

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Fair sampling of ground-state configurations of binary optimization problems

Zheng Zhu, Andrew J. Ochoa, and Helmut G. Katzgraber

Phys. Rev. E **99**, 063314 — Published 25 June 2019

DOI: [10.1103/PhysRevE.99.063314](https://doi.org/10.1103/PhysRevE.99.063314)

# Fair sampling of ground-state configurations of binary optimization problems

Zheng Zhu,<sup>1</sup> Andrew J. Ochoa,<sup>1</sup> and Helmut G. Katzgraber<sup>2,1,3</sup>

<sup>1</sup>*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA*

<sup>2</sup>*Microsoft Quantum, Microsoft, Redmond, WA 98052, USA*

<sup>3</sup>*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA*  
(Dated: June 11, 2019)

Although many efficient heuristics have been developed to solve binary optimization problems, these typically produce correlated solutions for degenerate problems. Most notably, transverse-field quantum annealing—the heuristic employed in current commercially-available quantum annealing machines—has been shown to often be exponentially biased when sampling the solution space. Here we present an approach to sample ground-state (or low-energy) configurations for binary optimization problems. The method samples degenerate states with almost equal probability and is based on a combination of parallel tempering Monte Carlo with isoenergetic cluster moves. We illustrate the approach using two-dimensional Ising spin glasses, as well as spin glasses on the D-Wave Systems Inc. quantum annealer chimera topology. In addition, a simple heuristic to approximate the number of solutions of a degenerate problem is introduced.

PACS numbers: 75.50.Lk, 75.40.Mg, 05.50.+q, 64.60.-i

## I. INTRODUCTION

Quantum annealing [1–9] and, in particular, quantum annealing machines have ignited an ever-increasing interest in algorithms used in statistical physics to solve hard combinatorial industrial optimization problems, as well as related applications. While there has been an extensive body of work attempting to discern if the D-Wave Systems Inc. special-purpose quantum annealing machine can outperform algorithms on conventional CMOS hardware [10–14, 14–32], there have only been few studies [13, 33–36] attempting to characterize the sampling ability of quantum annealing. Initial studies [35, 37] suggested that transverse-field quantum annealing with stoquastic drivers result in biased solution distributions for degenerate problems. However, more recently, it was shown [36] that even with high-order drivers the sampling bias can only be removed in special cases.

Many industrial applications rely more on a broad solution pool than on the minimum of the cost function, with some prominent examples being propositional model counting and related problems [38–40], SAT-based probabilistic membership filters [41–44], machine learning applications [45, 46], or simply estimating the ground-state entropy of a degenerate system. In addition, having multiple solutions to a given problem might allow for the inclusion of constraints in a post-processing step. Here we demonstrate that Monte Carlo methods paired with cluster updates can result in algorithms that asymptotically sample ground-states fairly.

Classical Monte Carlo heuristics based on thermal annealing are known to almost uniformly sample all ground-state and low-lying excited state configurations [47, 48]. Studies on three-dimensional diluted Ising antiferromagnets in a field and three-dimensional Ising spin glasses show that parallel tempering Monte Carlo [49] is more efficient than simulated annealing [50] at finding spin-glass ground-state configurations with near-equal probability [47, 51]. Isoenergetic cluster moves (ICM) [52], related to Houdayer’s cluster updates [53], introduced for Ising spin glasses significantly speed up thermalization on quasi-two-dimensional topolo-

gies, such as D-Wave’s Chimera graph. The combination of low-temperature parallel tempering (PT) Monte Carlo and the rejection-free isoenergetic cluster moves (PT+ICM) allow for a wide-spread sampling of search space and help escape local minima separated by large energy barriers. Here we demonstrate that isoenergetic cluster moves paired with parallel tempering Monte Carlo (i.e., PT+ICM) enhance the fair sampling of ground-state configurations for spin-glass problems better than the previous PT gold standard. We illustrate the approach using two-dimensional Ising spin glasses on a square lattice, as well as the Chimera graph. Higher-dimensional problems can be embedded in lower-dimensional graphs where PT+ICM is more efficient via, e.g., minor embedding [54, 55].

The paper is organized as follows. In Sec. II we introduce the a quality metric for fair sampling, as well as a detailed description of a fair-sampling algorithm using ICM. Following that, we present numerical results in Sec. III for both PT, as well as PT+ICM, and introduce an algorithm to approximate the number of degenerate states for highly-degenerate problems. We conclude with a discussion of our results.

## II. MODEL AND ALGORITHM

To illustrate the improved sampling of PT+ICM over PT, we start with an Ising spin-glass model on a nonplanar Chimera graph [56]. Its nonplanar topology makes finding ground states of random Ising spin glasses worst-case NP-hard. The Hamiltonian for the spin-glass model is given by

$$\mathcal{H} = - \sum_{i < j}^N J_{ij} s_i s_j, \quad (1)$$

where  $s_i \in \{\pm 1\}$  are Ising spins and the couplers  $J_{ij}$  are drawn for this study from three discrete distributions, namely  $\{\pm 1, \pm 2, \pm 4\}$ ,  $\{\pm 5, \pm 6, \pm 7\}$  and  $\{\pm 1\}$ . The couplers are selected based on the range of ground-state degeneracy we can handle with our high-performance computing cluster, i.e.,

the less symmetries between the different coupler values, the smaller the ground-state degeneracy.

### A. Assessing optimal sampling

Suppose  $n$  is the total number of times that ground states are found for an instance with ground-state degeneracy  $G$ . The probability distribution for finding any particular ground-state configuration follows a *binomial* distribution. For the theoretically perfect sampling, if  $p = 1/G$  is the probability of finding a state and  $q = 1 - p$  is the probability of failure in a given trial, then the expected number of successes in  $n$  trials is  $e = np$  and the variance of the binomial distribution is  $\sigma^2 = npq$ . Therefore, the theoretical relative standard deviation given by sampling a finite set of random uncorrelated numbers  $Q_{\text{th}}$  is given by

$$Q_{\text{th}} = \sigma/e = \sqrt{(1-p)/np} = \sqrt{(G-1)/n}. \quad (2)$$

Assuming that the states are uncorrelated (which is a safe assumption for large  $G$ ), an algorithm is said to be optimal (sampling fairly) if the numerical relative standard deviation of the frequency of ground-state configurations  $Q_{\text{num}}$  determined experimentally is close or equal to the theoretical value  $\sqrt{(G-1)/n}$  (or  $Q_{\text{num}}/Q_{\text{th}} = 1$ ). In practice,  $Q_{\text{num}}$  for any algorithm is almost always greater than the theoretical value  $Q_{\text{th}}$ , due to a limited number of measurements via e.g., limited computing resources.

### B. PT+ICM for fair sampling

Our implementation of PT+ICM for sampling purposes can be summarized as follows:

1. Run  $N_T$  replicas of the system at a range of temperatures  $\{T_1, T_2, \dots, T_{N_T}\}$ , with each set consisting of  $M = 4$  copies of the system at the same temperature, thus  $4 \times N_T$  copies of the system with the same disorder are randomly initialized.
2.  $N_{\text{sw}}$  iterations are performed, each iteration consisting of one Monte Carlo sweep, a parallel tempering update, and an isoenergetic cluster move (for the lowest  $N_{\text{hc}}$  temperatures).
3. For the first  $N_{\text{sw}}/2$  iterations, keep track of the lowest energies for the 4 replicas at the lowest temperatures.
4. After  $N_{\text{sw}}/2$  iterations, the lowest energies  $E_1, E_2, E_3$ , and  $E_4$  for the 4 replicas with the lowest temperatures are compared, and if  $E_1 = E_2 = E_3 = E_4$ , the ground-state energy has been found with high confidence. Once this is the case, configurations at this energy are recorded, as well as their frequency for the remaining  $N_{\text{sw}}/2$  updates.

There is no guarantee that any solution obtained by this heuristic method is the true optimum, or that we have found

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

Topology	Couplers	$N$	$N_{\text{sa}}$	$b$	$T_{\text{min}}$	$T_{\text{max}}$	$N_T$	$N_{\text{hc}}$
2D	$\{\pm 1, \pm 2, \pm 4\}$	144	360	24	0.05	3.05	35	35
2D	$\{\pm 1, \pm 2, \pm 4\}$	256	360	24	0.05	3.05	35	35
2D	$\{\pm 1, \pm 2, \pm 4\}$	576	322	24	0.05	3.05	35	35
2D	$\{\pm 1, \pm 2, \pm 4\}$	784	232	24	0.05	3.05	35	35
2D	$\{\pm 1, \pm 2, \pm 4\}$	1024	370	24	0.05	3.05	35	35
Chimera	$\{\pm 1, \pm 2, \pm 4\}$	128	360	24	0.05	3.05	35	20
Chimera	$\{\pm 1, \pm 2, \pm 4\}$	288	360	24	0.05	3.05	35	20
Chimera	$\{\pm 1, \pm 2, \pm 4\}$	512	360	24	0.05	3.05	35	20
Chimera	$\{\pm 1, \pm 2, \pm 4\}$	800	360	24	0.05	3.05	35	20
Chimera	$\{\pm 5, \pm 6, \pm 7\}$	800	976	24	0.10	1.55	30	23
Chimera	$\{\pm 1, \pm 2, \pm 4\}$	1152	223	24	0.05	3.05	35	20

all configurations that minimize the Hamiltonian. However, we choose to make sure each configuration achieves a minimum number of 50 hits in order to increase our confidence that all accessible ground states have been found. The simulation parameters are shown in Table I.

## III. NUMERICAL RESULTS

To test whether PT+ICM can sample ground-state configurations with near-equal probabilities, we multiply the numerical relative standard deviation  $Q_{\text{num}}$  by  $\sqrt{n}$  and plot  $Q_{\text{num}}\sqrt{n}$  as a function of the ground-state degeneracy  $G - 1$ . Note that  $Q_{\text{th}}\sqrt{n}$  is the square root of the ground-state degeneracy  $G - 1$ , and therefore the function  $Q_{\text{th}}\sqrt{n} = \sqrt{G - 1}$  is a straight line in logarithmic scale for both the horizontal axis ( $G - 1$ ) and the vertical axis ( $Q_{\text{th}}\sqrt{n}$ ).

Figure 1 shows  $Q_{\text{num}}\sqrt{n}$  and  $Q_{\text{th}}\sqrt{n}$  as a function of the ground-state degeneracy  $G - 1$  for different spin-glass instances on a Chimera graph. As mentioned in the previous paragraph, the quantity  $Q_{\text{num}}\sqrt{n}$  is almost always greater than  $Q_{\text{th}}\sqrt{n}$  due to limited computational resources [57]. However, an algorithm samples optimally if the data from the numerical relative standard deviation are close to the theoretical line. It is clear that the data for PT+ICM (blue/darker color) are closer to a straight line than the data for PT (red/lighter color), and the discrepancy between PT+ICM and PT seems to become greater as the system size increases.

In Fig. 2 we plot the median ratio  $Q_{\text{num}}/Q_{\text{th}}$  as a function of the system size  $N$  for spin-glass problems on a Chimera lattice. We emphasize that when the ratio becomes unity an algorithm samples optimally. The data show that PT+ICM (blue squares) performs better than PT (red circles) and that the improvement is more significant with increasing system size. In this work the temperature set for the simulation is specifically optimized for  $N = 1152$ . Large median ratios  $Q_{\text{num}}/Q_{\text{th}}$  for smaller system sizes are due to the choice of temperature set. The statistical error bars are determined by a bootstrap analy-

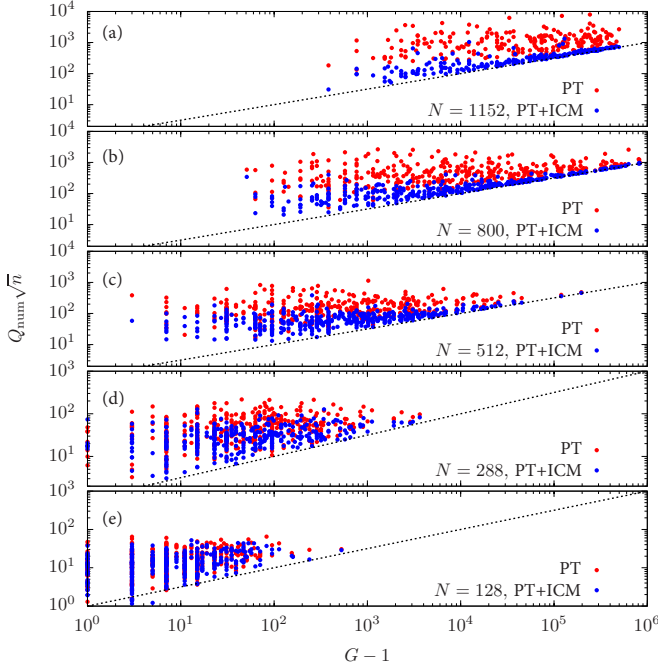


FIG. 1: Scatter plot of  $Q_{\text{num}}\sqrt{n}$  as a function of the ground-state degeneracy  $G - 1$  for different spin-glass instances with different system sizes  $N$  on a Chimera graph. The data points for PT+ICM (blue/darker dots) are closer to the theoretical limit than those from PT (red/lighter dots), and the improvement improves as the system size increases. The dotted line represents ideal uniform sampling of ground-state configurations, i.e.,  $Q_{\text{num}}/Q_{\text{th}} = 1$ . Note that any other heuristic, such as simulated or quantum annealing would perform worse than PT [35, 47]. Data for (a)  $N = 1152$ , (b)  $N = 800$ , (c)  $N = 512$ , (d)  $N = 288$ , and (e)  $N = 128$ .

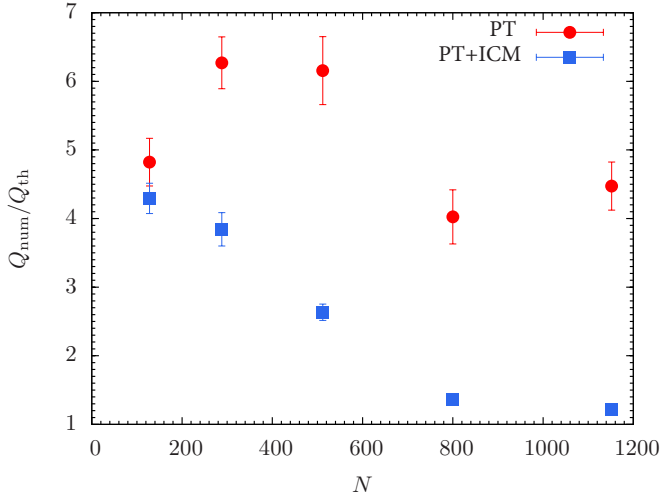


FIG. 2: Median ratio  $Q_{\text{num}}/Q_{\text{th}}$  for spin-glass instances on Chimera as a function of the system size  $N$ . The data points show that PT+ICM (blue squares) performs better than PT (red circles) for all system sizes and the gain is more significant with increasing system size. Statistical error bars are determined by a bootstrap analysis.

sis using the following procedure: For each system size  $N$  and  $N_{\text{sa}}$  disorder realizations, a randomly selected bootstrap sam-

ple of the  $N_{\text{sa}}$  disorder realizations is generated. The median ratio  $Q_{\text{num}}/Q_{\text{th}}$  is computed with this random sample. We repeat this procedure  $N_{\text{boot}} = 1000$  times for each system size to obtain an average and error bar using these  $N_{\text{boot}} = 1000$  data points.

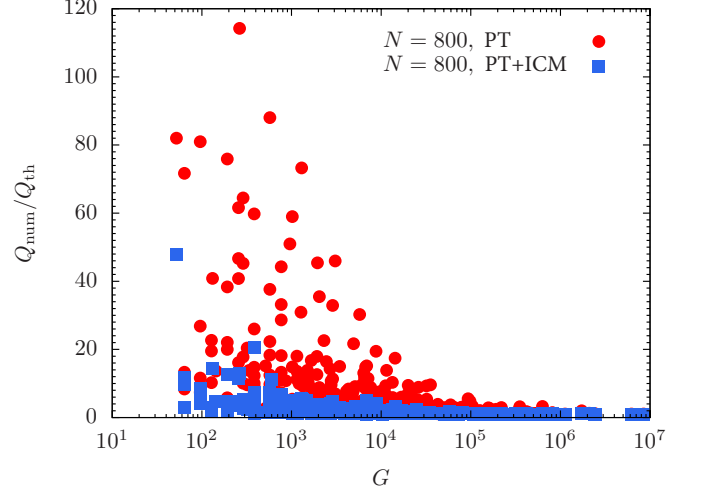


FIG. 3: Scatter plot of  $Q_{\text{num}}/Q_{\text{th}}$  as a function of the estimated ground-state degeneracy  $G$  for different spin-glass instances with system size  $N = 800$  on a Chimera graph. Both data for PT and PT+ICM suggest that the more ground-state configurations, the easier to sample all ground-state configurations with near-equal probabilities using these heuristics.

In addition to studying how fair sampling behaves with increasing system size, we also investigate how the quality of fair sampling is related to ground-state degeneracy and plot  $Q_{\text{num}}/Q_{\text{th}}$  as a function of ground-state degeneracy for  $N = 800$  variables. Figure 3 suggests that the more ground-state configurations, the easier to sample all ground-state configurations with near-equal probabilities. This is not surprising because a large ground-state manifold makes the algorithm easier to explore the configuration space. We do emphasize, however, that in cases where the ground-state degeneracy is exponentially large and with limited resources only a subset of minimizing configurations is accessible, for these the sampling improves, the more configurations are present. Furthermore, careful examination of instances with the same system size and ground-state degeneracy suggests that the  $Q_{\text{num}}/Q_{\text{th}}$  ratio is closely related to the Hamming distances between ground-state configurations. It is shown in Fig. 4 that the Sidon instances [23, 29] where  $J_{ij} \in \{\pm 5, \pm 6, \pm 7\}$  with large Hamming distances between the ground-state configurations tend to have a high  $Q_{\text{num}}/Q_{\text{th}}$  ratio compared to those with small Hamming distances between the states. Here, PT+ICM achieves more equiprobable sampling with large Hamming distances. Figure 5 shows two examples of ground-state configurations with different Hamming distances on a Chimera graph with  $N = 128$ . PT+ICM's cluster updates allow nonlocal moves in the energy landscape, therefore reducing  $Q_{\text{num}}/Q_{\text{th}}$  for instances with large Hamming distances between the ground-state configurations.

Figure 6 shows  $Q_{\text{num}}\sqrt{n}$  and  $Q_{\text{th}}\sqrt{n}$  as a function of the

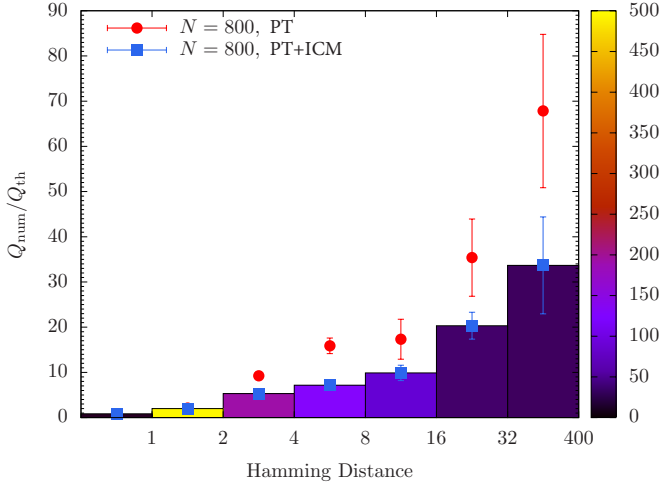


FIG. 4: Median ratio  $Q_{\text{num}}/Q_{\text{th}}$  as a function of Hamming distance for different spin-glass Sidon instances ( $J_{ij} \in \{\pm 5, \pm 6, \pm 7\}$ ) [23] with system size  $N = 800$  and degeneracy  $G = 2$  (up to spin reversal symmetry) on a Chimera graph. Data for PT and PT+ICM suggest that the larger the Hamming distance between ground-state configurations, the harder it is to sample all ground-state configurations with near-equal probability and the more PT+ICM improves fair sampling over PT. Note that the bar chart represents median ratios  $Q_{\text{num}}/Q_{\text{th}}$  between Hamming distance 1 – 2, 2 – 4, 4 – 8, 8 – 16, 16 – 32, and 32 – 400, respectively. The statistical error bars are determined by a bootstrap analysis. Bars are color coded with the number of instances that have a particular Hamming distance.

ground-state degeneracy  $G - 1$  for different spin-glass instances on a two-dimensional square lattice. Similar to the Chimera graph case, the data using PT+ICM (blue/dark color) are closer to the theoretical optimality line than the data using PT (red/light color), and the discrepancy between PT+ICM and PT becomes larger as the system size increases. In Fig. 7, the median ratio  $Q_{\text{num}}/Q_{\text{th}}$  again demonstrates that PT+ICM superior to PT in this case for square lattices.

#### A. Estimating the ground-state degeneracy

We also develop an approximate method to count the number of ground-state configurations based on the fair sampling capabilities of PT+ICM and compare the results to exact methods [58] for a handful of configurations. Counting problems [39] typically ask how many solutions exist for a given instance and belong to complexity class of #P. This approximate method exploits the fact that if one can sample ground states uniformly then one can obtain a reasonable order-of-magnitude estimate of the ground-state degeneracy. Our renormalization-inspired approach works as follows:

1. Compute the ground-state energy  $E_0$  for a fixed number of Monte Carlo sweeps (see above).
2. Sample the number of ground states  $G_0$  for the full system for a fixed number of Monte Carlo sweeps.

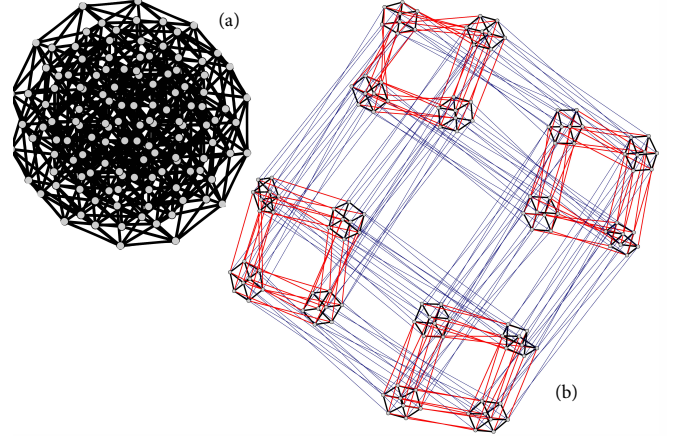


FIG. 5: Two examples of ground-state configurations with different Hamming distances on a Chimera graph for system size  $N = 128$ . The lines represent the distance between two binary strings (ground-state configurations). Each dot in the figure represents a ground-state configuration, black (thick) lines are 1-bit differences, red lines (medium shade) are 2-bit differences, and lighter colors (light gray or blue) indicate an even greater difference. In the first example (a) all ground-state configurations are related by 1-bit differences, while in the second example (b) the Hamming distances between certain ground-state configurations can be large—which means that it takes longer for the system to move from one ground-state configuration to another, therefore causing larger fluctuations in the ground-state frequency. Larger Hamming distances have been omitted for better visibility.

3. Iteratively restrict the number of free variables (i.e., those that are not restricted) and estimate the ratio

$$R_{i-1} = G_{i-1}/G_i$$

for a fixed number of Monte Carlo sweeps.

4. Repeat until the system size is small enough to be able to compute the number of ground state configurations  $G_{\text{final}}$  exactly, e.g., via enumeration.
5. Multiply the product of ratios by the exact count of ground-state configurations to estimate the number of ground states for the full system via

$$G^{\text{RG}} = G_{\text{final}} \prod_i R_{i-1}.$$

We compare results of this approximate method to exact counts on a two-dimensional square lattice with bimodal coupling constants  $J_{ij} \in \{\pm 1\}$ . Simulation parameters and results are shown in Table II. The renormalization-based estimates agree with the exact ground-state degeneracy within error bars.

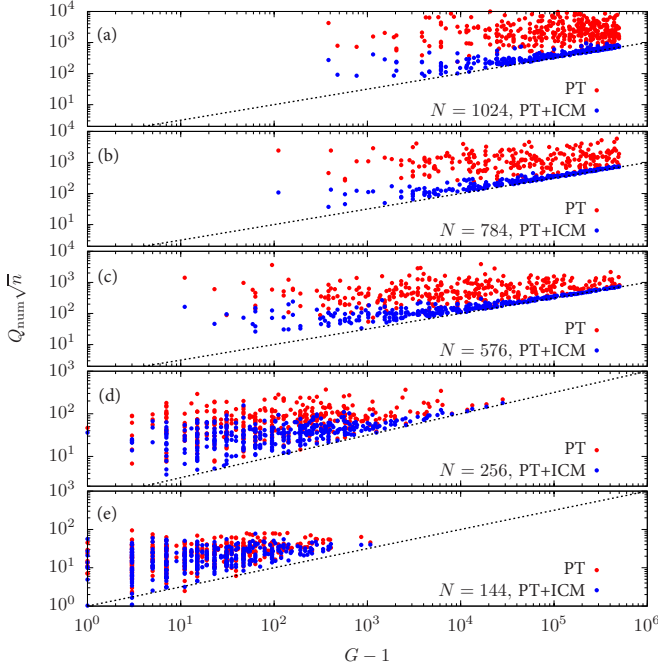


FIG. 6: Scatter plot of  $Q_{\text{num}}\sqrt{n}$  as a function of the ground-state degeneracy  $G - 1$  for different spin-glass instances with different system sizes  $N$  on a two-dimensional lattice. The data points for PT+ICM (blue/dark color) are closer to the theoretical limit than those for PT (red/light color), and this improvement gets better as the system size increases. Data for (a)  $N = 1024$ , (b)  $N = 784$ , (c)  $N = 576$ , (d)  $N = 256$ , and (e)  $N = 144$ .

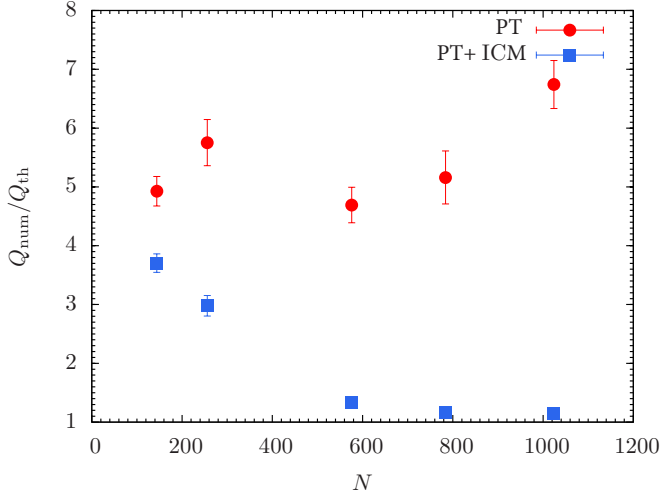


FIG. 7: Median ratio  $Q_{\text{num}}/Q_{\text{th}}$  for spin-glass instances as a function of the system size  $N$  on a two-dimensional lattice. The data points show that PT+ICM (blue squares) performs better than PT (red circles) for all system sizes and the gain is more significant with increasing system size. Statistical error bars are determined by a bootstrap analysis.

TABLE II: For each instance with system size  $N = 1024$ , we run  $N_{\text{sw}} = 2^{23}$  Monte Carlo sweeps for each of the  $4N_T = 4N_{hc} = 120$  replicas with lowest temperature  $T_{\text{min}} = 0.17$  and highest temperature  $T_{\text{max}} = 1.3$ . The fixed number of Monte Carlo sweeps and free variables for each iteration are  $1/24N_{\text{sw}}$  and  $1/24N$ , respectively. The median estimate of the degeneracy  $G^{\text{RG}}$  is averaged over 10 independent runs and error bars are computed using the jackknife method.

Instance	$G^{\text{exact}}$	$G^{\text{RG}}$	error	% error
0	$2.0094 \times 10^{29}$	$1.9415 \times 10^{29}$	$\pm 6.04 \times 10^{27}$	3.00%
1	$9.7771 \times 10^{34}$	$1.0081 \times 10^{35}$	$\pm 3.58 \times 10^{33}$	3.66%
2	$3.3778 \times 10^{27}$	$3.3188 \times 10^{27}$	$\pm 1.18 \times 10^{26}$	3.50%
3	$1.2826 \times 10^{32}$	$1.3041 \times 10^{32}$	$\pm 2.57 \times 10^{30}$	2.00%
4	$1.8613 \times 10^{39}$	$1.9317 \times 10^{39}$	$\pm 7.59 \times 10^{37}$	4.08%
5	$1.4104 \times 10^{40}$	$1.4515 \times 10^{40}$	$\pm 7.28 \times 10^{38}$	5.16%
6	$9.6510 \times 10^{29}$	$9.6105 \times 10^{29}$	$\pm 1.83 \times 10^{28}$	1.90%
7	$2.3699 \times 10^{38}$	$2.3543 \times 10^{38}$	$\pm 1.66 \times 10^{37}$	7.04%
8	$1.4168 \times 10^{31}$	$1.3527 \times 10^{31}$	$\pm 1.16 \times 10^{30}$	8.58%
9	$1.1265 \times 10^{34}$	$1.0789 \times 10^{34}$	$\pm 5.06 \times 10^{32}$	4.69%

#### IV. CONCLUSIONS

We have demonstrated that PT+ICM – parallel tempering Monte Carlo with isoenergetic cluster moves – samples ground-state configurations fairly and is an ideal method for applications where a pool of diverse solutions is needed. We also find that degeneracy and Hamming distances between different ground-state configurations are closely related to the relative standard deviation of frequency with which the ground states are found, namely: ground states with large degeneracy and small Hamming distances have a lower relative standard deviation of frequency. It will be interesting to exploit near-uniform sampling for model counting [59] and SAT filter construction [41, 43] in the future.

#### Acknowledgments

We would like to thank Ruben S. Andrist, Hamid Khoshbakht, Martin Weigel and Salvatore Mandrà for fruitful discussions. We especially thank Martin Weigel for providing the exact number of ground-state configurations on two-dimensional square lattices using the Vondrak code, and Ruben S. Andrist for rendering Fig. 5. H.G.K. acknowledges support from the National Science Foundation (Grant No. DMR-1151387) and would like to thank Binh Mi for providing the necessary motivation for this research. We thank the Texas Advanced Computing Center (TACC) at The University of Texas at Austin and Texas A&M University for providing HPC resources. Part of this research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI,

IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental

purpose notwithstanding any copyright annotation thereon.

- 
- [1] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, *Chem. Phys. Lett.* **219**, 343 (1994).
  - [2] T. Kadowaki and H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).
  - [3] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli, *Science* **284**, 779 (1999).
  - [4] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, *Science* **292**, 472 (2001).
  - [5] G. Santoro, E. Martonák, R. Tosatti, and R. Car, *Science* **295**, 2427 (2002).
  - [6] A. Das and B. K. Chakrabarti, *Quantum Annealing and Related Optimization Methods* (Edited by A. Das and B.K. Chakrabarti, Lecture Notes in Physics 679, Berlin: Springer, 2005).
  - [7] G. E. Santoro and E. Tosatti, *J. Phys. A* **39**, R393 (2006).
  - [8] A. Das and B. K. Chakrabarti, *Rev. Mod. Phys.* **80**, 1061 (2008).
  - [9] S. Morita and H. Nishimori, *J. Math. Phys.* **49**, 125210 (2008).
  - [10] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez, et al., *Nat. Commun.* **4**, 1903 (2013).
  - [11] K. L. Pudenz, T. Albash, and D. A. Lidar, *Nat. Commun.* **5**, 3243 (2014).
  - [12] G. Smith and J. Smolin, *Physics* **6**, 105 (2013).
  - [13] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar, *Nat. Commun.* **4**, 2067 (2013).
  - [14] T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, *Eur. Phys. J. Spec. Top.* **224**, 111 (2015).
  - [15] 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).
  - [16] H. G. Katzgraber, F. Hamze, and R. S. Andrist, *Phys. Rev. X* **4**, 021008 (2014).
  - [17] T. Lanting, A. J. Przybysz, A. Y. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, et al., *Phys. Rev. X* **4**, 021041 (2014).
  - [18] S. Santra, G. Quiroz, G. Ver Steeg, and D. A. Lidar, *New J. Phys.* **16**, 045006 (2014).
  - [19] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani (2014), (arXiv:1401.7087).
  - [20] W. Vinci, T. Albash, A. Mishra, P. A. Warburton, and D. A. Lidar (2014), (arXiv:1403.4228).
  - [21] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, *Nat. Phys.* **10**, 218 (2014).
  - [22] T. Albash, W. Vinci, A. Mishra, P. A. Warburton, and D. A. Lidar, *Phys. Rev. A* **91**, 042314 (2015).
  - [23] H. G. Katzgraber, F. Hamze, Z. Zhu, A. J. Ochoa, and H. Munoz-Bauza, *Phys. Rev. X* **5**, 031026 (2015).
  - [24] V. Martin-Mayor and I. Hen, *Nature Scientific Reports* **5**, 15324 (2015).
  - [25] K. L. Pudenz, T. Albash, and D. A. Lidar, *Phys. Rev. A* **91**, 042302 (2015).
  - [26] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, *Phys. Rev. A* **92**, 042325 (2015).
  - [27] D. Venturelli, S. Mandrà, S. Knysh, B. O’Gorman, R. Biswas, and V. Smelyanskiy, *Phys. Rev. X* **5**, 031040 (2015).
  - [28] W. Vinci, T. Albash, G. Paz-Silva, I. Hen, and D. A. Lidar, *Phys. Rev. A* **92**, 042310 (2015).
  - [29] Z. Zhu, A. J. Ochoa, F. Hamze, S. Schnabel, and H. G. Katzgraber, *Phys. Rev. A* **93**, 012317 (2016).
  - [30] S. Mandrà, Z. Zhu, W. Wang, A. Perdomo-Ortiz, and H. G. Katzgraber, *Phys. Rev. A* **94**, 022337 (2016).
  - [31] S. Mandrà and H. G. Katzgraber, *Quantum Sci. Technol.* **2**, 038501 (2017).
  - [32] S. Mandrà and H. G. Katzgraber, *Quantum Sci. Technol.* **3**, 04LT01 (2018).
  - [33] T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar (2014), (arXiv:1409.3827).
  - [34] A. D. King, E. Hoskinson, T. Lanting, E. Andriyash, and M. H. Amin, *Phys. Rev. A* **93**, 052320 (2016).
  - [35] S. Mandrà, Z. Zhu, and H. G. Katzgraber, *Phys. Rev. Lett.* **118**, 070502 (2017).
  - [36] M. S. Könz, G. Mazzola, A. J. Ochoa, H. G. Katzgraber, and M. Troyer (2018), (arXiv:quant-ph/1806.06081), 1806.06081.
  - [37] Y. Matsuda, H. Nishimori, and H. G. Katzgraber, *New J. Phys.* **11**, 073021 (2009).
  - [38] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, *Theoretical Computer Science* **43**, 169 (1986).
  - [39] C. P. Gomes, A. Sabharwal, and B. Selman, in *Handbook of Satisfiability*, edited by A. Biere, M. Heule, H. van Maaren, and T. Walsch (IOS Press, 2008).
  - [40] P. Gopalan, A. Klivans, R. Meka, D. Stefankovic, S. Vempala, and E. Vigoda, in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on* (IEEE, Palm Springs CA, 2011), p. 817.
  - [41] S. A. Weaver, K. J. Ray, V. W. Marek, A. J. Mayer, and A. K. Walker, *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)* **8**, 129 (2014).
  - [42] T. J. Schaefer, in *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, USA, 1978), STOC ’78, p. 216.
  - [43] A. Douglass, A. D. King, and J. Raymond, in *Theory and Applications of Satisfiability Testing – SAT 2015* (Springer, Austin TX, 2015), pp. 104–120.
  - [44] D. Herr, M. Troyer, M. Azinović, B. Heim, and E. Brown, *SciPost Physics* **2**, 013 (2017).
  - [45] G. E. Hinton, *Neural Comput.* **14**, 1771 (2002).
  - [46] S. M. A. Eslami, N. Heess, C. K. I. Williams, and J. Winn, *Int. J. of Computer Vision* **107**, 155 (2014).
  - [47] J. J. Moreno, H. G. Katzgraber, and A. K. Hartmann, *Int. J. Mod. Phys. C* **14**, 285 (2003).
  - [48] W. Wang, J. Machta, and H. G. Katzgraber, *Phys. Rev. E* **92**, 013303 (2015).
  - [49] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
  - [50] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
  - [51] Note that population annealing and parallel tempering Monte Carlo are comparably efficient [48] solvers.
  - [52] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber (2015), (cond-mat/1501.05630).
  - [53] J. Houdayer, *Eur. Phys. J. B.* **22**, 479 (2001).
  - [54] V. Choi, *Quantum Inf. Process.* **7**, 193 (2008).
  - [55] V. Choi, *Quantum Inf. Process.* **10**, 343 (2011).
  - [56] P. Bunyk, E. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, and



J. Whittaker, IEEE Trans. Appl. Supercond. **24**, 1 (2014).

- [57] Note that for most instances the ratio  $Q_{\text{num}}/Q_{\text{th}}$  is indeed greater than one. However, there are a few exceptions with small system size and ground-state degeneracy where  $Q_{\text{num}}/Q_{\text{th}}$  is less than one. These disparities are due to the limited CPU resources, because in theory all ground-state configurations are supposed to be found with the same probabilities at long-time limit.
- [58] A. Galluccio, M. Loebl, and J. Vondrak, Phys. Rev. Lett. **84**, 5924 (2000).
- [59] W. Wei and B. Selman, in *Theory and Applications of Satisfiability Testing* (Springer, 2005), p. 324.