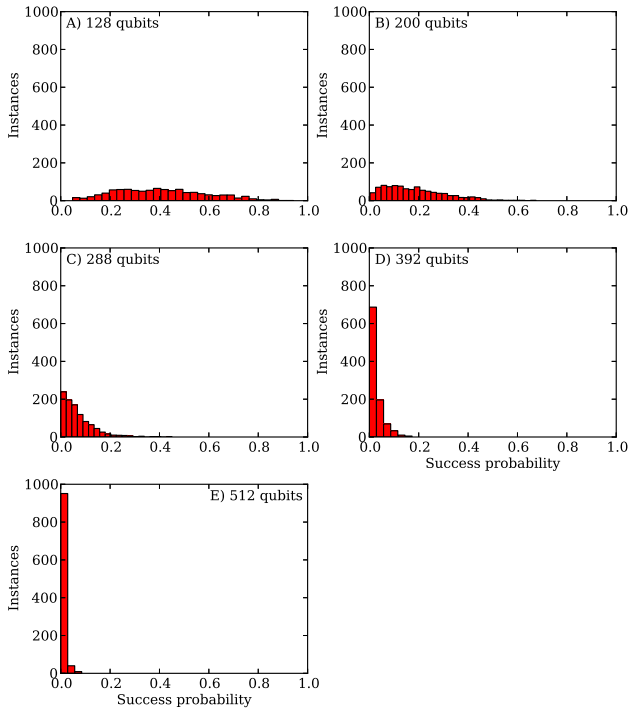

FIG. 9: **Scaling to larger problem sizes for the simulated classical annealer.** Histograms of the hardness distribution obtained by simulated annealing with $N_{\text{updates}} = 10000$ Monte Carlo updates per spin.

C. Hardness, gaps and free qubits

Figures 3 and 4 of the main text show that the "hard" instances typically exhibit small gaps during the evolution and often get trapped in excited states with a large Hamming distance to the closest ground state. The correlation between small gaps and large Hamming distance can be understood with a simple perturbative argument in the regime of small transverse fields $-\Gamma \sum \sigma^x$, where most of the small gap avoided level crossings appear. At an avoided level crossing between two states with Hamming distance d, d spins need to be flipped to adiabatically follow the ground state. In perturbation theory, the tunneling matrix element between the two states is of order Γ^d, exponentially small in the Hamming distance. The small matrix element not only poses problems for adiabatic evolution but also suppresses quantum Monte Carlo updates that connect the two states. This common origin of the hardness in both quantum annealing and simulated quantum annealing explains the observed correlations despite the different underlying dynamics (deterministic *vs* stochastic).

An intuitive understanding of how such small gap avoided level crossings can arise can be obtained by considering degenerate states that are connected by single spin flips. The *free* qubits of a state are the qubits that can be flipped without changing the energy. From per-

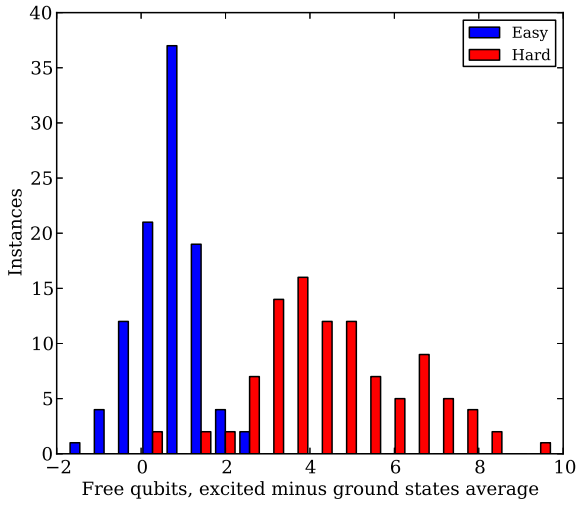

FIG. 10: **Degeneracies and hardness.** To illustrate correlations between hardness and degeneracies we calculate the average number of free qubits in excited states and ground states found by the quantum annealer. The histogram shows ⟨free qubits in excited states⟩ - ⟨free qubits in ground states⟩. Difficult instances (the 10% hardest over 1000 instances) are typically trapped in excited states with many free qubits, which lead to avoided level crossings. The easier decile of glasses, on the contrary, typically does not have more free qubits in excited states than in ground states. Results from the simulated quantum annealer are very similar.

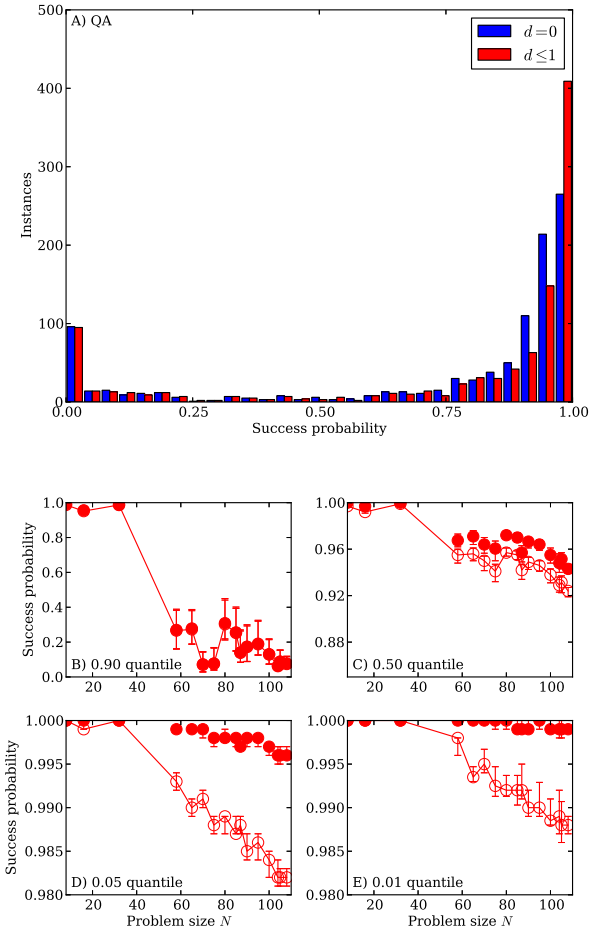

FIG. 11: **Effect of error correction.** Shown are results for $N = 108$ spin instances with local fields: A) Histogram of the success probability s with and without single spin error correction. The dependence of the effect of error correction on problem size N can be seen by following the quantiles of success probability as a function of system size for B) the 90% quantile, C) 50% quantile D) 5% quantile and E) 1% quantile.

turbation theory, a small transverse field Γ breaks the degeneracy of each free qubit [14, 34, 35]. In the simplest case, degenerate states form a hypercube and have the same free qubits. If the unperturbed energy was E_0, the lowest energy state of a hypercube with F free qubits will then have energy $E_0 - \Gamma F$ to first order in perturbation theory. If the low energy excited states have more free qubits than the ground states, their energy will be lowered more, resulting in avoided level crossings and small gaps. As seen in Fig. 10, hard instances tend to have more free qubits in low energy excited states. This phenomenon has been previously observed in the random subcubes model, a specially constructed toy model [36].

In the spin glasses with ± 1 couplings chosen here, we can expect to find a significant number of free qubits per state. This is because many qubits have six couplings, that can cancel each other. We also expect more low energy excited states than ground states, and consequently some low energy excited states will have more free qubits than most ground states. As argued above, both physical and simulated quantum annealing are sensitive to this problem: many spins updates are needed to move between the states at both sides of the avoided crossing. On the other hand, classical simulated annealers, not having a transverse field, do not suffer from these avoided crossings. This might explain why simulated quantum annealing scales slightly worse than classical annealing for the ± 1 spin glasses considered in this work.

D. Correcting single spin errors

Single spin flip errors can be corrected with a linearly scaling effort by checking once for each spin whether the total energy can be lowered by flipping it, and then flipping the spin that lowers the energy the most. To illustrate the effect of this procedure we show in figure 11A), the success probability histogram for $N = 108$ spin instances with local fields, with and without error correction. Without error correction only ground states (Hamming distance $d = 0$) count as a successful annealing run, while with error correction also the runs giving states a Hamming distance $d = 1$ away from the ground state

will, after error correction, end up in the ground state. It is clear from figure 11A) that while "hard" instances do not benefit from this error correction scheme, such single spin flip errors are a dominant error source for the "easy" instances and their failure rate is substantially reduced using error correction. The most likely source for such errors are thermal excitations or readout errors. To understand the dependence on problem size N and hardness, we consider the quantiles of the success probability distribution as a function of N in figure 11B) to E). We see that the success probability decreases upon increasing N, i.e., the probability of such errors increases for larger problems. Independently of N, single-bit error correction is ineffective in increasing the success probability for the "hard" instances (high quantiles). It becomes more and more effective as problems become easier (low quantiles), improving the success probability to very nearly 1, independently of N, at the 1% quantile.

Variants of this simple error correction scheme can be derived that, also with linear scaling, correct not only one single spin errors, but all disconnected single spin errors. However, we saw only minimal improvements using these alternative schemes, because independent single spin errors are not common on the studied problem sizes. They may become more common on larger problems.

IV. CORRELATIONS

A. Copulae

To better understand the correlations we here show copulae in addition to the correlations shown in the main text. Copulae

$$c(x_1, x_2) = \frac{f(x_1, x_2)}{f_1(x_1) f_2(x_2)}. \quad (1)$$

factor out the marginal distributions $f_1(x_1)$ and $f_2(x_2)$ (*e.g.*, the strong bimodal distribution) from the joint probability function $f(x_1, x_2)$ describing the correlation density. For independent sets the copula density is $c(x_1, x_2) = 1$ while for perfect correlations it is a delta function $c(x_1, x_2) = \delta(x_1 - x_2)$.

To plot copula densities of success probabilities we replace each success probability by its rank after sorting the success probabilities divided by the number of values, and then plot the correlation densities of these normalised ranks. In figure 12 we plot copulae corresponding to the correlations shown in figure 5 of the main text, enhancing the visibility of the correlations especially in the corners.

In figure 13 we show the copulae between the hardness for akmaxsat and the D-Wave device and the simulated classical and quantum annealers. We do not observe any correlations.

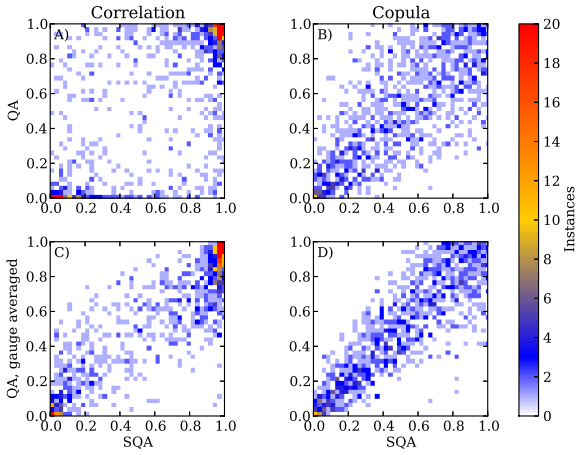

FIG. 12: **Correlations and copulae** between the simulated quantum annealer and the D-Wave device. Axes corresponds to success probabilities and pixels are color-coded according to the number of instances.

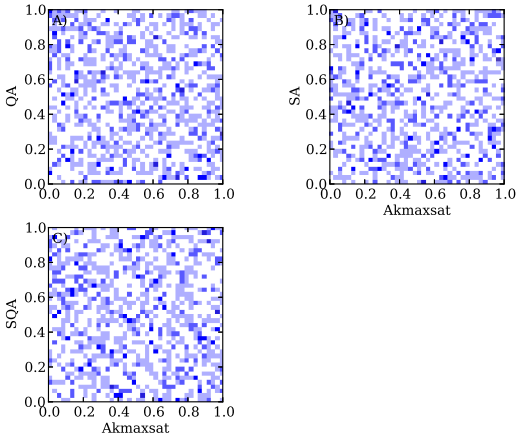

FIG. 13: **Copulae of the hardness between exact optimizers akmaxsat and the annealers.** A) for the D-Wave device, B) the simulated annealer and C) the simulated quantum annealer, showing the absence of correlations.

B. Reproducibility

To verify reproducibility of the QA data we performed experiments on $N = 108$ spin instances three times, with a month between the first and the second two repetitions. The correlations (see figure 14) show that the device is stable over the time of a month.

Strong deviations are seen for a small fraction of the instances. These are most likely due to 1/f noise or "programming errors" when flux quanta are loaded into the programmable magnetic memories to program a specific set of couplings into the device. Since these programming errors will limit the correlations between any model

and the device, no better correlations than shown in this figure can be expected between our simulations and the device.

C. Gauge averaging

The spectrum of an Ising spin glass is invariant under a gauge transformation that changes spins $\sigma_i^z \leftarrow a_i \sigma_i^z$, with $a_i = \pm 1$, when at the same time changing the couplings $J_{ij} \leftarrow a_i a_j J_{ij}$ and the local fields as $h_i \leftarrow a_i h_i$. While the simulated annealers are invariant under such a gauge transformation, the calibration of the D-Wave device is not perfect and breaks the gauge symmetry. Different gauges hence realise slightly different physical systems with different success probabilities. We average over gauges to reduce calibration uncertainties.

In figure 15 we show correlation between the arithmetic average over eight gauges compared against the average over eight other gauges. Gauge averaging significantly increases correlations compared to single gauge choices. For the marginal distributions (see figure 16), the number of hard instances (weight of the peak close to zero) is reduced by gauge averaging, but the distribution remains bimodal even after averaging over many gauge choices. While calibration errors enhance the bimodality, the convergence to a bimodal distribution after averaging many gauges shows that the bimodality is intrinsic and not solely due to calibration errors.

For scaling plots of the total effort we use geometric means of failure rates instead of arithmetic means of success probabilities. If the probability for finding a ground state for a specific instance and gauge choice g is denoted by s then the probability of achieving a ground state at

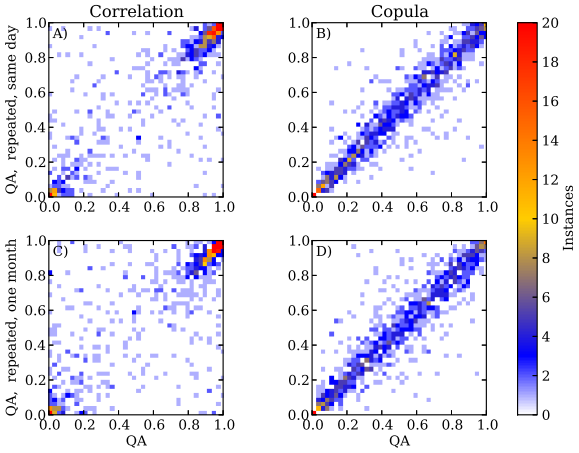

FIG. 14: **Reproducibility of the experiments**. We show correlations and copulae for repetitions of experiments: A) and B) on the same day, and, C) and D) one month apart. A) and C) show correlations between success probabilities, while B) and D) show the corresponding copulae.

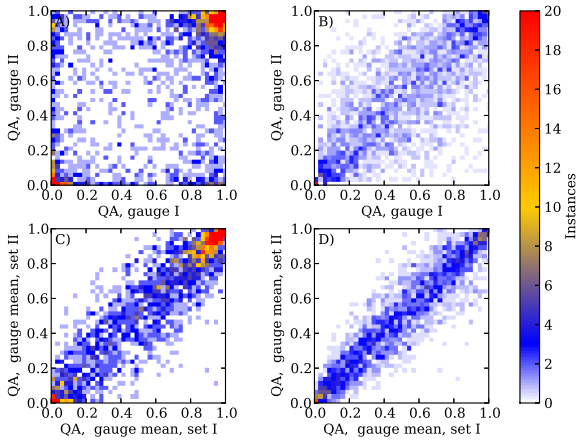

FIG. 15: **Effect of gauge averaging on correlations**. Shown are in A) and B) are the correlations and copulae between two different gauge choices of 1000 instances with $N = 108$ spins. In C) and D) we show correlations and copulae of results averaged over eight gauge choices against averages over eight different gauge choices. Gauge averaging significantly increases correlations.

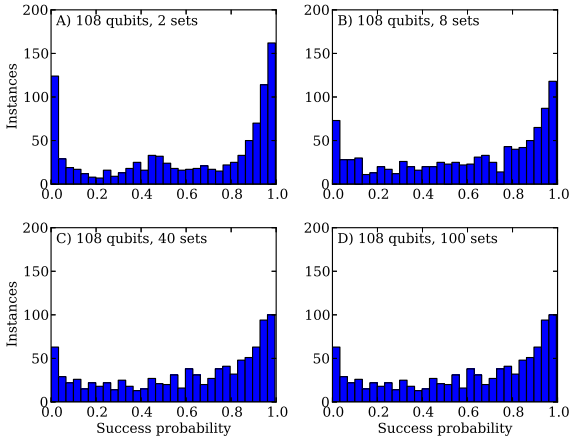

FIG. 16: **Gauge averaged histograms.** Arithmetic averaging over success probabilities obtained using different gauge transformations, using A) 2 gauges, B) 8 gauges, C) 40 gauges and D) 100 gauges. One sees that the marginal distribution converges after roughly 40 gauges.

least once in R repetitions of the annealing is

$$P = 1 - (1 - s)^R. \tag{2}$$

Splitting the R repetitions into R/G repetitions for each of G gauge choices and denoting the success probabilities for a specific gauge choice by s_g the total success probability becomes

$$P^{(G)} = 1 - \prod_{g=1}^{G} (1 - s_g)^{R/G}, \tag{3}$$

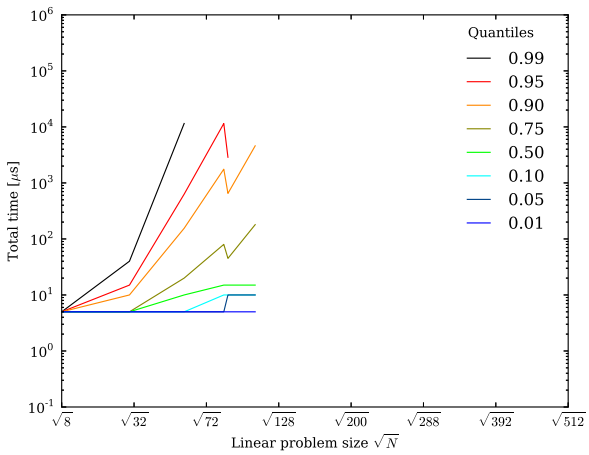

FIG. 17: **Scaling without gauge averaging**. Shown is the total effort on the D-Wave quantum annealer for a single gauge choice. The higher quantiles are not shown for large systems since 1000 repetitions of the annealing failed to find the ground states for these hardest instances.

which can be written in a form similar to equation (2)

$$P = 1 - (1 - \overline{s})^R, \tag{4}$$

by using the *geometric mean* of the failure rates to define $\overline{s}$ as

$$\overline{s} = 1 - \prod_{g=1}^{G} (1 - s_g)^{1/G}. \tag{5}$$

The higher quantiles grow much more rapidly in the absence of gauge averaging, as can be seen by comparing figure 17 to figure 6B) in the main text. Without gauge averaging the device actually fails to find the ground states for the 5% hardest instances in 1000 repetitions and those quantiles are thus not shown.

V. SCALING

A. Scaling of the exact solvers

Figure 6A) of the main text shows the time scaling of the exact solvers for instances without fields. Two of the algorithms, belief propagation using bucket sort [26] and the divide and conquer algorithm presented in section II C do not depend on the specific instance. The time to find a solution is the same for all instances with and without field. These algorithms make use of the structure of the chimera graph and thus show the expected scaling $\mathcal{O}(\exp(c\sqrt{N}))$.

The time to solution of the other two solvers, akmaxsat [23], and the biqmac algorithm [24] used in the spin glass server [25] depends on the specific instance. The scaling for the various quantiles from the easiest (1% quantile) to the hardest (99% quantile) is shown in figures 18 and 19 for instances with and without local fields.

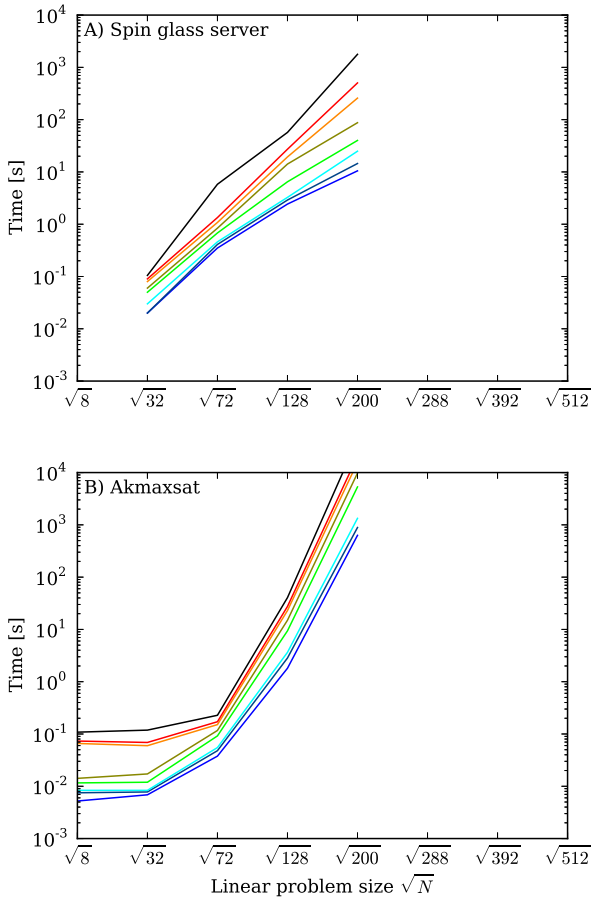

FIG. 18: **Scaling plot for exact solvers and instances without fields** for A) the spin glass server and B) akmaxsat.

While the biqmac algorithm scales as $\mathcal{O}(\exp(c\sqrt{N}))$, akmaxsat scales worse and – unlike the other solvers – does not benefit from the limited tree width of the chimera graph. Unfortunately, the spin glass server constrained us to instances of up to 200 spins. It would be interesting to see if a crossover is present, i.e., to check whether the best exact code scales better than $\mathcal{O}(\exp(c\sqrt{N}))$.

B. Optimising the total annealing time

For the purpose of scaling comparisons we consider the total annealing time needed to find a ground state with a probability of 99%. Using Eq. (2) we find that the number of repetitions to find the ground state at least once with probability $p = 0.99$ is

$$R = \left\lceil \frac{\log(1-p)}{\log(1-s)} \right\rceil, \tag{6}$$

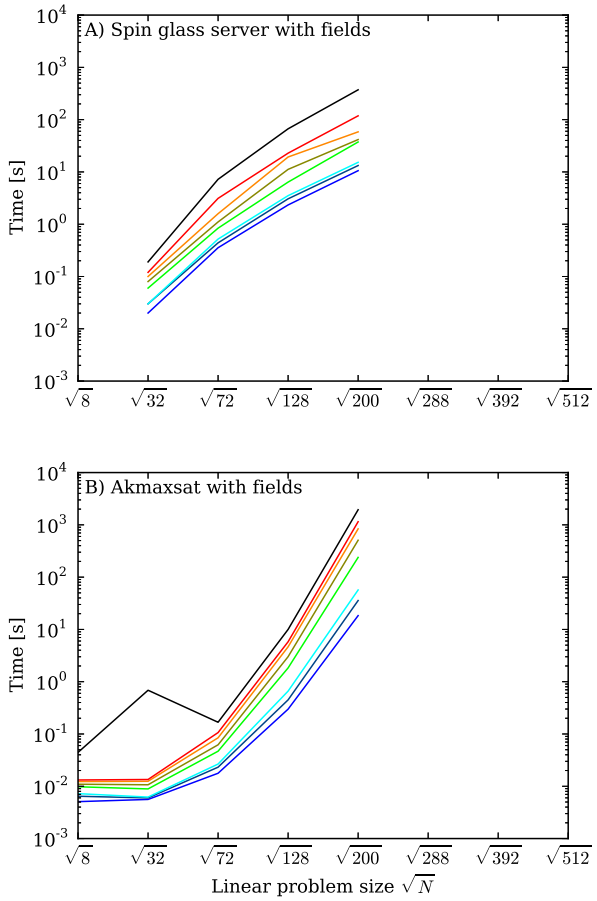

FIG. 19: **Scaling plot for exact solvers and instances with fields** for A) the spin glass server and B) akmaxsat.

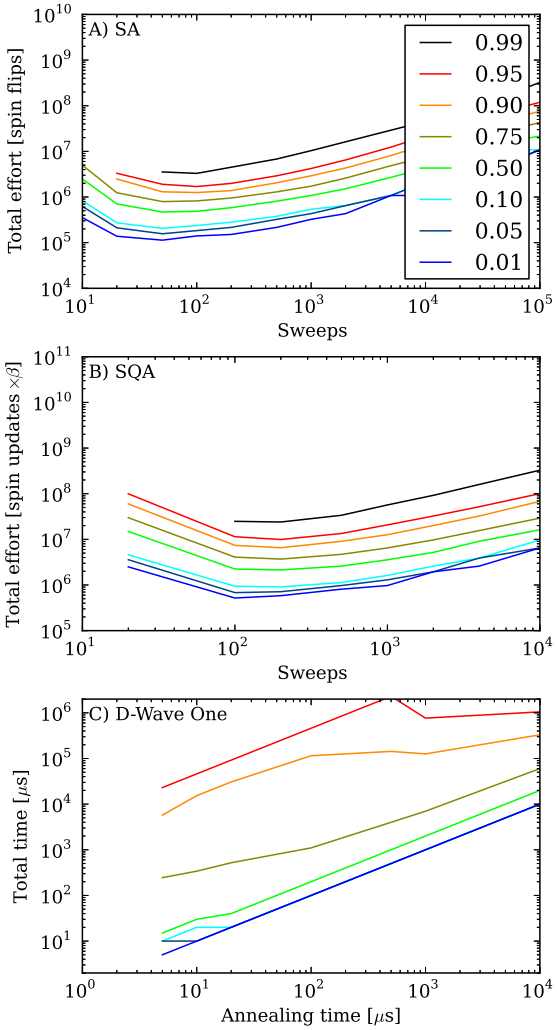

FIG. 20: **Optimisation of total annealing time for $N = 108$ spin instances without local fields**. Shown is the total annealing time run time Rt_{anneal} time as function of the annealing time t_{anneal} for a single annealing run.

and the total annealing time is Rt_{anneal}. Note that this expression assumes uncorrelated repetitions. On the D-Wave device we have seen statistically significant correlations between repetitions in 15% of the instances. In those instances the observed positive autocorrelations increase the length of "runs" of consecutive failures or successes on average by about a factor of two. This will result in an increase of the number of repetitions R required on the device, as given by equation (6) above. The simulated annealers, on the other hand, show no detectable correlations.

To optimise the total annealing time we vary t_{anneal}, measure the success probability p, calculate the required number of repetitions R, and plot the total annealing time Rt_{anneal} as a function of t_{anneal}. The required number of repetitions R diverges when the annealing time t_{anneal} is too short, causing diabatic transitions. When the annealing time is too long it becomes disadvantageous to perform repetitions since a single run already optimizes the success probability; however, Eq. (6) always yields $R \geq 1$. We thus pick the optimal t_{anneal} as the time that minimises Rt_{anneal} and use it in the scaling plots.

In figure 20 we show typical data for a simulated classical and quantum annealing (where time is measured in number of spin flips for SA and spin updates for SQA) and for the D-Wave device (measured in μs). For the simulated classical and quantum annealer we observe that both the optimal time and the required number of repetitions R increase exponentially. Note that a minimum is not found for the D-Wave device but instead a monotonic increase, indicating that even the fastest annealing time of the hardware of $t_{\text{anneal}} = 5\mu$s is slower than the optimal time. All timings reported for the device are thus only *upper bounds* on the optimal time.

While the total annealing time for the D-Wave device can simply be given in microseconds, the annealing times

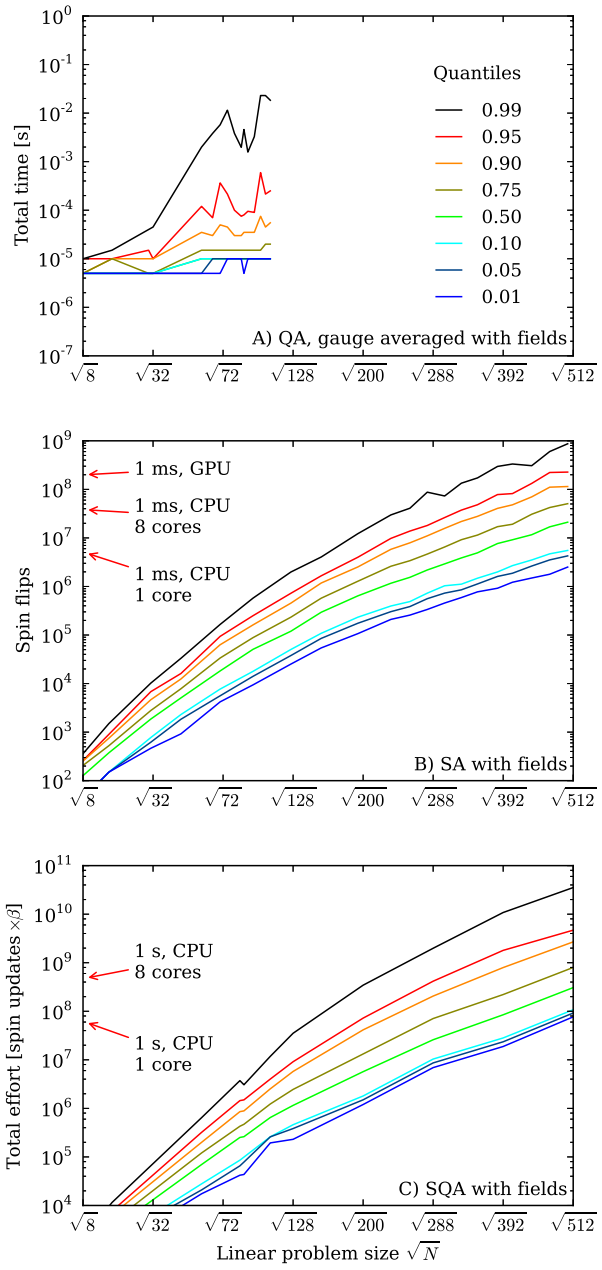

FIG. 21: **Scaling plot with fields for A) QA, B) SA and C) SQA.** Comparing with figure 6, B), D) and E) in the main text, which shows the same plots for problem instances without fields, we see that problems with fields are easier for SA and SQA.

of the simulated classical and quantum annealers depend on compiler options and the specific machine used. We thus give the total effort in terms of the number of attempted spin flips for the simulated classical annealer. For the simulated quantum annealer we specify the total effort as the number of spins updated multiplied with the inverse temperature β, to account for the complexity of updating the imaginary time world lines of the spins of length β.

Scaling plots of the total effort for instances with fields, complementing the results for instances without fields shown in figure 6 of the main text, are shown in figure 21. In panel A, we show the scaling of the D-Wave device. Again simulated classical annealing scales slightly better than simulated quantum annealing. Comparing the scaling for instances without fields (figure 6 of the main text) and instances with local random fields (figure 21), we see that problems with fields are not only easier for small problem sizes but the total annealing time also scales better when going to larger problem sizes.

C. Detecting quantum speed up

Quantum speedup of a hardware quantum annealer can be detected by comparing the scaling of the total annealing times to that of the simulated classical or quantum annealer. To draw valid conclusions about a speedup one must ensure that the experimental annealing times are optimal. This is not the case for the current device as the annealing time of 5μs is suboptimal, as demonstrated in figure 20. It follows that the inferred total annealing times are only upper bounds, and those bounds are worse for smaller problem sizes, which leads to scaling plots that might tempt one to draw the misleading conclusion that the scaling is better on the device. To see the pitfall it suffices to consider extremely long fixed annealing times where a single repetition $R = 1$ might be enough. We would then see a *constant* time needed to find the ground state.

The optimal annealing time defined in the previous section (see Fig. 20) increases exponentially with problem size for SA. It is expected to increase exponentially with problem size also for QA. Therefore we expect that going to $N = 512$ spins, or at most $N = 2048$ spins, the optimal annealing time for a single run will be larger than the minimum programmable annealing time on the D-Wave device. This will allow us to determine the optimal total annealing times for QA. To detect quantum speedup one should then compare to the total effort of the simulated annealers *divided by the number of spins* N. This division is necessary to compensate for the trivial parallelism of the hardware annealer, which updates N spins in parallel, since an analog classical annealing device would have the same parallelism. For completeness we provide the scaling plots of total effort in units of sweeps (spin flips divided by N) in figure 22.

VI. GAP CALCULATION

We finally describe how the excitation gaps are obtained. For simplicity we consider the transverse field

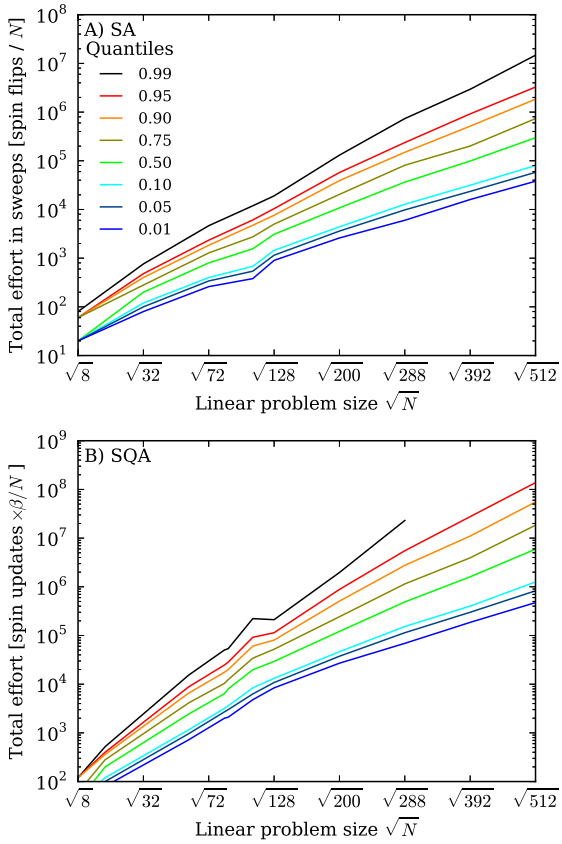

FIG. 22: **Scaling of total annealing effort assuming parallel updates of spins for instances without fields**. Since the quantum annealer (and a classical analog annealer) have intrinsic parallelism of updating all spins simultaneously, we need to divide the total effort by the number of spins N to obtain scaling plots against which quantum speedup can be detected.

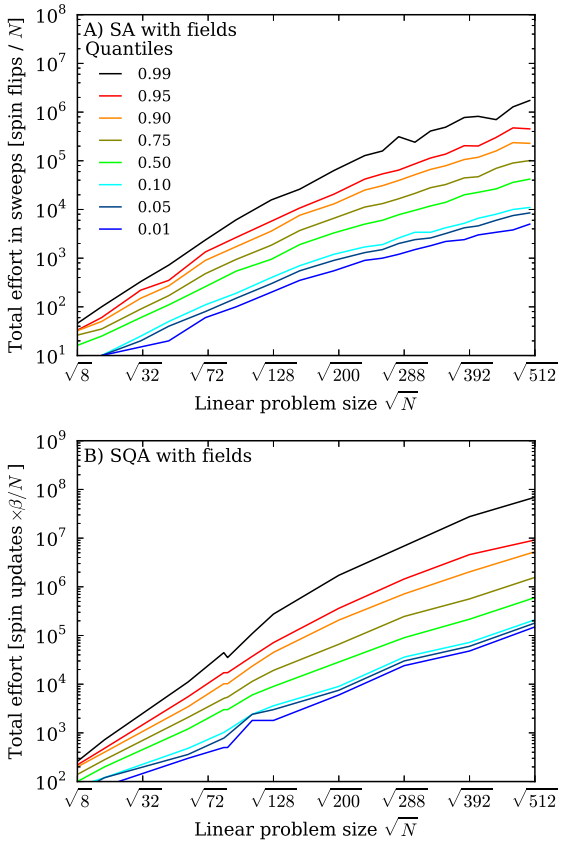

FIG. 23: **Scaling of total annealing effort assuming parallel updates of spins for instances with fields**. As in figure 22 the total effort is divided by the system size due to the intrinsic parallelism of the quantum annealer.

Ising spin glass Hamiltonian without fields,

$$H = -\sum_{i<j} J_{ij}\sigma_i^z\sigma_j^z - \Gamma\sum_i \sigma_i^x, \tag{7}$$

where Γ is the strength of a transverse magnetic field and for our instances $J_{ij} = \pm 1$.

The excitation gap can be obtained from the connected correlation function in imaginary time

$$C(\tau) = \langle\hat{O}(\tau)\hat{O}(0)\rangle - \langle\hat{O}\rangle^2, \tag{8}$$

where $\hat{O}$ is an observable with non-vanishing matrix element between the ground state and first excited state. The correlation function is given by a sum:

$$C(\tau) = \sum_i c_i \exp(-\Delta_i\tau), \tag{9}$$

where Δ_i is a gap to the ith eigenvalue above the ground state; $i = 0$ corresponds to the lowest gap. Only one term survives when τ is large enough:

$$C(\tau) = c_0 \exp(-\Delta_0\tau). \tag{10}$$

Thus $\Delta \equiv \Delta_0$ can be obtained by fitting $C(\tau)$ at large values of τ. We use periodic boundary conditions in the imaginary time direction. In this case, $C(\tau)$ can be efficiently calculated at discrete points using the Fast Fourier Transformation.

The observable $\hat{O}$ must be chosen carefully because for a poorly chosen observable, c_0 can be much smaller than $c_1, c_2, \ldots$ leading to very small values of $C(\tau)$ (comparable to statistical noise) at large τ and making the gap extraction very difficult. This issue is discussed in Ref. 37. We use a simple observable, the local magnetisation $\sigma_j^z(\tau)$ and its correlation function, given by

$$C(\tau) = \frac{1}{N}\sum_{j=0}^{N}\langle\sigma_j^z(\tau)\sigma_j^z(0)\rangle - \langle\sigma_j^z(0)\rangle^2, \tag{11}$$

where the sum runs over all the spins N.

We use the continuous time algorithm described in section II B and "anneal" the system from $\Gamma = 3$ to small values of Γ in steps of 0.1. The simulation at a transverse field Γ_i is started from the final configuration obtained in the previous simulation $\Gamma_{i-1} = \Gamma_i - 0.1$, except for the initial easy simulation at a strong transverse field

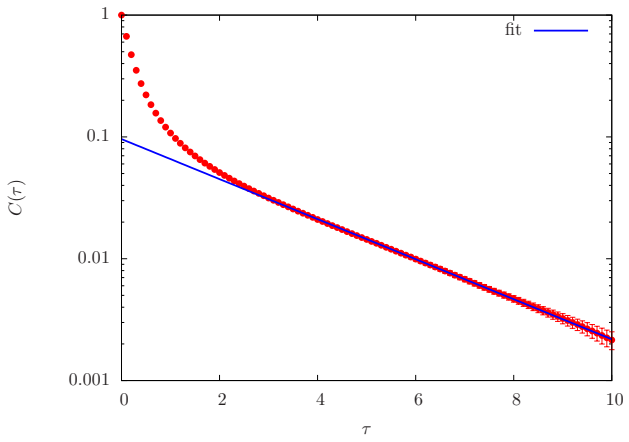

FIG. 24: **Correlation function** $C(\tau)$. The line shows an exponential fit. The gap can be obtained from the slope of the line.

$\Gamma = 3$, where simulation is started from a random configuration. $200000/\Gamma$ Monte Carlo sweeps (one Monte Carlo sweep consists of $2N$ site updates) are performed for equilibration before measurements. Simulations are done at inverse temperatures $\beta = 100$ and $\beta = 200$.

Extraction of the gap for many instances and many values of Γ can be a cumbersome task. To automate the process, we use the following approach. We measure $C(\tau_k)$ at discrete equidistant points τ_k and calculate the "running" gap:

$$\Delta_{\mathrm{r}}(\tau_k) = -\frac{1}{\Delta\tau}\left[\ln C(\tau_k) - \ln C(\tau_{k-1})\right], \qquad (12)$$

where $\Delta\tau = \tau_k - \tau_{k-1}$. Obviously $\Delta_{\mathrm{r}}(\tau_k)$ is equal to the gap Δ if there is no noise, $\Delta\tau$ is small and τ_k is large enough. The estimate $\Delta_{\mathrm{r}}(\tau_k)$ for small τ_k is larger than the true gap Δ because of higher excited states. We observe τ_m at which $\Delta_{\mathrm{r}}(\tau_m)$ becomes a constant within some threshold. Then we obtain the gap by fitting the correlation function to the exponential function given by Eq. (10) in the range from τ_m to $\tau_m + 5/J$. This procedure can be fully automated. In figure 24, we show an example of such a fit.

The excitation gap as a function of Γ is shown in figure 25 for eight additional instances, complementing the two instances shown in figure 3 of the main text. Note that the gap closes trivially around $\Gamma = 2.2$, related to a global Z_2 symmetry breaking. Once the gap becomes very small it can no longer be detected by our procedure since the decay of $C(\tau)$ becomes too slow and is indistinguishable from a constant. Our procedure then picks up the gap to the next excited state, which results in an apparent jump of the gap to a bigger value. Generally, all gaps shown here are upper bounds for the gap to the lowest excited state.

The gap always closes also in the limit of weak transverse field $\Gamma \longrightarrow 0$, where multiple ground states become degenerate. Some of the instances, however, have an additional small gap that can be associated with an avoided level crossing. These instances are "hard" for both the D-Wave device and SQA. The instances that do not have such a gap are "easy" for both the D-Wave device and SQA.

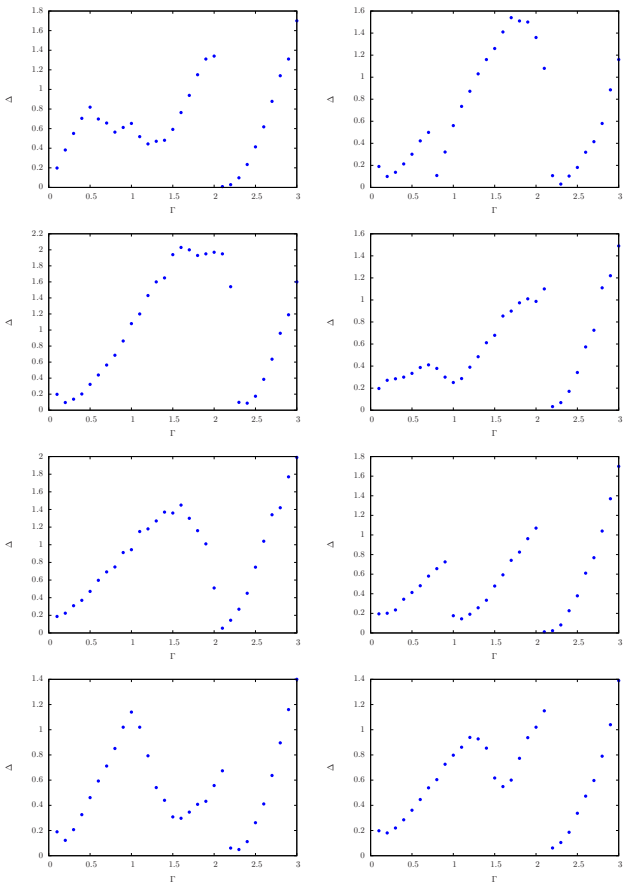

FIG. 25: **Evolution of the excitation gap**. Shown is the lowest excitation gap (in units of $|J| = 1$) for eight instances, to supplement the two instances shown in the main text. "Easy" instances are shown on the left (success probabilities from top to bottom: 0.986, 0.995, 0.99, 0.932) and 'hard' instances are shown on the right (success probabilities from top to bottom: 0.001, 0.063, 0.000625, 0.0000625).

[1] Scarani, V. *et al.* The security of practical quantum key distribution. *Rev. Mod. Phys.* **81**, 1301–1350 (2009). URL `http://link.aps.org/doi/10.1103/RevModPhys.81.1301`.

[2] Jennewein, T., Achleitner, U., Weihs, G., Weinfurter, H. & Zeilinger, A. A fast and compact quantum random

number generator. *Review of Scientific Instruments* **71**, 1675–1680 (2000). URL http://dx.doi.org/10.1063/1.1150518.

[3] Bloch, I., Dalibard, J. & Nascimbene, S. Quantum simulations with ultracold quantum gases. *Nat Phys* **8**, 267–276 (2012). URL http://dx.doi.org/10.1038/nphys2259.

[4] Ray, P., Chakrabarti, B. K. & Chakrabarti, A. Sherrington-Kirkpatrick model in a transverse field: Absence of replica symmetry breaking due to quantum fluctuations. *Phys. Rev. B* **39**, 11828–11832 (1989). URL http://link.aps.org/doi/10.1103/PhysRevB.39.11828.

[5] Finnila, A., Gomez, M., Sebenik, C., Stenson, C. & Doll, J. Quantum annealing: A new method for minimizing multidimensional functions. *Chemical Physics Letters* **219**, 343 – 348 (1994). URL http://www.sciencedirect.com/science/article/pii/0009261494001170.

[6] Kadowaki, T. & Nishimori, H. Quantum annealing in the transverse Ising model. *Phys. Rev. E* **58**, 5355–5363 (1998). URL http://link.aps.org/doi/10.1103/PhysRevE.58.5355.

[7] Muhly, J. The beginnings of metallurgy in the old world. In Maddin, R. (ed.) *The beginning of the use of metals and alloys*, 2–20 (Cambridge, MA: MIT Press., 1988).

[8] Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).

[9] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953). URL http://link.aip.org/link/?JCP/21/1087/1.

[10] Martoňák, R., Santoro, G. E. & Tosatti, E. Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model. *Phys. Rev. B* **66**, 094203 (2002). URL http://link.aps.org/doi/10.1103/PhysRevB.66.094203.

[11] Santoro, G. E., Martoňák, R., Tosatti, E. & Car, R. Theory of quantum annealing of an Ising spin glass. *Science* **295**, 2427–2430 (2002). URL http://dx.doi.org/10.1126/science.1068774.

[12] Brooke, J., Bitko, D., F., T., Rosenbaum & Aeppli, G. Quantum annealing of a disordered magnet. *Science* **284**, 779–781 (1999). URL http://www.sciencemag.org/content/284/5415/779.abstract. http://www.sciencemag.org/content/284/5415/779.full.pdf.

[13] Johnson, M. W. *et al.* Quantum annealing with manufactured spins. *Nature* **473**, 194–198 (2011). URL http://dx.doi.org/10.1038/nature10012.

[14] Boixo, S., Albash, T., Spedalieri, F. M., Chancellor, N. & Lidar, D. A. Experimental signature of programmable quantum annealing (2012). URL http://arxiv.org/abs/1212.1739. 1212.1739.

[15] Barahona, F. On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical and General* **15**, 3241 (1982). URL http://stacks.iop.org/0305-4470/15/i=10/a=028.

[16] Choi, V. Minor-embedding in adiabatic quantum computation: I. the parameter setting problem. *Quantum Information Processing* **7**, 193–209 (2008). URL http://dx.doi.org/10.1007/s11128-008-0082-9.

[17] Choi, V. Minor-embedding in adiabatic quantum computation: Ii. minor-universal graph design. *Quantum Information Processing* **10**, 343–353 (2011). URL http://dx.doi.org/10.1007/s11128-010-0200-3.

[18] Hen, I. & Young, A. P. Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems. *Phys. Rev. E* **84**, 061152 (2011). URL http://link.aps.org/doi/10.1103/PhysRevE.84.061152.

[19] Farhi, E. *et al.* Performance of the quantum adiabatic algorithm on random instances of two optimization problems on regular hypergraphs. *Phys. Rev. A* **86**, 052334 (2012). URL http://link.aps.org/doi/10.1103/PhysRevA.86.052334.

[20] Bian, Z., Chudak, F., Macready, W. G., Clark, L. & Gaitan, F. Experimental determination of Ramsey numbers with quantum annealing (2012). URL http://arxiv.org/abs/1201.1842. 1201.1842.

[21] Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, G. & Aspuru-Guzik, A. Finding low-energy conformations of lattice protein models by quantum annealing. *Scientific Reports* **2** (2012).

[22] Harris, R. *et al.* Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B* **82**, 024511 (2010). URL http://link.aps.org/doi/10.1103/PhysRevB.82.024511.

[23] Kuegel, A. Improved exact solver for the weighted max-sat problem. In Berre, D. L. (ed.) *POS-10*, vol. 8 of *EPiC Series*, 15–27 (EasyChair, 2012).

[24] Rendl, F., Rinaldi, G. & Wiegele, A. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Math. Programming* **121**, 307 (2010).

[25] Spin glass server. URL http://www.informatik.uni-koeln.de/spinglass/.

[26] Dechter, R. Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence* **113**, 41–85 (1999).

[27] Battaglia, D. A., Santoro, G. E. & Tosatti, E. Optimization by quantum annealing: Lessons from hard satisfiability problems. *Phys. Rev. E* **71**, 066707 (2005).

[28] Dayal, P. *et al.* Performance limitations of flat-histogram methods. *Phys. Rev. Lett.* **92**, 097201 (2004). URL http://link.aps.org/doi/10.1103/PhysRevLett.92.097201.

[29] Alder, S., Trebst, S., Hartmann, A. K. & Troyer, M. Dynamics of the Wang-Landau algorithm and complexity of rare events for the three-dimensional bimodal Ising spin glass. *Journal of Statistical Mechanics: Theory and Experiment* **2004**, P07008 (2004). URL http://stacks.iop.org/1742-5468/2004/i=07/a=P07008.

[30] Bhanot, G., Duke, D. & Salvador, R. A fast algorithm for the Cyber 205 to simulate the 3d Ising model. *J. Stat. Phys.* **44**, 985–1002 (1986).

[31] Heuer, H.-O. A fast vectorized Fortran 77 program for the Monte Carlo simulation of the three-dimensional Ising system. *Comput. Phys. Commun.* **59**, 387 – 398 (1990). URL http://www.sciencedirect.com/science/article/pii/0010465590901865.

[32] Rieger, H. & Kawashima, N. Application of a continuous time cluster algorithm to the two-dimensional random quantum Ising ferromagnet. *Eur. Phys. J. B* **9**, 233–236 (1999).

[33] Martoňák, R., Santoro, G. E. & Tosatti, E. Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model. *Phys. Rev. B* **66**, 094203 (2002). URL http://link.aps.org/doi/10.1103/PhysRevB.66.094203.

[34] Moessner, R. & Sondhi, S. L. Ising models of quantum frustration. *Phys. Rev. B* **63**, 224401 (2001). URL http://link.aps.org/doi/10.1103/PhysRevB.63.224401.

[35] Amin, M. H. S. & Choi, V. First order quantum phase transition in adiabatic quantum computation. *Phys. Rev. A* **90**, 062326 (2009).

[36] Bapst, V., Foini, L., Krzakala, F., Semerjian, G. & Zamponi, F. The quantum adiabatic algorithm applied to random optimization problems: The quantum spin glass perspective. *Physics Reports* **523**, 127 – 205 (2013). URL http://www.sciencedirect.com/science/article/pii/S037015731200347X. ¡ce:title¿The Quantum Adiabatic Algorithm Applied to Random Optimization Problems: The Quantum Spin Glass Perspective¡/ce:title¿.

[37] Hen, I. Excitation gap from optimized correlation functions in quantum Monte Carlo simulations. *Phys. Rev. E* **85**, 036705 (2012). URL http://link.aps.org/doi/10.1103/PhysRevE.85.036705.