

Nonflat histogram techniques for spin glassesFabio Müller^{⊗,*}, Stefan Schnabel^{⊗,†} and Wolfhard Janke^{⊗,‡}*Institut für Theoretische Physik, Universität Leipzig, IPF 231101, 04081 Leipzig, Germany*

(Received 7 June 2019; revised 2 July 2020; accepted 20 August 2020; published 11 November 2020)

We study the bimodal Edwards-Anderson spin glass comparing established methods, namely the multicanonical method, the $1/k$ ensemble, and parallel tempering, to an approach where the ensemble is modified by simulating power-law-shaped histograms in energy instead of flat histograms as in the standard multicanonical case. We show that by this modification a significant speed-up in terms of mean round-trip times can be achieved for all lattice sizes taken into consideration.

DOI: [10.1103/PhysRevE.102.053303](https://doi.org/10.1103/PhysRevE.102.053303)**I. INTRODUCTION**

Simulations of systems with rugged free-energy landscape [1] suffer from massive slowing down of the dynamics in the low-temperature phase. This problem encountered in many physical systems, e.g., folding polymers or spin glasses, renders the investigation of the thermodynamical properties of such systems in the low-temperature phase a very challenging task.

The Metropolis algorithm [2] is designed to sample configurations according to their statistical weight in the canonical ensemble. At low temperatures it fails dramatically for systems with rugged free-energy landscape because of the effectively broken ergodicity. The simulations get stuck in local minima (metastable states) and the thermal energy is not sufficient to overcome the huge free-energy barriers. There has been a wide range of algorithmic developments tackling this problem that can all be subsumed in the term of broad-energy ensembles.

One commonly employed method is the parallel tempering (PT) [3–5] method where Metropolis simulations of copies of the system (replicas) at different temperatures are performed. After certain time intervals exchanges of the replicas between the different temperatures are attempted. This procedure enables the replicas at low temperature to fully explore deep free-energy valleys and at high temperature to travel freely through the phase space and thus to decorrelate. Thereby, the different replicas can explore the rugged structure of the free-energy landscape much more efficiently than in a simple Metropolis simulation. Using this method at temperatures close to the transition, studies of spin-glass systems of sizes up to 40^3 spins have been reported [6,7]. For ground-state searches, systems of about 10^3 are feasible [8]. The great advantage of PT is its simplicity: The algorithm only needs a suitable temperature set and exhibits good performance,

which makes it probably the most employed method in the investigation of systems with rugged free-energy landscape.

Another recent development is the population annealing Monte Carlo [9–13] method, which proceeds similarly to simulated annealing [14] as the system is gradually cooled down according to an annealing schedule. The annealing is performed on a big population of replicas and by introducing intermediate resampling of the population of replicas after lowering the temperature the simulation is kept at thermal equilibrium. This permits the evaluation of thermodynamic observables in contrast to simple simulated annealing. Despite the attempts of optimizing the method for spin glasses [10,12,15] it is not able to outperform PT. Its optimization, however, remains more cumbersome due to the additional complexity. The main advantage of this algorithm is its suitability for massively parallel implementation. For disordered systems, however, this advantage does not come into play, because the necessity of simulating many different disorder realizations allows the efficient use of parallel computing for any method.

The multicanonical (MUCA) method [16–18] is another well-established algorithm designed for the simulation of systems with first-order phase transitions which performs well in the simulation of systems with rugged free-energy landscape, too. It has already been applied to spin glasses in Refs. [19,20]. In this method the simulation is set up to visit all possible energies with the same probability yielding a flat histogram in energy. However, it has been noted by different researchers that this ensemble is not optimal. One suggested improvement is the $1/k$ ensemble by Hesselbo and Stinchcombe [21], where the sampling distribution is the inverse of the integrated density of states. As the authors point out, this description samples the low-energy region more often than the high-energy region, resulting in energy histograms which grow toward the low-temperature phase.

Another nonparametric optimization of the MUCA algorithm was proposed in Ref. [22]. The method uses an estimator of the local diffusivity in order to maximize the number of performed round trips in energy. The method is among others applied to the ferromagnetic Ising model for which it improves the scaling behavior of the round-trip

*fmuller@itp.uni-leipzig.de

†schnabel@itp.uni-leipzig.de

‡janke@itp.uni-leipzig.de

times in energy. The improved performance for the models considered in that work and the nature of the algorithm of automatically identifying the bottlenecks of the simulation and concentrating the simulation effort on this region suggest that the round-trip times of the simulations should diminish independently of the underlying system. However, in our implementation, in the case of the three-dimensional (3D) bimodal Edwards-Anderson (EA) spin glass [23], the round-trip times did not systematically improve with this method. Instead, the simulation got stuck for some of the considered samples, rendering a comparison to the other methods impossible.

In this work we present a different approach: We prescribe parametric profiles for the histograms of the simulation and adjust the simulation weights accordingly. As for the three previous MUCA variants, it requires the knowledge of the underlying density of states, but it is much more flexible. The profiles are all chosen to be shifted power laws having two free parameters.

As an example we consider the 3D bimodal EA spin glass. This is one of the simplest models exhibiting a rugged free-energy landscape and is also interesting from the point of view of an optimization problem where finding ground states of hard disorder realizations is NP-hard [24]. Despite the exponential growth of the computational resources fundamental questions regarding the nature of the spin-glass phase still remain. For the progress in understanding the open questions, the development of new methods and an improvement of the existing methods is crucial.

The rest of the paper is organized as follows. In Sec. II the spin-glass model and the simulation methods are explained. The direct comparison of the round-trip times of the individual methods is performed in Sec. III. The framework of extreme-value statistics is introduced in Sec. IV. In Sec. V benchmarks for the global comparison are discussed and the different methods are compared in terms of those benchmarks. The results are summarized in Sec. VI.

II. MODEL AND EMPLOYED METHODS

We take into consideration the 3D bimodal EA model whose Hamiltonian takes the form

$$H = - \sum_{\langle ij \rangle} J_{ij} S_i S_j, \quad (1)$$

where the bonds J_{ij} and the spins S_i can take values ± 1 . The sum runs over all neighboring spins in the simple-cubic lattice with periodic boundary conditions.

Due to the disordered nature of spin glasses the study has to take into account a sufficiently large set of disorder realizations on which the averaged quantities can be computed. In this case one disorder realization consists of a set of $3V$ couplings J_{ij} which are either positive or negative unity with a probability of 50%, where $V = L^3$ is number of spins in a lattice of linear lattice size L . The disorder realizations are generated prior to the simulation and then kept fixed for all times (quenched disorder). As an adequate set of disorder realizations, 4000 samples with $L = 3$ and $L = 4$ are generated and 5000, 6000, and 4000 samples of size $L = 5, 6$, and 8 , respectively.

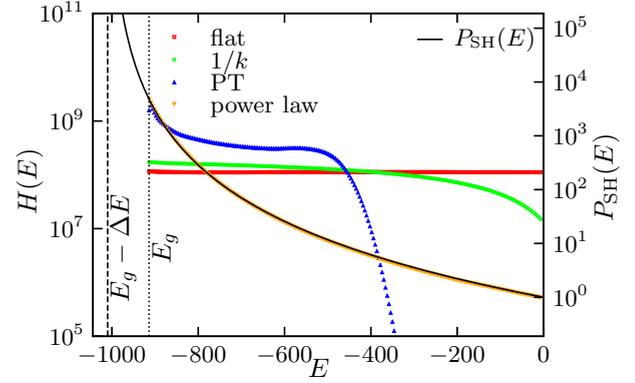


FIG. 1. The recorded histograms $H(E)$ of the different methods and the profile function $P_{\text{SH}}(E)$ for one disorder realization of linear lattice size $L = 8$. The dotted and the dashed vertical lines indicate the position of the ground-state energy E_g and the position of the pole of the power law (5), respectively.

The method which we adapted is the well-established MUCA method [17] employing a generalized Metropolis criterion with an energy-dependent weight function,

$$P_{\text{acc}} = \min \left[1, \frac{W(E_{\text{new}})}{W(E_{\text{old}})} \right], \quad (2)$$

where the weight function is proportional to the inverse of the density of states $\Omega(E)$,

$$W(E) \propto \Omega^{-1}(E). \quad (3)$$

For the MUCA simulations $\Omega(E)$ has to be sufficiently well-known *a priori* for each disorder realization. An estimator for it can, for instance, be obtained by means of the Wang-Landau algorithm [25] or, as in this work, by other iterative procedures which are explained, e.g., in Ref. [26]. This ensemble produces histograms which are flat in energy and is, therefore, often also referred to as “flat histogram method.”

A straightforward generalization of the flat histogram methods are the nonflat histogram methods. If the simulation weights for the flat MUCA method are multiplied with the desired energy-dependent shape (or profile) function $P_{\text{SH}}(E)$,

$$W(E) \propto \Omega^{-1}(E) P_{\text{SH}}(E), \quad (4)$$

then the resulting histograms will be shaped according to $P_{\text{SH}}(E)$. In this work all the profiles are shifted power laws of the form

$$P_{\text{SH}}(E, \Delta E, \alpha) = \left(\frac{E}{\Delta E - E_g} + 1 \right)^\alpha, \quad (5)$$

where the exponent $\alpha < 0$ and $\Delta E > 0$ is the position of the pole relative to the ground-state energy E_g of the respective spin-glass realization. In this parametrization the power laws are normalized to unity at $E = 0$.

In Fig. 1 the recorded histograms of the different methods are displayed on a logarithmic y scale for one disorder realization with $L = 8$. In contrast to flat MUCA, all methods have in common that the distribution of sampled states grows toward the ground-state energy. The recorded histogram of nonflat MUCA matches perfectly the imposed profile and

its histogram in the ground-state region is similar to that of PT. We are convinced that this feature, which among the existing methods is strongest for PT, enhances the ability of sampling the low-energy region and especially the ability of finding low-energy states of investigated systems. There are different possible choices of functional forms which enhance the sampling of the low-energy region and even a stepwise defined function could be employed and might even yield better results. We chose a power law because the two involved parameters allow for a good adaptation but the tuning of the parameters in the two-dimensional parameter space remains feasible.

For the above parametrization we found a fixed parameter set, namely $\alpha = -3.6$ and $\Delta E = 96$, which independently of the lattice size yielded the shortest mean round-trip times among the considered profiles. Subsequently, we will refer to the nonflat MUCA setting with the power-law shape belonging to this parameter set just as the power-law (PL) setting or nonflat MUCA method. While the overall best results are obtained with this parameter set, we want to point out that an improvement compared to flat MUCA was visible for each of the considered parameter sets. The parametrization with a fixed offset from the ground-state energy yields different relative distributions depending on the ground-state energy encountered in the respective disorder realization. The value of the profile function at the ground-state energy is given by

$$P_{\text{SH}}(E_g, \Delta E, \alpha) = \left(\frac{1}{1 - \frac{E_g}{\Delta E}} \right)^\alpha. \quad (6)$$

The sampling at the ground-state energy compared to zero energy is thus enhanced by a factor of ≈ 13 for a disorder realization with $L = 4$ and a typical ground-state energy of ≈ -100 . For a sample with $L = 8$ and typical ground-state energy of ≈ -900 instead it is enhanced by a factor of ≈ 4500 . Due to this feature this parametrization of the profile function does not require any adjustments of the parameters in the system sizes which we considered. Presumably such a profile will also yield good results for larger systems, although we cannot be certain.

Next, the $1/k$ ensemble [21] is considered, which is defined by setting the simulation weights equal to the inverse of the integrated density of states up to the energy of the respective bin,

$$W_{1/k}(E) \propto 1/k = \left[\int_{E_g}^E dE' \Omega(E') \right]^{-1}. \quad (7)$$

Here, a first-order Taylor expansion of $\ln \Omega$ at E leads to $W(E) \approx W_{1/k}(E)$ if $P_{\text{SH}}(E) = d \ln \Omega(E')/dE'|_{E'=E}$. This prescription again relies on the knowledge of the density of states. The authors of Ref. [21] stress its robust ergodicity and apply it to spin glasses and the traveling salesman problem [27].

Since for the above-mentioned methods the density of states is the only needed input, it was determined only once to high accuracy employing the iterative procedure adapted from Ref. [26] but with power-law-shaped distributions in energy. In this case, and generally when the ground-state energy of the system is not known, *a priori* the profile function has to be adapted whenever a lower energy is found.

Last, the PT method, being probably the most employed algorithm for spin-glass simulations, is included in the comparison. The ensemble in this case is defined by a set of M temperatures $\{T_i, i = 1, \dots, M\}$. For each temperature T_i a Metropolis simulation of a copy of the system (replica) is performed. The temperatures of the replicas i and j are allowed to exchange configuration according to

$$P_{ij}^{\text{ex}} = \min \left[1, e^{\left(\frac{1}{T_j} - \frac{1}{T_i}\right)(E_j - E_i)} \right], \quad (8)$$

where E_i and E_j are the energies of replica i and j and $k_B = 1$. This prescription allows for fast decorrelation when a replica travels to high temperature and the exploration of the local minima at low temperatures. Among the vast choice of different PT protocols available [28] we opted for the constant exchange rate protocol with acceptance rates between 40% and 60% [29]. For all simulations the maximal temperature was chosen to be well above the critical temperature, $T_{\text{max}} > 3 > T_c \approx 1$. The exchange rates were imposed on each individual disorder realization in an initial equilibration run during which the temperatures were modified accordingly. The number of replicas was set to $M = 7, 7, 12, 14$, and 20 for $L = 3, 4, 5, 6$, and 8, respectively. We note that the choice of the temperature set is crucial for the PT algorithm and also provides the possibility of optimizations as, for example, in Ref. [30]. However, in this work we rather limit ourselves to a well-established protocol for PT focusing on the optimization of the nonflat histogram technique.

III. COMPARISON OF THE ROUND-TRIP TIMES

The observable taken into account for this study is the round-trip time. For all methods except PT and each disorder realization it is defined as the time needed by the simulation to travel from the highest energy ($E \approx 0$) to the ground-state energy and back. For PT, instead, the round trip is measured between the ground-state energy and an energy typical for a canonical ensemble with a temperature well above the freezing point of the disorder realization [31,32]. This time can be taken as an upper bound of the autocorrelation time of the energy of the respective disorder realization at the ground state. We want to stress that the energies we refer to as ground-state energies are the lowest encountered energies and may not be the true ground states. However, the round-trip times were always measured performing at least 100 round trips for each individual sample and method so that several hundred round trips have been performed on each disorder realization. In case lower energies were measured during this process, the disorder realization was requeued and simulated again until the desired number of round trips was achieved. This procedure renders the discovery of the true ground state very probable. After at least 100 round trips the relative statistical error in the round-trip time τ_i is of the order of $\Delta \tau_i / \tau_i \approx 0.1$.

The first property we want to look at is the dependence of the round-trip times for the individual disorder realizations on the employed method. The scatter plots in Fig. 2 show the round-trip times of the same disorder realization for two different methods for all the simulated disorder realizations of size $L = 4$ and $L = 8$ on a log-log scale. The strong correlation of the round-trip times for each single disorder realization

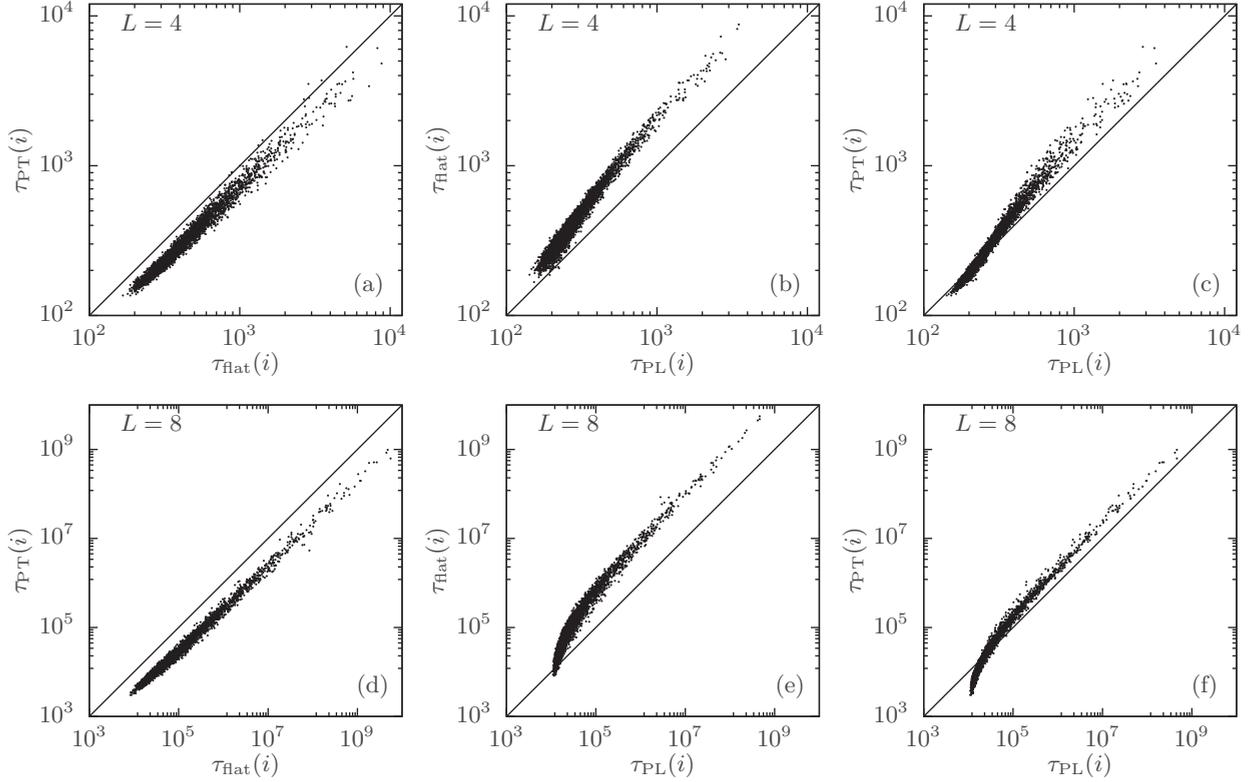


FIG. 2. Scatter plots of the round-trip times comparing the nonflat power-law histogram technique (PL) to the standard flat MUCA (flat) and parallel tempering (PT) methods for sizes $L = 4$ (upper panels) and $L = 8$ (lower panels). All points scattered above the identity line have longer round-trip times for the method on the y axis.

should be noted, indicating that the hardness of the underlying optimization problem is primarily a characteristic of the disorder realization and mostly independent of the employed method. This fact allows us to categorize the disorder realizations and speak of easy and hard instances. Comparing the round-trip times τ_i for the flat MUCA method and the parallel tempering method (left panels) for both lattice sizes $L = 4$ and $L = 8$, the τ_i are systematically lower for PT, indicating its superior performance for the whole classes of the bimodal EA spin glasses of the respective lattice sizes.

When comparing the performance of the nonflat histogram method to the flat MUCA method (central panels), the surrounding area of the scattered round-trip times shows a bending, i.e., for $L = 4$ the flat histogram method displays only slightly higher round trip times for the easy disorder realizations. With increasing hardness the round-trip times for the flat histogram method grow faster than for the PL setting. This effect gets enhanced with a further increase of the lattice size (see lower panel) where for the case of $L = 8$ the round-trip times of the easiest samples for the flat MUCA method are similar to those for PL. However, as will become apparent in the next section, the hard samples contribute most to the mean round-trip time so that even a slightly weaker performance for the easier samples would hardly contribute to the total computation time.

The right panels show the comparison of PL to PT. For $L = 4$ PT outperforms the nonflat histogram method for the easy disorder realizations, while for the hard ones PL displays shorter round-trip times. For $L = 8$ a large fraction of the

disorder realizations is characterized by shorter round-trip times for PT, but the tail of the distribution describing the hard samples exhibits shorter round-trip times for PL.

IV. ROUND-TRIP TIME DISTRIBUTIONS

In order to quantify the observations of the previous section the distributions of the round-trip times can be examined. Round trips in energy include the visit of the ground state of the respective disorder realization which is an extreme event. Their statistics must thus be described in the framework of extreme-value statistics. One of the main results in this field is given by the Fisher-Tippet-Gnedenko theorem [33], which characterizes the type of distributions to which extreme-value distributions can converge. The round-trip time distributions of the bimodal EA spin glass all seem to converge to Fréchet distributions independently of the method and the system size. This has already been suggested in Ref. [34] for the round-trip time distributions of the 3D EA model employing the flat histogram ensemble.

One parametrization of the cumulative distribution function (CDF) of the Fréchet distribution is given by

$$F(\tau) = \exp \left[- \left(1 + \xi \frac{\tau - \mu}{\beta} \right)^{-1/\xi} \right], \quad (9)$$

with $\tau \in [\mu - \beta/\xi, \infty)$. The location of the distribution along the τ axis is determined by μ , β is the scale parameter, and the shape parameter ξ describes the decay of the tail of the

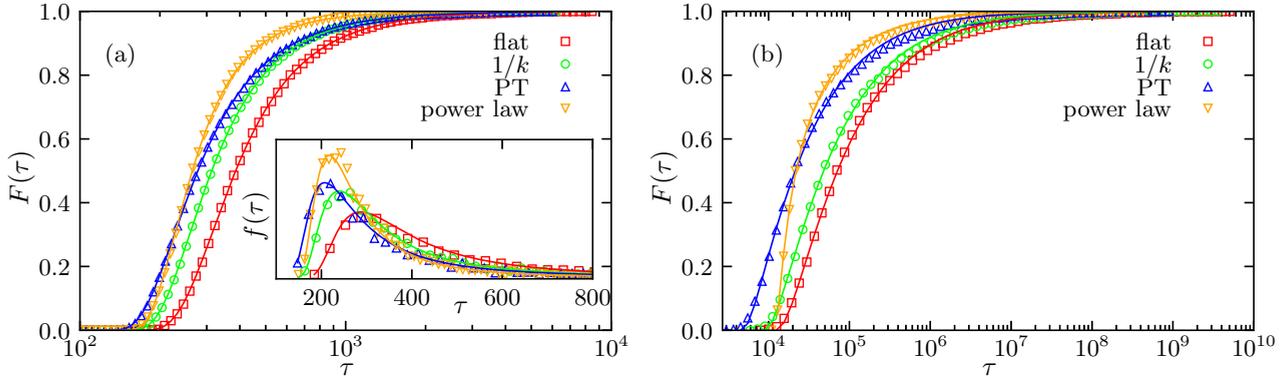


FIG. 3. Round-trip time distributions (symbols) and best fitting cumulative distribution functions (lines) for the different methods and lattice size $L = 4$ (a) and $L = 8$ (b). The inset of the left panel shows the PDF form of the distribution.

distribution, i.e., the occurrence of rare events. The CDF is the integrated form of the probability density function (PDF) $f(\tau)$. The round-trip time distributions are thus all defined by sets of parameters μ , β , ξ which are determined by fitting the CDF to the recorded round-trip times.

The measured round-trip times and the respectively best-fitting Fréchet distribution for lattice sizes $L = 4$ and $L = 8$ are plotted in Fig. 3. The points represent the measured data and the solid lines are the best-fitting Fréchet distributions. The varying performance of the methods in dependence on the difficulty of the disorder realizations which became visible in the last section also reflects in the distribution of the round-trip times. For both lattice sizes the CDF belonging to the flat MUCA method is lower for all τ than the one belonging to PT. The maximum increase which corresponds to the bulk of the distribution is shifted to higher τ for MUCA as compared to PT.

Comparing PT instead to the PL setting yields a different picture: The cumulative distribution functions for lattice size $L = 4$ cross at $F(\tau) \approx 1/3$, corresponding to a round-trip time $\tau \approx 2 \times 10^2$. This means that for the PT algorithm the easiest one-third of all samples has smaller round-trip times than the easiest third for the PL method, while PL is faster for the harder two-thirds. For $L = 8$ the PL round-trip times are larger for the easier half of the samples and smaller for the harder half. The round-trip times for the hard disorder realizations have most influence on the decay of the distribution and thus on the shape parameter ξ . In Fig. 4 the scaling of the shape parameter for the different methods is displayed, where the errors of the best fitting parameters are estimated via jackknifing [35]. For the considered lattice sizes the shape parameter scales similarly for all the different methods. However, the values for PL are systematically lower for $L \geq 4$. This is in good agreement with its superior performance for the difficult disorder realizations.

V. ASSESSING THE PERFORMANCE OF THE DIFFERENT METHODS

Next, we want to compare the performance of the different simulation methods. The most intuitive observable would be the disorder average of the round-trip times over the set of

considered disorder realizations. However, as it turns out, the rare-state events which have a dominating influence on the distribution mean are not within the set of simulated disorder realizations. This effect is accounted for by considering distribution means up to large quantiles of the underlying extreme-value distributions, yielding a more reliable measure of the *real* performance of the different methods.

A. Finding a benchmark

In principle, the *real* performance could be determined by measuring the round-trip time of every possible disorder realization. This procedure is discarded due to the enormous number of possible disorder realizations [36]. Instead, we generate a subset of all possible disorder realizations and from those we try to infer the expected mean round-trip time of all the disorder realizations belonging to the same problem class by means of the *population mean* τ_{pop} . This is a standard approach in all Monte Carlo studies and the law of large numbers assures its convergence for all random variables from distributions with a well-defined mean. However, this prerequisite is not fulfilled for all of the round-time distributions encountered in this work.

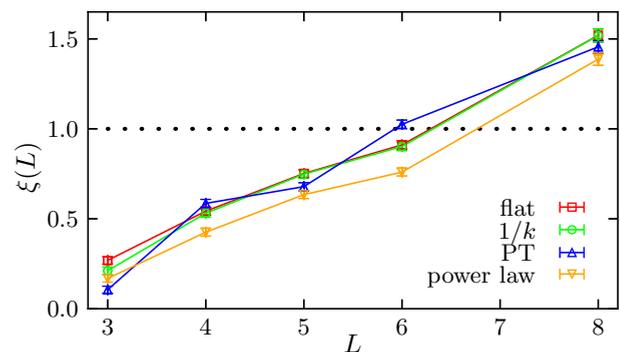


FIG. 4. The figure shows the shape parameter of the best fitting Fréchet distribution in dependence of the lattice size for the different employed simulation methods. The dotted line indicates the threshold value from which on the distribution mean diverges.

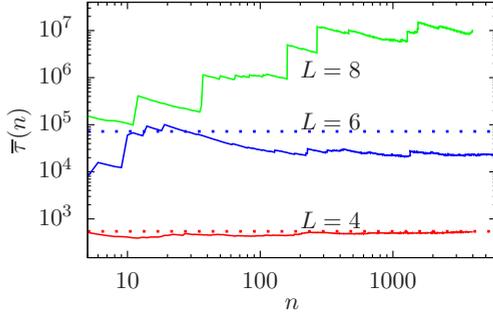


FIG. 5. Illustration of the convergence of the population mean of the round-trip times for the flat MUCA method in dependence of the population size to the distribution mean of the underlying distribution. The solid lines are the running mean including the first n samples and the dotted lines in the respective color are the means of the underlying distribution.

The expected mean round-trip time resulting from the underlying probability density could be estimated by the *distribution mean*,

$$\langle \tau \rangle = \int_{\mu - \beta/\xi}^{\infty} d\tau \tau f(\tau). \quad (10)$$

This integral can be computed analytically, yielding

$$\langle \tau \rangle = \begin{cases} \mu + \frac{\beta}{\xi} [\Gamma(1 - \xi) - 1] & \text{for } \xi < 1 \\ \infty & \text{otherwise} \end{cases}, \quad (11)$$

with $\Gamma(x)$ being the gamma function. The distribution mean is, therefore, only defined as long as the shape parameter ξ is smaller than 1 [37]. To illustrate this difficulty one can consider the *running mean* which is defined as the population mean over the first n generated disorder realizations keeping them in a fixed order,

$$\bar{\tau}(n) = \frac{1}{n} \sum_{i=1}^n \tau_i, \quad (12)$$

implying $\tau_{\text{pop}} = \bar{\tau}(N)$, where N is the number of all simulated disorder realizations. In Fig. 5 the running mean for the flat MUCA method and different system sizes is plotted together with the respective distribution mean if it is defined. For $L = 4$ ($\xi \ll 1$) the running mean quickly converges to the distribution mean indicated by the dotted line. For $L = 6$ ($\xi \approx 1$) the jumps due to rare events in the tail of the distribution become

more pronounced. The running mean is still expected to approach the distribution mean for a finite number of disorder realizations. For the 6000 samples considered in our work this is still not the case. For $L = 8$ ($\xi > 1$) the distribution mean is not defined. In the picture of the running mean, jumps represent round-trip times in the tail of the distribution. In the case of $\xi > 1$ those jumps τ_n/n in the running mean are clearly visible in Fig. 5 and will lead to a divergence of the population mean the more disorder realizations are taken into account and hence the more rigorously the tail of the distribution is explored. This illustrates that the population mean as a measure for the performance of the different methods must be taken with a grain of salt.

In order to retain the characteristics of the underlying round-trip time distribution into the estimator of the performance of the different methods the distribution mean up to a certain quantile can instead be taken into account. The quantile function is the inverse of the CDF (9),

$$Q(p) = F^{-1}(p) = \mu + \frac{\beta}{\xi} [(-\ln p)^{-\xi} - 1], \quad p \in (0, 1), \quad (13)$$

yielding the round-trip time $\tau_p = Q(p)$ at which a certain fraction p of the distribution is accumulated. For each $\epsilon < 1$ we define the *quantile mean* $\langle \tau \rangle_{\epsilon}$ disregarding a fraction ϵ of the tail of the distribution as the integral (10) with the upper bound replaced by $Q(1 - \epsilon)$,

$$\langle \tau \rangle_{\epsilon} = \int_{\mu - \beta/\xi}^{Q(1 - \epsilon)} d\tau \tau f(\tau). \quad (14)$$

The integral is evaluated with the parameters of the best-fitting distributions to the measured round-trip times; see Fig. 3. This enables a well-defined extrapolation beyond the measured round-trip times of the simulated disorder realizations of the underlying study and thus a comparison of the different methods beyond the mere population mean, which may be strongly dependent on the set of disorder realizations taken into account for the study.

B. Comparison of the different methods

Finally, for the comparison of the mean round-trip times only the population mean τ_{pop} and the quantile mean $\langle \tau \rangle_{\epsilon=10^{-4}}$ neglecting a fraction $\epsilon = 10^{-4}$ of the tail of the distribution

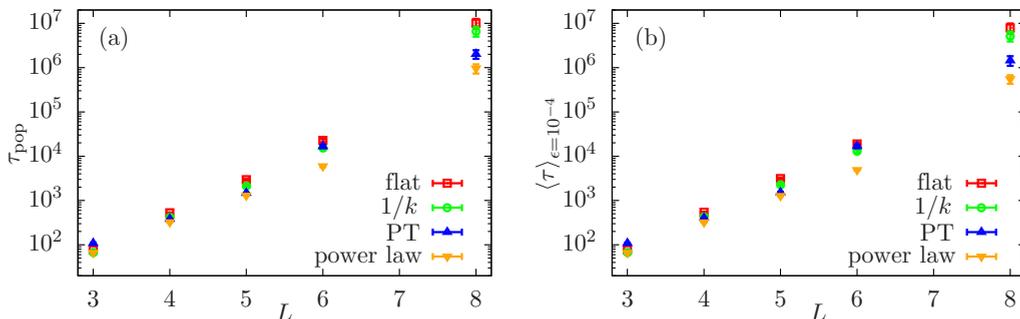


FIG. 6. Population mean τ_{pop} (a) and quantile mean $\langle \tau \rangle_{\epsilon=10^{-4}}$ (b) of the round-trip times for the different methods as a function of system size. The latter is the more reliable statistical quantity.

TABLE I. Ratios of the population mean τ_{pop} and the quantile mean $\langle \tau \rangle_{\epsilon=10^{-4}}$ of the round-trip times for flat MUCA, the $1/k$ ensemble, and parallel tempering with respect to the same quantities for the power-law MUCA method.

L	Flat MUCA		$1/k$ ensemble		Parallel tempering	
	r_{pop}	$r_{\epsilon=10^{-4}}$	r_{pop}	$r_{\epsilon=10^{-4}}$	r_{pop}	$r_{\epsilon=10^{-4}}$
3	1.160(2)	1.174(3)	1.0146(6)	1.0193(9)	1.637(4)	1.640(5)
4	1.622(8)	1.68(2)	1.288(5)	1.328(10)	1.175(6)	1.25(2)
5	2.28(5)	2.44(6)	1.63(3)	1.75(4)	1.136(5)	1.185(8)
6	3.8(2)	3.9(2)	2.59(9)	2.6(2)	2.8(2)	3.4(3)
8	10.5(2)	14.2(6)	6.9(3)	9.4(4)	2.1(2)	2.62(7)

are taken into account as the distribution mean for the parallel tempering method is already ill defined for $L = 6$.

The two definitions are evaluated for all simulated lattice sizes and plotted in Fig. 6. Both definitions of the mean grow exponentially up to linear system size $L = 6$ until which the mean is defined, while for $L > 6$, where the distribution means diverge, they seem to be growing faster than exponentially. We have also looked at the scaling of the more commonly used quantiles including the median [38,39], which are derived directly from the τ values without the intermediate step of fitting to a statistical model. These quantiles behave similarly to the quantile means (14) being, however, less stable for small ϵ .

For the direct comparison of PL with the existing methods we introduce the relative performance r which we define as the fraction of the mean of the respective method and the one of PL. In Table I, the relative performance for all different system sizes is listed. The errors in r are estimated using the Jackknife resampling technique. It consists of generating a set of ratios $\{r_i\}$, where for the calculation of each r_i only a subset of all disorder realizations is taken. The error in r is derived from the variance of the so-generated Jackknife sample.

The speedup of PL compared to flat MUCA increases with system size, reaching a factor of more than 10 for $L = 8$ for both definitions of the mean, while compared to the $1/k$ ensemble the speedup for the biggest system size is still a factor of $r \approx 7-9$. Compared to PT the speedup is less pronounced

and not steadily growing with system size, reaching a factor of $r \approx 2-3$ for our largest system sizes.

VI. CONCLUSION

Setting up multicanonical simulations such that the outgoing histograms are shaped according to power laws instead of being flat is trivially achievable. Nevertheless, this simple approach enables us to gather significantly more independent statistics at the ground-state energy, which is important because the thermodynamic contributions of the ground state of spin glasses are believed to be significant. It is likely that similar techniques will also improve the sampling of the ground state of other systems with complex free-energy landscape such as polymers and in particular proteins, for which the importance of the native state is well known.

While PT has been the most employed method in the simulation of spin glasses probably also due to its good ability to investigate the ground-state region, we were able to show that the power-law setting considerably improves the performance of multicanonical simulations in this respect, rendering them at least comparable to PT.

The overall gain in performance grows with increasing lattice size and reaches a factor of up to 10–15 in comparison to flat MUCA and still a factor of up to 2–3 compared to PT. In terms of round-trip time distributions, the heaviness of the tails is reduced by its superior ability to deal with the hard disorder realizations.

This improved ability of the here proposed power-law MUCA method of finding ground states for the hard instances implies its usefulness in the application to general optimization problems. This is particularly useful because many other optimization problems can be rephrased in terms of spin-glass Hamiltonians [40] and thus solved employing the same methodology.

ACKNOWLEDGMENTS

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project No. 189 853 844–SFB/TRR 102 (project B04). It was further supported by the Deutsch-Französische Hochschule (DFH-UFA) through the Doctoral College “ \mathbb{L}^4 ” under Grant No. CDFA-02-07.

- [1] *Rugged Free Energy Landscapes: Common Computational Approaches to Spin Glasses, Structural Glasses and Biological Macromolecules*, edited by W. Janke, Lecture Notes in Physics, Vol. 736 (Springer, Berlin, 2008).
- [2] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- [3] K. Hukushima and K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [4] R. H. Swendsen and J.-S. Wang, Replica Monte Carlo Simulation of Spin-Glasses, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [5] C. J. Geyer and E. A. Thompson, Annealing Markov chain Monte Carlo with applications to ancestral inference, *J. Am. Stat. Assoc.* **90**, 909 (1995).
- [6] M. Hasenbusch, A. Pelissetto, and E. Vicari, Critical behavior of three-dimensional Ising spin glass models, *Phys. Rev. B* **78**, 214205 (2008).
- [7] M. Baity-Jesi, R. A. Baños, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, D. Iniguez, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muñoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, M. Pivanti, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccone, and D. Yllanes (Janus

- Collaboration), Critical parameters of the three-dimensional Ising spin glass, *Phys. Rev. B* **88**, 224416 (2013).
- [8] W. Wang, J. Machta, and H. G. Katzgraber, Comparing Monte Carlo methods for finding ground states of Ising spin glasses: Population annealing, simulated annealing, and parallel tempering, *Phys. Rev. E* **92**, 013303 (2015).
- [9] Y. Iba, Population Monte Carlo algorithms, *Trans. Jpn. Soc. Artif. Intell.* **16**, 279 (2001).
- [10] K. Hukushima and Y. Iba, Population annealing and its application to a spin glass, *AIP Conf. Proc.* **690**, 200 (2003).
- [11] J. Machta, Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes, *Phys. Rev. E* **82**, 026704 (2010).
- [12] W. Wang, J. Machta, and H. G. Katzgraber, Population annealing: Theory and application in spin glasses, *Phys. Rev. E* **92**, 063307 (2015).
- [13] L. Y. Barash, M. Weigel, M. Borovský, W. Janke, and L. N. Shchur, GPU accelerated population annealing algorithm, *Comput. Phys. Commun.* **220**, 341 (2017).
- [14] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science* **220**, 671 (1983).
- [15] A. Barzegar, C. Pattison, W. Wang, and H. G. Katzgraber, Optimization of population annealing Monte Carlo for large-scale spin-glass simulations, *Phys. Rev. E* **98**, 053308 (2018).
- [16] B. A. Berg and T. Neuhaus, Multicanonical algorithms for first order phase transitions, *Phys. Lett. B* **267**, 249 (1991).
- [17] B. A. Berg and T. Neuhaus, Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions, *Phys. Rev. Lett.* **68**, 9 (1992).
- [18] W. Janke, Multicanonical simulation of the two-dimensional 7-state Potts model, *Int. J. Mod. Phys. C* **03**, 1137 (1992).
- [19] B. A. Berg and T. Celik, New Approach to Spin-Glass Simulations, *Phys. Rev. Lett.* **69**, 2292 (1992).
- [20] B. A. Berg, T. Celik, and U. Hansmann, Multicanonical study of the 3d Ising spin glass, *Europhys. Lett.* **22**, 63 (1993).
- [21] B. Hesselbo and R. B. Stinchcombe, Monte Carlo Simulation and Global Optimization Without Parameters, *Phys. Rev. Lett.* **74**, 2151 (1995).
- [22] S. Trebst, D. A. Huse, and M. Troyer, Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations, *Phys. Rev. E* **70**, 046701 (2004).
- [23] S. F. Edwards and P. W. Anderson, Theory of spin glasses, *J. Phys. F: Met. Phys.* **5**, 965 (1975).
- [24] F. Barahona, On the computational complexity of Ising spin glass models, *J. Phys. A: Math. Gen.* **15**, 3241 (1982).
- [25] F. Wang and D. P. Landau, Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [26] W. Janke, Histograms and all that, in *Computer Simulations of Surfaces and Interfaces*, edited by B. Dünweg, D. P. Landau, and A. I. Milchev (Springer, Dordrecht, the Netherlands, 2003), pp. 137–157.
- [27] The aim of the study was not maximizing the number of round trips in energy but rather the amount of statistically independent data in an uncorrelated Monte Carlo simulation.
- [28] T. Papakonstantinou and A. Malakis, Parallel tempering and 3D spin glass models, *J. Phys. Conf. Ser.* **487**, 012010 (2014).
- [29] E. Bittner, A. Nußbaumer, and W. Janke, Make Life Simple: Unleash the Full Power of the Parallel Tempering Algorithm, *Phys. Rev. Lett.* **101**, 130603 (2008).
- [30] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, Feedback-optimized parallel tempering Monte Carlo, *J. Stat. Mech.: Theory Exp.* (2006) P03018.
- [31] The estimated temperature was extracted from few single disorder realizations for each lattice size and then taken as constant for all samples with the same lattice size.
- [32] This different measuring prescription gives PT a slight advantage in comparison to the other methods which is, however, negligible in the authors' opinion.
- [33] M. Charras-Garrido and P. Lezard, Extreme value analysis: An introduction, *J. Soc. Fr. Stat.* **154**, 66 (2013).
- [34] S. Alder, S. Trebst, A. K. Hartmann, and M. Troyer, Dynamics of the Wang–Landau algorithm and complexity of rare events for the three-dimensional bimodal Ising spin glass, *J. Stat. Mech.: Theory Exp.* (2004) P07008.
- [35] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982).
- [36] The bimodal EA spin glass, having discrete randomness, has only a finite number of possible disorder realizations. Due to the symmetries in the absence of external fields this number can be estimated to be of the order of 2^V , where $V = L^3$ is the number of spins contained in the lattice.
- [37] Due to the finite number of possible disorder realizations the mean \bar{m} is actually defined. Its finite value is expected to be of the order of 10^{100} and can therefore in terms of computation time numerically not be distinguished from a real divergence.
- [38] B. A. Berg, A. Billoire, and W. Janke, Spin-glass overlap barriers in three and four dimensions, *Phys. Rev. B* **61**, 12143 (2000).
- [39] E. Bittner and W. Janke, Free-energy barriers in the Sherrington–Kirkpatrick model, *Europhys. Lett.* **74**, 195 (2006).
- [40] A. Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 5 (2014).