Dynamics of the Wang-Landau algorithm and complexity of rare events for the three-dimensional bimodal Ising spin glass

Simon Alder[1], Simon Trebst[1,2], Alexander K. Hartmann[3], Matthias Troyer[1,2]
[(1)] *Theoretische Physik, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland*
[(2)] *Computational Laboratory, Eidgenössische Technische Hochschule Zürich, CH-8092 Zürich, Switzerland and*
[3]*Institut für Theoretische Physik, Universität Göttingen, 37077 Göttingen, Germany*
(Dated: May 18, 2004)

We investigate the performance of flat-histogram methods based on a multicanonical ensemble and the Wang-Landau algorithm for the three-dimensional $\pm J$ spin glass by measuring round-trip times in the energy range between the zero-temperature ground state and the state of highest energy. Strong sample-to-sample variations are found for fixed system size and the distribution of round-trip times follows a fat-tailed Fréchet extremal value distribution. Rare events in the fat tails of these distributions corresponding to extremely slowly equilibrating spin glass realizations dominate the calculations of statistical averages. While the *typical* round-trip time scales exponential as expected for this NP-hard problem, we find that the *average* round-trip time is no longer well-defined for systems with $N \geq 8^3$ spins. We relate the round-trip times for multicanonical sampling to intrinsic properties of the energy landscape and compare with the numerical effort needed by the genetic Cluster-Exact Approximation to calculate the exact ground state energies. For systems with $N \geq 8^3$ spins the simulation of these rare events becomes increasingly hard. For $N \geq 14^3$ there are samples where the Wang-Landau algorithm fails to find the true ground state within reasonable simulation times. We expect similar behavior for other algorithms based on multicanonical sampling.

PACS numbers: 02.70.Rr, 75.10.Hk, 64.60.Cn

I. INTRODUCTION

The three-dimensional $\pm J$ Ising spin glass has been extensively studied[1] as a prototype system which exhibits a finite temperature second order phase transition to a slowly equilibrating glassy phase.[2,3] The simulation of such a system with conventional Monte Carlo methods is slowed down by long relaxation times in the spin-glass phase. This problem has been addressed by a number of algorithmic developments such as the multicanonical method,[4] simulated and parallel tempering,[5] broad histograms[6] and transition matrix Monte Carlo.[7] In order to speed up equilibration most of these methods aim at broadening the energy range sampled within the Monte Carlo (MC) simulations from the sharply peaked distribution of canonical sampling at a fixed temperature.

Recently, Wang and Landau introduced a new algorithm which systematically calculates an estimate of the density of states and iteratively converges to sampling a flat histogram in energy.[8] The Wang-Landau (WL) algorithm simulates a biased random walk in configuration space. The bias depends only on the total energy of a configuration and is defined by a statistical ensemble with weights $w(E)$. For this ensemble the transition probabilities are given by the Metropolis scheme

$$p(E_1 \to E_2) = \min\left(\frac{w(E_2)}{w(E_1)}, 1\right) . \tag{1}$$

The equilibrium distribution of the energy in this ensemble is $n_w(E) \propto w(E)g(E)$ where $g(E)$ is the density of states. By setting the weights $w(E) \propto 1/g(E)$ the WL algorithm aims at sampling a flat histogram in energy. The crucial feature of the WL algorithm however is that the simulated ensemble is *dynamically* modified during the course of the simulation: after every spin update the current estimate of the density of states is multiplied by a modification factor f and the ensemble weights are analogously updated. The modification factor is iteratively reduced to 1 whenever the sampled energy histogram is close to the expected equilibrium distribution, that is when the histogram is "flat" within a given range. The dynamic modification of the ensemble allows to push the random walker towards the low entropy states in the initial stages of the algorithm while ensuring that it converges to a flat-histogram/multicanonical ensemble in the final stages of the computation.

In this paper, we determine the performance of the Wang-Landau algorithm for the three-dimensional $\pm J$ Ising spin glass for both stages of the algorithm. First, we study the *dynamic behavior* of the algorithm in the initial stages by considering its ability to find the ground-state energy of a number of three-dimensional spin glass samples which is a well known NP-hard problem.[9] We compare the obtained ground state energies to exact results calculated with the genetic Cluster-Exact approximation (CEA)[14,15]. While the WL algorithm reproduces the exact ground-state energy for small systems, we find that for moderately large systems ($N \geq 14^3$) the WL algorithm does not find the exact ground-state energy for few spin-glass samples. Even when restricting the simulated energy bin around the known ground-state energy the algorithm does not find the lowest energy state within a reasonable number of sweeps ($N_{\text{sweeps}} \approx 10^7$). The genetic CEA gives a superior performance to find ground state energies, but does not give the full thermodynamic information. Second, we investigate the *asymptotic be-*

havior of the WL algorithm by measuring round-trip times in energy for the converged ensemble. The round-trip time gives a direct estimate of the equilibration time for the multicanonical ensemble. The asymptotic scaling of the Wang-Landau algorithm therefore also reflects the performance of other flat-histogram methods based on multicanonical sampling, such as the multicanonical method,[4] broad histograms[6] or transition matrix Monte Carlo.[7] We find large sample-to-sample variations of the round-trip times which can be described by fat-tailed Fréchet extremal-value distributions. We discuss important implications for statistical sample averaging which are caused by the rare events in the fat tails of these distributions. The intrinsic character of the observed extremal-value distributions is demonstrated. Finally it is shown that these distributions scale exponentially with the linear system size. In comparison, we find that the computational effort of the genetic CEA is also correlated to the density of states, but is less sensitive to the low-energy landscape than the WL algorithm. Finally, we discuss these limitations of multicanonical sampling by measuring the local diffusivity of the random walker in energy. We find a pronounced minimum of the diffusivity near the ground-state energy which is symptomatic for the entropic barrier which slows down the equilibration of the random walker.

II. DYNAMIC PERFORMANCE

To study the dynamic performance of the Wang-Landau algorithm we have tested its ability to find the ground-state energy of a number of spin glass samples where the exact ground state is known. Finding the ground-state energy of the $\pm J$ spin-glass is an extensively studied problem, both by physicists and computer scientists. For the two-dimensional $\pm J$ spin glass the *full* density of states can be calculated with polynomial effort.[10,11] However, for the three-dimensional $\pm J$ spin glass as well as the two-dimensional $\pm J$ spin glass with an external field, even the problem of finding the ground-state energy has been shown to be NP-hard.[9] The direct calculation of the ground-state energy of spin glass samples using sophisticated exact branch-and-cut algorithms[12,13] is therefore limited to rather small systems. In this study, we applied a combination of a genetic algorithm with Cluster-Exact Approximation (CEA) to determine ground state energies.[14,15] The algorithm, although based on heuristics, is able to find true ground states of samples up to size $N = 16^3 = 4096$ in reasonable, although exponentially growing time[15]

We use the results obtained from this genetic + CEA approach to check the accuracy of the ground-state energies found with the WL algorithm for samples up to size $N = 14^3$. For system size, $N \leq 6^3$, the density of states of each sample was calculated without energy binning by 50 independent runs. For all runs we find that the WL algorithm gives the exact ground-state energy. For $N = 8^3$, for some few samples (17 out of 1000), the exact ground-state energy was found only after running extensive runs after the comparison with the results from the heuristic approach revealed that the true ground state energy has not been found.

For larger systems with $N = 10^3$ and more spins we have restricted the WL simulations to the smallest allowed energy range around the ground-state energy calculated by the genetic CEA. In order to keep ergodicity the energy bin has to be larger than

$$\Delta E/J > 4L^{d-1} , \qquad (2)$$

to assure that two domain walls can be inserted which thus enables moves which subsequently flip all spins and ergodicity within the energy bin is given.

Each sample was simulated by six independent runs. For all samples with $N = 10^3$ spins the WL algorithm found true ground-state energies. However, for three samples not all runs gave the true ground-state energy, but sampled energies down to the first excited state only (within $(2.8 \pm 0.2) \times 10^7$ MC sweeps). For the samples with $N = 12^3$ spins we find similar results. For all samples there is at least one run which finds the true ground-state energy within some $(2.9 \pm 0.1) \times 10^7$ MC sweeps. However, for 9 out of 10 samples the WL algorithm does not converge to sampling a flat histogram in the full energy range. For the samples with $N = 14^3$ spins the WL algorithms finds the true ground-state energy for 4 samples only. For 6 out of 10 samples the algorithm did not sample the exact ground-state energy once within $(3.0 \pm 0.1) \times 10^7$ MC sweeps for all independent runs. Furthermore, there is one sample where even the first excited state is not found within the given number of sweeps. For all samples the simulated ensemble does not converge towards the multicanonical ensemble sampling a flat energy histogram and the WL algorithm gets stuck.

III. ASYMPTOTIC PERFORMANCE

We now turn to the asymptotic behavior of the WL algorithm and flat-histogram methods in general which we determine by measuring the round-trip times in energy of the simulated random walker for the converged flat-histogram ensemble. Here the round-trip time corresponds to the number of single-spin flips needed to get from a configuration with the ground-state energy to a configuration with highest energy (the anti ground state). The round-trip time thus gives an estimate of the equilibration time for the flat-histogram ensemble. For the Ising model the number of energy levels scales linearly with system size N. While the round-trip time of an unbiased random walker scales like $\tau \sim N^2$, it was recently shown that for various two-dimensional Ising models the growth with the number of spins is significantly stronger for the biased flat-histogram random walker.[16] For the ferromagnetic and fully frustrated Ising models polynomial scaling, $\tau \sim N^{2.4}$ and $\tau \sim N^{2.9}$, was found,

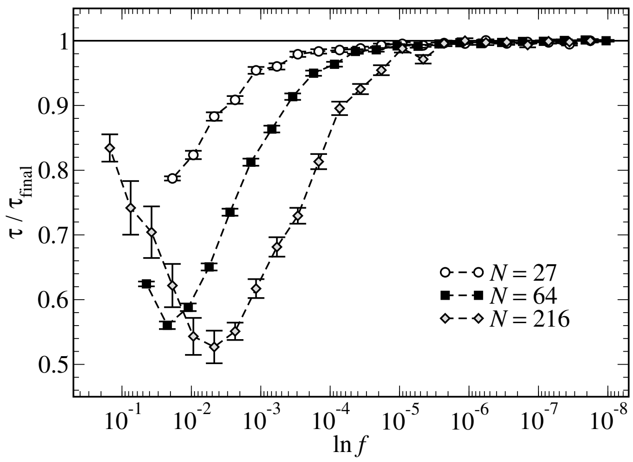

FIG. 1: Convergence of the round-trip time τ versus the modification factor f in the Wang-Landau algorithm. Shown are the results for three randomly generated three-dimensional $\pm J$ spin glass samples with $N = 3^3 = 27, 4^3, 6^3 = 216$ spins. The measurements are averaged over 500 independent runs for $N = 27$, 64 and over 30 runs for $N = 216$, respectively.

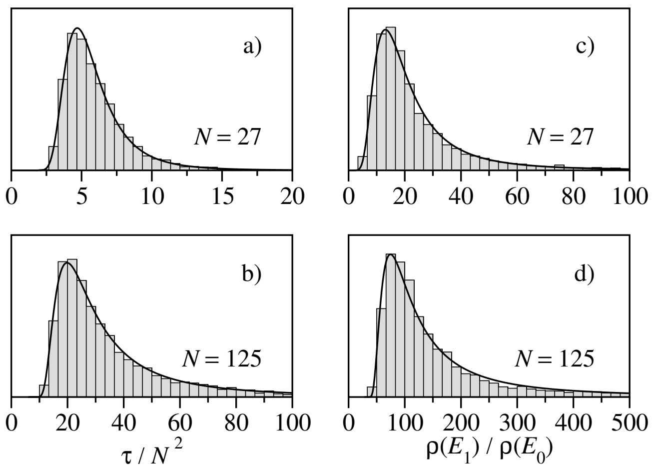

FIG. 2: Left panels: Distribution of round-trip times τ for 5000 randomly generated spin glass samples of size $N = 3^3 = 27$ and $N = 5^3 = 125$ respectively. Right panels: Distribution of the ratio of the number of first excited states to the number of ground states $g(E_1)/g(E_0)$ for the same system sizes. In all panels the solid lines indicate fits to fat-tailed Fréchet extremal-value distributions.

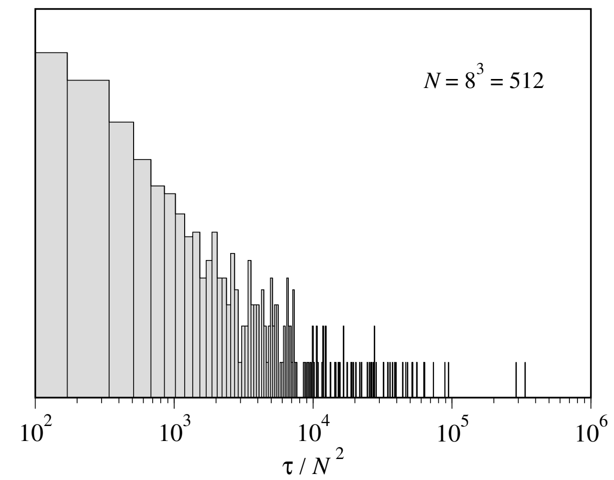

FIG. 3: Log-log plot of the distribution of round-trip times τ for the three-dimensional $\pm J$ spin glass with $N = 8^3 = 512$ spins. The round-trip times τ were measured for the converged flat-histogram ensemble in the final step of the Wang-Landau algorithm. Data from 1000 randomly generated spin glass samples are shown.

and exponential growth for the two-dimensional $\pm J$ spin glass.[16]

For a given sample we find that the round-trip time measured during the iterations of the WL algorithm converges as the simulated ensemble approaches the flat-histogram ensemble. The convergence of round-trip times is illustrated for three randomly generated spin glass samples in Fig. 1. Since correct convergence to the round-trip times of the *exact* flat-histogram ensemble was shown for the two-dimensional $\pm J$ Ising spin glass[16] we assume correct convergence for the 3D case as well and thereby justify that our results for the asymptotic round-trip times hold for *any* flat-histogram method.

A. Sample-to-sample variation

To study the sample dependence of the round-trip times we have analyzed 5000 randomly generated spin glass samples for $N = 3^3$, 4^3, 5^3, 6^3 and 1000 samples for $N = 8^3 = 512$, respectively. To assure convergence of the measured round-trip times we restrict the measurement to the final step in the Wang-Landau algorithm. We find strong sample-to-sample variations over several orders of magnitude for fixed system size N which is shown in the left panels of Fig. 2 ($N = 3^3$ and $N = 5^3$). For the spin glass of size $N = 8^3 = 512$ the full distribution of round-trip times is shown in Fig. 3. The distribution covers some 4 orders of magnitude and contains spin glass samples which were simulated between some minutes and about a month on a 500 MHz Pentium III CPU.

To quantitatively analyze the extremal events in the tails of the distributions, we use extremal-value theory.[17] The central limit theorem for extremal-value states that the extrema of random subsets of any distribution follow a generalized extremal-value distribution.[18] This distribution takes one of three characteristic forms: Fréchet (algebraic decay, fat-tailed), Weibull (exponential decay) or Gumbel (faster than exponential decay, thin-tailed). Here we find that *all measured round-trip times* seem to follow a fat-tailed Fréchet extremal-value distribution, similar to the two-dimensional $\pm J$ spin glass.[16] This implies that the central limit theorem applies even to the smallest possible subset – a *single* round-trip time. As a consequence, every single three-dimensional $\pm J$ spin glass sample constitutes an extremal event. The inte-

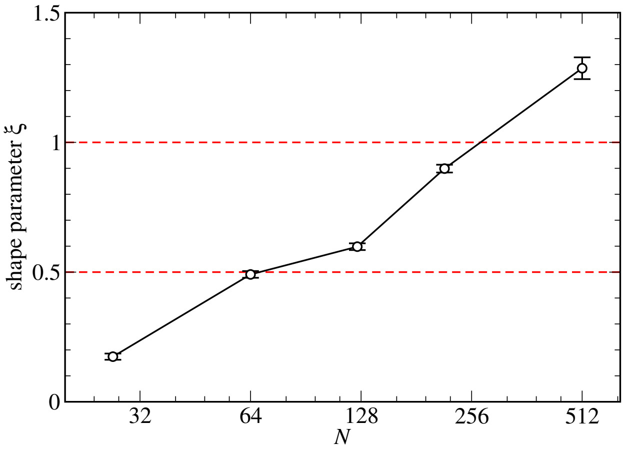

FIG. 4: Scaling of the shape parameter ξ of the fitted Fréchet distributions versus system size. The m-th moment of a fat-tailed Fréchet distribution is well-defined only if $\xi < 1/m$. The dashed lines indicate where variance and mean of the respective distributions become ill-defined.

grated form of the Fréchet distribution is given by

$$H_{\xi,\mu,\beta}(\tau) = \exp\left(-\left(1+\xi\frac{\tau-\mu}{\beta}\right)^{-1/\xi}\right) . \qquad (3)$$

The parameter μ indicates the location of the distribution, that is the most probable round-trip time, and the parameter β defines the scale of the distribution, e.g. the height of the peak. The shape parameter ξ, which is positive for fat-tailed distributions, describes the decay of the tail of the distribution. We have determined the three parameters μ, β and ξ of the fitted Fréchet distributions with a maximum likelihood estimator. The resulting fits are shown as solid lines in Fig. 2.

We now turn to the scaling of the shape parameter with system size N which is shown in the inset of Fig. 4. With increasing system size the shape parameter monotonically increases. This strongly affects the fat tails of the distributions which in the limit $\tau \to \infty$ exhibit a power-law decay of the form

$$\frac{\mathrm{d}}{\mathrm{d}\tau} H_{\xi,\mu,\beta}(\tau) \xrightarrow{\tau\to\infty} \tau^{-(1+1/\xi)} . \qquad (4)$$

From this asymptotic behavior we can see that the m-th moment of a fat-tailed Fréchet distribution is well-defined only if $\xi < 1/m$. The scaling in the inset of Fig. 4 suggests that for $N > 4^3 = 64$ ($N \geq 8^3 = 512$) the shape parameter becomes larger than 0.5 (1) and thus the variance (mean) of the distribution is no longer well-defined.

To illustrate this unusual behavior we can calculate running moments of the distribution, e.g. by only considering subsets of the first n round-trip times of all measured round trip times $\{\tau\}$ when calculating the moments of the distribution. The running mean of round-trip

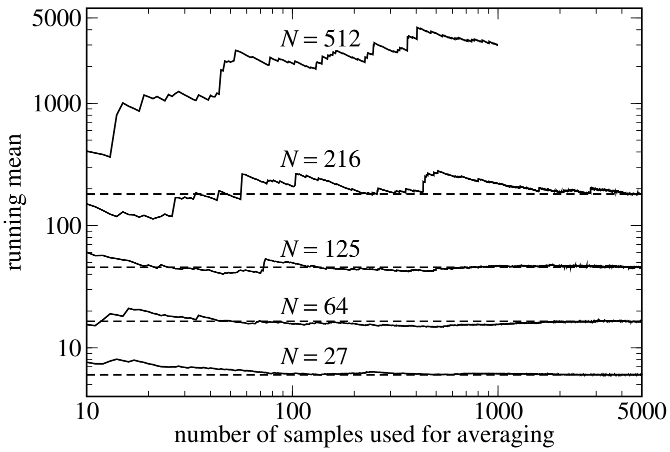

FIG. 5: Log-log plot of the running mean of round-trip times defined in Eq. (5) versus the number of samples used for averaging. Results for various system sizes are shown. For $N \leq 6^3 = 216$ the mean is well-defined and the running mean seems to converge (dashed line). For $N = 8^3 = 512$ the mean becomes ill-defined and the running mean diverges[20].

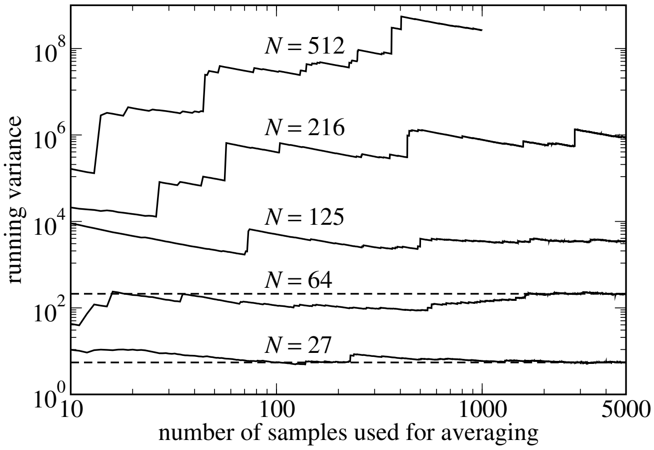

FIG. 6: Log-log plot of the running variance of the distribution of round-trip times defined in Eq. (6). Results for various system sizes are shown. For $N \leq 4^3 = 64$ the variance is well-defined and the running variance seems to converge (dashed line). For $N \geq 5^3 = 125$ the variance becomes ill-defined and the running variance diverges[20].

times is then defined by

$$\mathrm{Mean}_{\{\tau\}}(n) = \frac{1}{n}\sum_{i=1}^{n} \tau_i , \qquad (5)$$

and the running variance by

$$\mathrm{Var}_{\{\tau\}}(n) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\tau_i - \mathrm{Mean}_{\{\tau\}}(n)\right)^2 . \qquad (6)$$

The running mean and variance are calculated for a fixed random order of the set of round-trip times $\{\tau\} \equiv \{\tau_1, \tau_2, \ldots, \tau_{5000}\}$. Fig. 5 shows the running mean for

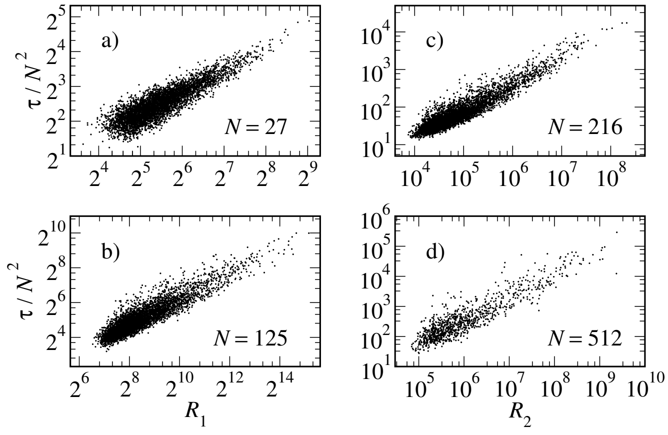

FIG. 7: Correlation of the round-trip time τ and ratios in the density of states for various system sizes. In the left panels the ratio $R_1 = g(E_1)/g(E_0)$ is shown. In the rights panels transitions to higher excited states are also included, $R_2 = g(E_1)/g(E_0) + g(E_2)/g(E_1) + g(E_2)/g(E_0)$. Shown are data from 5000 randomly generated spin glass samples for $N = 27$, 125, 216 and 1000 for $N = 8^3 = 512$, respectively.

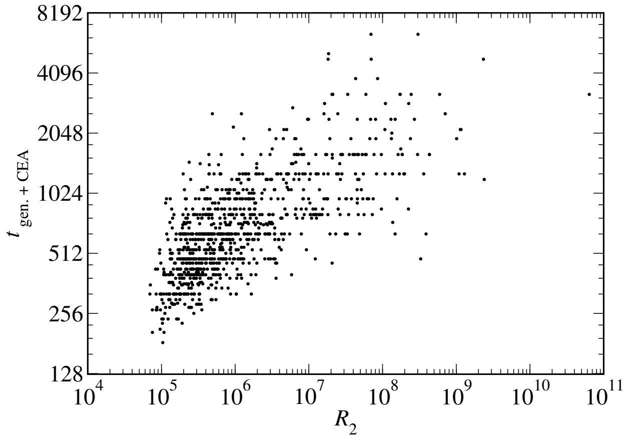

FIG. 8: Correlation of the computational effort, $t_{\text{gen.+CEA}}$, of the genetic CEA (see text) and the ratio in the density of states for $N = 8^3$ (1000 samples). Here the ratio $R_2 = g(E_1)/g(E_0) + g(E_2)/g(E_1) + g(E_2)/g(E_0)$ is shown. The correlation to $R_1 = g(E_1)/g(E_0)$ looks similar.

various system sizes. For small system sizes, $N \leq 64$, both mean and variance are well defined and the running mean and its error converges. For larger systems the variance becomes ill-defined as the shape parameter ξ turns larger than 0.5. Irregular "jumps" indicating rare events occur in the calculation of the running mean. For systems with $N \leq 216$ spins the running mean still converges, but the error of the running mean does not reduce even for large sample sets. For systems with more than $N = 8^3 = 512$ spins the shape parameter becomes larger than 1 and the mean thus ill-defined. This divergence of the mean round-trip time also becomes apparent in the calculation of the running mean where the irregular jumps occur so frequently that the mean round-trip time no longer converges[20].

This behavior becomes even more evident for the variance, the second moment of the distribution, which according to the scaling of the shape parameter illustrated in the inset of Fig. 4 is no longer well-defined for systems larger than $N > 4^3 = 64$. As illustrated in Fig. 6 the running variance diverges for larger systems[20]. Again, frequent "jumps" in the running variance indicate the occurrence of rare events in the fat tails of the distribution which dominate the calculation of the running variance.

B. Intrinsic correlations for the WL algorithm

In order to test whether the occurrence of the Fréchet extremal-value distributions reflect an intrinsic property of the energy landscape three-dimensional spin glass we have analyzed the calculated density of states near the ground-state energy. The energy landscape near the ground state dominates to a large extent the measured round-trip times which is further discussed in the context of diffusivity measurements in section IV. Here we study ratios in the density of states, such as the number of first excited states to the number of ground states, $g(E_1)/g(E_0)$. From the ground state a single-spin-flip update can connect at most $Ng(E_0)$ states of energy E_1 and thus the ratio $g(E_1)/g(E_0)$ gives a qualitative measure of the number of local minima, i.e. which are not reachable from a ground state via a single spin flip. In the right panels of Fig. 2 the distribution of this ratio $g(E_1)/g(E_0)$ is shown for 5000 spin glass samples of size $N = 3^3 = 27$ and $N = 5^3 = 125$ respectively. Again we find that these distributions follow fat-tailed Fréchet extremal-value distributions. For these system sizes a strong correlation to the distribution of round-trip times is found which spans over several orders of magnitude as demonstrated in the left panels of Fig. 7. For larger systems these correlations become less pronounced. However, if we consider additional transitions, such as the transition from the second to first excited state, $E_2 \to E_1$ and $E_2 \to E_0$, and calculate the sum of the respective ratios in the density of states we can recover a correlation over several orders of magnitude as it is shown for systems with $N = 6^3 = 216$ and $N = 8^3 = 512$ spins in the right panels of Fig. 7.

C. Intrinsic correlations for the heuristic approach

The strong correlation between intrinsic features of the energy landscape and the measured round-trip times for flat-histogram sampling naturally leads to the question whether the computational effort of other algorithms also comply with these intrinsic features. To this end, we compare the computational effort of the alternative heuristic approach with the density of states and the round-trip times measured in the WL algorithm for 1000 samples with $N = 8^3 = 512$ spins.

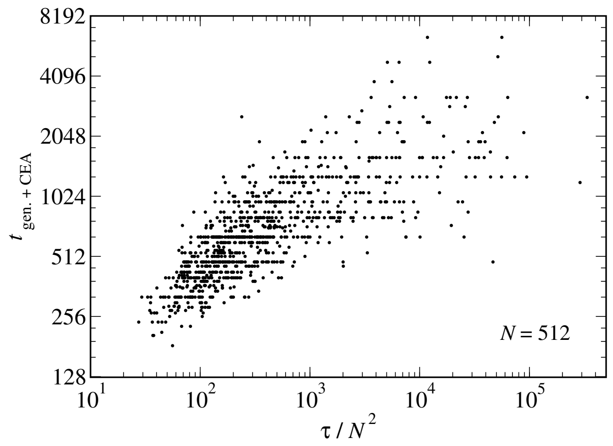

FIG. 9: Correlation of the computational effort, $t_{\text{gen.+CEA}}$, of the genetic CEA (see text) and the round-trip time τ/N^2 measured for multicanonical sampling in the final step of the WL algorithm for a system with $N = 8^3$ spins (1000 samples).

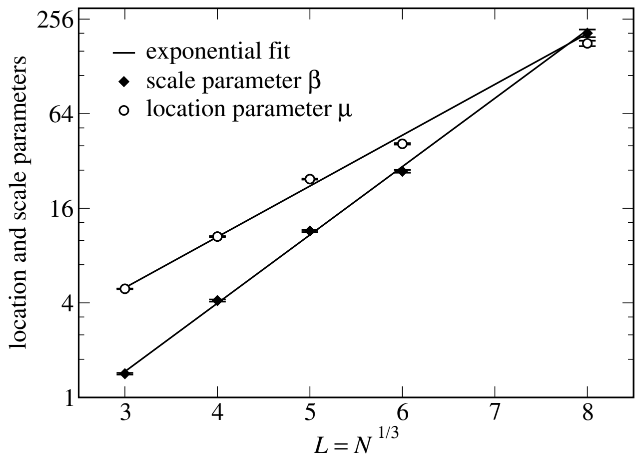

FIG. 10: Scaling of the location (μ) and scale (β) parameter of the Fréchet distribution versus the linear system size $L = N^{1/3}$. The parameters were fitted by a maximum-likelihood estimator. The solid lines indicate exponential fits of the respective data points.

We define the computational effort of the genetic CEA as the running time of the algorithm which strongly depends on the parametrization of the underlying genetic algorithm, namely

- M_{i} = initial size of population, i.e. how many configurations are optimized in parallel
- n_{o} = average number of offspring per configuration, i.e. how many iterations of the algorithm are run
- n_{min} = number of CEA minimization steps per configuration and per iteration

The running time then depends linearly on the product $M_{\text{i}} n_{\text{o}} n_{\text{min}}$. If the algorithm is run independently several times, one finds that not all runs result in ground states and we denote the fraction of runs which find the true ground state as f_{GS}. In general one can observe that this fraction increases with the running time.

We now define the computational effort for a given sample and parameter set $\{M_{\text{i}}, n_{\text{o}}, n_{\text{min}}\}$ as $t(M_{\text{i}}, n_{\text{o}}, n_{\text{min}}) = M_{\text{i}} n_{\text{o}} n_{\text{min}} / f_{\text{GS}}$ (i.e. $t = \infty$, if no ground state is found). For "simple" samples, many ground states are found for most parameter combinations, while "hard" samples need large sizes of populations and/or many iterations and/or many minimization steps, thereby increasing the computational effort. For a given spin glass sample we perform multiple simulations with different combinations of parameters, and define the overall computational effort of a sample as the minimum over all parameter combinations considered (using a fixed value $n_{\text{R}} = 20$):

$$t_{\text{gen.+CEA}} = \min_{(M_{\text{i}}, n_{\text{o}}, n_{\text{min}})} \left(M_{\text{i}} n_{\text{o}} n_{\text{min}} / f_{\text{GS}} \right) . \quad (7)$$

In Fig. 8 the correlation of the computational effort of the genetic CEA algorithm is shown versus the ratio R_2 of the density of states as defined above for 1000 samples with $N = 8^3 = 512$ spins. There is only a weak correlation indicating that the heuristic algorithm is less sensitive to the energy-landscape close to the ground state than the WL algorithm. A direct comparison between the genetic CEA and the WL algorithm is shown in Fig. 9. We find that the computational effort for the heuristic algorithm spreads over only 2 orders of magnitude in comparison to some 4 orders of magnitude for the WL algorithm. The correlation between the two algorithms is weak, and less pronounced for those samples which are especially hard to equilibrate using multicanonical sampling.

D. Scaling of typical round-trip times

Although the mean round-trip times are no longer well-defined for larger systems, the location and scale parameter of the Fréchet distribution stay well-defined and can be used to further characterize the scaling of these fat-tailed distributions with system size. Fig. 10 shows that these parameters exponentially diverge with linear system size like

$$\begin{aligned} \mu &\propto \exp\left(L/(1.34 \pm 0.03)\right) , \\ \beta &\propto \exp\left(L/(1.00 \pm 0.03)\right) . \end{aligned} \quad (8)$$

The study of the equilibrium behavior of larger system size is thus not only limited by the occurrence of rare events, but also by the exponential growth of the round-trip times in the "bulk" of the distribution which renders a comprehensive study of the sample-to-sample variations impossible.

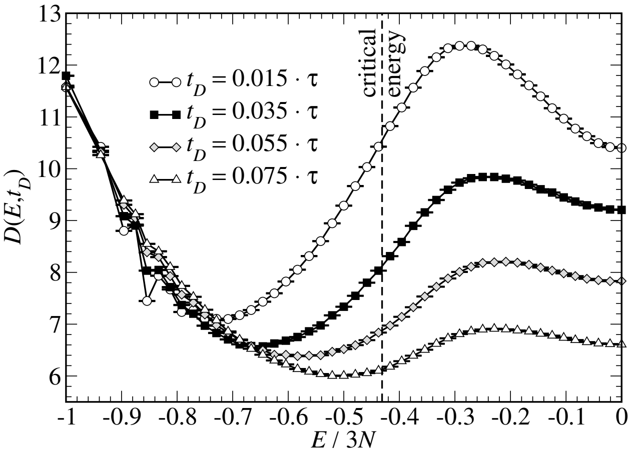

FIG. 11: Local diffusivity $D(E, t_D)$ in energy of a flat-histogram random walker for the three-dimensional Ising ferromagnet. Data for various diffusion times are shown for a system with $N = 4^3 = 64$ spins.

IV. DIFFUSIVITY MEASUREMENTS

We have seen that both the asymptotic and the dynamic performance are limited by the rough energy landscape close to the ground-state. To study these limitations in more detail we measure the local diffusivity of the random walker in energy. Recently, it was shown that the simulated statistical ensemble can be optimized by a feedback loop which reweights the ensemble based on preceeding measurements of the local diffusivity.[22] Although we do not follow up on this idea in the present study, we can measure the diffusivity to reveal the "bottlenecks" of the biased random walk in energy as local minima in the diffusivity. Here we use a time-dependent definition of the diffusivity, $D(E, t_D)$, in energy space

$$D(E, t_D) = \left\langle (E(t) - E(t + t_D))^2 \right\rangle / t_D \ , \qquad (9)$$

where t_D is the diffusion time. The relevant time scale for the diffusion time is set by the round-trip time in energy.

As a first example we study the local diffusivity of the three-dimensional ferromagnetic Ising model which undergoes a second-order phase transition from a magnetically ordered to a disordered phase at finite energy $E_c/3N \cong -0.43$.[23] The measured diffusivity $D(E, t_D)$ of the flat-histogram random walker is shown in Fig. 11. We find that the diffusivity is not constant as it is expected for an unbiased random walk, but there is a broad minimum below the critical energy. In this energy region the random walker is slowed down due to the slow dynamics of domain walls which separate droplets of magnetically ordered phases.

Next we turn to the three-dimensional spin glass. Measurements of the diffusivity of the flat-histogram random walker for two randomly generated samples are shown in Fig. 12. For both samples we find a minimum of the diffusivity at the ground-state energy. We further note

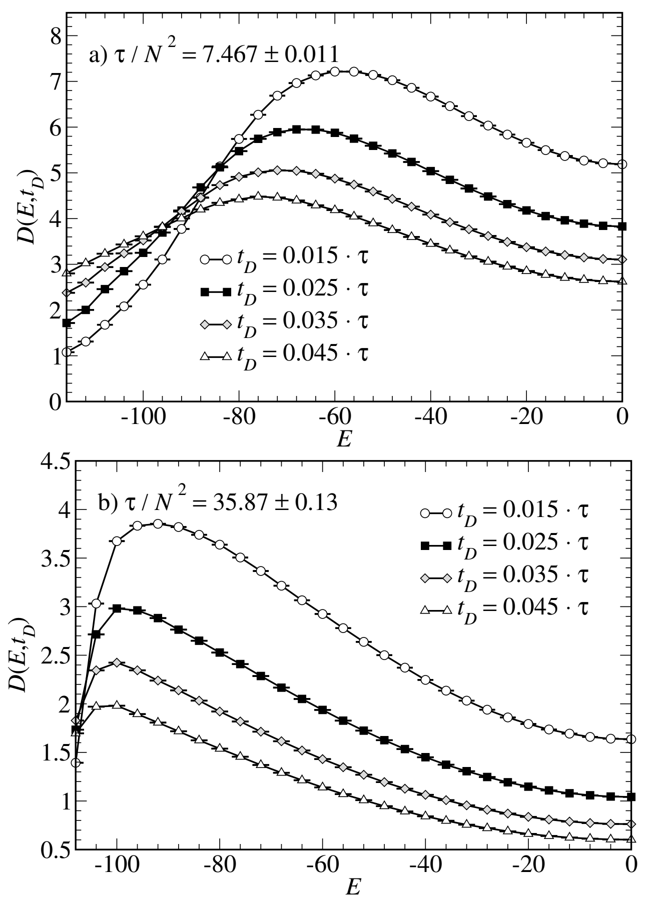

FIG. 12: Local diffusivity $D(E, t_D)$ in energy of a flat-histogram random walker for two three-dimensional $\pm J$ Ising spin glass samples. Data for various diffusion times are shown for systems with $N = 4^3 = 64$ spins.

that with increasing round-trip times the minimum in the diffusivity becomes more pronounced. These diffusivity measurements further underline that the bottleneck of the flat-histogram random walker are at the ground-state energy. The suppressed diffusivity in this energy region gives rise to an entropic barrier which results in long round-trip times and aggravates equilibration of the random walker in the low energy phase. For larger systems the suppressed local diffusivity of the flat-histogram random walker renders a comprehensive study of the glassy phase impossible as it becomes computationally too expensive to equilibrate a large number of samples needed to calculate statistical averages. The recent development of a systematic means to optimize the simulated statistical ensemble[22] holds promise to overcome some of these limitations.

V. CONCLUSIONS

To summarize, we have studied the performance of the Wang-Landau algorithm for the three-dimensional

$\pm J$ Ising spin glass. The asymptotic performance — which corresponds to the performance of any flat-histogram method sampling a multicanonical ensemble such as the multicanonical method,[4] simulated and parallel tempering,[5] broad histograms[6] and transition matrix Monte Carlo[7] — is found to be dominated by strong sample-to-sample variations. The measured round-trip times follow fat-tailed Fréchet extremal-value distributions. The typical round-trip times in the bulk of these distributions obey exponential scaling, as expected for this NP-hard problem. The fat tails of the distributions dominate the calculation of statistical averages. A careful statistical analysis is needed which goes beyond the calculation of the moments of a finite sample distribution as it was done in previous studies. The intrinsic character of the Fréchet extremal distributions becomes evident in strong correlations over several orders of magnitude between the round-trip time and the behavior of the density of states near ground-state energy. The origin of the extremal character of every single spin glass sample remains an open question which deservers further investigations.

Our measurements of the dynamic performance and comparison with ground states obtained by genetic CEA showed that for samples with up to $N = 8^3$ spins the Wang-Landau algorithm always finds the correct ground-state energies. For sizes 10^3, 12^3, one can find true ground states, if one restricts the random walk to a small energy bin around the exact ground-state energy calculated by the heuristic CEA approach. For samples with $N = 14^3$ spins we identified samples for which the Wang-Landau algorithm does not find the true ground-state energy within reasonable simulation times ($\approx 10^7$ MC sweeps) and does not converge towards the multicanonical ensemble.

The entropic barrier near the zero-temperature ground state can be revealed as a pronounced minimum in the local diffusivity of the flat-histogram random walker in energy. The recent development of optimized statistical ensembles based on feedback of the local diffusivity holds promise to overcome the observed slow down of the flat-histogram random walker and thereby enhancing equilibration in these systems.[22]

VI. ACKNOWLEDGMENTS

We acknowledge discussions with H. Gould, D. A. Huse, J. Machta, and S. Wessel. SA and ST acknowledge support by the Swiss National Science Foundation. AKH was supported by the *VolkswagenStiftung* (Germany) within the program "Nachwuchsgruppen an Universitäten".

1 K. Binder and A.P. Young, Rev. Mod. Phys. **58**, 801 (1986); M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond*, World Scientific, Singapore (1987); K.H. Fisher and J.A. Hertz, *Spin Glasses*, (Cambridge University Press, Cambridge 1991); A.P. Young (ed.), *Spin glasses and random fields*, (World Scientific, Singapore 1998).

2 H. G. Ballesteros, A. Cruz, L. A. Fernández, V. Martín-Mayor, J. Pech, J. J. Ruiz-Lorenzo, A. Tarancón, P. Téllez, C. L. Ullod, and C. Ungil, Phys. Rev. B **62**, 14237 (2000).

3 M. Palassini and S. Caracciolo, Phys. Rev. Lett. **82**, 5128 (1999).

4 B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992); Phys. Lett. B. **267**, 249 (1991).

5 K. Hukushima and K. Nemoto, J. Phys. Soc. Japan **65**, 1604 (1996); 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).

6 P. M. C. de Oliveira, T. J. P. Penna, and H. J. Herrmann, Braz. J. Phys. **26**, 677 (1996).

7 J. S. Wang, T. K. Tay, and R. H. Swendsen, Phys. Rev. Lett. **82**, 476 (1999); J. S. Wang and R. H. Swendsen, J. Stat. Phys. **106**, 245 (2001).

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

9 F. Barahona, J. Phys. A **15**, 3241 (1982).

10 I. Bieche, R. Maynard, R. Rammal, and J. P. Uhry, J. Phys. A **13**, 2553 (1980).

11 L. K. Saul and M. Kardar, Nuclear Physics B **432**, 641 (1994).

12 C. De Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, and G. Rinaldi, *J. Stat. Phys.* **80**, 487 (1995).

13 C. De Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, and G. Rinaldi, *J. Stat. Phys.* **84**, 1363 (1996).

14 A. K. Hartmann, Physica A **224**, 480 (1996); Phys. Rev. E **59**, 84 (1999).

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

16 P. Dayal, S. Trebst, S. Wessel, D. Würtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, Phys. Rev. Lett. **92**, 097201 (2004).

17 P. Embrechts, C. Klüpelberg, and T. Mikosch, *Modelling Extremal Events*, Springer, Berlin (1997).

18 R. A. Fisher and L. H. C. Tippett, Proc. Cambridge Philos. Soc. **24**, 180 (1928).

19 R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values*, Birkhäuser, Basel (2001).

20 For the *finite* set of 2^{2N} distinct spin glass realizations of a given system size N the Fréchet distribution is cutoff at some finite value. The average round-trip time does not diverge, but is dominated by this cutoff value. For the system with $N = 8^3$ spins we can estimate the average round-trip time form the fitted Fréchet distributions to be $\tau_{\mathrm{average}} > 10^{100}$ which for all practical purposes cannot be distinguished from a divergence.

21 B. J. Schulz, K. Binder, M. Müller, and D. P. Landau, Phys. Rev. E **67**, 067102 (2003).

22 D. A. Huse, S. Trebst, and M. Troyer, cond-mat/0401195.

23 H. G. Ballesteros, L. A. Fernández, V. Martín-Mayor,

A. Muñoz Sudupe, G. Parisi, and J. J. Ruiz-Lorenzo, J. Phys. A **32**, 1 (1999).