

Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes

J. Machta*

*Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003, USA
and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA*

(Received 4 June 2010; published 18 August 2010)

The population annealing algorithm introduced by Hukushima and Iba is described. Population annealing combines simulated annealing and Boltzmann weighted differential reproduction within a population of replicas to sample equilibrium states. Population annealing gives direct access to the free energy. It is shown that unbiased measurements of observables can be obtained by weighted averages over many runs with weight factors related to the free-energy estimate from the run. Population annealing is well suited to parallelization and may be a useful alternative to parallel tempering for systems with rough free-energy landscapes such as spin glasses. The method is demonstrated for spin glasses.

DOI: [10.1103/PhysRevE.82.026704](https://doi.org/10.1103/PhysRevE.82.026704)

PACS number(s): 05.10.Ln, 61.43.Bn, 75.10.Nr

I. INTRODUCTION

One of the most difficult challenges in computational statistical physics is sampling the low-temperature equilibrium states of systems with complex free-energy landscapes such as those that occur, for example, in spin glasses, configurational glasses, and biomolecules. Standard Markov chain Monte Carlo (MCMC) methods such as the Metropolis algorithm become stuck in local minima and are unable to correctly sample the equilibrium distribution. Attempts to solve this difficulty generally involve simulating a broadened distribution that smooths the free-energy landscape. One of the most widely used of these methods is parallel tempering or replica exchange Monte Carlo [1–3]. In parallel tempering, many replicas of the system are simultaneously simulated using a standard MCMC method with each replica at a different temperature. The sequence of temperatures spans a range from high temperatures where the free-energy landscape is smooth to the low temperatures of interest where it is rough. Replica exchanges are allowed that permit replicas to move between low and high temperatures. The hope is that visits to high temperatures allow more rapid mixing of the Markov chain and equilibration between different minima.

Closely related to the problem of sampling equilibrium is the problem of finding ground states of systems with complex energy landscapes. In computer science the same question arises for NP-hard combinatorial optimization problems. One generic method for finding ground states or at least low-lying states is simulated annealing [5]. In simulated annealing the system is subject to an equilibrating MCMC procedure at a sequence of temperatures. Following this “annealing schedule” the system is gradually cooled through the sequence of temperatures from a high temperature where the free-energy landscape is smooth to a low temperature where it is rough. The hope is that the system will explore many minima while the temperature is high and settle into the deepest minima as the temperature is gradually lowered.

The population annealing algorithm introduced by Hukushima and Iba [4] is based on simulated annealing [5] but also shares features with parallel tempering, histogram reweighting [6], and diffusion Monte Carlo [7]. Like simulated annealing, it is a single pass algorithm with an annealing schedule. A population of replicas of the system is simultaneously cooled following the annealing schedule. Unlike simulated annealing, the population is maintained in equilibrium throughout the annealing process and an equilibrium sample is generated at every temperature. Equilibrium is maintained by differential reproduction (resampling) of replicas according to their relative Boltzmann weights. A useful by-product of the calculation of relative Boltzmann weights is direct access to the equilibrium free energy at every temperature. Methods for extracting free-energy differences from population annealing are related to Jarzinski’s equality [8]. Population annealing is an example of a sequential Monte Carlo scheme [9] and is related to nested sampling methods [10].

In [4], Hukushima and Iba compared population annealing to parallel tempering and showed that they produce results with comparable efficiencies. However, they noted that population annealing suffers biases for small population size. The main contribution of this work is to show that these biases, together with statistical errors, can be made arbitrarily small by using appropriate weighted averages over many independent runs of the algorithm. The weight factor is obtained from the free-energy estimator from each run and the variance of this estimator serves as a useful diagnostic for the algorithm. As a proof of principle of the method, we apply it to one-dimensional (1D) and three-dimensional (3D) spin glasses.

In the next section of the paper population annealing is described. In Sec. III the weighted averaging procedure is introduced. Section IV describes simulations of the Ising spin glass and the paper closes with a discussion.

II. ALGORITHM

We now describe our implementation of the algorithm in detail. We start with a population of R_0 replicas of the sys-

*machta@physics.umass.edu

tem. For disordered spin systems, each replica has the same set of couplings. Each replica is initialized at infinite temperature, $\beta=0$; thus, each replica is in an independent microstate chosen at random with equal probability from among the $\Omega=\sum_{\gamma}1$ microstates of the system. For Ising systems with N spins, the microstates γ are described by N binary variables and $\Omega=2^N$. Without loss of generality, suppose that the average energy of the system at infinite temperature is zero, $\langle E_{\gamma} \rangle_{\beta=0}=0$, as is the case for the Ising model. The algorithm can also be used with a finite initial temperature but then absolute measurements of the free energy are not available.

The temperature of the set of replicas is now lowered while keeping the ensemble in equilibrium. In order to accomplish this, the population is resampled—some replicas are reproduced and others are eliminated with the expected number of replicas held fixed. Suppose we have a population of \tilde{R}_{β} replicas chosen from an equilibrium ensemble at temperature $1/\beta$ and we want to lower the temperature to $1/\beta'$, with $\beta' > \beta$. For a replica j with energy E_j the ratio of the statistical weights at β and β' is $\exp[-(\beta' - \beta)E_j]$. This reweighting factor is typically large, so that a normalization is needed to keep the population size reasonable. We compute normalized weights $\tau_j(\beta, \beta')$ whose sum over the ensemble is \tilde{R}_{β} ,

$$\tau_j(\beta, \beta') = \frac{\exp[-(\beta' - \beta)E_j]}{Q(\beta, \beta')}, \quad (1)$$

where Q is the normalization given by

$$Q(\beta, \beta') = \frac{\sum_{j=1}^{\tilde{R}_{\beta}} \exp[-(\beta' - \beta)E_j]}{\tilde{R}_{\beta}}. \quad (2)$$

The new population of replicas is generated by resampling the original population. The number of copies of state j in the new population is $\mathcal{N}[(R_{\beta'}/\tilde{R}_{\beta})\tau_j(\beta, \beta')]$, where $\mathcal{N}[a]$ is a Poisson random variate with expected value a . If $\mathcal{N}[(R_{\beta'}/\tilde{R}_{\beta})\tau_j(\beta, \beta')]=0$, the configuration j is eliminated. The size $\tilde{R}_{\beta'}$ of the population after the temperature step has expectation $R_{\beta'}$. In our implementation, the size of the population is variable but stays close to the set of target values $\{R_{\beta}\}$ in contrast to the fixed population method introduced in [4].

Of course, the new population is now more correlated than before since several replicas may be in the same microstate. Furthermore, to the extent that the energy distributions at the two temperatures do not overlap, the equilibrium distribution may no longer be adequately sampled at the lower temperature. In order to mitigate both of these difficulties, all replicas are now subject to an equilibrating MCMC process at the new temperature.

The algorithm thus consists of a sequence of temperature steps with differential reproduction within the population according to the relative Boltzmann weights followed by additional equilibration at the new temperature using a standard MCMC procedure. The ordered sequence of $K+1$ inverse

temperatures, $\beta_0 > \beta_1 > \beta_2 \cdots > \beta_K=0$, determines the annealing schedule. The population of \tilde{R}_{β_k} replicas is equilibrated at β_k using θ_k sweeps of the MCMC process. The temperature is now lowered to β_{k-1} with differential reproduction as described above followed by θ_{k-1} MCMC sweeps. At each temperature, observables are measured after the MCMC sweeps and $\ln Q$ is recorded. In the present implementation, observables are measured *after* resampling, so that no weight factors are needed in taking averages. The temperature is lowered until β_0 is reached. In the simplest version of the algorithm the target population sizes and number of MCMC sweeps are independent of temperature, $R_{\beta}=R$, and $\theta_k=\theta$ for all β .

Either the MCMC process or the differential reproduction steps are in principle sufficient to produce equilibrium at the new temperature although neither one individually is efficient. However, the combined processes are substantially more efficient than either alone.

One advantage of the algorithm is that it gives direct access to the free energy at each temperature. The normalization factor $Q(\beta, \beta')$ is an estimator of the ratio of the partition functions at the two temperatures,

$$\begin{aligned} \frac{Z(\beta')}{Z(\beta)} &= \frac{\sum_{\gamma} e^{-\beta' E_{\gamma}}}{\sum_{\gamma} e^{-\beta E_{\gamma}}} = \sum_{\gamma} e^{-(\beta' - \beta)E_{\gamma}} \left(\frac{e^{-\beta E_{\gamma}}}{Z(\beta)} \right) = \langle e^{-(\beta' - \beta)E_{\gamma}} \rangle_{\beta} \\ &\approx \frac{1}{\tilde{R}_{\beta}} \sum_{j=1}^{\tilde{R}_{\beta}} e^{-(\beta' - \beta)E_j} = Q(\beta, \beta'). \end{aligned} \quad (3)$$

Thus, the estimator of the free energy \tilde{F} at each simulated temperature $1/\beta_k$ is

$$-\beta_k \tilde{F}(\beta_k) = \sum_{\ell=K}^{k+1} \ln Q(\beta_{\ell}, \beta_{\ell-1}) + \ln \Omega. \quad (4)$$

This section concludes with pseudocode for the algorithm.

Population annealing

```

Initialize  $R_0$  replicas at  $\beta=0$ 
for  $k=K$  to 1 step  $-1$  do
  Compute the partition function ratio  $Q(\beta_k, \beta_{k-1})$ 
  {Eq. (2)}
  for all  $j \leq \tilde{R}_{\beta_k}$  do
    Compute the relative weight  $\tau_j(\beta_k, \beta_{k-1})$  {Eq. (1)}
    Resample: make  $\mathcal{N}[(R_{\beta_{k-1}}/\tilde{R}_{\beta_k})\tau_j(\beta_k, \beta_{k-1})]$  copies
    of replica  $j$ 
    { $\mathcal{N}(a)$  is a Poisson random variate with mean  $a$ }
  end for
  Compute the population size  $\tilde{R}_{\beta_{k-1}}$ 
  for all  $j \leq \tilde{R}_{\beta_{k-1}}$  do
    Equilibrate replica  $j$  for  $\theta_{k-1}$  Monte Carlo sweeps
  end for
  Compute observables and the free energy at  $\beta_{k-1}$ 
  {Eq. (4)}
end for

```

III. WEIGHTED AVERAGES

Unless the population size, number of temperature steps, and number of MCMC sweeps or step are sufficiently large, the result of a single run of population annealing is significantly biased because the equilibrium distribution is not fully sampled. Simple averaging over an ensemble of independent runs does not reduce this bias. However, an appropriate *weighted average* over an ensemble of independent runs effectively reduces both statistical errors and biases. The free-energy estimator from a given run of the algorithm represents the weight that should be assigned to the observations made during that run. In particular, suppose that an observable A is estimated in run r to be $\tilde{A}_r(\beta)$ at temperature $1/\beta$. An unbiased estimate of $\bar{A}(\beta)$ from the entire simulation is the weighted average over the ensemble of M independent runs of population annealing,

$$\bar{A}(\beta) = \sum_{r=1}^M \tilde{A}_r(\beta) \omega_r(\beta), \quad (5)$$

where the weight for run r is given by

$$\omega_r(\beta) = \frac{e^{-\beta \tilde{F}_r(\beta)}}{\sum_{r=1}^M e^{-\beta \tilde{F}_r(\beta)}}, \quad (6)$$

and $\tilde{F}_r(\beta)$ is the free-energy estimator (4) for the r th run of the algorithm at temperature $1/\beta$. The weight factors $e^{-\beta \tilde{F}_r(\beta)}$ are needed because they represent the total weight at that temperature that would have been associated with the r th run if the normalization factor in Eq. (1) had not been used to keep the population size near the target size. Runs with different but fixed target populations sizes $R^{(r)}$ can be combined using Eq. (5) with $e^{-\beta \tilde{F}_r(\beta)}$ replaced with $R^{(r)} e^{-\beta \tilde{F}_r(\beta)}$ in both the numerator and denominator of Eq. (6).

The free energy appears to require a more complicated weighting formula because it is the sum of terms from all temperatures. However, as shown below it may be expressed as a simple average. The quantity that is directly averaged at each temperature is Q , not $\ln Q$, so the logarithm must be taken after the weighted average. The unbiased estimate of the free energy $\bar{F}(\beta_k)$ is given by

$$\begin{aligned} -\beta_k \bar{F}(\beta_k) &= \sum_{\ell=K}^{k+1} \ln \left[\sum_{r=1}^M Q_r(\beta_\ell, \beta_{\ell-1}) \omega_r(\beta_\ell) \right] + \ln \Omega \\ &= \sum_{\ell=K}^{k+1} \ln \left[\frac{\sum_{r=1}^M \exp[-\beta_{\ell-1} \tilde{F}(\beta_{\ell-1})]}{\sum_{r=1}^M \exp[-\beta_\ell \tilde{F}(\beta_\ell)]} \right] \\ &= \ln \left[\frac{1}{M} \sum_{r=1}^M \exp[-\beta_k \tilde{F}(\beta_k)] \right]. \end{aligned} \quad (7)$$

The inner summation indexed by r is over the ensemble of independent runs while the outer summation indexed by ℓ is

over the temperature steps between the highest temperature $\ell=K$ and the temperature of interest $\ell=k$. The second expression follows from the first via Eqs. (4) and (6). The last expression shows that $\exp[-\beta \tilde{F}(\beta)]$ may be directly averaged.

Errors in the weighted averages can be obtained by resampling. Here, we use the bootstraps method [11] where the error is the standard deviation of a large number ($\sim 10^2$) of resampled ensembles, each of size M .

Statistical and systematic errors in the method are determined by the probability distribution of the dimensionless free energy $-\beta \tilde{F}$. If this distribution has a variance that is much smaller than unity and tails that decay faster than exponential then the important weight factors (6) do not vary much and the tails of the distribution are unimportant. On the other hand, if the variance is larger than 1, \bar{A} is controlled by the upper tail of the distribution and the ensemble size M must be large to reduce errors. In the case that $-\beta \tilde{F}$ is normally distributed, the typical value of $-\beta \tilde{F}$ that dominates the sum in Eq. (5) is one variance (not one standard deviation) above the mean. The variance of $-\beta \tilde{F}$ is thus a criterion for the performance of the algorithm. If it is smaller than about 0.5 then a modest ensemble of independent runs will be sufficient to yield accurate results. Optimizing population annealing should be guided by minimizing the variance of the free-energy estimator for a fixed amount of computational effort. Of course, it is always possible that the actual $-\beta \tilde{F}$ distribution has significant weight far above its mean because of important low-energy states that have never been explored. This diabolical situation would similarly fool equilibration tests for parallel tempering [12].

Population annealing using M independent runs with population size R is less accurate than a single run with population size MR . If $\text{Var}(\beta \tilde{F})$ is small, the loss of accuracy is small, and vice versa. Nonetheless, an ensemble of independent runs has several important advantages. First, independent runs with moderate R require only a single processor or simple parallelism without communication between processors. Large values of R are only practical with true parallelism involving communication between processors. Second, the measurement of the variance of $\beta \tilde{F}$ over the ensemble provides a diagnostic for the method. Third, error bars are simply obtained from an ensemble using resampling.

IV. APPLICATION TO ISING SPIN GLASSES

To illustrate the effectiveness of population annealing for high-precision simulations we apply it to the Ising spin glass in one and three dimensions. Ising spin glasses are extremely difficult to equilibrate at low temperature. The Ising spin glass is defined by the energy

$$E = \sum_{(i,j)} J_{i,j} \sigma_i \sigma_j, \quad (8)$$

where $\sigma_i = \pm 1$ is the Ising spin at site i and the summation is over the nearest-neighbor bonds (i, j) of the lattice. We consider the simple cubic lattice (3D) with periodic boundary

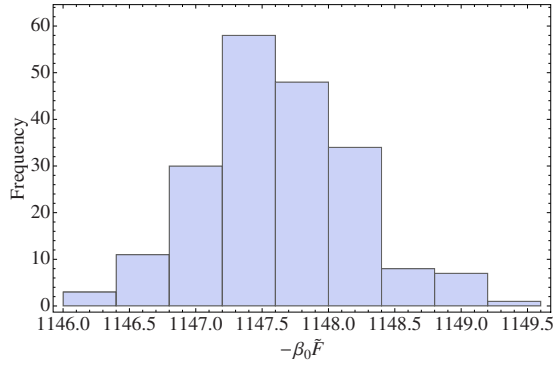


FIG. 1. (Color online) The histogram of the dimensionless free energy, $-\beta_0 \tilde{F}$, for population annealing applied to a single realization of a 1D Ising spin glass. The exact value for this realization is $-\beta_0 F^{\text{exact}} = 1147.87$.

conditions and a periodic chain of spins (1D). The couplings $J_{i,j}$ are quenched random variables chosen independently for each bond from a Gaussian distribution with mean zero and variance one. The 3D Ising spin glass with Gaussian couplings has a continuous phase transition at transition temperature $T_c = 1.052$ [13]. The low-temperature phase of this model is still the subject of controversy. On the other hand, the 1D Ising spin glass is a trivial model. Nonetheless, it has a rough free-energy landscape and is difficult to equilibrate using generic Monte Carlo methods including parallel tempering [14,15]. The 1D spin glass has the advantage that the Monte Carlo results can be compared with exact transfer-matrix calculations.

For the 1D spin glass we simulated 18 disorder realizations for a chain with $N=256$ spins using population annealing. We report results for the lowest temperature, $1/\beta_0=0.2$. We used a mean population size $R=1000$ with $K+1=100$ temperatures and $\theta=50$ Metropolis sweeps at each temperature. Of the 99 temperature steps, 19 are equally spaced in inverse temperature between $\beta=0$ and $\beta=0.5$, while the remaining 80 are equally spaced in temperature. Each run used 5×10^6 Metropolis sweeps and for each realization the ensemble consisted of $M=200$ independent runs.

First we report results for a single 1D disorder realization at the lowest temperature, $1/\beta_0=5$. Figure 1 shows the histogram of the dimensionless free energy $-\beta_0 \tilde{F}$. The variance of $-\beta_0 \tilde{F}$ is 0.34 and the distribution does not appear to have fat tails. Therefore, we expect that weighted averages should be useful in calculating observables. Using Eq. (5) to compute \bar{E} and the bootstraps method to find the one standard deviation error, we obtain $\bar{E} = -224.175 \pm 0.037$ compared to the exact value, $E^{\text{exact}} = -224.1896$, obtained from numerically summing the transfer matrices. On the other hand, the unweighted average energy of the ensemble is -223.987 ± 0.024 . Thus, the unweighted average is significantly biased while the weighted average shows no bias to within the statistical errors. The ground-state energy for this disorder realization is -226.861 so at $1/\beta_0=5$ the system is quite close to its ground state. For the free energy of the same realization we have $-\beta_0 F^{\text{exact}} = 1147.8697$, while the weighted average is $-\beta_0 \tilde{F} = 1147.82 \pm 0.04$ and the un-

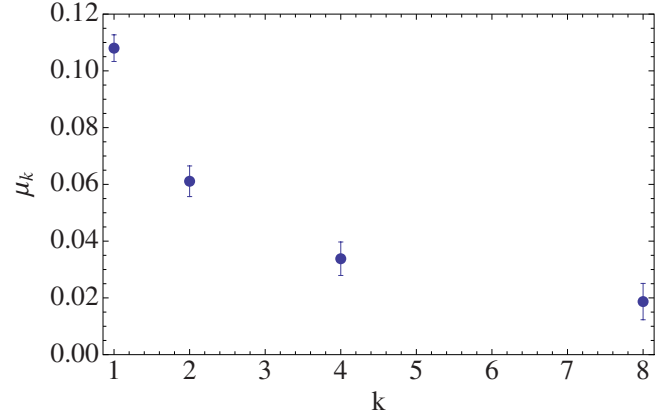


FIG. 2. (Color online) The bias for the weighted energy estimator μ_k for ensemble size k averaged over 18 disorder realizations for the 1D Ising spin glass.

weighted average is 1147.65 ± 0.04 . Again, the weighted estimator is not significantly biased while the unweighted estimator is biased.

Next we consider the set of 18 disorder realizations and investigate how the bias decreases as the size of the ensemble of runs increases. We divide the 200 independent runs for each disorder realization into B blocks of size $k=200/B$ and then perform weighted averages for the energy within each of these blocks. Next we take the mean over the set of blocks and finally average these means over the 18 disorder realizations. Let μ_k be this disorder averaged bias,

$$\mu_k = \frac{1}{18} \sum_{j=1}^{18} \left[\left(\frac{1}{B} \sum_{b=1}^B \bar{E}_j^b \right) - E_j^{\text{exact}} \right], \quad (9)$$

where \bar{E}_j^b is the weighted energy estimator in block b with block size $k=200/B$. We can similarly obtain the standard error of the mean from the B blocks. Figure 2 shows the bias μ_k as a function of block size k . The case $k=1$ corresponds to unweighted averages. By the time an ensemble of eight runs is used, the bias is reduced by approximately a factor of 6. The weighted energy estimator for the full ensemble of $M=200$ runs for each disorder realization yields a mean bias of -0.0013 ± 0.0069 , where the error estimate is obtained using the bootstraps method. This weighted average displays no bias within the error bars. Similar results hold for the free energy. The $M=200$ weighted estimate of $-\beta \bar{F}$ averaged over the 18 disorder realizations deviates from the exact value by 0.0018 ± 0.0083 , which is again indistinguishable from zero, while the unweighted free energy averaged over the same 18 realizations deviates from the exact answer by -0.1013 ± 0.0077 .

For the 3D Ising spin glass we studied 25 disorder realizations using population annealing for a system size of 8^3 . The target population size is $R=2000$, and $K+1=100$ temperatures are evenly spaced between $1/\beta_{K+1}=2$ and $1/\beta_0=0.2$ with $\theta=100$ Metropolis sweeps performed after every temperature step. Each run requires 2×10^7 Metropolis sweeps. In addition to the energy and free energy, we considered the observable Γ , the probability that the overlap is

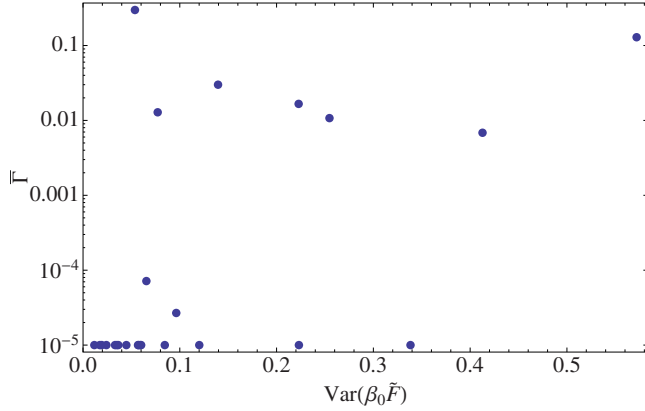


FIG. 3. (Color online) The probability of a small overlap $\bar{\Gamma}$ vs variance of the free energy $\text{Var}(\beta_0\tilde{F})$, for a sample of 25 disorder realizations of the 3D Ising spin glass.

less than 0.2, $\Gamma = \text{Prob}(|q| < 0.2)$. The overlap is defined as $q = (1/N) \sum_{i=1}^N \sigma_i^{(1)} \sigma_i^{(2)}$, where the superscripts refer to two independent spin configurations of the system. The probability distribution for the overlap reflects features of the free-energy landscape. When there is substantial weight for small q , there are two or more free-energy minima that are widely separated but with comparable free energies. For parallel tempering this situation is expected to lead to long equilibration times [12]. Thus, it is interesting to look for correlations between the variance of $\beta\tilde{F}$ and Γ . In the study of spin glasses, Γ is important in distinguishing the competing pictures of the low-temperature phase of the 3D Ising spin glass [16–19]. Since Γ involves the correlation between two independent replicas, we use pairs of independent runs to construct q . For each pair of runs, we compute q by averaging over pairs of replicas, one from each run, and then evaluate $\bar{\Gamma}$ for that pair. The appropriate weighted estimator $\bar{\Gamma}$ is

$$\bar{\Gamma} = \sum_{r=1}^M \bar{\Gamma}_r \omega'_r, \quad (10)$$

with

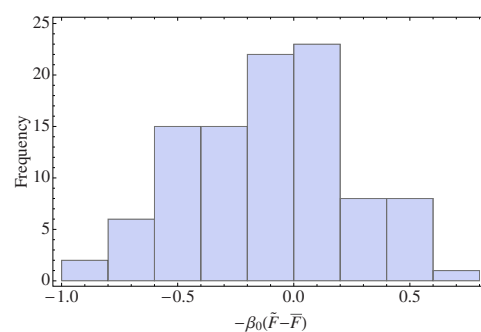
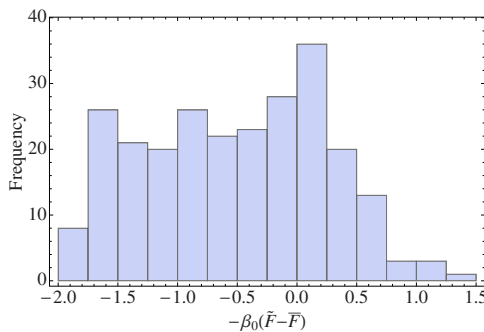


FIG. 4. (Color online) The histogram of the dimensionless free-energy deviation from its weighted average, $-\beta_0(\tilde{F} - \bar{F})$, for one realization of the 3D Ising spin glass. In the left panel, the mean population size is $R=2000$ and in the right panel $R=10000$.

$$\omega'_r = \frac{e^{-\beta(\tilde{F}_r^{(1)} + \tilde{F}_r^{(2)})}}{\sum_{r=1}^M e^{-\beta(\tilde{F}_r^{(1)} + \tilde{F}_r^{(2)})}}, \quad (11)$$

where the superscripts refer to the two of independent runs from which $\tilde{\Gamma}_r$ is constructed. We did $M=50$ pairs of independent runs to compute $\bar{\Gamma}$.

Figure 3 shows a scatter plot of $\text{Var}(\beta_0\tilde{F})$ vs $\bar{\Gamma}$ for the 25 disorder realizations. The values of Γ along the line 10^{-5} are estimated as zero by the algorithm on the basis of 10^5 measurements of the overlap. Most realizations have very small values of Γ , with only a few realizations dominating the disorder average. The variance of the free-energy estimator for the 25 realizations never exceeds 0.6. There appears to be a correlation between a large variance of the free energy and a large value of Γ . Realizations with free-energy variances less than 0.05 are all associated with very small values of the overlap near zero. This correlation can be understood by realizing that $\Gamma > 0$ implies that there are at least two free energy minima with comparable free energies. If R is not sufficiently large, the population of replicas may be dominated by one or another free-energy minima in a single run and thus display a large variance in free energy from one run to another.

Of the 25 disorder realizations we now look more closely at the realization with the largest variance of $-\beta_0\tilde{F}$. This “worst” realization has $\text{Var}(\beta_0\tilde{F})=0.572$. For this realization we obtained data from an ensemble of 125 runs (with $R=2000$, $K+1=100$, and $\theta=100$). The histogram of $-\beta_0(\tilde{F} - \bar{F})$ is shown in Fig. 4 (left panel). It should be noted that for the 3D simulations the highest temperature is not infinite; however, as we shall see below, a negligible contribution to $\text{Var}(\beta_0\tilde{F})$ arises from high temperatures. The mean of $-\beta_0\tilde{F}$ is 0.521 below the best estimate $-\beta_0\bar{F}$, and this deviation is roughly equal to $\text{Var}(\beta_0\tilde{F})$ as expected. The value of Γ from these runs is $\bar{\Gamma}=0.129 \pm 0.011$ and the value of the energy is $\bar{E}=-847.294 \pm 0.015$. For this realization we also did simulations with a larger target population size without changing K or θ . The histogram of $-\beta_0(\tilde{F} - \bar{F})$ for population size $R=10000$ is shown in Fig. 4 (right panel). For the larger popu-

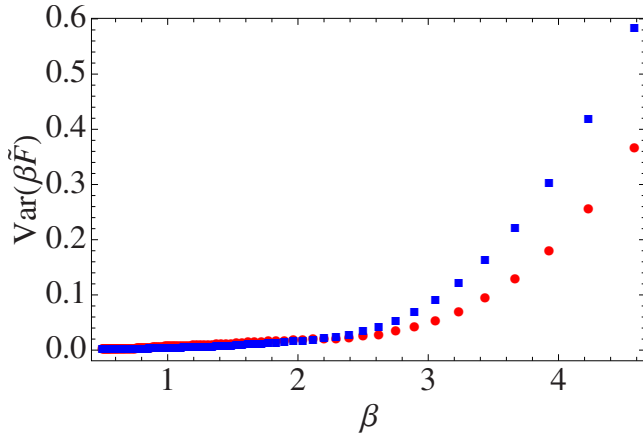


FIG. 5. (Color online) The variance of the free-energy estimator $\text{Var}[\beta \tilde{F}(\beta)]$ vs β for one disorder realization for the 3D Ising spin glass. The upper (blue squares) points show the result for a fixed population size ($R_\beta=2000$ for all β) while the lower (red circles) points are for a population that grows for low temperature ($R_\beta=1000$ for $\beta<0.9$ and $R_\beta=3000$ for $\beta\geq 0.9$). Both simulations involve the same amount of computational work.

lution size $\text{Var}(\beta_0 \tilde{F})=0.122$, so increasing the population size by a factor of 5 has reduced the variance by nearly the same factor. The estimate for Γ from this larger population size is $\bar{\Gamma}=0.145 \pm 0.005$, which is consistent with the results from the smaller population size. The estimate for the energy is $\bar{E}=-847.317 \pm 0.009$, which is marginally consistent with the smaller run. For comparison we simulated the same disorder realization using parallel tempering with two sets of 18 replicas each equally spaced in temperature with the same high and low temperatures. The system was equilibrated for 10^6 sweeps and data were collected for 10^6 sweeps. We ran this simulation 200 times to obtain means and errors. Thus, the total number of Metropolis sweeps for the parallel tempering runs was approximately 1.5×10^{10} compared to approximately 1×10^{10} for population annealing. Parallel tempering runs yielded $\Gamma=0.150 \pm 0.002$ and $E=-847.307 \pm 0.004$. Results are marginally consistent for both Γ and E .

Next we consider an initial attempt to improve the performance of the algorithm. The variance of the free-energy estimator should be minimized to optimize the algorithm. From Eq. (4) we see that $-\beta \tilde{F}(\beta)$ at each temperature is a sum of positive terms, so that successively lower temperatures have successively larger variances. Figure 5 shows $\text{Var}(\beta \tilde{F})$ as a function of temperature for the worst disorder realizations with the largest variance discussed above. The variance at each temperature is obtained from 50 iterations of the algorithm. The upper points (blue squares online) show the results from runs where the average population size is fixed,

$R_\beta=2000$ for all β . The variance increases slowly for small β and then much more rapidly for large β . The upward curvature of $\text{Var}(\beta \tilde{F})$ suggests that more replicas should be used at low temperatures. The total computational work remains nearly fixed if $R_\beta=1000$ for $\beta<0.9$ and $R_\beta=3000$ for $\beta\geq 0.9$. The lower points (red circles online) in Fig. 5 show the results for this choice of target population size and demonstrates that, indeed, a significant reduction in $\text{Var}(\beta \tilde{F})$ from 0.57 to 0.36 is achieved at the lowest temperature. This attempt at optimizing the algorithm shows that there is room for improving its performance by careful choice of the parameters.

V. DISCUSSION

Population annealing is a promising tool for measuring the free energy and other observables in spin glasses and perhaps other systems with rough free-energy landscapes. By using weighted averages over an ensemble of runs, biases inherent in a single run can be made small and high-precision results can be obtained. If the variance of the dimensionless free-energy estimator $\text{Var}[\beta \tilde{F}(\beta)]$ is less than about 0.5, high-precision results can be obtained from a relatively small ensemble of runs. This variance thus serves as a measure of the convergence of the algorithm.

The method can be optimized by minimizing the variance of the free-energy estimator. In future work it would be important to optimize the method and study its efficiency and scaling with the system size in comparison with well-established methods such as parallel tempering.

Compared to parallel tempering, population annealing is better suited to parallelization. Parallel tempering is optimized with a relatively small number of replicas. Increasing the number of replicas in parallel tempering improves the acceptance rate of replica exchange moves but also lengthens the round trip distance between the warmest and coldest replicas. Thus, the optimum number of replicas is relatively small in practice. On the other hand, population annealing is monotonically improved by increasing the population size. For example, for the $L=8$, 3D Ising spin glasses studied here population annealing makes effective use of thousands of replicas while parallel tempering is optimized with tens of replicas. Population annealing may find use in quickly equilibrating large systems using a very large population spread over many processors.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grant No. DMR-0907235. I thank Helmut Katzgraber, Burcu Yucesoy, and Yukita Iba for helpful discussions. I thank the Santa Fe Institute for their hospitality and support.

- [1] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [2] C. Geyer, in *Computing Science and Statistics: 23rd Symposium on the Interface*, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, VA, 1991), p. 156.
- [3] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [4] 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, New York, 2003), Vol. 690, pp. 200–206.
- [5] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [6] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [7] J. B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975).
- [8] C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
- [9] *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon (Springer-Verlag, New York, 2001).
- [10] J. Skilling, *Bayesian Analysis* **1**, 833 (2006).
- [11] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, Oxford, 1999).
- [12] J. Machta, *Phys. Rev. E* **80**, 056706 (2009).
- [13] H. G. Katzgraber, M. Körner, and A. P. Young, *Phys. Rev. B* **73**, 224432 (2006).
- [14] H. G. Katzgraber and A. P. Young, *Phys. Rev. B* **67**, 134410 (2003).
- [15] H. G. Katzgraber, M. Körner, F. Liers, M. Jünger, and A. K. Hartmann, *Phys. Rev. B* **72**, 094421 (2005).
- [16] E. Marinari and F. Zuliani, *J. Phys. A* **32**, 7447 (1999).
- [17] M. Palassini and A. P. Young, *Phys. Rev. Lett.* **85**, 3017 (2000).
- [18] F. Krzakala and O. C. Martin, *Phys. Rev. Lett.* **85**, 3013 (2000).
- [19] H. G. Katzgraber and A. P. Young, *Phys. Rev. B* **65**, 214402 (2002).