From local to global ground states in Ising spin glasses

Ilia Zintchenko,[1] Matthew B. Hastings,[2,3] and Matthias Troyer[1]
[1]*Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland*
[2]*Station Q, Microsoft Research, Santa Barbara, California 93106-6105, USA*
[3]*Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA*
(Received 13 August 2014; published 5 January 2015)

We consider whether it is possible to find ground states of frustrated spin systems by solving them locally. Using spin glass physics and Imry-Ma arguments in addition to numerical benchmarks we quantify the power of such local solution methods and show that for the average low-dimensional spin glass problem outside the spin glass phase the exact ground state can be found in polynomial time. In the second part we present a heuristic, general-purpose hierarchical approach which for spin glasses on chimera graphs and lattices in two and three dimensions outperforms, to our knowledge, any other solver currently around, with significantly better scaling performance than simulated annealing.

DOI: 10.1103/PhysRevB.91.024201 PACS number(s): 05.10.−a, 75.10.Nr, 75.50.Lk

I. INTRODUCTION

The combination of disorder and frustration in spin glasses [1] creates a complex energy landscape with many local minima that makes finding their ground states a formidable challenge. In particular finding the assignments of spins $s_i = \pm 1$ which minimizes the total energy of an Ising spin glass with Hamiltonian

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

where $s_i = \pm 1$ and $J_{ij}, h_i \in \mathbb{R}$, is nondeterministic polynomial (NP) hard [2] and no polynomial time algorithm is known for the hardest instances. NP-hardness also means that any problem in the complexity class NP can be mapped to an Ising spin glass with only polynomial overhead. This includes the traveling salesman problem, satisfiability of logical formulas, and many other hard optimization problems. Explicit mapping for a number of these problems have recently been given in Ref. [3]. Efficient solvers for Ising spin glass problems hence can have an impact far beyond spin glass physics.

This broad spectrum of applications has also motivated the development of the devices by the Canadian company D-Wave Systems [4–7]. These devices have been designed to employ quantum annealing [8] for Ising spin glass problems using superconducting flux qubits. However, it has not yet been shown that they can outperform classical devices [9,10]. Determining the complexity of solving the spin glass problems on the so-called "chimera graph," which is implemented by the hardware of the D-Wave devices, and finding the best classical algorithms for them is important in the search for quantum speedup on these devices [10].

Motivated by these comparisons and the importance of efficiently solving Ising spin glass problems, here we consider the complexity of solving such problems for random spin glass instances on finite-dimensional lattices, including the chimera graph. In Sec. II we discuss the effects of nonzero temperature and magnetic field on Ising spin glasses and argue that the absence of correlations outside the spin glass phase allows for polynomial time algorithms. Section III presents an exact solver based on this idea which solves the system quasilocally by considering finite patches of the lattice. Finally, in Sec. IV we present a hierarchical heuristic approach, which recursively solves groups of spins by splitting each group into smaller subgroups. For our benchmark problems on two- and three-dimensional periodic lattices and chimera graphs with random disorder this approach outperforms, to our knowledge, any other solver currently available and scales significantly better than simulated annealing. While we give a qualitative explanation of the advantage of the hierarchical solver, it remains an open theoretical question to give a quantitative argument for its improved scaling. The interested reader can skip directly to this section as it can be understood independently of the scaling analysis earlier in the paper.

II. BOUNDARY CONDITION DEPENDENCE IN FRUSTRATED SPIN SYSTEMS

It is evident that if the fields h_i in Eq. (1) are very large, the problem can be solved by simply aligning each spin relative to the field. The problem becomes more difficult at smaller h_i, and the meaningful question is whether a phase transition intervenes at some nonzero value of the field strength, where the difficulty increases greatly. In this section, we argue that the relevant transition indeed is already known in the literature, where it is referred to as the de Almeida–Thouless line. We argue that above this transition (which happens for *any* nonzero random choice of h_i, J_{ij} in two dimensions), the problem can be solved by considering larger patches of spins, with the patch size diverging as the field strength goes to zero; the spins in the middle of these large patches become independent of those outside the patch and can be fixed using a local algorithm. We first review the relevant literature at $h_i = 0$ and the scaling theory at small h_i.

A. Review

In the particular ensemble where fields vanish ($h_i = 0$), the behavior of the model depends strongly on both the dimensionality of the system and upon the choice of the ensemble for the couplings between spins. In this discussion, we will focus on the case of a continuous distribution, e.g., a Gaussian one, with vanishing mean.

We will also refer to results in the literature that study nearest-neighbor couplings on a square or cubic lattice, rather than the chimera graph. One important distinction between the two-dimensional square lattice and the chimera graph is that for the square lattice, as for any planar graph, if the magnetic fields vanish there are efficient polynomial time matching algorithms for finding exact ground states [11], while on nonplanar graphs, such as the chimera, it is NP-hard. We discuss this further below.

In two dimensions it is accepted that there is no spin glass phase at temperature $T > 0$ [12]. To quantify this, consider a pair of sites i, j. Let $\langle \ldots \rangle$ denote the thermal average of an operator at temperature T and let $[\ldots]_H$ denote the disorder average over Hamiltonians H. Since the couplings are chosen with zero mean, we have that $[\langle s_i s_j \rangle]_H = 0$ exactly. However, generically the ground state is unique and hence $[(\langle s_i s_j \rangle)^2]_H = 1$ at $T = 0$, and this average is expected to be positive at $T > 0$; however, the average vanishes in the limit of large distances between i, j.

The reason for the absence of a spin glass phase is that it costs very little energy to flip a domain of spins. Consider flipping a cluster of spins of linear size ℓ. In a ferromagnetic state, this costs energy proportional to ℓ. In a spin glass ground state, it is possible, however, that a cluster can be found which costs very low energy to flip. Using various methods of generating flipped patches (by for example boundary condition changes), it is found that the energy of the domain wall scales proportional to ℓ^θ with $\theta \approx -0.282(2)$ [12]. Thus, it costs *less* energy to flip larger clusters, and no matter how small T is, for $T > 0$ there eventually will be some ℓ such that flipping clusters at that scale costs energy smaller than T. Hence there will be many thermally excited domain walls. On the other hand, for three dimensions and higher, there is believed to be a phase transition temperature $T_c > 0$ with a domain wall exponent $\theta > 0$ for excitations above the ground state [13].

Similarly, we can consider random models with nonzero fields [14–16] and denote standard deviation of the field magnitude by h. In this case, we consider the quantity

$$[(\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle)^2]_H.$$

If this quantity tends to a nonzero limit at large distance between i, j, then we term this a spin glass phase. It has been shown that such a spin glass phase can exist in a mean-field model at $h \neq 0$ [17]; the line in the h-T plane separating the spin glass from the paramagnetic phase is termed the de Almeida–Thouless line. However, it is unclear whether such a spin glass phase at $h \neq 0$ can persist in a local finite-dimensional model. Numerical work [18,19] suggests that it exists for dimension $d > d_c = 6$. However, it is accepted that the spin glass phase at $h \neq 0$ does not persist in dimension $d = 2$ and in the next subsection we will explain why this is expected given the exponent θ discussed above.

It should be emphasized that it is not necessarily difficult to find ground states in a spin glass phase, as exemplified by the matching algorithm for the planar case in $d = 2$ at $h = 0$. Conversely, even if a random ensemble is not in the spin glass phase, particular instances may be difficult, as exemplified by the fact that in $d = 2$ at $h \neq 0$ the model is not in a spin glass phase, but finding the ground state of arbitrary instances is still NP-hard.

B. Weak-field scaling in $d = 2$

We now consider the effect of a weak magnetic field $h \neq 0$ in $d = 2$. Our general goal is to show that in this case, we expect that the value of a given spin in the ground state can often be fixed using a purely *local* calculation. The argument is a version of the Imry-Ma argument applied to disordered systems [20] and in the specific application to spin glasses is an example of the droplet picture [21]. We conjecture that a similar argument (with different exponents) will work if there is no de Almeida–Thouless line (i.e., whenever there is no spin glass phase at nonzero magnetic field).

Consider a spin s_{cent} at the center of a patch of size ℓ inside a larger system of linear size L. Suppose that we have found some configuration of spins which is a ground state. At $h = 0$, it is impossible to know whether $s_{\text{cent}} = +1$ or $s_{\text{cent}} = -1$ without knowing the value of the boundary spins because there is a Z_2 symmetry. However, at $h \neq 0$, it may be possible to determine the value of the spin s_i independently of the value of the boundary spins. That is, there may be some choice (either $s_{\text{cent}} = +1$ or $s_{\text{cent}} = -1$) that minimizes the energy inside the patch for *all* choices of boundary spins. In this case, we know that in the global ground state the spin s_{cent} will take the given value.

To analyze the ability to fix the spin independently of boundary conditions, we again begin with the case $h = 0$ to develop a scaling argument that will apply at small h. Consider a given configuration of boundary spins, which we write as $\vec{s}_{\text{bdry}}$, where we write this as a vector to emphasize that there are many boundary sites. At $h = 0$, we can minimize the energy inside the patch for this choice of boundary spins, uniquely fixing all spins inside the patch. Suppose that this minimization gives $s_{\text{cent}} = +1$. Now consider the case in which we force $s_{\text{cent}} = -1$, defining a new configuration of spins inside the patch which minimizes the energy subject to the given boundary conditions $\vec{s}_{\text{bdry}}$ and given that $s_{\text{cent}} = -1$. Forcing s_{cent} to take the opposite value will flip also a cluster of spins around the central spin, creating a domain wall around that cluster of spins, as shown in Fig. 1. The energy of this domain wall will be proportional to ℓ^θ which therefore decreases with increasing ℓ.

The number of spins in the cluster scales also as a power of ℓ, with the power slightly less than [22] 2; in that reference, the exponent 1.80(2) was found for one specific method of constructing droplets. Our numerical studies, shown in Fig. 2, indicate that the number scales as $\ell^{d_{\text{clust}}}$, with a fractal dimension $d_{\text{clust}} \approx 1.84$; while this dimension might revert to 2 for larger system sizes, we use the fractal dimension extracted at these system sizes to facilitate comparison with our complexity analysis below.

This cluster then defines a larger effective spin. The cost to flip this effective spin relative to the rest of the patch is proportional to ℓ^θ. We now consider the case that $h \neq 0$, and analyze the effect of the nonzero h on this effective spin. Given that the magnetic fields acting on the spins in the cluster are chosen randomly, we expect that the cluster will experience an effective magnetic field

$$h_{\text{eff}} \propto h \ell^{d_{\text{clust}}/2}. \tag{2}$$

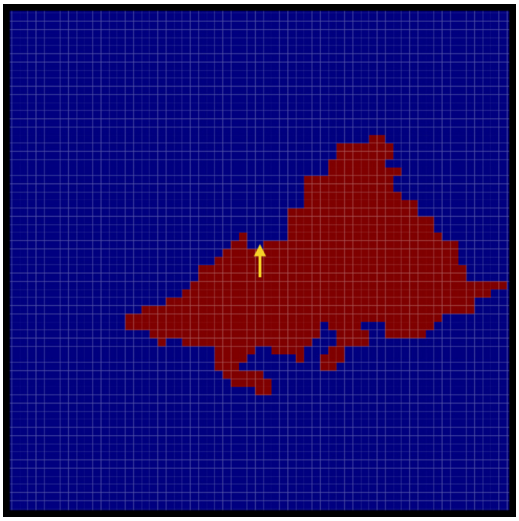

FIG. 1. (Color online) If central spin is forced to be opposite to its optimal orientation while keeping the spins on the boundary of the patch fixed, a cluster of spins around it will also flip. Central spin marked in yellow, flipped cluster marked in red, and the boundary in black.

Balancing these energy scales, we find that

$$h_{\text{eff}} \sim \ell^{\theta} \rightarrow \ell \propto h^{-\frac{1}{-\theta+d_{\text{clust}}/2}}. \tag{3}$$

In Fig. 3, we show our estimate for $\theta \approx -0.33$ obtained from the defect energy E_d gained when the central spin is forced in opposite direction to its optimal with a fixed boundary configuration around a patch. While this differs slightly from the numbers quoted above, we remark that many different ways of forcing domain walls in have been considered in the literature, such as flipping a central spin as here or changing global boundary conditions and these may give rise to different values, especially for finite sizes; see Refs. [22–24] for various possibilities. Thus, we get that

$$\ell \propto h^{-0.8}. \tag{4}$$

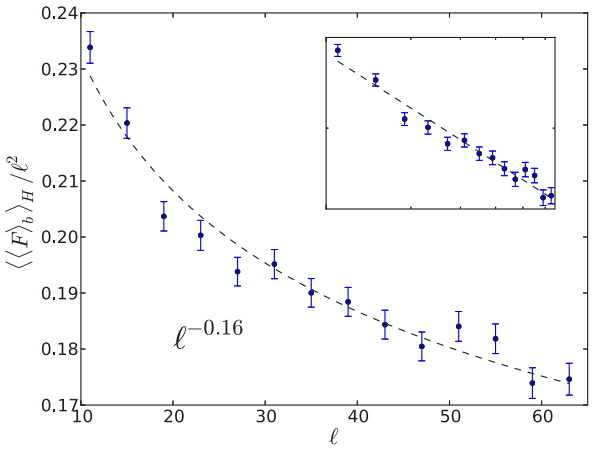

FIG. 2. (Color online) Number of spins in the cluster F averaged over different boundary configurations and Hamiltonians vs linear dimension of patch ℓ. The data are fitted to a power law (dashed line). Inset shows log-log scale.

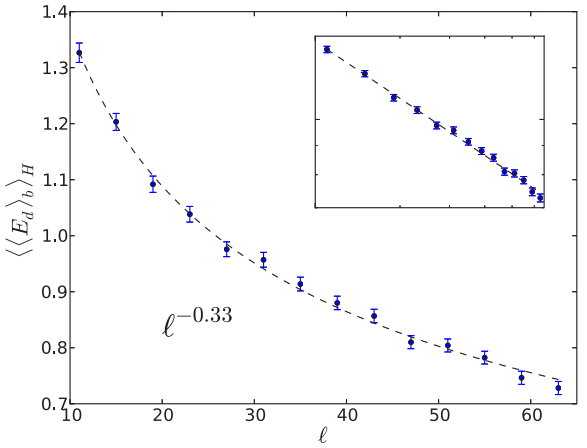

FIG. 3. (Color online) Defect energy E_d averaged over different boundary configurations and Hamiltonians vs linear dimension of patch ℓ. The data are fitted to a power law (dashed line). Inset shows log-log scale.

For ℓ larger than this number, the coupling of the cluster to the effective field exceeds its coupling to the rest of the patch, so that the value of the central spin can be fixed *independently* of boundary conditions. Note that this analysis focuses on one possible way to fix in which the central spin can become independent of boundary conditions; others may be possible.

C. Boundary condition dependence

The above scaling analysis gives an estimate of the length scale at which we can fix the central spin in a patch. The total number of spins which can be fixed in the system depends on the local fields and patch size and can be estimated from the probability of fixing a single spin. To quantify this probability we define

$$\chi_B(h,\ell) = 1 - [([s_c]_B)^2]_H, \tag{5}$$

where s_c is the central spin and where $[\ldots]_B$ denotes the average over boundary conditions. We term this quantity χ_B as it measures the response of the central spin to change in boundary conditions. If this quantity is equal to 0, then the spin can be fixed independently of boundary conditions as it assumes the same value for all choice of boundary.

For this averaged quantity, we find a scaling collapse as shown in Fig. 4. The scaling collapse onto a single curve is implemented by defining the scaling variable $\xi = h\,\ell^{1.19}$. This implies a scaling

$$\ell \sim h^{1/1.19} = h^{0.840\ldots}.$$

This should be compared with the estimate in Eq. (4); the agreement of exponents is reasonable, and if we use $\theta = -0.28$ instead of our measured $\theta = -0.33$ the agreement becomes more accurate. We find that the scaling collapse can be approximately fitted by the form

$$\chi_B(h,\ell) = \exp\left[-\text{poly}(h,\ell)\right]. \tag{6}$$

To obtain statistical information about whether we can fix a spin independently of boundary, it suffices to determine

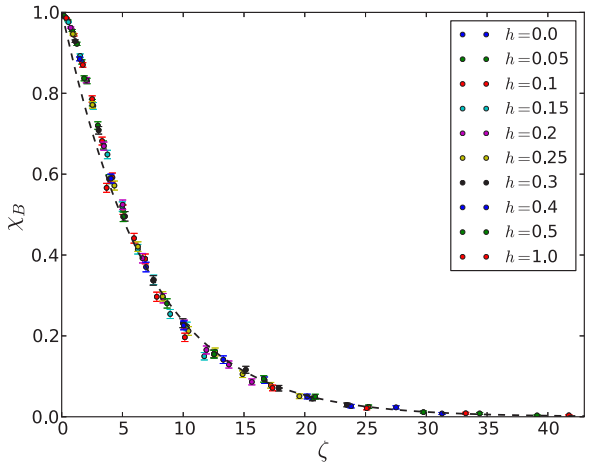

FIG. 4. (Color online) Correlation of central spin with the boundary for different fields and patch sizes. $\zeta = h\,\ell^{1.19}$. The fit of $\chi_B(h,\ell)$ is to $2^{-a\,\zeta^b}$, where $a = 0.198$ and $b = 1.02$ (dashed line).

the behavior of χ_B in the tail; see Fig. 4, where we fit $\chi_B = 2^{-a\,\zeta^b}$ with the constants a and b. We cannot be completely confident about the tail behavior of χ_B at large h,ℓ from these simulations, but let us use this estimate to try to determine the complexity of a simple solver which tries to solve each spin by taking a sufficiently large patch that $\chi_B = 0$. The complexity of the solver will depend on the scaling of χ_B, but we will estimate that it takes a polynomial time (in N) for any nonzero h. We will find in the next section that we can improve on this, by using the fact that once a single spin is fixed it simplifies the fixing of other spins.

Since there are only $2^{4\ell}$ possible boundary conditions, the minimum nonzero value of χ_B is of order $2^{-4\ell}$. Considering the N possible choices of central spin, only $O(1)$ spins correlate with the boundary if $\exp[-\mathrm{poly}(h,\ell)] = O(1/N)2^{-4\ell}$. Equivalently, this holds if $2^{4\ell}\exp[-\mathrm{poly}(h,\ell)] = O(1/N)$; since $\zeta^b > 1$, the scaling of the left-hand side of this equation is dominated by the second term. Hence, the equation will hold when

$$h\,\ell^{1.19} \sim \log_{10}(N)^{1/b}. \tag{7}$$

Thus, we expect that for ℓ larger than this, it will be possible to fix all spins.

Since each patch can be solved exactly with complexity $\exp\{\ell\}$ using a dynamic programming method [25], at a fixed h the whole system can be solved with complexity

$$\mathrm{poly}(L)\exp\{h^{-1/1.19}(\log_{10} N)^{1/(1.19\cdot b)}\}. \tag{8}$$

Since $b > 1$ and $1/(1.19\,b) < 1$, the exponential term is sub-linear and the total complexity is therefore polynomial; however, it diverges as $h \to 0$. The exact estimate may depend sensitively upon the tail of the curve which we cannot determine with full confidence.

It should, however, be emphasized that the data in Fig. 4 arise only from an average over a finite number (in this case, 1000) of boundary conditions. This finite number was chosen to enable rapid sampling of the curve. To exactly solve a specific sample, we need to consider *all* possible boundary conditions, as discussed in the next section.

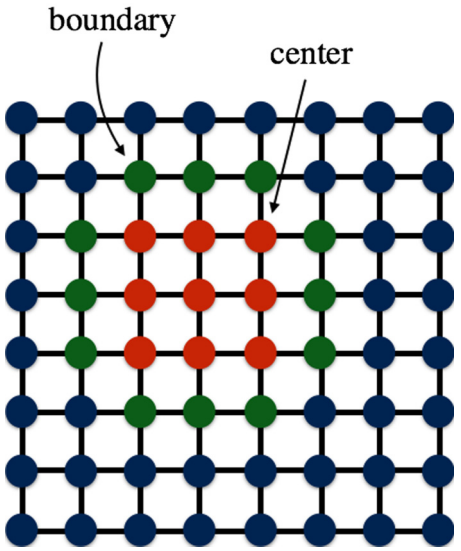

FIG. 5. (Color online) Illustration of a patch. Central spins are marked in red, boundary spins in green, and the rest of the spins in the system are blue.

III. FINDING THE EXACT GLOBAL GROUND STATE

Following the argument above, correlations in a typical finite-dimensional lattice decay exponentially if $h > 0$ and the ground state for such a system can therefore be found in polynomial time as the optimal orientation of single spins can be determined with high probability by considering only finite regions of the system. Furthermore, even for zero fields we present strong numerical evidence that the typical two-dimensional case can be solved in polynomial time with a more general approach which we describe below.

A. Single-spin reduction

Let us consider a spin in the center of a patch in our system. If for all boundary configurations of the patch the optimal orientation of the central spin is the same, then it is independent of the boundary and can thus be fixed to that value. Based on this idea, the simplest way to find the ground state is by determining the optimal orientation of each spin independently by building a patch around it and checking whether the optimal orientation of the central spin is independent of the boundary. If this is not the case, we increase the patch size and check again until the spin becomes independent of the boundary. When all spins are fixed the system is solved.

This approach can be further improved by solving the system similar to a crossword puzzle rather than considering each spin independently. If a spin gets fixed, this will reduce the number of possible configurations for patches containing that spin, which in return may allow more spins to get fixed without increasing the patch sizes.

Fixing single spins is a simple algorithm which can be very efficient for systems with large fields. In the limit of very large fields the complexity approaches $O(N)$ as each spin

becomes independent of its neighbors. However, for small fields the computational effort increases as the correlation length diverges when the field approaches zero requiring patches comparable to the total system size. A more general approach, discussed next, remains effective in that limit.

B. Patch reduction

Instead of only attempting to fix the central spin, correlations between spins inside a patch can be captured by considering all possible configurations of a patch that minimize the energy for a given choice of boundary conditions (see Fig. 5). These configurations are then constrained by requiring consistency between overlapping patches. We find numerically that this approach is significantly more efficient than the single-spin algorithm.

The algorithm starts with a small patch size (e.g., a single spin in the center) and sequentially builds patches around each spin. For each boundary configuration of a given patch we store the configuration of the boundary together with the corresponding optimal configuration of the center spins. If the local ground state of a patch turns out to be degenerate for a given boundary condition, we arbitrarily pick any of these configurations if our aim is to obtain just one of the potentially degenerate global ground states. Note that if instead we are interested in finding all ground states, then for each boundary configuration all degenerate interior configurations need to be stored.

The number of potential ground state configurations within a patch (boundary and interior) is then further reduced by removing those configurations which are inconsistent with the constraints imposed by overlapping patches.

After a pass through all spins we increase the patch size and repeat the above steps with larger patches until only a single configuration remains or all remaining configurations have the same energy. As the patch size increases, the set of configurations which satisfy all constraints is strongly reduced and typically scales much better than the exponential worst case.

C. Improved patch reduction

One way to significantly reduce the cost of storing configurations is by removing some spins from the system. If for a pair of neighboring spins s_i and s_j, their product $s_i s_j$ is constant in all configurations, they can be replaced by a single spin. If only one ground state is targeted, this procedure will finally eliminate all but one spin. More generally, any arbitrary spin can be removed by replacing it with multispin interactions such that for each configuration of the neighboring spins the local energy is conserved given that the spin to be removed aligns optimally with respect to its neighbors.

D. Empirical scaling

As shown in Figs. 6 and 7 the median time to solution appears to scale polynomially in the number of spins for all values of the field h, including zero field [26]. While faster specialized exact solvers are available [27], this algorithm is not necessarily intended as a general purpose optimizer, but rather to demonstrate polynomial scaling in the number of spins at all values of h for typical low-dimensional spin glass instances.

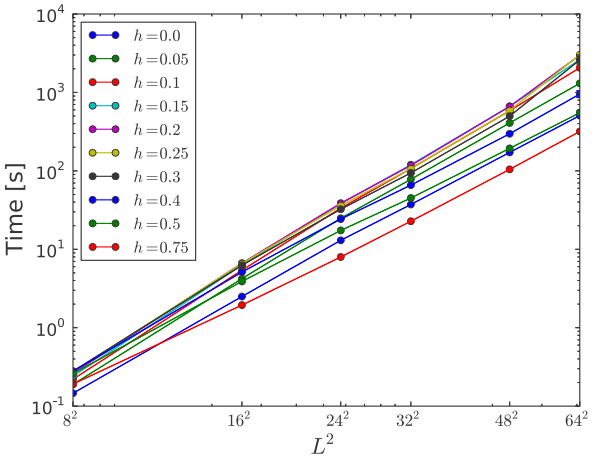

FIG. 6. (Color online) Median wall clock time (in seconds) for different system sizes and various fields.

IV. HIERARCHICAL SEARCH

A. Motivation

In this section we present a general purpose heuristic hierarchical algorithm for finding the ground state of Ising spin glasses based on recursively optimizing groups of variables. Before describing the algorithm we motivate why solving groups of variables is significantly more efficient than solving the whole system at once.

The arguably simplest heuristic algorithm for finding the ground state is by generating random spin configurations and recording the energy, in other words random guessing. The probability to find the global ground state of N spins this way is trivially 2^{-N} per guess, assuming for simplicity a nondegenerate ground state in the discussion here and below. A more sophisticated way to guess the solution is to generate random configurations of only $N_r = N - N_g$ spins and for

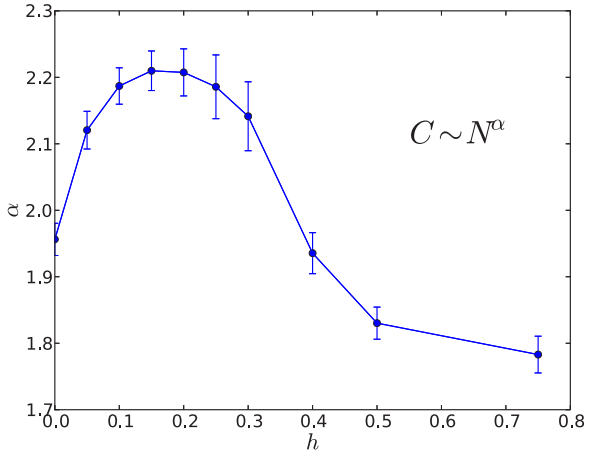

FIG. 7. (Color online) Scaling exponent for the runtime shown in Fig. 6, obtained from fitting the data to a power law, for different system sizes and various fields.

each configuration find the lowest energy of the remaining N_g spins by some other algorithm, e.g., by enumerating all possible combinations or any other more optimized algorithm. This improves the probability of guessing the correct solution, but as the cost of finding the optimal orientation of the remaining N_g variables may be as much as 2^{N_g}, we might not have gained much. This idea can, however, be extended to solving multiple groups. Let us consider two groups with N_1 and N_2 spins respectively, chosen such that spins in one group do not couple to any of the spins in the other group. For each random guess of the remaining $N_r = N - N_1 - N_2$ spins, the complexity of finding the optimal configuration of both of them with respect to the rest of the system is $2^{N_1} + 2^{N_2}$, thus reducing the total complexity by an exponential amount from $2^N = 2^{N_r+N_1+N_2}$ to $2^{N_r}(2^{N_1} + 2^{N_2})$. In our algorithm, described below, we find a significant reduction in complexity even if spins in subgroups are coupled and overlap with each other.

B. Optimization of groups

The above argument provides a basis for a simple algorithm to find the global ground state by iteratively optimizing groups of spins. We start with a random global state, sequentially pick M groups with N_g spins each, and optimize their configurations by calling some—as yet unspecified—solver as follows:

procedure SOLVE ()
  initialize random spin configuration
  for $j \in \{1 \dots M\}$ **do**
    pick a random spin i
    build group G of size N_g around spin i
    $\vec{\sigma} \leftarrow$ SOLVE GROUP (G)
    UPDATE CONFIGURATION(G,$\vec{\sigma}$)
  end for
end procedure

Here, SOLVE GROUP(G) is a solver that solves the group G (taking into account the interaction with spins outside G to produce an effective field) and returns an optimized configuration $\vec{\sigma}$ for the spins in that group. The procedure UPDATE CONFIGURATION(G,$\vec{\sigma}$) updates the spins inside G to configuration $\vec{\sigma}$; if the solver SOLVE GROUP is a heuristic solver, then UPDATE CONFIGURATION (G,$\vec{\sigma}$) only makes this change if the energy is lowered. Alternatively one may also consider an algorithm which replaces the group configuration probabilistically with a Metropolis-type criterion or similar.

If we pick trivial groups of size $N_g = 1$, consisting of a single spin, the group solver just returns the spin direction which minimizes its energy with respect to its neighbors. For larger groups—as will usually be the case—we can use any arbitrary exact or heuristic solver, including potentially special purpose classical or quantum hardware. We note in passing that in the case of $N_g = 1$, if the new configuration is accepted probabilistically depending on its energy this algorithm reduces to simulated annealing.

C. Hierarchical recursive algorithm

If solving a given system in groups is more efficient than solving the whole system at once, performance can be increased even further by solving each group by subdividing it recursively into subgroups, thus giving a hierarchical version of the algorithm. That is, in the pseudocode written above, we could use the function SOLVE(), restricted to the spins in a group, as the solver SOLVE GROUP(). The recursion terminates at some (small) group size, which is solved by another algorithm.

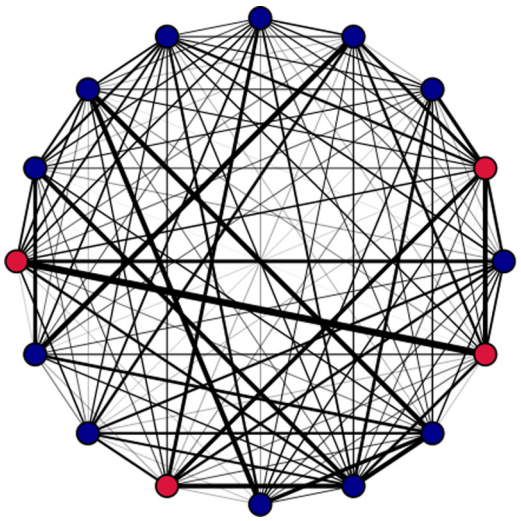

FIG. 8. (Color online) Optimal group, marked in red, of 4 spins in complete graph of 16 spins with Gaussian disorder.

Note that the hierarchical scheme randomizes the configuration of each group before solving it by optimizing subgroups, thus implementing random local restarts without affecting the global spin configuration. This randomization also implies that it makes no sense to solve a particular group more than once in a row, but rather a new group should be chosen after one group has been optimized. It should be emphasized that random restarting is just one possible way to initialize the state of a group and the one we used here. Other ways are possible and could be more efficient.

The total complexity of the hierarchical algorithm is dominated by the number of calls to the solver for the bottom level group rather than by the group size at each level. This is because for a given group of size N_g, the effort to calculate the local energy and randomize spins is at most $O(N_g^2)$ for dense

TABLE I. Optimal parameters for chimera graphs with random bimodal disorder. N is the system size, M is the number of groups, N_g is the group size, S_g is the number of simulated annealing sweeps per group, S is the number of sweeps for plain simulated annealing.

N	M	N_g	S_g	S
32	78	9	1	8
72	80	37	5	24
128	100	60	7	64
200	349	41	4	192
288	408	68	6	400
392	500	105	13	1024
512	642	129	14	2048

TABLE II. Optimal parameters for chimera graphs with cluster bimodal disorder. The parameters have the same meaning as in Table I.

N	M	N_g	S_g	S
32	28	8	1	4
72	237	9	1	4
128	388	9	1	4
200	197	26	3	1281
288	454	29	3	4800
392	470	27	3	24576
512	718	28	3	131072

graphs, which is typically negligible relative to the effort of finding a lower energy configuration of that group.

D. Selecting groups

Up to now, we have ignored the hard problem of how to best pick groups. Here we provide a simple strategy that turned out to work well. Intuitively, in a well chosen group spins are strongly coupled to each other and more weakly coupled to the rest of the system; see Fig. 8. We thus build a group G by starting from one spin and greedily adding spins until the group G has reached the desired size. We add the spin i that maximizes $W_i = \sum_{j\in G} |J_{ij}| - \sum_{j\notin G} |J_{ij}|$, if this maximum is positive, and a random neighbor of one of the spins in G otherwise.

Other ways of building a group may be more effective. For example, single spins could be added probabilistically, or instead of single spins we could consider sets of spins which can be added to the group. Such improvements will be discussed in follow-up work.

E. Results

To test the performance of our algorithm we compare it to simulated annealing, which is currently one of the most versatile and efficient solvers for finding ground states of spin glasses. As mentioned above, simulated annealing is a special case of our algorithm. For our benchmarks we perform a hierarchical search with two levels, using simulated annealing to solve groups of size N_1 with the optimized configuration being accepted if its energy is lower or the same as the current configuration.

As a measure of complexity we use the median total number of spin updates required to find the ground state with a target probability $p_0 = 0.99$. Since a heuristic algorithm will find the ground state with some probability $p_s < 1$ we may have to repeat the optimization multiple times if $p_s < p_0$. Assuming independent repetitions, the required number of repetitions is $R = \lceil \log_{10}(1-p_0)/\log_{10}(1-p_s) \rceil$. For each set of parameters the probability p_s was estimated by performing 1024 repetitions from random initial states.

For both algorithms and each class and size of problems we optimize the simulation parameters to minimize the median effort in terms of single-spin updates. For simulated annealing the total effort for a single repetition is SN, where S is the number of sweeps and N is the system size. We used a linear schedule in $\beta = 1/T$ where the initial and final values of inverse temperature, β_0 and β_1, respectively, as well as the number of sweeps S are chosen to minimize the total effort. We list the parameters used in Tables I–V.

For the hierarchical approach a single repetition requires a total effort MS_gN_g, where M is the number of groups, S_g is the number of simulated annealing sweeps per group, and N_g is the group size. The same annealing schedule is used for each group. The values of M, S_g, and N_g are chosen to minimize the total effort and are listed in Tables I–V.

As benchmark problems we used typical spin glass problems on two- and three-dimensional lattices: two-dimensional square lattices, three-dimensional simple cubic lattices [28], and so-called two-dimensional chimera graphs. The unit cell of the chimera graph [29], shown in Fig. 9, is a complete bipartite graph with eight vertices and is coupled to the neighboring unit cells with four edges each. Hence, each vertex has either five or six edges corresponding to four edges to spins within the unit cell and one or two edges to neighboring unit cells depending on whether it is on the edges of the graph or in the interior, respectively.

One choice of benchmark problems is spin glasses with bimodal disorder, i.e., couplings $J_{ij} = \pm 1$, and another choice will be Gaussian disorder with couplings drawn from a normal distribution with zero mean and unit variance. In all benchmarks we choose zero local fields $h = 0$.

TABLE III. Optimal parameters for two-dimensional lattices with Gaussian disorder. The parameters have the same meaning as in Table I.

N	M	N_g	S_g	S
16	72	8	1	4
64	314	9	1	48
144	273	33	21	891
256	573	47	43	30189

TABLE IV. Optimal parameters for three-dimensional lattices with Gaussian disorder. The parameters have the same meaning as in Table I.

N	M	N_g	S_g	S
27	39	10	1	5
64	118	22	4	45
125	232	30	5	512
216	271	58	23	2700
343	562	86	42	13056
512	614	113	101	61440

TABLE V. Optimal parameters for two-dimensional lattices with bimodal disorder. The parameters have the same meaning as in Table I.

N	M	N_g	S_g	S
256	235	56	10	64
400	224	119	23	227
576	384	103	17	768
784	686	208	31	3506
1024	656	151	29	7680

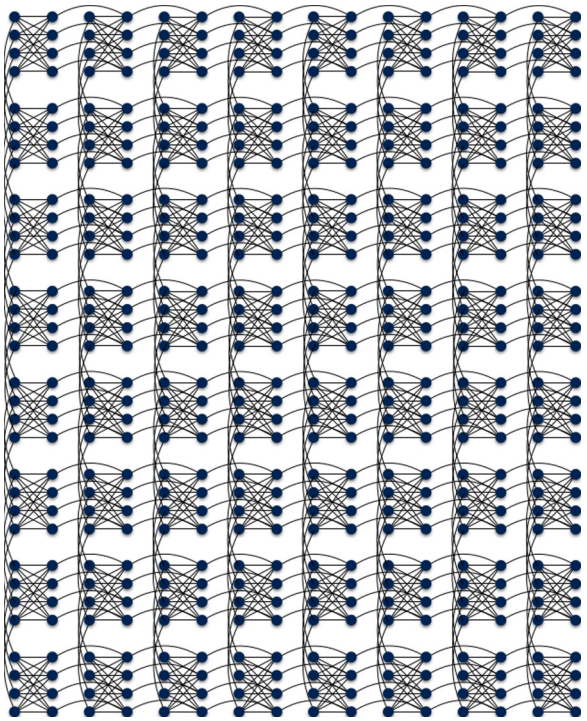

FIG. 9. (Color online) Chimera graph with 512 spins composed of an 8×8 grid of unit cells. Each unit cell is a complete bipartite graph with 8 spins.

A special benchmark problem is chimera graphs with cluster structure, which has recently been proposed as a class of problems to explore an advantage of quantum annealing over simulated annealing [30]. In these problems the spins within each unit cell are coupled ferromagnetically with $J_{ij} = -1$. Of the four edges connecting neighboring pairs of unit cells one randomly chosen edge is assigned a random coupling $J_{ij} = \pm 1$ and the rest is set to zero.

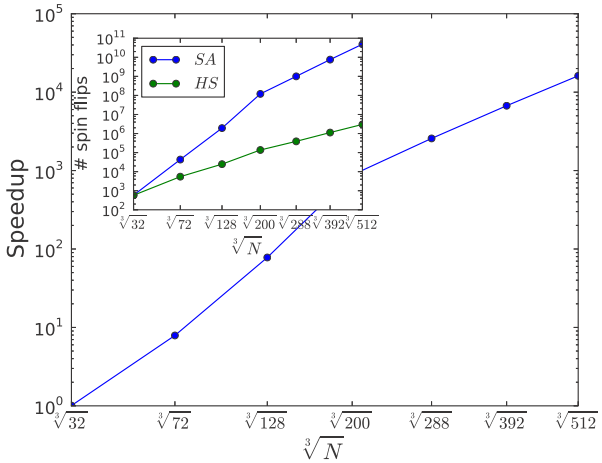

FIG. 10. (Color online) Chimera graphs with cluster bimodal disorder. Speedup and the inset are defined the same as in Fig. 13. For both plain simulated annealing and for each group, $\beta_0 = 0.1$, $\beta_1 = 3$. Optimal parameters for both algorithms are listed in Table II.

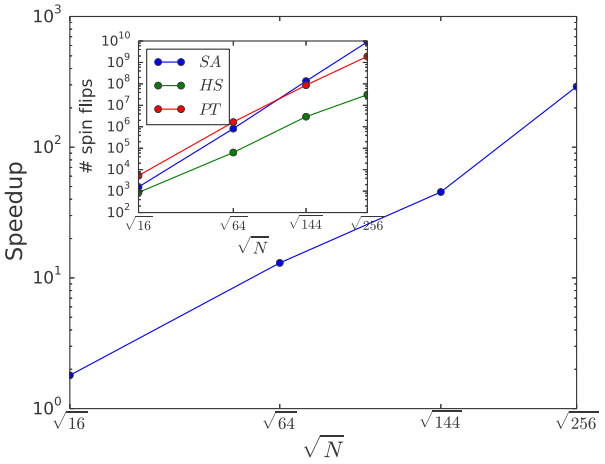

FIG. 11. (Color online) Two-dimensional square lattices with Gaussian disorder. Speedup and the inset are defined the same as in Fig. 13. As the energy gap between the ground state and first excited state decreases linearly with system size, the final temperature is also reduced with the number of spins. $\beta_0 = 0.014$, $\beta_1 = 0.037N + 2.5$ for plain simulated annealing (SA) and parallel tempering (PT) and $\beta_1 = 0.037N_g + 2.5$ for each group of the hierarchical algorithm (HS). Optimal parameters for SA and HS are listed in Table III.

In all benchmarks we find that hierarchical search performs significantly better than simulated annealing. The gain is evidently more significant for problems that are harder for simulated annealing, such as cluster chimera graphs and systems with Gaussian disorder; see Figs. 10, 11, and 12, respectively. Random bimodal disorder is significantly easier for simulated annealing and hence the speedup on those problems is smaller, although still substantial; see Figs. 13 and 14, respectively.

A further comparison was also made with parallel tempering [31], another state-of-the-art method for finding ground

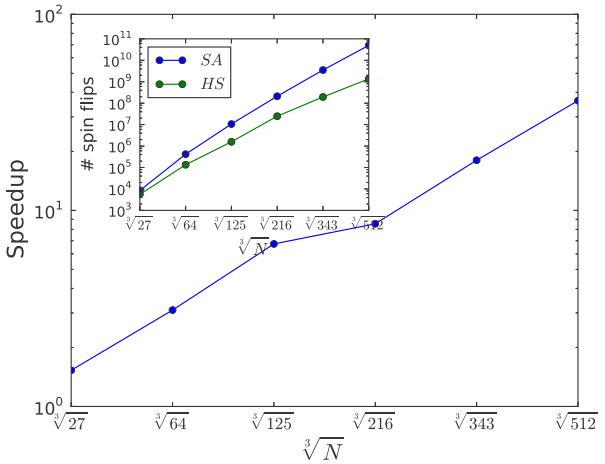

FIG. 12. (Color online) Three-dimensional cubic lattices with Gaussian disorder. Speedup and the inset are defined the same as in Fig. 13. $\beta_0 = 0.05$, $\beta_1 = 0.028N + 5.68$ for plain simulated annealing and $\beta_1 = 0.028N_g + 5.68$ for each group of the hierarchical algorithm. Optimal parameters for both algorithms are listed in Table IV.

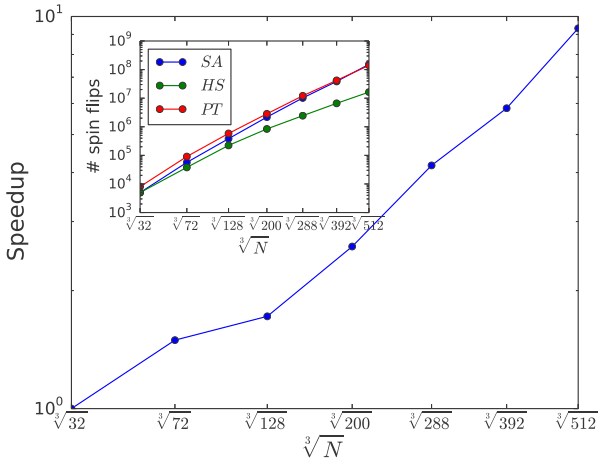

FIG. 13. (Color online) Speedup (ratio of the median number of spin updates) of hierarchical search relative to simulated annealing for chimera graphs with random bimodal disorder. The inset shows the total number of spin updates for simulated annealing (SA), parallel tempering (PT), and hierarchical search (HS). For SA, PT, and each group of HS, $\beta_0 = 0.1$, $\beta_1 = 3$. Optimal parameters for SA and HS are listed in Table I.

states of Ising spin glasses. For each class and problem size, the total number of replicas and sweeps per replica was optimized minimizing the median total number of spin updates. For a single repetition the total effort is $N_R S N$, where N_R is the number of replicas, S is the number of sweeps per replica, and N is the system size. On chimera graphs its performance is very similar to simulated annealing; see Fig. 13. On two-dimensional lattices with Gaussian disorder it performs slightly better; see Fig. 11. However, analogously to simulated annealing, its performance can be significantly improved by optimizing groups of spins rather than the whole system at once; see Fig. 15. Note that although in all cases the advantage of hierarchical search over plain simulated

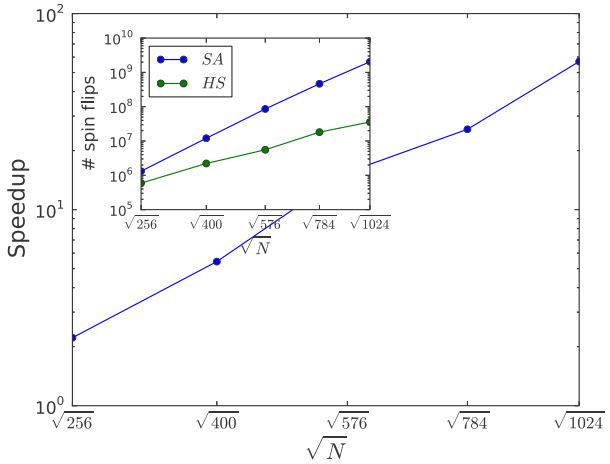

FIG. 14. (Color online) Two-dimensional square lattices with bimodal disorder. Speedup and the inset are defined the same as in Fig. 13. For both plain simulated annealing and for each group, $\beta_0 = 0.2$, $\beta_1 = 3$. Optimal parameters for both algorithms are listed in Table V.

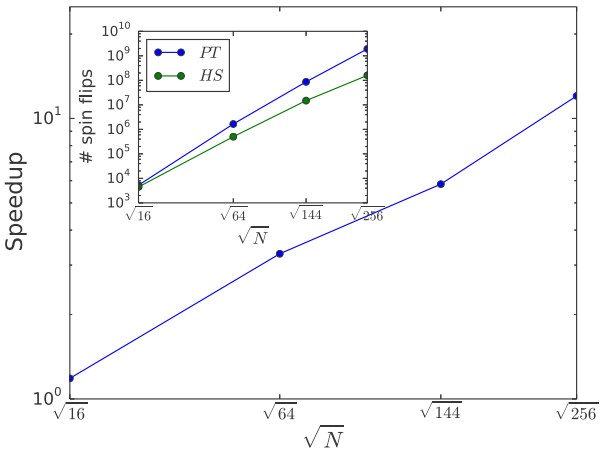

FIG. 15. (Color online) Speedup (ratio of the median number of spin updates) of hierarchical search (HS) with parallel tempering as a group solver to plain parallel tempering (PT) for two-dimensional square lattices with Gaussian disorder. $\beta_0 = 0.014$, $\beta_1 = 0.037N + 2.5$ for plain parallel tempering and $\beta_1 = 0.037N_g + 2.5$ for each group of the hierarchical algorithm.

annealing and parallel tempering grows with problem size, a spin update is effectively more costly due to the additional overhead of randomizing the spins and computing the energy of a group. However, the difference is typically insignificant. For example, the wall clock time per spin update is only about 7% higher than plain simulated annealing for $8 \times 8 \times 8$ 3D lattices with Gaussian disorder run with optimal parameters.

V. CONCLUSION

It has long been established that the complexity of finding grounds states of spin glasses is strongly dependent on the ensemble of couplings and is in the worst case NP-hard. However, while the most trivial cases, such as the ferromagnetic Ising model, are relatively evident, the hardest problems are far more elusive [32].

One way to look for hard problems is by sampling randomly distributed couplings. Although this approach certainly includes such problems, in this work we presented strong numerical evidence that the average complexity of low-dimensional spin glasses with randomly distributed couplings is actually polynomial in the number of spins and looking for hard problems in such a large ensemble might be next to futile. Another way to generate hard cases is to map 3-SAT problems at the critical clause to variable ratio [33], where previous studies have shown evidence of a universal peak in complexity, to the Ising model. Further studies are to be done in this direction.

Our most significant result reported here is a hierarchical approach as a way to potentially improve the performance of a given algorithm for finding ground states of Ising spin glasses. With simulated annealing as a reference solver, on all our benchmark instances we find that optimizing groups of spins is significantly more efficient than solving the whole system at once.

It should be noted that approaches other than simulated annealing can be used at the bottom level of the hierarchical

solver. Suppose that for some class of problems another algorithm (or special purpose classical or quantum device) outperforms simulated annealing. In that case, we can use that algorithm or device at the lowest level. Let T_0 denote the time annealing takes to optimize the bottom level. If it is now replaced by a device which takes time, including communication overhead, $T_1 \ll T_0$, we expect the potential speedup to the whole algorithm to be $\sim T_0/T_1$. As the complexity of finding the ground state scales exponentially with the number of spins, this can be significant even for small groups.

Although we limited our investigation to spin glasses, similar ideas can be applied directly to other problems such as machine learning, protein folding, traveling salesman, etc., by constraining groups of variables independently relative to the rest of the system. We will address such applications in a follow-up work.

ACKNOWLEDGMENTS

This work was supported by Microsoft Research, the Swiss National Competence Center in Research NCCR QSIT, and the ERC Advanced Grant SIMCOFE. We thank M. Freedman, L. Gamper, I. Hen, J. Imriska, S. Isakov, H. G. Katzgraber, A. Kosenkov, J. Osorio, I. Pizorn, D. Poulin, T. Rønnow, A. Soluyanov, and D. Steiger for fruitful discussions. Simulations were performed on the Monch and Brutus clusters of ETH Zurich.

[1] K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
[2] F. Barahona, J. Phys. A: Math. Gen. **15**, 3241 (1982).
[3] A. Lucas, Front. Phys. **2** (2014), doi:10.3389/fphy.2014.00005.
[4] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose, Phys. Rev. B **82**, 024511 (2010).
[5] M. W. Johnson, P. Bunyk, F. Maibaum, E. Tolkacheva, A. J. Berkley, E. M. Chapple, R. Harris, J. Johansson, T. Lanting, I. Perminov, E. Ladizinsky, T. Oh, and G. Rose, Supercond. Sci. Technol. **23**, 065004 (2010).
[6] A. J. Berkley, M. W. Johnson, P. Bunyk, R. Harris, J. Johansson, T. Lanting, E. Ladizinsky, E. Tolkacheva, M. H. S. Amin, and G. Rose, Supercond. Sci. Technol. **23**, 105014 (2010).
[7] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, Nature (London) **473**, 194 (2011).
[8] T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355 (1998).
[9] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, arXiv:1401.7087.
[10] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Science **345**, 420 (2014).
[11] L. Bieche, J. P. Uhry, R. Maynard, and R. Rammal, J. Phys. A: Math. Gen. **13**, 2553 (1980).
[12] A. K. Hartmann and A. P. Young, Phys. Rev. B **64**, 180404 (2001).
[13] W. L. McMillan, Phys. Rev. B **30**, 476 (1984).
[14] L. Leuzzi, G. Parisi, F. Ricci-Tersenghi, and J. J. Ruiz-Lorenzo, Phys. Rev. Lett. **103**, 267201 (2009).
[15] R. A. Baos, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, M. Guidetti, D. Iiguez, A. Maiorano, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, P. Tellez, R. Tripiccione, and D. Yllanes, Proc. Natl. Acad. Sci. USA **109**, 6452 (2012).
[16] M. Baity-Jesi, R. A. Bańos, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, D. Iñiguez, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. M. Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, M. Pivanti, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccione, and D. Yllanes, J. Stat. Mech.: Theory Exp. (2014) P05014.
[17] A. P. Young and H. G. Katzgraber, Phys. Rev. Lett. **93**, 207203 (2004).
[18] H. G. Katzgraber, D. Larson, and A. P. Young, Phys. Rev. Lett. **102**, 177205 (2009).
[19] D. Larson, H. G. Katzgraber, M. A. Moore, and A. P. Young, Phys. Rev. B **87**, 024414 (2013).
[20] Y. Imry and S.-k. Ma, Phys. Rev. Lett. **35**, 1399 (1975).
[21] D. S. Fisher and D. A. Huse, Phys. Rev. Lett. **56**, 1601 (1986).
[22] N. Kawashima, J. Phys. Soc. Jpn. **69**, 987 (2000).
[23] N. Kawashima and T. Aoki, J. Phys. Soc. Jpn. **69**, 169 (2000).
[24] A. K. Hartmann and M. A. Moore, Phys. Rev. B **69**, 104409 (2004).
[25] U. Bertele and F. Brioschi, *Nonserial Dynamic Programming* (Academic Press, Inc., Orlando, FL, USA, 1972).
[26] The median was calculated from 5280 random instances for each system size and field strength. In our implementation we use a library [34] for binary decision diagrams to store these configurations which we found to be significantly more efficient than using simple Boolean tables.
[27] C. De Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, and G. Rinaldi, J. Stat. Phys. **80**, 487 (1995).
[28] We used the spin glass server [35] to compute the exact ground states for three-dimensional lattices.
[29] M. Denil and N. D. Freitas, Toward the Implementation of a Quantum RBM, http://www.cs.ubc.ca/~nando/papers/quantumrbm.pdf.
[30] Google Quantum A.I. Lab Team, https://plus.google.com/+QuantumAILab/posts/DymNo8DzAYi.
[31] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[32] H. G. Katzgraber, F. Hamze, and R. S. Andrist, Phys. Rev. X **4**, 021008 (2014).
[33] D. Mitchell, B. Selman, and H. Levesque, in *Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI '92* (AAAI Press, Menlo Park, CA, 1992), pp. 459–465.
[34] J. Lind-Nielsen, http://vlsicad.eecs.umich.edu/BK/Slots/cache/www.itu.dk/research/buddy/.
[35] Spin Glass Server, http://www.informatik.uni-koeln.de/spinglass/.