Performance Limitations of Flat-Histogram Methods

P. Dayal,[1] S. Trebst,[1,2] S. Wessel,[1] D. Würtz,[1] M. Troyer,[1,2] S. Sabhapandit,[3] and S. N. Coppersmith[3]
[1]*Theoretische Physik, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland*
[2]*Computational Laboratory, Eidgenössische Technische Hochschule Zürich, CH-8092 Zürich, Switzerland*
[3]*Department of Physics, University of Wisconsin, Madison, Wisconsin 53706, USA*
(Received 17 June 2003; published 2 March 2004)

We determine the optimal scaling of local-update flat-histogram methods with system size by using a perfect flat-histogram scheme based upon the exact density of states of 2D Ising models. The typical tunneling time needed to sample the entire bandwidth does not scale with the number of spins N as the minimal N^2 of an unbiased random walk in energy space. While the scaling is power law for the ferromagnetic and fully frustrated Ising model, for the $\pm J$ nearest-neighbor spin glass the distribution of tunneling times is governed by a fat-tailed Fréchet extremal value distribution that obeys exponential scaling. Furthermore, the shape parameters of these distributions indicate that statistical sample means become ill defined already for moderate system sizes within these complex energy landscapes.

DOI: 10.1103/PhysRevLett.92.097201 PACS numbers: 75.10.Hk, 02.70.Rr, 64.60.Cn

Monte Carlo methods are well suited for the simulation of large many body problems, since the complexity for a single Monte Carlo update step scales only polynomially and often linearly in the system size, while the configuration space grows exponentially with the system size. The performance of a Monte Carlo (MC) method is then determined by how many update steps are needed to efficiently sample the configuration space. For second order phase transitions in unfrustrated systems the problem of "critical slowing down"—a rapid divergence of the number of MC steps needed to obtain a subsequent uncorrelated configuration—was solved more than a decade ago by cluster update algorithms [1]. At first order phase transitions and in systems with many local minima of the free energy such as frustrated magnets or spin glasses, there is the similar problem of long tunneling times between local minima. With energy barriers ΔE scaling linearly with the linear system size L, the tunneling times τ at an inverse temperature $\beta = 1/k_B T$ scale exponentially with the system size, $\tau \sim \exp(\beta \Delta E) \propto \exp(\text{const} \times L)$. Several methods were developed to overcome this tunneling problem, such as the multicanonical method [2], simulated and parallel tempering [3], and Wang-Landau sampling [4]. The common aim of these methods is to broaden the range of energies sampled within MC simulations from the sharply peaked distribution of canonical sampling at fixed temperature in order to ease the tunneling through barriers. Ideally, all relevant energy levels are sampled equally often during a simulation, thus producing a "flat histogram" in energy space.

With a probability $p(E)$ for a single configuration of energy E, the probability of sampling an arbitrary configuration with energy E is given as $P_E = \rho(E)p(E)$, where the density of states $\rho(E)$ counts the number of states with energy E. Upon choosing $p(E) \propto 1/\rho(E)$ instead of the canonical weight $p(E) \propto \exp(-\beta E)$ one obtains a constant probability P_E for visiting each energy level E, and hence a perfectly flat histogram.

In this Letter we investigate the performance of flat-histogram algorithms for three systems for which the density of states $\rho(E)$ is known exactly on finite two-dimensional (2D) lattices: the Ising ferromagnet, the fully frustrated Ising model (see inset of Fig. 1) as a prototype for frustrated systems, and the $\pm J$ Ising spin glass. For each of these models we construct a *perfect flat-histogram method* by simulating a random walk in configuration space where we employ the *known density of states* for these models to set $p(E) \propto 1/\rho(E)$.

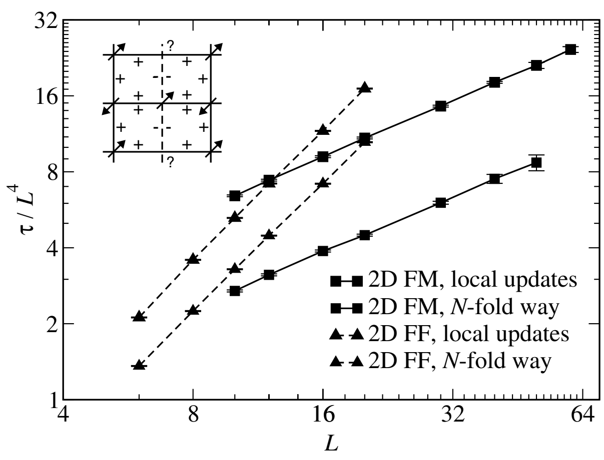

FIG. 1. Scaling of tunneling times τ from the ground state to the anti–ground state as a function of system size L for a perfect flat histogram method that samples using the exact density of states. Shown are results for the ferromagnet (FM) and fully frustrated (FF) 2D Ising models with both local and N-fold way updates; the inset illustrates the frustrated couplings. In all cases polynomial scaling $\tau \propto L^{2d+z}$ is found, with $z^{\text{FM}}_{\text{local}} = 0.743 \pm 0.007$, $z^{\text{FM}}_{N\text{-fold}} = 0.729 \pm 0.011$, $z^{\text{FF}}_{\text{local}} = 1.727 \pm 0.004$, and $z^{\text{FF}}_{N\text{-fold}} = 1.692 \pm 0.004$.

As a measure of performance we use the average tunneling time τ to get from a ground state (lowest energy configuration) to an antiground state (configuration of highest energy), which is the relevant time scale for sampling the whole phase space [5]. Since the number of energy levels in a d-dimensional system with linear size L scales with the number of spins $N = L^d$, the tunneling time for a pure random walk in energy space is $\tau \propto N^2 = L^{2d}$. None of the systems we study exhibit this scaling. While the ferromagnetic and fully frustrated models exhibit power-law scaling, for the spin glass the distribution of characteristic tunneling times is extremely broad and appears to diverge exponentially with system size.

We first look at the homogeneous systems for which the exact densities of states $\rho(E)$ were calculated using the program of Beale [6] for the ferromagnet and the algorithm of Saul and Kardar [7] for the fully frustrated Ising model. In Fig. 1 the measured tunneling times are plotted versus system size L for the perfect flat-histogram method using both local and N-fold way updates. Instead of the expected N^2 scaling, we find a more rapid increase following $\tau \propto L^{2d+z}$. Frustration significantly increases the scaling exponent but still conserves power-law scaling.

A previous study [8] suggests that N-fold way updates [9] speed up flat-histogram sampling. We find identical scaling exponents for N-fold way and local updates within our error bars, see Fig. 1, implying that any performance improvement remains constant with system size. N-fold way updates reduce the tunneling time by roughly a factor of 2, independent of system size. In practice, the reduction of tunneling times is offset by the added expense of N-fold way updates.

To determine the performance in more complex energy landscapes, we study the 2D $\pm J$ Ising spin glass, where we find exponential scaling. We measured tunneling times of the perfect flat-histogram method for 1000 realizations for the system sizes $L = 6, 8, \ldots, 18$, and 350 realizations for $L = 20$ using the exact density of states obtained by the algorithm of Saul and Kardar [7]. For fixed system size the tunneling times are scattered over several orders of magnitude for the various realizations. The lower end of this distribution is shown in Figs. 2(a) and 2(b), and the whole range of measured tunneling times is shown in Fig. 4.

To analyze the underlying distributions we use extremal value theory [10]. The central limit theorem for extremal values [11] states that the extrema of large samples are distributed according to one of only three distributions, depending on whether the tails of the original distribution are fat tailed (algebraic), exponential, or thin tailed (decaying faster than exponential). This theorem is successfully applied in the analysis of tails in diverse fields such as hydrology, insurance, and finance. Surprisingly, here we find that *not only the extrema*, but *all of the measured tunneling times* [see Figs. 2(a) and 2(b)] are distributed according to the Fréchet extremal value distribution for fat-tailed distributions which has the integrated probability density:

$$H_{\xi;\mu;\beta}(\tau) = \exp\left[-\left(1 + \xi\frac{\tau - \mu}{\beta}\right)^{-1/\xi}\right], \qquad (1)$$

with $\xi > 0$. The parameters of the distribution are determined by a maximum likelihood estimator. Figure 3

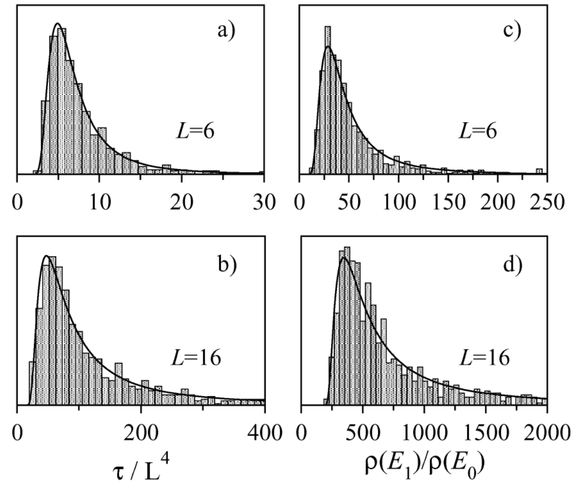

FIG. 2. Normalized distributions of tunneling times (left panels) and the ratio of the number of first excited states to the number of ground states $\rho(E_1)/\rho(E_0)$ (right panels). For both system sizes, $L = 6$ and $L = 16$, 1000 randomly generated 2D $\pm J$ spin glass realizations are sampled. The measured histograms of tunneling times (using local updates and the exact density of states) and $\rho(E_1)/\rho(E_0)$ both follow fat-tailed Fréchet distributions (solid lines).

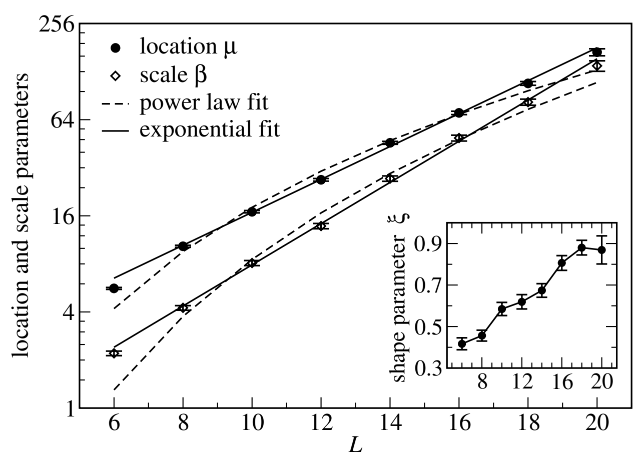

FIG. 3. Scaling of the parameters of Fréchet distribution of the tunneling times [see Eq. (1)] of the 2D $\pm J$ Ising spin glass using perfect flat histogram sampling as a function of system size L. Solid and dashed lines show least squares fits assuming exponential and algebraic (power law) scaling, respectively. The exponential fits of the data with $L \geq 8$ for μ and β give $\chi^2 = 12$ and 7.6, respectively, being significantly better than algebraic fits with $\chi^2 = 130$ and 61, respectively. The inset shows the scaling of the shape parameter.

shows that the location parameter μ specifying the maximum of the distribution and the scale parameter β determining the width of the distribution scale exponentially with linear system size L, $\mu \propto \exp[L/(4.21 \pm 0.04)]$, and $\beta \propto \exp[L/(3.37 \pm 0.05)]$.

The shape parameter ξ, shown in the inset of Fig. 3, determines the power-law decay of the fat tails of the distribution

$$\frac{dH_{\xi;\mu;\beta}}{d\tau} \overset{\tau\to\infty}{\longrightarrow} \tau^{-(1+1/\xi)}. \qquad (2)$$

From this asymptotic behavior one can see that the mth moment of a fat-tailed Fréchet distribution (with $\xi > 0$) is well defined only if $\xi < 1/m$. We find that $\xi > 1/2$ for $L \geq 10$, which implies that the variance ($m = 2$) does not exist and the central limit theorem for mean values does not apply. Any direct estimate of the mean tunneling time—as opposed to the most likely time given by the Fréchet location parameter—thus has an infinite error. This important fact was overlooked in earlier performance studies of flat-histogram methods, and calls into questions power-law scalings as obtained in Refs. [12,13], due to an insufficient statistical analysis.

It also looks plausible, although we cannot go to large enough systems, that ξ increases monotonically with system size and could become larger than 1, in which case even the mean tunneling time ($m = 1$) becomes ill defined. The most likely tunneling time, given by the location parameter μ, would still remain well defined and finite.

Since the tunneling time is to a large extent dominated by the energy landscape at low energies, we study, as a qualitative measure for the complexity of the energy landscape, the distribution of $\rho(E_1)/\rho(E_0)$ where $\rho(E_1)$ is the number of first excited states and $\rho(E_0)$ the number of ground states. Again we find fat-tailed Fréchet distributions, as shown in Figs. 2(c) and 2(d), suggesting that intrinsic properties of the 2D $\pm J$ spin glass account for the distribution of the measured tunneling times. In Fig. 4 we show the measured tunneling times versus the ratio $\rho(E_1)/\rho(E_0)$. A strong correlation over 5 orders of magnitude is found, supporting the argument.

The question arises of why the scaling behaviors of the fully frustrated model and the spin glass are different. Both models have an exponentially large number of ground states and an extensive ground state entropy, but the tunneling time scaling is algebraic for the fully frustrated model and exponential for the spin glass. We believe that the reason is the difference in complexity of the energy landscapes in the two models. While the energy landscape above the large number of ground states of a frustrated model can be simple, the energy landscape of the spin glass is more complex with an extremely large number of local minima: The number of first excited states $\rho(E_1)$ that can be reached from the $\rho(E_0)$ ground states by one single spin flip is at most $L^d\rho(E_0)$. The number of local minima that are not connected to the ground state by a single spin flip can thus be estimated as $\rho(E_1) - L^d\rho(E_0)$. For the 2D ferromagnetic model the ratio $\rho(E_1)/\rho(E_0)$ is exactly L^2, and $\rho(E_1)/\rho(E_0)$ is of order L^2 for the fully frustrated model. In contrast, in the 2D spin glass the ratio $\rho(E_1)/\rho(E_0)$ exceeds L^2 by several orders of magnitude for some realizations, as can be seen from Figs. 2(c), 2(d), and 4.

The behavior of the tunneling time is a bound for the performance of any flat-histogram algorithm. Here we show that it limits the convergence of the flat-histogram algorithm recently proposed by Wang and Landau [4] which has been applied to a large number of problems [8,14–16] and extended to quantum systems [17]. This algorithm iteratively improves approximations to the initially unknown density of states $\rho(E)$ by multiplying a current estimate at each visited level by a factor f that is reduced towards one over time as the algorithm converges.

We have compared tunneling times measured during the course of the Wang-Landau algorithm to those of the perfect flat-histogram method. In Fig. 5 tunneling times using Wang-Landau sampling are shown as a function of the correction factor f for the Ising ferromagnet (main panel) and spin glass (inset). Results for the fully frustrated model (not shown) are qualitatively similar. For homogenous systems we find that during the initial stage of the simulation ($\ln f \gtrsim 10^{-6}$) the tunneling times are shorter than for exact sampling. This indicates that in this case the random walk is biased as to always being driven away from the last region visited [due to the increased $\rho(E)$ there]. For the spin glass, we find samples for which this ease in tunneling only appears after a finite number of iteration steps (as the data for $L = 18$ in the inset of Fig. 5). In both cases, however, the tunneling times eventually converge to exactly the same times as for the

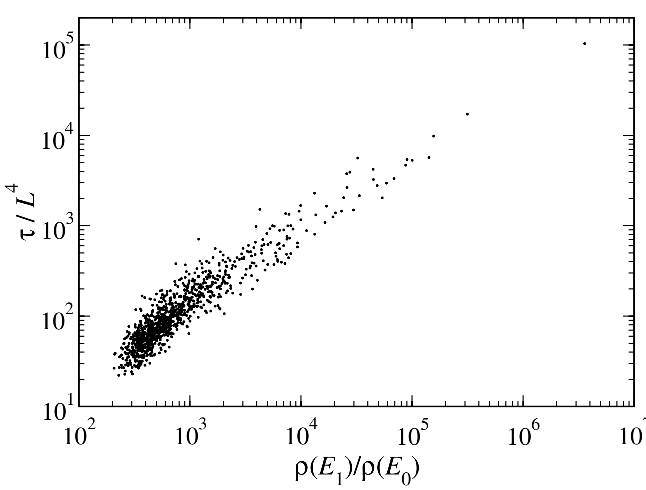

FIG. 4. Correlation between tunneling times and ratios of the number of first excited states to the number of ground states $\rho(E_1)/\rho(E_0)$. Shown are data from 1000 randomly generated 2D $\pm J$ spin glass realizations of fixed system size $L = 16$.

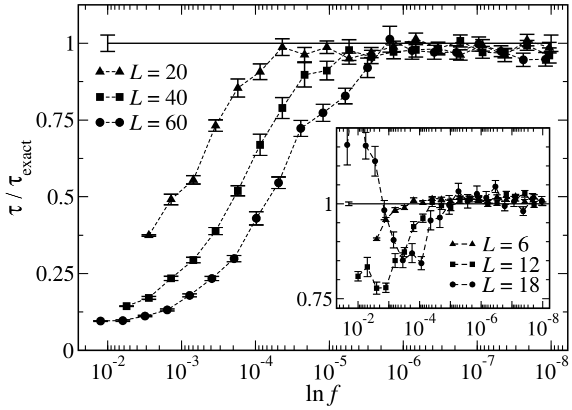

FIG. 5. Evolution of the tunneling times τ during the calculation of the density of states using the Wang-Landau algorithm. Main panel: 2D Ising ferromagnet; inset: typical samples for the 2D $\pm J$ spin glass. For each system size the tunneling times are averaged over 50 (FM), respectively 500 (SG) independent runs, and are given in units of the average tunneling time τ_{exact} measured using the exact density of states.

perfect flat-histogram method. We have checked that the computational time (sweeps) needed to reach convergence increases slower (homogenous systems) or proportional (spin glass) to the tunneling time with system size. This indicates that the Wang-Landau algorithm is eventually limited by the bound set by the perfect flat-histogram method and in this sense is optimal.

Our benchmarks of a perfect flat-histogram method set a lower bound for the tunneling times of all flat histogram methods such as multicanonical [2], tempering [3], or Wang-Landau sampling [4]. This lower bound is also given for related techniques such as transition matrix Monte Carlo [13] or broad histograms [18] if combined with multicanonical sampling. We find that although the Wang-Landau algorithm might initially ease tunneling due to a bias in the approximate density of states, it eventually approaches this lower bound. Whether one uses local or N-fold way updates, the scaling is not the N^2 scaling of a random walk in a system with N energy levels, but slower, namely N^2L^z for both the ferromagnetic ($z = 0.743 \pm 0.007$) and the fully frustrated Ising model ($z = 1.727 \pm 0.004$). The power-law scaling for the frustrated model is very encouraging and convincingly demonstrates that the flat-histogram algorithms are well suited for frustrated models. The observation of a power law with an exponent larger than 2 is not trivially explained in the context of a random walk in energy space, e.g., by mapping the random walk in configuration space to a one-dimensional Markov chain in energy space, and is the subject of further investigations.

The exponential scaling for the $\pm J$ spin glass already for the perfect method exhibits severe limitations of flat-histogram methods, which were overlooked before. Here the distribution of tunneling times follows a fat-tailed Fréchet extremal value distribution, the details of which call earlier performance analyses indicating power-law scalings into question. The origin of this extremal character of the 2D $\pm J$ Ising spin glass remains an interesting open question. Further studies are in progress to investigate this issue as well as three-dimensional classical spin and quantum spin glasses.

We thank F. Alet, N. Kawashima, W. Krauth, M. Pelikan, A. Sandvik, and B. J. Schulz for helpful discussions and acknowledge the support of the Aspen Center for Physics where this project was started. P. D., S. T., and S. W. acknowledge support by the Swiss National Science Foundation, and S. C. and S. S. acknowledge support from the Petroleum Research Fund of the American Chemical Society and by NSF-DMR.

[1] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987); U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[2] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992); Phys. Lett. B **267**, 249 (1991).
[3] K. Hukushima and Y. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996); E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992); A. P. Lyubartsev *et al.*, J. Chem. Phys. **96**, 1776 (1992).
[4] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001).
[5] We have checked that the behavior of τ is insensitive to details of its definition by, e.g., considering tunneling between zero energy and the ground state, or the first or second excited state instead.
[6] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1995).
[7] L. K. Saul and M. Kardar, Nucl. Phys. **B432**, 641 (1994).
[8] B. J. Schulz *et al.*, Int. J. Mod. Phys. C **13**, 477 (2002).
[9] A. B. Bortz *et al.*, J. Comput. Phys. **17**, 10 (1975).
[10] P. Embrechts, C. Klüpelberg, and T. Mikosch, *Modelling Extremal Events* (Springer, New York, 1997).
[11] R. A. Fisher and L. H. C. Tippett, Proc. Cambridge Philos. Soc. **24**, 180 (1928).
[12] B. A. Berg and T. Celik, Int. J. Mod. Phys. C **3**, 1251 (1992); Z. F. Zhan *et al.*, Physica (Amsterdam) **285A**, 239 (2000).
[13] J. S. Wang *et al.*, Phys. Rev. Lett. **82**, 476 (1999); J. S. Wang and R. H. Swendsen, J. Stat. Phys. **106**, 245 (2002).
[14] C. Yamaguchi and Y. Okabe, J. Phys. A **34**, 8781 (2001); F. Calvo, Mol. Phys. **100**, 3421 (2002); Q. Yan *et al.*, J. Chem. Phys. **116**, 8745 (2002); M. S. Shell *et al.*, Phys. Rev. E **66**, 056703 (2002); T. S. Jain and J. J. de Pablo, J. Chem. Phys. **116**, 7238 (2002); N. Rathore and J. J. de Pablo, J. Chem. Phys. **116**, 7225 (2002); T. J. H. Vlugt, Mol. Phys. **100**, 2763 (2002); E. B. Kim *et al.*, J. Chem. Phys. **117**, 7781 (2002).
[15] C. Yamaguchi and N. Kawashima, Phys. Rev. E **65**, 056710 (2002).
[16] Q. Yan and J. J. de Pablo, Phys. Rev. Lett. **90**, 035701 (2003); M. S. Shell *et al.*, J. Chem. Phys. **119**, 9406 (2003).
[17] M. Troyer *et al.*, Phys. Rev. Lett. **90**, 120201 (2003).
[18] P. M. C. de Oliveira *et al.*, Braz. J. Phys. **26**, 677 (1996).