

# Improving the efficiency of extended ensemble simulations: The accelerated weight histogram method

Jack Lidmar\*

*Theoretical Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden*

(Received 19 August 2011; revised manuscript received 1 May 2012; published 25 May 2012)

We propose a method for efficient simulations in extended ensembles, useful, e.g., for the study of problems with complex energy landscapes and for free energy calculations. The main difficulty in such simulations is the estimation of the *a priori* unknown weight parameters needed to produce flat histograms. The method combines several complementary techniques, namely, a Gibbs sampler for the parameter moves, a reweighting procedure to optimize data use, and a Bayesian update allowing for systematic refinement of the free energy estimate. In a certain limit the scheme reduces to the  $1/t$  algorithm of B. E. Belardinelli and V. D. Pereyra [*Phys. Rev. E* **75**, 046701 (2007)]. The performance of the method is studied on the two-dimensional Ising model, where comparison with the exact free energy is possible, and on an Ising spin glass.

DOI: [10.1103/PhysRevE.85.056708](https://doi.org/10.1103/PhysRevE.85.056708)

PACS number(s): 05.10.-a, 02.70.-c

## I. INTRODUCTION

The complex behavior of models with rough energy landscapes (such as spin glasses, biopolymers, etc.) is an important but challenging problem. In many situations progress is possible only using computer simulations, but this too is a notoriously difficult problem. In order to efficiently sample the equilibrium distribution of such models it is necessary to overcome the barriers separating different metastable minima, a process which can be very slow if the temperature is low. A particularly fruitful strategy to enhance the sampling is to enlarge the configuration space to include some well-chosen parameter(s) in the model. In simulated tempering [1], e.g., the temperature is promoted to a dynamical variable, whereby the system heats up and cools down randomly and gets a good chance to explore the energy landscape. Such *extended ensemble* or *generalized ensemble* methods have gained much attention recently and are routinely used in simulations of such diverse problems as spin glasses, biomolecules, and problems in statistics. The methods are also highly useful for free energy calculations and for the estimation of the probability of extreme events. An attractive feature is that they can easily be incorporated into existing simulation methods. The downside is, however, that in order to work properly they require fine tuning certain *a priori* unknown weights. The weights must be tuned to ensure that each parameter value (e.g., temperature) of the extended ensemble is visited equally often on average. They are simply related to the free energy at the given parameter value, and therefore they are a highly useful byproduct of the simulation, *if they can be estimated efficiently using some scheme*. While several such schemes have been constructed [1,2] there is a strong need for improvements. In this paper we propose one such scheme with a number of distinct advantages.

Section I gives a brief background on extended ensembles and discusses some shortcomings of existing methods. Section II introduces an improved method, the accelerated weight histogram method. In Sec. III the method is tested and

benchmarked on two model problems, the two-dimensional Ising model and a three-dimensional Ising spin glass.

### A. Extended ensembles

We consider a model described by a probability distribution  $\pi_\lambda(x)$ , which depends on one or more parameters  $\lambda$ . Typically we want to study the model for a whole range of parameter values. In an extended ensemble simulation, states are sampled according to a joint distribution  $P(x, \lambda)$ , which we express, without loss of generality, as

$$P(x, m) = \frac{1}{Z} e^{f_m - E_m(x)}, \quad (1)$$

where  $x \in \mathcal{X}$  denotes the configuration of the system and we assume a discrete set of preselected parameter values  $\lambda_m \in \mathcal{M} = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$ . The weights  $e^{f_m}$  introduced in Eq. (1) allow tuning the marginal distribution  $P(m)$  to approach any desired form. We assume that we have a way of generating samples from the conditional distribution

$$P(x|m) \equiv \pi_m(x) = e^{F_m - E_m(x)} \quad (2)$$

at fixed parameter  $\lambda_m$ , using, e.g., Markov chain Monte Carlo (MC) or molecular dynamics (MD) methods. Generally, this can be done without knowledge of the normalization constants  $e^{-F_m}$ . In physics applications Eq. (2) is often the ordinary canonical distribution  $\sim e^{-\mathcal{E}/T}$ , where we absorbed the temperature into the energy in order to treat it on equal footing as any other parameter of the system. Likewise,  $F_m$  denotes the dimensionless free energy [3]. In Bayesian statistics problems Eq. (2) is typically the *a posteriori* distribution for the model parameters and possibly missing data given a set of observations.

The ordinary (MC or MD) moves are then complemented with transitions in parameter space, which in most cases consist of a nearest neighbor random walk. The weights  $e^{f_m}$  need to be adjusted to make the marginal distribution

$$P(m) = \sum_x P(x, m) = \frac{1}{Z} e^{f_m - F_m} \quad (3)$$

\*jlidmar@kth.se

of  $m$  approximately flat [4]. This requires  $f_m \approx F_m$ , where  $F_m$  is the exact (dimensionless) free energy at  $\lambda_m$ , which is unknown at the beginning of the simulation.

Quite generally, existing methods to estimate the weights  $e^{f_m}$  can be divided into two different classes, *iterative* and *dynamic*. In an iterative method the weights are produced in a sequence of preliminary runs, each run giving a better estimate than the old, until sufficient accuracy is reached. On the other hand, in a dynamic scheme the weights are being continuously updated during a long simulation. The dynamic schemes have potential for faster convergence, but since the weights are constantly changing, detailed balance is violated and the samples collected cannot therefore be safely used to estimate average values of interest. In the iterative scheme the weights are fixed during each run, and they are only updated between the runs.

### B. The Wang-Landau and 1/t methods

One particularly elegant dynamic scheme is the Wang-Landau method [2], originally developed for simulations in the closely related multicanonical ensemble [5]. In this ensemble the state space is not extended, but instead one replaces the Boltzmann weights of the ordinary canonical ensemble with a different one aimed at producing a flat histogram of some quantity  $\lambda(x)$ , usually the energy. From an algorithmic point of view the main difference is that the elementary moves  $x \rightarrow x'$  also change the value of  $\lambda(x)$ , e.g., the energy, whereas in the extended ensemble method they can be performed at constant  $\lambda$ . The latter allows for more flexibility when choosing the parameter moves, something we exploit below. The Wang-Landau method is straightforward to adapt to extended ensemble simulations (as demonstrated in Ref. [6]). Each time the system visits a particular parameter  $m$ , the corresponding free energy parameter  $f_m$  is decreased by a certain amount,  $f_m \leftarrow f_m - \delta f$ . A histogram of visited parameter values is collected and  $\delta f$  is reduced by a factor  $\delta f \leftarrow \delta f/2$ , when the histogram meets a certain flatness criteria. Then the histogram is reset and the process starts over with the reduced modification constant. The scheme where  $\delta f$  is halved each iteration anticipates an exponential convergence of  $f_m$  to its true value  $F_m$ . Unfortunately, this is not the case. Instead, the error saturates at a level where reasonably flat histograms are produced, but the free energy estimate no longer improves since  $\delta f$  becomes too small [7,8]. It has been realized [7] that the modification factor should rather be decreased at a slow steady rate  $\delta f \sim 1/t$ , where  $t$  is the Monte Carlo time, without regard to any histograms, at least during the later stages of a simulation. The resulting 1/t method turns out to perform very well, both in multicanonical and extended ensemble simulations.

### C. Open issues

Nevertheless, there is still plenty of room for improvements. What is, for example, the most efficient way to move around in parameter space? How can the data collected during the simulations be used most effectively to produce an estimate of the free energy needed for uniform sampling? How should the estimates from different iterations, perhaps run in parallel, be

combined in an optimal way? How should the set of parameters  $\mathcal{M}$  be chosen? Most often the parameter moves form a nearest neighbor random walk, and then the choice of the spacing between adjacent values may be a critical issue. Having too large gaps between adjacent values may lead to small acceptance rates and therefore very slow dynamics along the parameter axis. Having too densely spaced parameter values, on the other hand, can make the dynamics of the random walk itself a limiting factor, again slowing down the dynamics.

## II. THE ACCELERATED WEIGHT HISTOGRAM METHOD

In this paper we propose an iterative scheme—the accelerated weight histogram method (AWH)—which combines several different complementary techniques to give a very efficient method which addresses the issues mentioned above. First of all, we allow large parameter steps by the use of a Gibbs sampler (also referred to as a heat bath algorithm). This is combined with a reweighting procedure which makes optimal use of the information collected during the moves. Together these make it possible to choose a rather densely spaced set of parameters, without being limited by slow diffusion. The free energy parameters are updated based on a histogram of *weights* (rather than a histogram of visited parameter values) combined with the information collected during previous iterations.

The parameter moves are carried out as follows. In the simplest case we allow transitions  $m \rightarrow m'$  to *any* new parameter value  $\lambda_{m'}$ , with a probability given simply by the conditional probability of  $m'$  given the current configuration  $x$ ,

$$w_{m'm}(x) = P(m'|x) = \frac{e^{-E_{m'}(x)+f_{m'}}}{\sum_{k \in \mathcal{M}} e^{-E_k(x)+f_k}}. \quad (4)$$

The transition probabilities just calculated are accumulated in a histogram of weights

$$W_k \leftarrow W_k + w_{km}(x), \quad \forall k. \quad (5)$$

Further, they can be used for on-the-fly reweighting of sampled observables,

$$\langle A \rangle_k = \frac{\sum_t A(x_t, k) w_{km_t}(x_t)}{\sum_t w_{km_t}(x_t)}, \quad (6)$$

where  $\{x_t, m_t\}$  denote the time series of visited configurations. The averages  $\langle A \rangle_k$  at a particular value  $\lambda_k$  thus get contributions from a whole range of parameter values. Note that the validity of Eq. (6) does not depend on the  $f_m$  being converged. This reweighting scheme is akin to the optimal multihistogram reweighting technique of Ferrenberg and Swendsen [9] (but with no need to solve a nonlinear equation system).

The update procedure continues in an iterative way. During each iteration a certain number, say  $N_I$ , of samples are collected and then the free energy parameters are updated as  $f_k \leftarrow f_k + \Delta f_k, \forall k$ , with

$$\Delta f_k = -\ln \left( \frac{W_k M}{N} \right), \quad (7)$$

where  $N$  is the total number of samples collected so far. The weight histogram is then updated to reflect this change:

$$W_k \leftarrow W_k e^{\Delta f_k} = N/M; \quad (8)$$

i.e., the total weight collected is distributed evenly among the  $M$  parameter values, and the next iteration starts. Note that the identity  $N = \sum_k W_k$  holds before and after the update. The histogram is thus *not* reset to zero after the iteration but continues to grow. This makes the updates from Eq. (7) become smaller and smaller and allows for finer and finer details of the free energy to be resolved.

Equations (4) to (8) form the core of the algorithm, which can be summarized as follows:

- (1) Perform  $N_x$  updates of the configurations  $x$  at fixed parameter value  $\lambda_m$ .
- (2) Perform a parameter move  $m \rightarrow m'$  using the Gibbs sampler, Eq. (4).
- (3) Update the weight histogram using Eq. (5) and sample any observables of interest using Eq. (6).
- (4) Repeat steps 1–3 until  $N_I$  samples have been obtained.
- (5) Update the free energy parameters  $f_m$  using Eq. (7) and the weight histogram using Eq. (8).
- (6) Start a new iteration from step 1 unless the desired accuracy has been reached.

One possible concern is that step 2 of the algorithm requires the computation of  $M = |\mathcal{M}|$  different quantities, which can become time consuming if the set  $\mathcal{M}$  is large (as can easily happen in the case of two- or higher-dimensional parameter spaces). In practice,  $w_{m'/m}(x)$  will be exponentially small except for a range of  $m'$  close to  $m$ . If this is the case one may limit the search for the new state to a neighborhood  $\Lambda \subset \mathcal{M}$  of  $m$  by replacing step 2 with

(2') Choose a subset of parameter values  $\Lambda$  with probability  $P(\Lambda|m)$ . Perform a parameter move  $m \rightarrow m' \in \Lambda$  using the Gibbs sampler [Eq. (4), but with the sum restricted to  $\Lambda$ ].

Detailed balance is maintained if  $P(\Lambda|m) = P(\Lambda|m')$  for all  $m, m' \in \Lambda$ . A simple choice (in the one-dimensional case) is to select a range of parameter values as an interval  $\Lambda = \{m - L, \dots, m - L + R\} \cap \mathcal{M}$ , where  $L$  is a random uniformly distributed integer in  $[0, R]$  and  $R$  is a predetermined range. The generalization to higher-dimensional parameter spaces is straightforward.

### A. Bootstrapping the simulation

Clearly, the update equation (7) requires an initial guess for  $f_k$  and a positive value of  $W_k = W_{\text{prior}}$  at the start of the simulation. This latter value can be seen as a Bayesian prior of our initial guess of  $f_k$ , which is later on updated as new data become available. If we have reason to believe that the starting estimate of the free energy is good (e.g., because the free energy is expected to have small variations), we can use a large  $W_{\text{prior}}$ . In many applications, however, our initial guess is going to be poor and we need some kind of bootstrap to get an acceptable prior. We propose the following heuristic scheme: Carry out the same steps in the simulation as above, but in addition check, after each iteration has completed (after step 5), whether all parameter values have been visited a certain fixed (usually small  $\sim 1$ –10) number of times. If not, reset the number of samples  $N \leftarrow M'$ , where  $M'$  is the number of parameters visited so far, and let  $W_k \leftarrow N/M = M'/M$ . In this way the weight histogram does not start to accumulate data

until  $M' = M$  whereby the free energy parameters will get relatively large updates at the initial stages. Also one should avoid sampling observables during this initial stage. Alternatively, one may use free energy perturbation or a few Wang-Landau iterations to get a reasonable initial estimate of  $f_m$ . After this, the simulation may proceed with an initial prior  $W_k = 1$ .

It is further recommended to make each iteration quite short, consisting of only  $N_I \sim 100$ –1000 parameter moves, during this initial stage. (Later on it may be increased.) It is also advisable to monitor the histogram  $H_m$  of visited parameter values, although it is not used directly to update the free energy. The robustness of the algorithm can then be increased by restarting the simulation if the histogram gets too skewed, e.g., if the minimum value  $H_{\min}$  is less than a certain fraction of the mean. This could be an indication that initial nonequilibrium transients have distorted the distribution of the collected samples, which would violate the main assumption of the algorithm, namely that the samples collected during each iteration follow Eq. (3). If this happens one should reset the weight histogram and the effective number of samples (e.g.,  $W_k \leftarrow H_{\min}$ ,  $N \leftarrow MH_{\min}$  or perhaps even  $W_k \leftarrow 1$ ,  $N \leftarrow M$ ), to allow the simulation to recover from that situation.

### B. Combining several simulations

Often it is advantageous to run simulations in parallel to make efficient use of computational resources. The scheme introduced above can easily be adapted to such situations. Each computing node ( $n$ ) runs an independent simulation (consisting of  $N_n$  samples), leading to an estimate  $f_m^{(n)}$  of the free energy parameters. These may then be combined into a best estimate  $\bar{F}_m$ ,

$$e^{-\bar{F}_m} = \mathcal{N} \sum_n N_n e^{-f_m^{(n)}} \mathcal{Z}^{(n)}, \quad (9)$$

where  $\mathcal{Z}^{(n)} = \sum_k e^{f_k^{(n)} - \bar{F}_k}$  and  $\mathcal{N}$  is an unimportant normalization constant. This equation is easily solved by iterating

$$\bar{F}_m \leftarrow \bar{F}_m - \ln \left( \frac{M \sum_n N_n e^{\bar{F}_m - f_m^{(n)}} \mathcal{Z}^{(n)}}{\sum_{m,n} N_n e^{\bar{F}_m - f_m^{(n)}} \mathcal{Z}^{(n)}} \right), \quad (10)$$

starting from one of the  $f_m^{(n)}$  (and this usually converges within 2–5 iterations). This way of organizing the simulation also has the advantage that statistical errors can be estimated using the standard jackknife method [10] applied to Eq. (10).

### C. Relation to the 1/t method

Many variations of the basic algorithm are possible and may be related to other methods. For example, the basic algorithm reduces to the 1/t method in the limit obtained by the following modifications: (1) Replace the Gibbs sampler by a simple nearest neighbor Metropolis step. (2) Replace the weight histogram  $W_m$  by a simple histogram  $H_m$  of visited  $m$ . (3) Update the free energy parameters after *every* step. Since the histogram after a visit to  $m$  is  $H_k = N/M + \delta_{km}$ , the free energy update becomes  $\Delta f_k = -\ln[H_k M / (N + 1)] = -\ln(1 + \delta_{km} M / N) + \ln(1 + 1/N) \approx -\delta_{km} M / N + 1/N$ , where the approximation holds when  $N \gg M$ . The last term represents a constant shift of all  $f_k$  and can be dropped. The resulting update rule is thus

simply  $f_m \leftarrow f_m - M/N$ , leaving all other  $f_k$  unmodified. This corresponds exactly to the  $1/t$  method [7] discussed earlier and provides a new perspective on and additional justification for that update scheme.

### III. BENCHMARKS OF THE METHOD

To study the performance of the method and compare it with other ones we apply it to the Ising model and a spin glass. We carry out a simulated tempering simulation; i.e., we choose as parameter  $\lambda$  the temperature. The algorithm alternates between ordinary canonical Metropolis MC updates in which randomly chosen spins are flipped with probability  $\min(1, e^{-\beta\Delta\mathcal{E}})$  and updates which change the temperature, leaving the spin configuration and the energy  $\mathcal{E}$  unchanged. In the latter ones a new temperature  $T_{m'} = 1/\beta_{m'}$  is chosen with the probability

$$w_{m'm}(\mathcal{E}) = \frac{e^{-\beta_{m'}\mathcal{E} + f_{m'}}}{\sum_k e^{-\beta_k\mathcal{E} + f_k}}. \quad (11)$$

#### A. Two-dimensional Ising model

The two-dimensional (2D) Ising model is a common test case, since its free energy can be calculated exactly [11]. We choose  $M = 128$  temperatures evenly spaced in the interval  $[1.8, 3]$ , which includes the critical temperature  $T_c = 2/\ln(1 + \sqrt{2}) \approx 2.27$ . The system size is  $L = 64$  and we use 100 000 iterations, each lasting for 1000 MC sweeps, in total  $10^8$  sweeps, where each MC sweep corresponds to one update trial per spin. A temperature move is attempted after each MC sweep.

During the initial stages we use the scheme discussed in Sec. II A to get an initial guess for the  $f_m$  and a prior weight  $W_{\text{prior}}$ : At the start of the simulation we set  $f_k = \beta_k \mathcal{E}_0$ , where  $\mathcal{E}_0 = -2L^2$  is the ground-state energy and  $W_k = 1/M$ . Then we check, after each iteration, whether all  $M$  temperatures have been visited at least twice during the simulation so far. If not, the effective number of samples is reset to  $N = M'/M$ , where  $M'$  is the number of temperatures which actually were visited twice. When all temperatures have been visited we have a reasonable initial guess of  $f_m$ , and we continue the simulation as described in Sec. II, with  $W_k \geq 1$ . Furthermore, we also monitor the histogram of visited temperatures to look for anomalous deviations, which would indicate that the initial guess was not so good after all. Thus, we restart the simulation (i.e., we set  $N = M$  and  $W_k = 1$  and reset the calculations of any observables, but we do not touch the  $f_k$ ) should the histogram of visited temperatures  $H_m$  at some point fall below 2% of its mean. This happened in about half of the simulation runs, typically within the first 50 iterations.

To benchmark the method we plot, in Fig. 1(a), the mean absolute deviation

$$\delta F = \frac{1}{M-1} \sum_{m=1}^{M-1} |f_{m+1} - f_m - F_{m+1} + F_m| \quad (12)$$

of consecutive free energy differences against the number of samples. Here  $F_m = \mathcal{F}(T_m)/T_m$  is the exact dimensionless

free energy. For comparison we also include results from simulations using Wang-Landau iterations (with flatness criteria  $H_{\text{min}} > 0.4H_{\text{mean}}$ ) and the  $1/t$  method. For large times, the error for both the  $1/t$  and our method decrease as  $1/\sqrt{N}$ , whereas it saturates for the Wang-Landau method. For a given number of samples, the accuracy of the AWH method is almost one order of magnitude better than that of the  $1/t$  method. The inset shows the difference between the final estimate, obtained by combining 40 independent simulations using Eq. (10), and the true free energy over the temperature range. The error bars are estimated using the jackknife method.

Another useful measure of the efficiency is the tunneling time, i.e., the time to go from the highest temperature to the lowest or *vice versa*. This time was significantly reduced, nearly by a factor of 2, from  $\sim 40\,000$  MC sweeps for the  $1/t$  to  $\sim 21\,000$  for the AWH method. It should be noted that the dynamics suffer severely from critical slowing down in the vicinity of the phase transition, which constitutes a bottleneck for the movement along the temperature axis. While the extended temperature ensemble methods are effective for crossing energy barriers, they do not overcome this slowing down by themselves. In this sense the 2D Ising model (using single spin flip dynamics) is not a particularly favorable test case. However, the methods can easily be combined with cluster methods, if available, which do overcome the critical slowing down. Replacing the single spin flip moves by, e.g., Wolff cluster updates [12] for  $|T_m - T_c| < 0.1$  (the cluster moves being most effective in the critical region) in the example above practically eliminates the bottleneck and further reduces the tunneling time by an additional factor  $\approx 10$  to about 2200, for the AWH method. The  $1/t$  method on the other hand only gained a factor of 2.

As discussed in Sec. II one of the advantages of the AWH method is the insensitivity to the spacing of parameter values  $\lambda_m$ . Indeed, varying the number of temperatures from  $M = 32, 64, 128$  up to 256 had negligible effect on the performance of the algorithm, in terms of both the accuracy of the final free energy estimate and the tunneling time, while the increase in the run time of the simulation was marginal (and could be practically eliminated using the update rule 2'). Upon decreasing  $M$  below 16, on the other hand, the performance quickly dropped.

#### B. Three-dimensional Ising spin glass

Next we apply the method to the three-dimensional Ising spin glass with Gaussian couplings. This model has a disorder-dominated glass phase at low temperatures  $T < T_g \approx 0.95$  [13], with a very rough energy landscape, making it extremely challenging to study using conventional simulations. The system size is  $L = 8$ , and we use  $M = 200$  temperatures logarithmically spaced in  $[0.7, 3.5]$ . Figure 1(b) compares the convergence of the different methods for one particular random realization of the couplings. As there is no exact solution to compare with we use as reference instead the best estimate obtained from 80 different runs (with an estimated standard error  $< 0.002$ ). Here, the gain in accuracy, compared to the  $1/t$  method, is more than an order of magnitude. The tunneling time, i.e., the time to go between the high- and low-temperature extremes, is also significantly shorter, by nearly a factor of 20.



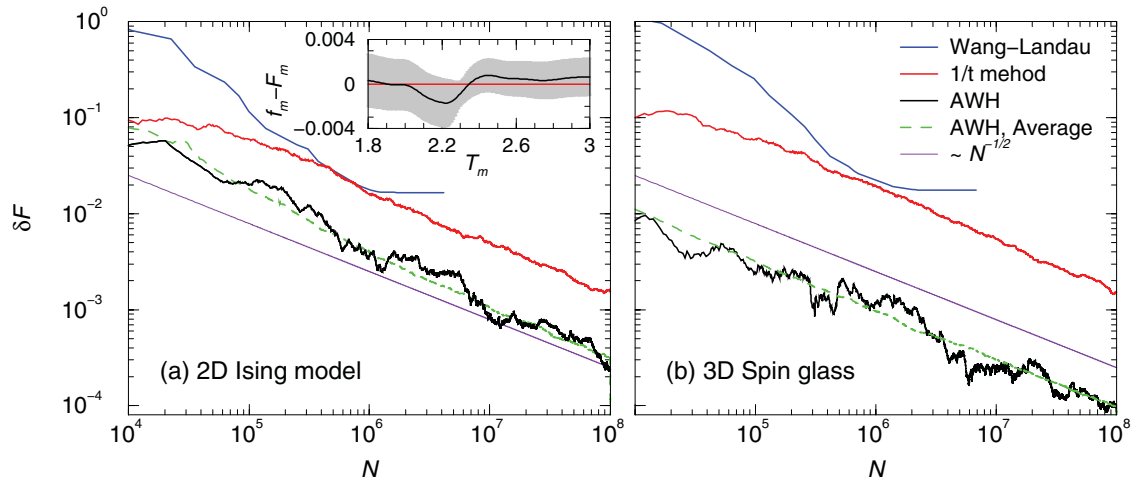


FIG. 1. (Colors online) (a) Mean absolute deviation of estimated and exact free energy differences  $\delta F$  of the  $64 \times 64$  Ising model as a function of number of samples,  $N$ . From top to bottom: results from Wang-Landau iterations, the  $1/t$  method, and the AWH method. Also included is the average behavior of the AWH method and a curve showing a  $1/\sqrt{N}$  dependence. Inset: Difference between estimated and exact free energy. The error bars (the shaded area) represent one standard deviation. (b) As in (a), but for an  $8 \times 8 \times 8$  Ising spin glass.

#### IV. SUMMARY AND CONCLUSIONS

Let us reiterate the advantages of the AWH method: Allowing for large steps gives a fast diffusion along the parameter axis. As a result, the spacing between neighboring values in the discretized parameter space is not critical as long as it is small enough and does not require any fine tuning to perform well. We make efficient use of the data collected at all stages of the simulation. This is done by reweighting on the fly the samples taken at the current parameter value to a whole range of different parameter values. The information needed for this reweighting procedure is essentially the same as what enables the large steps. The data taken at earlier iterations are not thrown away, but instead they are used together with the new data to refine the estimate of the free energy parameters. Since the weights are constant during each iteration, the data collected will, after an initial relaxation, be in equilibrium and can be used for the calculation of any desired averages.

Altogether, these properties make up a very convenient method for sampling models with rough energy landscapes, and for the calculation of free energy differences. It should be emphasized that it is the combination of the Gibbs sampler, the reweighting scheme, and the update rule using the weight histogram which leads to the dramatic improvements. The method is very general, is simple to implement, and can be applied to a broad range of problems in statistical physics, biophysics, statistics, etc. Further improvements are likely, especially when it comes to the heuristic scheme used during the early-stage bootstrap.

#### ACKNOWLEDGMENTS

This work was supported by the Swedish Research Council (VR) through Grant No. 621-2007-5138 and the Swedish National Infrastructure for Computing (SNIC 001-10-155) via PDC.

- 
- [1] E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992); A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).
  - [2] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
  - [3] A common example is the simulated tempering ensemble, where  $\lambda_m = T_m$  and  $F_m = \mathcal{F}(T_m)/T_m$ , with  $\mathcal{F}$  equal to the real free energy.
  - [4] We leave aside the question of whether a flat distribution is really optimal. This will typically depend on the particular problem at hand. One reasonable possibility would be, e.g., to make it proportional to the correlation time at  $\lambda_m$ . We just note that a slight modification of our formalism allows for any prescribed target distribution  $\pi_m$  (or even simpler let the density of parameter values be nonuniform).
  - [5] B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
  - [6] C. Zhang and J. Ma, *Phys. Rev. E* **76**, 036708 (2007).
  - [7] R. E. Belardinelli and V. D. Pereyra, *Phys. Rev. E* **75**, 046701 (2007); *J. Chem. Phys.* **127**, 184105 (2007).
  - [8] C. Zhou and J. Su, *Phys. Rev. E* **78**, 046705 (2008).
  - [9] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
  - [10] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).
  - [11] A. Ferdinand and M. Fisher, *Phys. Rev.* **185**, 832 (1969).
  - [12] U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989).
  - [13] H. G. Katzgraber, M. Körner, and A. P. Young, *Phys. Rev. B* **73**, 224432 (2006).