PHYSICAL REVIEW E **84**, 065702(R) (2011)

Convergence for the Wang-Landau density of states

G. Brown
Department of Physics, Florida State University, Tallahassee, Florida 32306, USA and
Computational Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA

Kh. Odbadrakh
Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA

D. M. Nicholson
Computational Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA

M. Eisenbach
National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA
(Received 23 May 2011; revised manuscript received 7 October 2011; published 20 December 2011)

The Wang-Landau method of estimating the density of states $g(E)$ has become a powerful tool in statistical mechanics. Here it is shown that the distribution of random walkers sampled using an estimated density of states can always be used to improve the estimate. Specifically, this can be done without resorting to an auxiliary modification factor f, which previously has been used to find $g(E)$ self-consistently through a procedure that reduces f incrementally toward unity. This straightforward approach is validated for multiple, independent random walkers.

DOI: 10.1103/PhysRevE.84.065702 PACS number(s): 05.10.Ln, 75.10.Hk, 02.70.Tt

Calculations in statistical mechanics usually involve a large number of degrees of freedom which define a high-dimensional phase space. Frequently, calculations over the points—or microstates—in this phase space involve a series of random steps from one microstate to another microstate, and a particular trajectory is thought of in terms of a simulation that involves a Markov random walker jumping from microstate to microstate following some probabilistic formula. The concept of detailed balance refers to the constraint that the net flux of random walkers between any two microstates in the phase space is zero, which in turn ensures the formula describes an equilibrium calculation. By extension, detailed balance implies the flux from the set of all microstates with energy E to the set of those with energy E' must be equal in magnitude to the flux in the other direction

$$\Phi(E \to E') = \Phi(E' \to E). \tag{1}$$

For importance-sampling Monte Carlo (MC) [1,2] the flux Φ is usually factored into the average density of walkers n, the probability P of proposing a move from a microstate with energy E to one with energy E', and A the probability of accepting that move. Detailed balance then implies

$$\begin{aligned} &n(E)P(E \to E')A(E \to E') \\ &\quad = n(E')P(E' \to E)A(E' \to E). \end{aligned} \tag{2}$$

Usually the method for proposing moves is symmetric in the sense $g(E)P(E \to E') = g(E')P(E' \to E)$, with $g(E)$ the density of states (DOS), i.e., the relative fraction of microstates with energy E. Accepting moves with the Metropolis [1] formula $A(E \to E') = \min[1, w(E')/w(E)]$, with $w(E)$ a function that weights the choices, will generate a sequence of microstates that satisfy detailed balance. Using the Metropolis formula, Eq. (2) can be rewritten

$$n_w(E)w(E')g(E') = n_w(E')w(E)g(E), \tag{3}$$

with the density of walkers associated with a particular choice of w written as $n_w(E)$. For any energy E or E' detailed balance thus enforces the ratio between n_w, w and g [3]

$$\frac{n_w(E)}{w(E)g(E)} = \frac{n_w(E')}{w(E')g(E')} = 1, \tag{4}$$

with the normalization chosen to be unity for convenience. If the Boltzmann factor is used as the weight function $w(E) = e^{-E/k_BT}$, the average walker density will have the same probability distribution as the canonical ensemble

$$\frac{n_w(E')}{n_w(E)} = e^{-(E'-E)/k_BT} e^{[S(E')-S(E)]/k_B} = e^{-(F'-F)/k_BT}, \tag{5}$$

with $S = k_B \ln[g(E)]$ the entropy and $F = E - TS$ the Helmholtz free energy. The expectation value of a quantity X in the canonical ensemble is

$$\langle X \rangle = \frac{\sum_E X(E)g(E)e^{-E/k_BT}}{\sum_E g(E)e^{-E/k_BT}}. \tag{6}$$

Using the Boltzmann weighting factor for the random walker reduces this, using Eq. (5), to

$$\langle X \rangle = \frac{\sum_E X(E)n_w(E)}{\sum_E n_w(E)}, \tag{7}$$

which is equivalent to the average value of X sampled over the random-walk simulation.

In contrast, the Wang-Landau approach to importance sampling [4] uses an estimated value for the density of states for the weighting function in the Metropolis acceptance formula $w(E) = g_i^{-1}(E)$. The estimated density of states $g_i(E)$ is calculated successively along with the associated density of

walkers $n_i(E)$ until it is converged to a self-consistent value, where i is used to index each stage in the sequence of estimates. While finding the self-consistent estimate,

$$g(E) = \lim_{i\to\infty} g_i(E) \tag{8}$$

takes significant computational resources, and once determined it can be used to evaluate Eq. (6) at all temperatures at essentially no cost. It is not surprising that understanding the convergence of $g_i(E)$ for the Wang-Landau method, and trying to optimize it, is a very active topic [5].

Detailed balance provides an avenue for understanding the convergence in the estimated density of states. At stage i, detailed balance formulated as Eq. (4) gives

$$\ln[g(E)] = \ln[g_i(E)] + \ln[n_i(E)]. \tag{9}$$

The value of $n_i(E)$ is estimated during the course of K sampling steps $n_i^K(E)$. There are two contributions to $n_i^K(E)$. The dominant contribution is the difference between the logarithm of the estimated density of states and the exact value; Eq. (9) is what provides a convergence procedure for the Wang-Landau method. There is a second contribution due to the random nature of the walkers even if the estimated density of states is exactly equal to the true value. In this case of unbiased diffusion $g_i(E) = g(E)$, sampling of the equilibrium distribution will contain statistical fluctuations and $n_i^K(E)$ will converge only for very large K. Determining convergence in the Wang-Landau method often depends on the behavior of these statistical fluctuations.

Wang and Landau implemented Eq. (9) by adding a factor $\ln(f_i)$ at each step of the random walker

$$\ln[g_{i,j+1}(E)] = \ln[g_{i,j}(E)] + \ln(f_i), \tag{10}$$

with j indexing the random-walk steps within a stage i of the calculation. To find a self-consistent estimate of g, the value f_i is gradually reduced from e to 1. Within one stage of constant f_i, the histogram $\ln[g_i(E)]$ is modified until $n_i(E)$ meets some flatness criterion which indicates that the unbiased, diffusive contribution to $n_i(E)$ dominates the deterministic part driving the convergence. Mathematically, the estimated density of states is expressed as

$$\ln[g_{i+1}(E)] = \ln[g_i(E)] + \ln(f_i)n_i(E), \tag{11}$$

but since $g_i(E)$ is modified at each step of the random walker, $n_i(E)$ will depend on the history of the random walk. This effect can be minimized by delaying the modification of the density of states, i.e., modifying it only after every K steps of the random walk [6].

The convergence of this traditional Wang-Landau approach is illustrated for the Ising model on a square lattice with Hamiltonian

$$E = 2NJ - J\sum_{\langle i,j\rangle} s_i s_j, \tag{12}$$

with $s_i = \pm 1$, the sum restricted to nearest-neighbor spins, and $E = 0$ when all spins are aligned. The results for the 6×6 Ising model are shown in Fig. 1. The true density of states found by exact enumeration [7] appear as the filled circles, with $g(E) = 2$ for the ground state ($E = 0$) and the antiferromagnetic configurations ($E = 144J$). The curves are

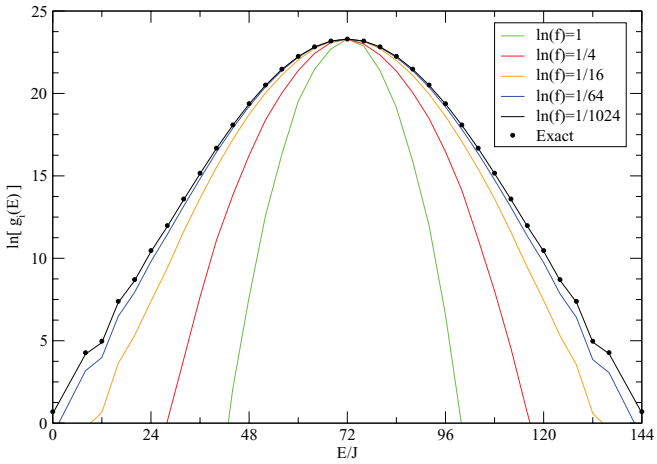

FIG. 1. (Color online) Convergence of the traditional Wang-Landau approach for the 6×6 Ising model using a flatness criterion of 0.5. The exact results appear as filled circles, and the estimates for progressively smaller values of the self-consistent parameter f are shown as curves, from narrowest to widest, respectively.

nested (color online), with $\ln(f) = 1$ the innermost curve and $\ln(f) = 1/1024$ the one in best agreement with the exact results. Each curve represents the results of 100 000 Monte Carlo steps of the random walker at fixed f_i (with $K = 1$), more than sufficient to meet the traditional flatness criterion of Wang and Landau with a value of 0.5. Increasing the number of Monte Carlo steps at each stage does not change the results. The results of 500 independent runs of the ith stage were averaged together to yield smooth results.

For large values of f_i, the estimated $g_i(E)$ is a parabola whose curvature is determined by a minimum curvature that appears to be inherent to the algorithm, and which can be motivated as follows. Numerically, the second derivative can be determined using the difference formula

$$\ln[g_i(E)]'' = \frac{1}{(\delta E)^2}(\ln[g_i(E+\delta E)] + \ln[g_i(E-\delta E)] - 2\ln[g_i(E)]), \tag{13}$$

with δE the difference between adjacent energies. Since neighboring values of $\ln[g_i(E)]$ can differ by only $\ln(f)$, the smallest nonzero value the second derivative can have is $\ln[g_i(E)]'' = \ln(f)/(\delta E)^2$. Qualitatively, this minimum-curvature effect can be seen in Fig. 1. In order to produce a good estimate of $\ln[g_i(E)]$, $\ln(f_i)$ must be small enough to avoid this effect. Further reductions in $\ln(f_i)$ below this level lead mostly to a smoothing of the estimated density of states.

A more obvious implementation of Eq. (9) is

$$\ln[g_{i+1}(E)] = \ln[g_i(E)] + \ln[n_i(E)], \tag{14}$$

with $n_i(E)$ determined while $\ln[g_i(E)]$ is held fixed. In this approach there is no convergence parameter $\ln(f_i)$. Repeated iterations of sampling for many Monte Carlo steps determine $n_i(E)$ and modify $g_i(E)$ according to Eq. (14). This approach is essentially the same as the $f = 1$ update of Lee, Okabe, and Landau [8], the entropic sampling of Lee [9], and the multicanonical approach of Berg and Celik [10]. The convergence of this approach are shown in Fig. 2 with $K = 50\,000$ samples in each stage indexed by i. In the figure, a thicker curve [orange

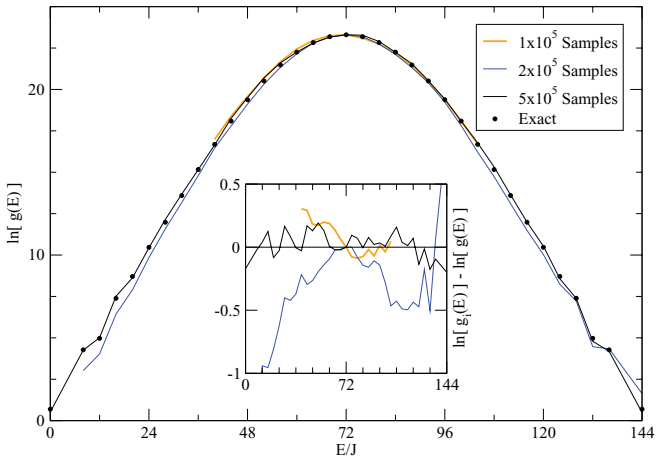

FIG. 2. (Color online) Convergence of the ln-modification scheme of Eq. (14). The estimated $\ln[g_j(E)]$ is modified every 5×10^4 samples, and the results shown cover progressively wider ranges of energy. The results are for the 6×6 Ising model, with the exact results appearing as filled circles. The difference between the estimate and the exact result are shown in the inset.

online (light gray)] is used for $\ln[g_2(E)]$, which only exists for $40 \leqslant E/J \leqslant 104$. The estimate $\ln[g_4(E)]$ [blue online (dark gray)] exists for all energies except the ground state, while the estimate for $\ln[g(E)]$ agrees visually with the exact result after 500 000 samples. The difference between the estimate and the exact result are shown in the inset. These results validate the approach of Eq. (14).

The fact that the estimates do not approach the exact value from below suggests better convergence than occurs for the traditional Wang-Landau approach, which is consistent with the results reported in Ref. [8]. However, the underestimation of $\ln[g_i(E)]$ keeps $n_i(E)$ large for the extreme values of E and enables the Wang-Landau algorithm to sample a wider range of energies faster than other methods. This makes the Wang-Landau algorithm useful in the initial stages of estimating $\ln[g_i(E)]$. Ultimately, it is unable to converge to the true value of $\ln[g_i(E)]$, as shown by Belardinelli and Pereyra [11]. The relative error between the Wang-Landau estimate and the exact value is shown in Fig. 3 as the dashed curve. The error is quite large for the first 10^5 MC steps when the minimum curvature effects dominate and then steadily decrease until becoming constant for all times $t > 10^6$ MC steps. Belardinelli and Pereyra suggested an approach that overcomes this limitation, namely,

$$\ln[g_{i,j+1}(E)] = \ln[g_{i,j}(E)] + \max[\ln(f_i), N_E/t], \quad (15)$$

with t the Monte Carlo "time" of the random walker and N_E the number of discrete energy states. This has become known as the $1/t$ Wang-Landau method, and its relative error is shown as the solid curve in Fig. 3. Naturally, it behaves the same as the traditional Wang-Landau approach at early times, and it continues to converge to the exact value at larger times. This makes sense with the observation that $\ln n \approx 1 + 1/2 + 1/3 + \cdots + 1/n$ if the random walker visits each energy on the order of every N_E steps. In terms of Eq. (14), the results for using Wang-Landau for the first 10^5 Monte Carlo steps, at which point all the energies have been visited, followed by sampling

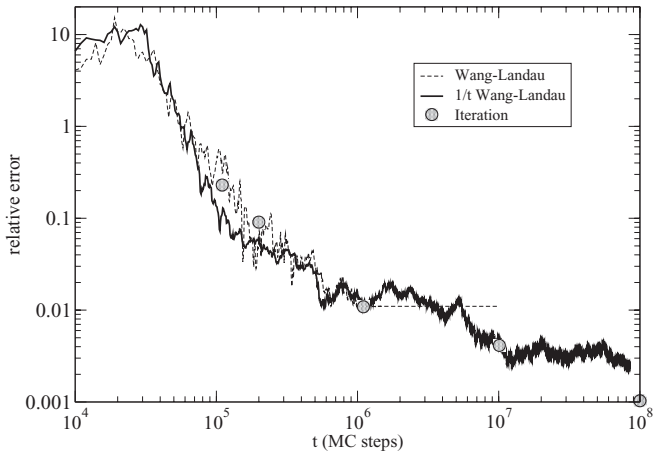

FIG. 3. The energy-averaged relative error vs Monte Carlo steps for the 6×6 Ising model using all three methods discussed here. The traditional Wang-Landau algorithm is shown as the dashed curve, and the convergence is observed to be constant at a finite value after $\sim 10^6$ steps. The results for the $1/t$ Wang-Landau algorithm are shown as the solid curve, and continued convergence is seen throughout the sampling. A flatness criterion of 0.5 was used for both. The results for the iterative approach, starting from the Wang-Landau density of states after 10^5 steps, is shown as the circles.

for $K = 10^5, 10^6, 10^7$, and 10^8 steps are shown as the circles in Fig. 3. This performs at least as well as the $1/t$ Wang-Landau method. Successively longer sampling times are required since sampling errors will dominate $n_i(E)$ after a few updates with fixed K.

The fixed density of states approach of Eq. (14) is especially powerful when utilizing many random walkers linked by slow communication [12]. This is a very real advantage in computational environments with many processing units linked by high-latency communication, such as GPU-based architectures or massively parallel supercomputers. The $\ln[g_i(E)]$ for the 32×32 Ising model calculated using 100 independent random walkers is shown in Fig. 4, with the difference between

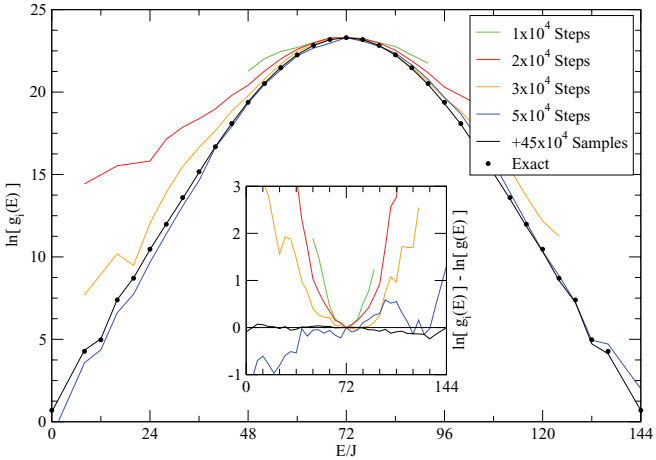

FIG. 4. (Color online) Wang-Landau density of states $g(E)$ for the 32×32 Ising Model for 100 independent Markov walkers. The initial estimate is shown as the dashed curve, and the estimated DOS was modified every $K = 10^7$ samples. The difference between the estimate and the exact result are shown in the inset.

the estimate and the exact result shown in the inset. This calculation was started with a $\ln[g_0(E)]$ found by fitting a parabola to the density of states near the maximum in $g(E)$. All walkers started from the ferromagnetic configuration, with $K = 10^7$, i.e., 10^5 steps per walker between modifications. The difference between the estimate and the exact values are shown in the inset.

To summarize, the expression governing convergence of the Wang-Landau method, Eq. (9), can be derived from the concept of detailed balance. The traditional Wang-Landau algorithm is observed to generate estimated density of states that approach the exact value from below, which enables it to sample a wider range of energies than other algorithms. The algorithm is unable to converge to arbitrarily precision, and another method must be used to refine the estimated density of states. While the $1/t$ Wang-Landau method achieves this in a statistical sense, a more natural approach that can be easily extended to parallel computations is suggested here. This method appears to converge at least as fast as the $1/t$ Wang-Landau approach.

We would like to acknowledge computer resources provided by Florida State University and review of the mathematics by P. A. Rikvold. This work was performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725, and sponsored by the Laboratory Directed Research and Development Program (ORNL), by the Mathematical, Information, and Computational Sciences Division; Office of Advanced Scientific Computing Research (US DOE), and by the Division of Materials Sciences and Engineering; Office of Basic Energy Sciences (US DOE).

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[2] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986).

[3] S. Trebst, D. A. Huse, and M. Troyer, Phys. Rev. E **70**, 046701 (2004).

[4] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001).

[5] C.-G. Zhou and J. Su, Phys. Rev. E **78**, 046705 (2008).

[6] C. Zhou and R. N. Bhatt, Phys. Rev. E **72**, 025701 (2005).

[7] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).

[8] H. K. Lee, Y. Okabe, and D. P. Landau, Comput. Phys. Commun. **175**, 36 (2006).

[9] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).

[10] B. A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).

[11] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).

[12] L. Zhan, Comput. Phys. Commun. **79**, 339 (2008); M. Eisenbach, C.-G. Zhou, D. M. Nicholson, G. Brown, J. Larkin, and T. C. Schulthess, *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (ACM, New York, 2009).