Comparing Monte Carlo methods for finding ground states of Ising spin glasses: Population annealing, simulated annealing, and parallel tempering

Wenlong Wang, Jonathan Machta, and Helmut G. Katzgraber

Comparing Monte Carlo methods for finding ground states of Ising spin glasses: population annealing, simulated annealing and parallel tempering

Wenlong Wang,[1, *] Jonathan Machta,[1, 2, †] and Helmut G. Katzgraber[3, 4, 2]

[1]*Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003 USA*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501 USA*
[3]*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA*
[4]*Materials Science and Engineering, Texas A&M University, College Station, Texas 77843, USA*

Population annealing is a Monte Carlo algorithm that marries features from simulated annealing and parallel tempering Monte Carlo. As such, it is ideal to overcome large energy barriers in the free-energy landscape while minimizing a Hamiltonian. Thus, population annealing Monte Carlo can be used as a heuristic to solve combinatorial optimization problems. We illustrate the capabilities of population annealing Monte Carlo by computing ground states of the three-dimensional Ising spin glass with Gaussian disorder, whilst comparing to simulated annealing and parallel tempering Monte Carlo. Our results suggest that population annealing Monte Carlo is significantly more efficient than simulated annealing but comparable to parallel tempering Monte Carlo for finding spin-glass ground states.

I. INTRODUCTION

Spin glasses present one of the most difficult challenges in statistical physics [1]. Finding spin-glass ground states is important in statistical physics because some properties of the low-temperature spin-glass phase can be understood by studying ground states. For example, ground-state energies in different boundary conditions have been used to compute the stiffness exponent of spin glasses [2–4]. More generally, the problem of finding ground states of Ising spin glasses in three and more dimensions is an NP-hard combinatorial optimization problem [5] and is thus closely related to other hard combinatorial optimization problems [6], such as protein folding [7] or the traveling salesman problem. As such, developing efficient algorithms to find the ground state of a spin-glass Hamiltonian—as well as related problems that fall into the class of "quadratic unconstrained binary optimization problems"—represents an important problem across multiple disciplines.

Many *generally applicable* computational methods have been developed to solve hard combinatorial optimization problems. Exact algorithms that efficiently explore the tree of system states include branch-and-cut [8] algorithms. Heuristic methods include genetic algorithms [9, 10], particle swarm optimization [11] and extremal optimization [12, 13]. The focus of this paper is on heuristic Monte Carlo methods based on thermal annealing approaches. In particular, we studied simulated annealing [14], parallel tempering Monte Carlo [15–17] and population annealing Monte Carlo [18]. The first two methods are well-known and have been successfully applied to minimize Hamiltonians, while the third has been much less widely used in statistical physics and a primary purpose of this paper is to introduce population

annealing as an effective method for finding ground states of frustrated disordered spin systems.

Population annealing Monte Carlo [18–22] is closely related to simulated annealing and also shares some similarities with parallel tempering. Both simulated annealing and population annealing involve taking the system through an annealing schedule from high to low temperature. Population annealing makes use of a *population* of replicas of the system that are simultaneously cooled and, at each temperature step, the population is resampled so that it stays close to the equilibrium Gibbs distribution. The resampling step plays the same role as the replica exchange step in parallel tempering Monte Carlo. On the other hand, population annealing is an example of a sequential Monte Carlo algorithm [23], while parallel tempering is a Markov chain Monte Carlo algorithm. In a recent large-scale study [22], we have thoroughly tested population annealing Monte Carlo against the well-established parallel tempering Monte Carlo method. Simulations of thousands of instances of the Edwards-Anderson Ising spin glass show that population annealing Monte Carlo is competitive with parallel tempering Monte Carlo for doing large-scale spin-glass simulations at low but nonzero temperatures where thermalization is difficult. Not only is population annealing Monte Carlo competitive in comparison to parallel tempering Monte Carlo when it comes to speed and statistical errors, it has the added benefits that the free energy is readily accessible, multiple boundary conditions can be simulated at the same time, the position of the temperatures in the anneal schedule does not need to be tuned with as much care as in parallel tempering, and it is trivially parallelizable on multi-core architectures.

It is well known that parallel tempering is more efficient at finding spin-glass ground states than simulated annealing [24, 25] because parallel tempering is more efficient at overcoming free-energy barriers. Here we find that population annealing is comparably efficient to parallel tempering Monte Carlo and, thus, also more efficient than simulated annealing. Nonetheless, because of

---

the strong similarities between population annealing and simulated annealing, a detailed comparison of the two algorithms is informative and sheds light on the importance of staying near equilibrium, even for heuristics designed to find ground states.

The outline of the paper is as follows. We first introduce our benchmark problem, the Edwards-Anderson Ising spin glass, and the population annealing algorithm in Sec. II. We then study the properties of population annealing for finding ground states of the Edwards-Anderson model and compare population annealing with simulated annealing in Section III B. We conclude by comparing the efficiency of population annealing and parallel tempering in Section IV and present our conclusions in Section V.

II. MODELS AND METHODS

A. The Edwards-Anderson model

The Edwards-Anderson (EA) Ising spin-glass Hamiltonian is defined by

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} s_i s_j, \tag{1}$$

where $s_i = \pm 1$ are Ising spins on a d-dimensional hypercubic lattice with periodic boundary conditions of size $N = L^3$. The summation $\langle ij \rangle$ is over all nearest neighbor pairs. The couplings J_{ij} are independent Gaussian random variates with mean zero and variance one. For Gaussian disorder, with probability one, there is a unique pair of ground states for any finite system. We call a realization of the couplings $\{J_{ij}\}$ a sample. Here we study the three-dimensional (3D) EA model.

B. Population Annealing and Simulated Annealing

Population annealing (PA) and simulated annealing (SA) are closely related algorithms that may be used as heuristics to find ground states of the EA model. Both algorithms change the temperature of a system through an annealing schedule from a high temperature where the system is easily thermalized to a sufficiently low temperature where there is a significant probability of finding the system in its ground state. At each temperature in the annealing schedule N_S sweeps of a Markov chain Monte Carlo algorithm are applied at the current temperature. The annealing schedule consists of N_T temperatures. The ground state is identified as the lowest energy configuration encountered during the simulation.

Population annealing differs from simulated annealing in that a population of R replicas of the system are cooled in parallel. At each temperature step there is a resampling step, described below.

The resampling step in PA keeps the population close to the Gibbs distribution as the temperature is lowered by differentially reproducing replicas of the system depending on their energy. Lower energy replicas may be copied several times and higher energy replicas eliminated from the population. Consider a temperature step in which the temperature $T = 1/\beta$ is lowered from from β to β', where $\beta' > \beta$. The re-weighting factor required to transform the Gibbs distribution from β to β' for replica i with energy E_i is $e^{-(\beta'-\beta)E_i}$. The expected number of copies $\tau_i(\beta, \beta')$ of replica i is proportional to the re-weighting factor

$$\tau_i(\beta, \beta') = \frac{e^{-(\beta'-\beta)E_i}}{Q(\beta, \beta')}, \tag{2}$$

where Q is a normalization factor that keeps the expected population size fixed,

$$Q(\beta, \beta') = \frac{1}{R} \sum_{i=1}^{R_\beta} e^{-(\beta'-\beta)E_i}. \tag{3}$$

Here R_β is the actual population size at inverse temperature β. A useful feature of PA is that the absolute free energy can be easily and accurately computed from the sum of the logarithm of the normalizations Q at each temperature step. Details can be found in Ref. [19].

Resampling is carried out by choosing a number of copies to make for each replica in the population at β'. There are various ways to choose the integer number of copies $n_i(\beta, \beta')$ having the correct (real) expectation $\tau_i(\beta, \beta')$. The population size can be fixed ($R_\beta = R$) by using the multinomial resampling [20] or residual resampling [26]. Here we allow the population size to fluctuate slightly and use *nearest integer* resampling. We let the number of copies be $n_i(\beta, \beta') = \lfloor \tau_i(\beta, \beta') \rfloor$ with probability $\lceil \tau_i(\beta, \beta') \rceil - \tau_i(\beta, \beta')$ and $n_i(\beta, \beta') = \lceil \tau_i(\beta, \beta') \rceil$, otherwise. $\lfloor x \rfloor$ is the greatest integer less than x and $\lceil x \rceil$ is least integer greater than x. This choice insures that the mean of $n_i(\beta, \beta')$ is $\tau_i(\beta, \beta')$ and the variance is minimized.

The resampling step ensures that the new population is representative of the Gibbs distribution at β' although for finite population R, biases are introduced because the low-energy states are not fully sampled. In addition, the population is now correlated due to the creation of multiple copies. Both of these problems are partially corrected by carrying out Metropolis sweeps and the undersampling of low energy states is reduced by increasing R. Indeed, for PA, which is an example of a sequential Monte Carlo method [23], systematic errors are eliminated in the large R limit. By contrast, for PT, which is a Markov chain Monte Carlo method, such systematic errors are eliminated in the limit of a large number of Monte Carlo sweeps.

In all our PA and SA simulations, the annealing schedule consists of temperatures that are evenly spaced in $\beta = 1/T$ with the highest temperature $1/T = \beta = 0$ and

the lowest temperature $1/T = \beta = 5$. The Markov chain Monte Carlo is the Metropolis algorithm and, unless otherwise stated, there are $N_S = 10$ Metropolis sweeps at each temperature.

For both PA and SA the ground state is presumed to be the lowest energy spin configuration encountered at the end of the simulation. For SA it is most efficient to do *multiple* runs and choose the lowest energy from among the runs rather than do one very long run. Thus, the SA results are typically stated as a function of the number of runs R. Population annealing is inherently parallel and we report results for a *single* run with a population size R. Indeed, choosing the minimum energy configuration among R runs of SA is equivalent to running PA with the same population size but with the resampling step turned off, which justifies using the same symbol R to describe the population size in PA and the number of runs in SA.

While population annealing is primarily designed to sample from the Gibbs distribution at nonzero temperature, here we are interested in its performance for finding ground states. We test the hypothesis that the resampling step in PA improves ground-state searches as compared to SA. The motivation for this hypothesis is that the resampling step removes high-energy spin configurations and replaces them with low-energy configurations, thus potentially increasing the probability of finding the ground state for a given value of R.

The equilibration of population annealing can be quantified using the *family entropy* of the simulation. A small fraction of the initial population has descendants in the final population at the lowest temperature. Let ν_i be the fraction of the final population at the lowest temperature descended from replica i in the initial population. Then the family entropy S_f is given by

$$S_f = -\sum_i \nu_i \log \nu_i. \tag{4}$$

The exponential of the family entropy is an effective number of surviving families. A high family entropy indicates smaller statistical and systematic errors and can be used as a thermalization criterion for the method.

C. Measured Quantities

To compare PA and SA we investigated the following quantities. For PA let $g(R)$ be the fraction of the population in the ground state for a run with population size R. It is understood that g is measured at the lowest simulated temperature. Clearly, the quantity $g(1)$ is simply the probability of finding the ground state in a single run of SA. Let $\mathcal{P}(R)$ be the probability of finding the ground state in a run with population size R. For SA, $\mathcal{P}_{SA}(R)$ is the probability of finding the ground state in R independent runs, i.e.,

$$\mathcal{P}_{SA}(R) = 1 - [1 - g(1)]^R. \tag{5}$$

However, for PA the resampling step tends to reproduce discoveries of the ground state so that the probability $\mathcal{P}_{PA}(R)$ is less than the result for R independent searches. What we actually measured is N_0, the number of occurrences of the ground state in the population from which we obtained $g(R) = N_0/R$ in the case of PA and $g(1)$ in the case of SA.

In the limit of large R, PA generates an equilibrium population described by the Gibbs distribution so

$$\lim_{R \to \infty} g(R) = g_0, \tag{6}$$

where g_0 is the fraction of the ensemble in the ground state,

$$g_0 = \frac{1}{Z(\beta)} 2e^{-\beta E_0} = 2e^{-\beta E_0 + \beta F(\beta)}, \tag{7}$$

where E_0 is the ground-state energy, $Z(\beta)$ is the partition function and $F(\beta) = -\log[Z(\beta)]/\beta$ is the Helmholtz free energy. As explained in Refs. [19, 22, 27], PA yields accurate estimates for the Helmholtz free energy based on the normalization factors $Q(\beta, \beta')$ defined Eq. (3). Thus, we have an independent prediction for the limiting value of $g(R)$.

We considered two disorder averaged quantities as well. The first is the probability of finding the ground state, averaged over disorder samples,

$$\eta = \overline{\mathcal{P}}, \tag{8}$$

where the overbar indicates a disorder average. The quantity η is the primary measure we used to compare the three algorithms.

The second quantity, α is a disorder-averaged measure of accuracy of finding the ground-state energy, i.e.,

$$\alpha = 1 - \overline{(E_{\min}/E_0)}, \tag{9}$$

where $E_{\min}$ is the minimum energy found in the simulation, which might *not* be the true ground-state energy E_0.

III. RESULTS

A. Finding Ground States with Population Annealing

To compare population annealing and simulated annealing, we need a collection of samples with known ground-state energies. In Ref. [22] we reported on a simulation of approximately 5000 samples of the 3D EA spin glass for size $L = 4$, 6, 8, and 10 using large population runs of PA. Note that these sizes are typical of recent ground-state studies of spin glasses. Ground-state energies were obtained from these runs by taking the lowest energy encountered in the population at the lowest temperature, $\beta = 5$, using more than adequate resources.

We used the data from this large-scale simulation as the reference ground-state energy for each sample and compared the same set of samples for smaller PA runs and for SA runs. The population size and number of temperature steps in the reference data set are shown in Table I. Our PAimplementation uses OpenMP and each simulation runs on eight cores.

TABLE I: Simulation parameters of the reference simulations of Ref. [22] from which ground states were obtained. L is the linear system size, R is the population size, N_T is the number of temperatures in the annealing schedule, $\min(N_0)$ is the minimum with respect to samples of the number of replicas in the ground state.

L	R	N_T	$\min(N_0)$
4	5×10^4	101	3370
6	2×10^5	101	1333
8	5×10^5	201	172
10	1×10^6	301	2

Population annealing, like simulated annealing and parallel tempering, is a heuristic method and it is not guaranteed to find the ground state except in the limit of an infinite population size. Nonetheless, we have confidence that we have found the ground state for all or nearly all samples. For an algorithm like PA that is designed to sample the Gibbs distribution at low temperature, the question of whether the true ground state has been found is closely related to the question of whether equilibration has been achieved at the lowest simulated temperature. The candidate ground state is defined as the minimum energy state in the population at the lowest temperature β. For an equilibrium ensemble, the fraction of the ensemble in the ground g_0 is given by the Gibbs distribution, Eq. (7). If the number of copies of the found ground state in the low-temperature population N_0 is large and if the population is in equilibrium, then it is unlikely that the true ground-state energy has not been found. Because, if we have not found the true ground state, the number of copies of the true ground state, Rg_0 would be expected to be even larger than N_0. Thus, if we believe the population is in equilibrium at low temperature and if the candidate ground state is found many times in the low-temperature population, then we have high confidence that the candidate is the true ground state.

Of course, it cannot be guaranteed that the population generated by PA is in equilibrium at low temperature. However, the production runs from which we measured ground-state energies passed a stringent thermalization test. We required a large effective number of independent families based on the family entropy, defined in Eq. (4). We required $e^{S_f} \geq 100$; additional runs were done for those samples that did not meet these criteria. We also compared our results for the same set of samples to results reported in Ref. [28] using parallel tempering Monte Carlo. We found good agreement for both averaged quan-
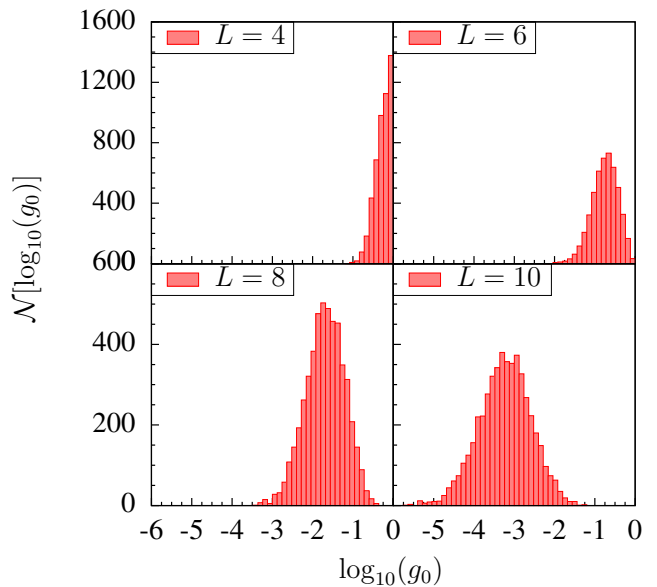

FIG. 1: (Color online) Histogram of the number of samples with fraction in the ground state g_0 at $\beta = 5$ for various sizes L, estimated from the reference runs described in Table 1. $\mathcal{N}[\log_{10}(g_0)]$ is the number of samples in the logarithmic bin centered at $\log_{10}(g_0)$. There are a total of 50 bins. Note that as L increases, the histograms shift rapidly to smaller values.

tities and the overlap distribution for individual samples.

In addition to the equilibration test, we recorded the number of copies of the ground state in the population at the lowest temperature and found that for most samples this number is large. A histogram of $N_0/R = g(R) \approx g_0$ of all samples is given in Fig. 1 for each system size L. The minimum value of N_0 for each system size is shown in Table I. For the small fraction (0.7%) of $L = 10$ samples with $N_0 < 10$ we re-ran PA with a ten-fold larger population, $R = 10^7$. In no case did the ground-state energy change. In addition, for the one sample with $N_0 = 2$ we confirmed the ground state using an exact branch and cut algorithm run on the University of Cologne Spin Glass Server [29]. Based on the strict equilibration criteria and the large number of ground states reported in Table I, we are confident that we have found true ground states for all samples.

As an additional check, we compared the disorder averaged ground-state energy per spin against values in the literature using the hybrid genetic algorithm [30] and parallel tempering (PT) [24]. The comparison is shown in Table II and reveals that all three methods yield the same average energy within statistical errors.

A striking feature of Fig. 1 is that the fraction of the ensemble in the ground state g_0 decreases rapidly as L increases. Thus, for any temperature-based heuristic, including PA, SA, and PT, it is necessary to simulate at lower temperatures and/or use larger populations (or for PT, longer runs) as L increases. To understand this requirement more formally we re-write Eq. (7) in terms of

TABLE II: Comparison of the disorder averaged ground-state energy per spin for the EA model with those obtained from the hybrid genetic algorithm [30] and PT [24].

L	PA	Hybrid genetic	PT
4	-1.6639(14)	-1.6655(6)	-1.6660(2)
6	-1.6899(7)	-1.6894(5)	-1.6891(4)
8	-1.6961(5)	-1.6955(4)	-1.6955(6)
10	-1.6980(3)	-1.6975(5)	-1.6981(7)

intensive quantities

$$g_0 = 2\exp[-N\beta(e_0 - f(\beta))], \qquad (10)$$

where e_0 and $f(\beta)$ are the ground-state energy and free energy per spin, respectively, and $N = L^3$ is the number of spins. In the thermodynamic limit, $(e_0 - f(\beta))$ is expected to converge to a positive number that is independent of the disorder realization. Thus, for fixed β, the fraction of the ensemble in the ground state decreases exponentially in the system size.

As discussed in Sec. II B, population annealing gives a direct estimator of the free energy, thus we can independently measure all of the quantities in Eq. (7) and carry out a disorder average. Because the observables on the right-hand side of Eq. (7) appear in the exponent, it is convenient to take the logarithm and then carry out the disorder average. Table III compares $\overline{\log_{10} g_0}$ and $\log_{10} 2 - \beta\overline{(E_0 - F)}/\log(10)$ at $\beta = 5$. The table confirms the expected equilibrium behavior of the fraction in the ground state. Note that the observables g_0, E_0 and F are not entirely independent quantities, which explains why the statistical errors are significantly larger than the difference in the values. On the other hand, if the simulation was not in thermal equilibrium, these quantities would not agree.

TABLE III: Comparison of the disorder average of the log of the two sides of Eq. (7) at $\beta = 5$.

L	$\overline{\log_{10} g_0}$	$\log_{10} 2 - \beta\overline{(E_0 - F)}/\log(10)$
4	-0.2644(28)	-0.2643(28)
6	-0.7573(46)	-0.7563(46)
8	-1.6933(77)	-1.6925(67)
10	-3.2358(104)	-3.2297(91)

For all the reasons discussed above we believe that we have found the true ground state for all samples. However, our main conclusions would not be affected if a small fraction of the reference ground states are not true ground states.

B. Comparison between Population Annealing and Simulated Annealing

1. Detailed Comparison for a Single Sample

In this section we present a comparison of population annealing and simulated annealing for a single disorder realization. This sample was chosen to be hardest to equilibrate for $L = 8$ based on having the smallest family entropy [see Eq. (4)], however it has a probability of being in the ground state at the lowest temperature near the average for size $L = 8$. For this sample we confirmed the ground-state energy found in the reference PA run using the University of Cologne Spin Glass Server [29].

Figure 2 shows the fraction of the population in the ground state $g(R)$ as a function of population size R for PA. The result for the probability that SA has found the ground state in a single run is simply the value at $R = 1$. In this simulation, we used $N_T = 101$ temperatures with $N_S = 10$ sweeps per temperature for both algorithms. It is striking that the fraction of ground states in the population increases by about four orders of magnitude from the small value for SA, $g(1)$ to the limiting value for PA for large R, $g(10^6) \approx g_0$. This result shows that resampling greatly increases the probability that a member of the PA population is in the ground state. It suggests that even though equilibration is not required for finding ground states, the probability of finding the ground state is improved when the simulation is maintained near thermal equilibrium. Of course, remaining near equilibrium as the temperature is lowered is also a motivation for SA but lacking the resampling step, SA falls out of equilibrium once the free-energy landscape roughness significantly exceeds $k_\mathrm{B}T$. However, the ratio of $g(R)/g(1)$ is an overestimate of the ratio the probabilities for actually finding the ground state for a fixed R because once the ground state is discovered in PA, it is likely to be reproduced many times.

The probability of finding the ground state $\mathcal{P}$ for a given amount of computational work is an appropriate metric to compare the two algorithms. We measured the amount of work W in Metropolis sweeps, $W = RN_TN_S$. In most of our comparisons we used the same value of N_T and N_S for both PA and SA. However, it is not clear whether the two algorithms are optimized with the same values of N_T and N_S. We performed additional optimization of SA varying N_T and N_S. We used the computational work divided by the probability of finding the ground state in a single SA run, N_TN_S/g as a figure of merit. Note that in the relevant large-R regime, minimizing N_TN_S/g is equivalent to maximizing $\mathcal{P}$ for a fixed amount of work. Figure 3 shows N_TN_S/g vs N_TN_S and reveals a broad minimum near $N_TN_S \approx 5 \times 10^3$. We therefore performed SA simulations at the same value used for PA, $N_TN_S = 1010$ and a more nearly optimal value, $N_TN_S = 5000$. Note that for SA it is only the product, N_TN_S that determines the efficiency, not N_T and N_S separately. Note also that the efficiency decreases
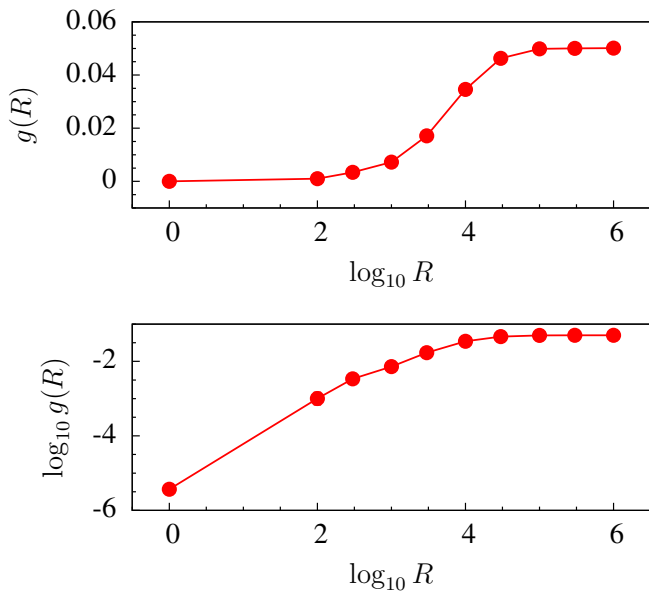
FIG. 2: (Color online) The fraction of the population in the ground state $g(R)$ as a function of population size R for a single sample using PA with $N_T = 101$ and $N_S = 10$. The point at $\log_{10} R = 0$ corresponds to the probability that a single run of SA will yield the ground state. The upper panel is a log-linear plot and the lower panel is a log-log plot of the same data. Error bars are smaller than the symbols.
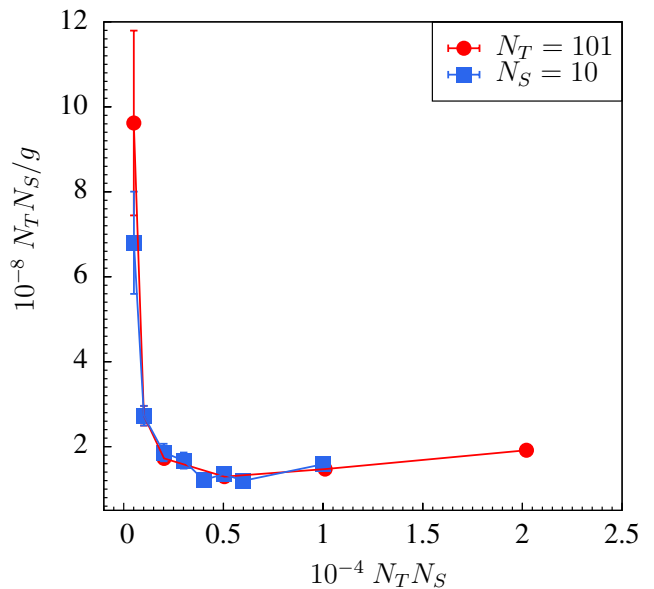

FIG. 3: (Color online) The computational work divided by the probability of finding the ground state in a single SA run, $N_T N_S/g$ vs the computational work $N_T N_S$ for a single sample. The two curves correspond to holding $N_S = 10$ fixed and varying N_T (blue squares) and holding $N_T = 101$ fixed and varying N_S (red circles). Smaller values of $N_T N_S/g$ correspond to more efficient simulations.

when $N_T N_S$ is too large suggesting that it is better to do many shorter SA runs rather than a single long run.

Figure 4 compares $\mathcal{P}_{SA}$, obtained from Eq. (5), and $\mathcal{P}_{PA}(R)$, obtained from multiple runs of PA as a function of the computational work W. In this simulation we used $N_T = 101$ temperatures with $N_S = 10$ for PA and the lower SA curve. The upper SA curve corresponds to the optimal value $N_T N_S = 5000$. Computational work was varied by changing R holding N_T and N_S fixed. For intermediate values of R, corresponding to realistic simulations, $\mathcal{P}_{PA}$ exceeds $\mathcal{P}_{SA}$ by one or two orders of magnitude and the amount of work needed to be nearly certain of finding the ground is also more than an order of magnitude less for PA than SA. Note that the effect of optimizing SA is only about a factor of 2. We conclude that for this sample, there is a large difference in efficiency between PA and SA and this difference cannot be explained by a difference in the optimization of the two methods. To see whether this difference is typical and how it depends on system size, in Sec. III B 2 we consider averages over disorder realizations.

2. Disorder-Averaged Comparison

We compared population annealing and simulated annealing for approximately 5000 disorder realizations for each of the four system sizes, $L = 4$, 6, 8, and 10, and for several different population sizes. For SA the pop-

ulation size refers to the number of independent runs. Both algorithms use the same annealing schedule with evenly spaced inverse temperatures starting with infinite temperature and ending at $T_0 = 0.2$. The number of sweeps per temperature is $N_S = 10$. The population sizes R, number of temperatures in the annealing schedule N_T, the number of disorder realizations M and the corresponding parameters for the reference runs are given in Table IV.

TABLE IV: Parameters of the numerical simulations for comparison between PA and SA. R is the population size, N_T is the number of temperatures, and M is the number of samples studied. The reference parameters are for the PA runs used to estimate the ground-state energy for each sample.

L	$\log_{10} R$	N_T	M	Ref. R	Ref. N_T
4	{1,2,3,4}	101	4941	5×10^4	101
6	{1,2,3,4,5}	101	4959	2×10^5	101
8	{1,2,3,4,5}	101	5099	5×10^5	201
10	{1,2,3,4,5}	201	4945	1×10^6	301

Figure 5 shows α, the disorder averaged error in finding the ground state [see Eq. (9)], as function of population size R for SA and PA. For small systems neither algorithm makes significant errors even for small populations but as the system size increases, PA is significantly more accurate.

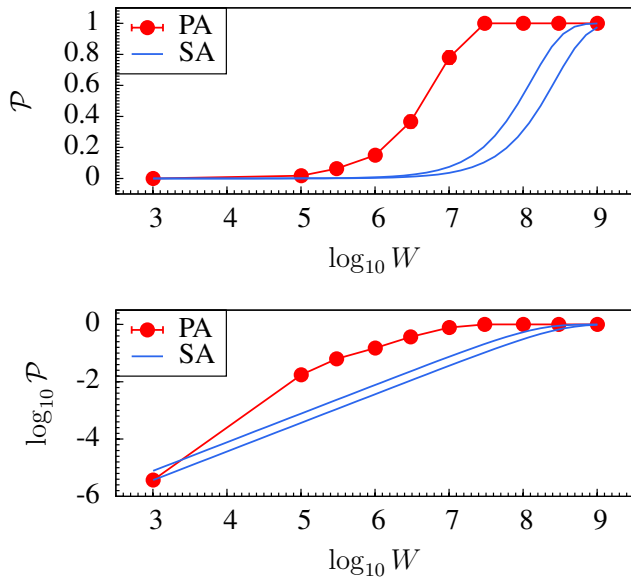Figure 6 shows η, the disorder-averaged fraction of

FIG. 4: (Color online) Probability of finding the ground state $\mathcal{P}$ as a function of the computational work $W = RN_T N_S$ for a single sample for both SA and PA. The computational work is varied by changing population size R, holding $N_T N_S$ fixed. For PA and the lower SA curve, $N_T N_S = 1010$ while for the upper SA curve, $N_T N_S = 5000$, which is near the optimum value for SA. The upper panel is a log-linear plot and the lower panel is a log-log plot. Error bars for PA are smaller than the symbols. The SA curves are obtained from Eq. (5).

samples for which the ground state is found [see Eq. (8)], as a function of population size R. Again, we see that for $L = 4$ and 6, the two algorithms are comparable but for $L = 8$ and 10, population annealing is far more likely to find the ground state for the same population size. It is clear from Figs. 5 and 6 that population annealing is both more accurate and more efficient at finding ground states than simulated annealing and that as system size increases, the relative advantage of PA over SA increases.

IV. COMPARISON BETWEEN POPULATION ANNEALING AND PARALLEL TEMPERING

In this section, we compare the efficiency of population annealing (PA) and parallel tempering (PT) when finding ground states. We first briefly describe parallel tempering Monte Carlo. Parallel tempering simultaneously simulates N_T replicas of the system at N_T different temperatures. In addition to single-temperature Metropolis sweeps, PA uses *replica exchange* moves in which two replicas at neighboring temperatures swap temperatures. To satisfy detailed balance, the swap probability p_{swap} is given by

$$p_{\text{swap}} = \min \left[1, \exp \left[(\beta' - \beta)(E' - E) \right] \right], \quad (11)$$
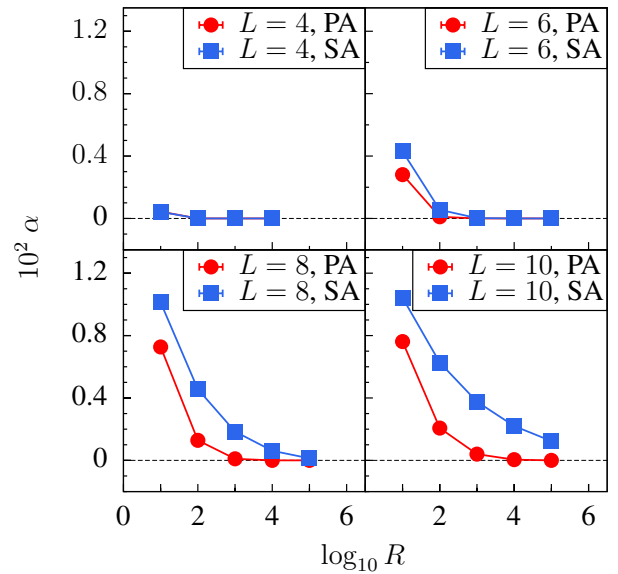

FIG. 5: (Color online) Error in approximating the the ground-state energy (α) vs log population size, $\log_{10}(R)$.
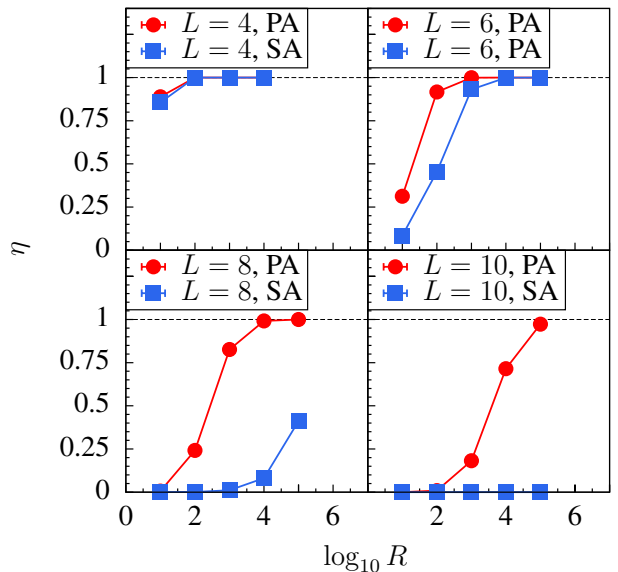

FIG. 6: (Color online) Fraction of samples for which the ground state is found (η) vs log population size, $\log_{10}(R)$ for population annealing and simulated annealing.

where E and E' are the energies of the replicas proposed for exchange at temperatures β and β', respectively.

Results for PT are taken from Romá *et al.* [24], who studied the disorder-averaged probability of finding the ground state η for the 3D EA model for the same sizes considered here. They gave an empirical fit of their data of the form,

$$\eta = \frac{e^{qx}}{1 + e^{qx}}, \quad (12)$$

where q is a fitting parameter and x is a function of the computational work W and system size L defined as

$$x = [\log(W/2) - (bL^c - a)]/L^d, \quad (13)$$

and the work is calculated in units of Monte Carlo sweeps. For PT, the computational work is given by, $W = N_T N_S$, while for PA, it is given by, $W = R N_T N_S$. We assume that the work involved in replica exchange moves for PT and in population resampling for PA is negligible compared to the work associated with the Metropolis sweeps. The fitting parameters for the 3D EA model reported in [24] are $a = -0.05$, $b = 1.55$, $c = 1$, $d = 0.2$, and $q = 2$.

Figure 7 shows η, the fraction of samples for which the ground state is correctly found, vs the scaled work x for our PA data (points) and the fit for PT from [24] (solid curve). It is striking that both algorithms perform nearly identically over the whole range of sizes and amounts of computational work.
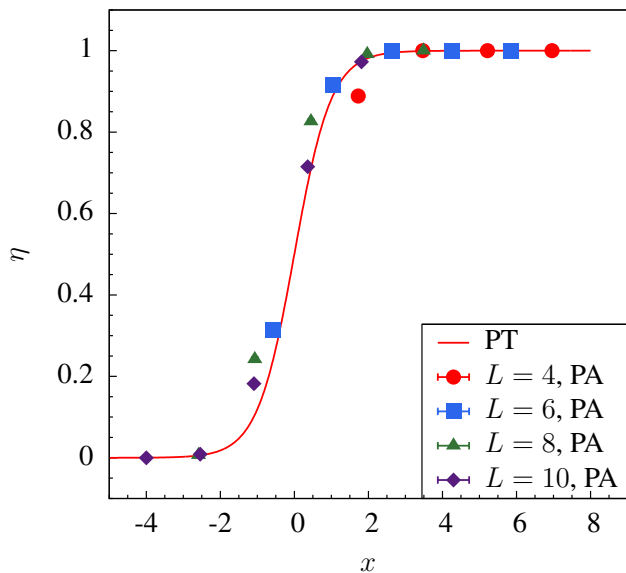

FIG. 7: (Color online) Fraction of samples for which the ground state is found (η) as a function of the scaled computational work x [Eq. (13)] for both population annealing and parallel tempering. The curve is taken from the empirical fit [Eq. (12)] of Ref. [24].

V. CONCLUSION

We have carried out a detailed comparison of three Monte Carlo heuristics based on thermal annealing for finding ground states of spin-glass Hamiltonians. The algorithms compared are population annealing, simulated annealing and parallel tempering Monte Carlo. We find that population annealing is more efficient than simulated annealing and has better scaling with the system size. In particular, the CPU time needed for resampling the population is negligible. Thus, with a similar numerical effort as for simulated annealing, population annealing provides a sizable performance improvement.

We find that population annealing and parallel tempering are comparably efficient for finding spin-glass ground states. Population annealing, however, is much better suited to a massively parallel implementation and would be the preferred choice for large systems or when ground states are required quickly. A general conclusion is that Monte Carlo heuristics based on thermal annealing are enhanced by mechanisms that improve thermalization at every temperature. In population annealing this mechanism is resampling and in parallel tempering it is replica exchange. Simulated annealing depends entirely on local Monte Carlo moves and fails to remain close to equilibrium at low temperatures where the free-energy landscape is rough. Furthermore, we observed that the ensemble defined by simulated annealing has far less weight in the ground state than the equilibrium ensemble for realistic computational effort. This deficiency results in a significantly lower probability of finding the ground state for a given amount of computational effort as compared to either population annealing or parallel tempering, which stay close to thermal equilibrium.

There is no obvious reason to suppose that the temperature-dependent Gibbs distribution is the best target distribution for improved heuristics such as population annealing or parallel tempering Monte Carlo. Distributions other than the Gibbs distribution that concentrate on the ground state as "temperature" is decreased might perform even better than the Gibbs distribution and should be investigated.

[1] D. L. Stein and C. M. Newman, *Spin Glasses And Complexity* (Princeton University Press, 2013).

[2] A. Carter, A.J.Bray, and M. Moore, *Aspect-Ratio Scal-*

ing and the Stiffness Exponent θ for Ising Sping Glasses, Phys. Rev. Lett. **88**, 077201 (2002).

[3] A. K. Hartmann, *Scaling of stiffness energy for three-dimensional ±J Ising spin glasses*, Phys. Rev. E **59**, 84 (1999).

[4] K. Hukushima, *Domain-wall free energy of spin-glass models: Numerical method and boundary conditions*, Phys. Rev. E **60**, 3606 (1999).

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

[6] C. Moore and S. Mertens, *The Nature of Computation* (Oxford University Press, 2011).

[7] S. Trebst, M. Troyer, and U. H. Hansmann, *Optimized parallel tempering simulations of proteins*, J. Chem. Phys. **124**, 174903 (2006).

[8] C. D. Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, and G. Rinaldi, *Exact ground states of Ising spin glasses: New experimental results with a branch and cut algorithm*, J. Stat. Phys. **80**, 487 (1995).

[9] A. K. Hartmann and H. Rieger, *Optimization Algorithms in Physics* (Wiley-VCH, Berlin, 2001).

[10] A. K. Hartmann and H. Rieger, *New Optimization Algorithms in Physics* (Wiley-VCH, Berlin, 2004).

[11] A. Băutu and E. Băutu, *Searching ground states of Ising spin glasses with genetic algorithms and binary particle swarm optimization*, in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2007)*, edited by N. Krasnogor, G. Nicosia, M. Pavone, and D. Pelta (Springer Berlin Heidelberg, 2008), vol. 129 of *Studies in Computational Intelligence*, pp. 85–94.

[12] S. Boettcher and A. G. Percus, *Optimization with extremal dynamics*, Phys. Rev. Lett. **86**, 5211 (2001).

[13] A. A. Middleton, *Improved extremal optimization for the Ising spin glass*, Phys. Rev. E **69**, 055701 (2004).

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

[15] R. H. Swendsen and J.-S. Wang, *Replica Monte Carlo simulations of spin glasses*, Phys. Rev. Lett. **57**, 2607 (1986).

[16] C. Geyer, in *Computing Science and Statistics: 23rd Symposium on the Interface*, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, 1991), p. 156.

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

[18] K. Hukushima and Y. Iba, in *The Monte Carlo Method In The Physical Sciences: Celebrating the 50th Anniversary of the Metropolis Algorithm*, edited by J. E. Gubernatis (AIP, 2003), vol. 690, pp. 200–206.

[19] J. Machta, *Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes*, Phys. Rev. E **82**, 026704 (2010).

[20] J. Machta and R. Ellis, *Monte Carlo methods for rough free energy landscapes: Population annealing and parallel tempering*, J. Stat. Phys. **144**, 541 (2011).

[21] E. Zhou and X. Chen, in *Proceedings of the 2010 Winter Simulation Conference (WSC)* (2010), pp. 1211–1222.

[22] W. Wang, J. Machta, and H. G. Katzgraber, *Evidence against a mean-field description of short-range spin glasses revealed through thermal boundary conditions*, Phys. Rev. B **90**, 184412 (2014).

[23] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice* (Springer-Verlag, New York, 2001).

[24] F. Romá, S. Risau-Gusman, A. J. Ramirez-Pastor, F. Nieto, and E. E. Vogel, *The ground state energy of the Edwards-Anderson spin glass model with a parallel tempering Monte Carlo algorithm*, Physica A: Statistical Mechanics and its Applications **388**, 2821 (2009).

[25] J. J. Moreno, H. G. Katzgraber, and A. K. Hartmann, *Finding low-temperature states with parallel tempering, simulated annealing and simple monte carlo*, International Journal of Modern Physics C **14**, 285 (2003).

[26] R. Douc and O.Cappé, in *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA)* (IEEE, 2005), pp. 64–69.

[27] W. Wang, J. Machta, and H. G. Katzgraber, *Population annealing for large scale spin glass simulations* ((2014), in preparation).

[28] B. Yucesoy, Ph.D. thesis, University of Massachusetts Amherst (2013).

[29] *University of Cologne spin glass server*, `http://www.informatik.uni-koeln.de/spinglass/`.

[30] M. Palassini and A. P. Young, *Triviality of the ground state structure in Ising spin glasses*, Phys. Rev. Lett. **83**, 5126 (1999).