Distribution of metastable states of Ising spin glasses

Stefan Schnabel* and Wolfhard Janke†

Institut für Theoretische Physik, Universität Leipzig, Postfach 100920, 04009 Leipzig, Germany

Local minima, also known as inherent structures, are expected to play an essential role in the behavior of spin glasses. Here, we propose techniques to efficiently sample these configurations in Monte Carlo simulations. For the Sherrington-Kirkpatrick and the three-dimensional Edwards-Anderson model their spectra are determined and compared to analytical results.

I. INTRODUCTION

Structurally or magnetically disordered glassy systems show special characteristics such as memory effects or replica symmetry breaking as a result of a complex energy landscape. A large number of local minima or metastable states exist and can trap the system at low temperatures. Therefore, these local energy minima have been of interest for some time. For mean-field spin glasses their distribution as function of energy was determined as early as 1981 [1]. The one-dimensional system with short-range interactions can also be treated analytically [2] and some general properties are known for higher-dimensional spin glasses [3]. A completely analytical treatment for the cubic lattice is, however, not available and the numerical investigation of metastable states is even more demanding than standard spin-glass simulations. Previous studies were restricted to exact enumeration of small systems [4–6] or quenches from random configurations [7]. In the latter work as well as in studies on structural glasses [8] the local minima are seen in relation to the equilibrium configurations from which they are derived by steepest descent (greedy algorithm) and are referred to as inherent structures. Recently we introduced a technique [9] for the Edwards-Anderson model that efficiently derives inherent structures from a sequence of spin configurations by means of a dynamic greedy algorithm. In this study we extend this approach and propose a method that samples all metastable states with equal probability. We use this and a more traditional method to measure the distributions of metastable states as a function of energy for the Sherrington-Kirkpatrick (SK) [10] and the three-dimensional Edwards-Anderson (EA) [11] model.

The rest of the paper is organized as follows: We discuss the models in Sec. II and briefly review the analytical solution of Bray and Moore in Sec. III. In Sec. IV we introduce our methods and Sec. V contains the results. The paper finishes with some concluding remarks in Sec. VI.

II. MODEL

We consider the Ising Hamiltonian

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} s_i s_j, \qquad s_i \in \{-1, 1\}, \tag{1}$$

where the sum runs over all pairs of spins s_i interacting via bonds J_{ij}. The latter are randomly chosen according to a Gaussian distribution:

$$P(J_{ij}) = \frac{1}{\sqrt{2\pi J^2}} e^{-J_{ij}^2 / 2J^2}. \tag{2}$$

In the case of the SK model every spin interacts with every other while for the 3d EA model spins are placed on the sites of a cubic lattice and only adjacent ones contribute to the energy. If we consider single-site energies, i.e., the sum of all terms to which an individual spin s_k contributes,

$$e_k = -\sum_{\langle ij \rangle} J_{ij} s_i s_j (\delta_{ik} + \delta_{jk}), \tag{3}$$

one can express the Hamiltonian as

$$\mathcal{H} = \frac{1}{2} \sum_k e_k = E, \tag{4}$$

and the energy change associated with a single spin flip,

$$\mathbf{S} = (s_1, \dots, s_N) \to \mathbf{S}' = (s_1, \dots, s_{k-1}, -s_k, s_{k+1}, \dots), \tag{5}$$

$$e_k \to e_k' = -e_k, \tag{6}$$

as functions of it,

$$\mathcal{H}(\mathbf{S}') - \mathcal{H}(\mathbf{S}) = -2e_k. \tag{7}$$

Hence, a metastable state or more precisely a single-flip stable state, i.e., a spin configuration for which every single spin flip causes an increase in energy, can be asserted if $e_k < 0$ for all k. It is the distribution $\Omega(E)$ of these metastable states that we are interested in.

III. BRAY AND MOORE'S SOLUTION

In 1981 Bray and Moore [1] derived an analytic expression for the distribution of metastable states for the SK model. They used the dimensionless normalized energy

$$\varepsilon = \frac{E}{NJz^{\frac{1}{2}}}, \tag{8}$$

where N is the total number of spins and z the coordination number of the lattice, i.e., $z = N - 1$ for the SK model. For

*schnabel@itp.uni-leipzig.de
†Wolfhard.Janke@itp.uni-leipzig.de
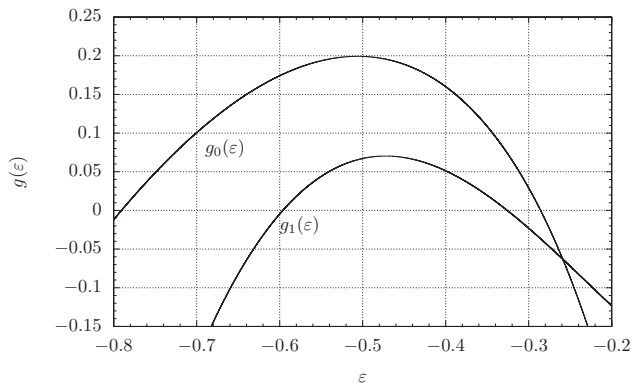
FIG. 1. The functions g_0 and g_1.

the limit

$$g_0(\varepsilon) := \lim_{N \to \infty} N^{-1} \langle \ln \Omega(\varepsilon) \rangle_J, \qquad (9)$$

where $\langle \dots \rangle_J$ denotes the disorder average, they obtained

$$g_0(\varepsilon) = \varepsilon^2 + 2\varepsilon\tau - \ln[\sqrt{\pi/2}(-2\varepsilon - \tau)], \qquad (10)$$

where the function $\tau = \tau(\varepsilon)$ is implicitly defined by

$$0 = 2\varepsilon + \tau + \Phi'(\tau)/\Phi(\tau) \qquad (11)$$

with

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy. \qquad (12)$$

However, they state that this solution is only valid for $\varepsilon > \varepsilon_c \approx -0.672$. Nonetheless, it follows from the position of the maximum of $g_0(\varepsilon)$ shown in Fig. 1 that for large systems the number of metastable states is given by

$$\langle \ln N_S \rangle_J / N = 0.199228, \qquad (13)$$

and that the average energy of a local minimum becomes

$$\epsilon_{\mathrm{av}} = -0.5061. \qquad (14)$$

Bray and Moore then proceeded with an expansion in $1/z$ and obtained an approximation for non-mean-field spin glasses:

$$N^{-1} \langle \ln \Omega(\varepsilon) \rangle = g_0(\varepsilon) + z^{-1} g_1(\varepsilon) + O(z^{-2}) \qquad (15)$$

with

$$g_1(\varepsilon) = -\varepsilon^2(\tau^2 - 2\varepsilon^2), \qquad (16)$$

also displayed in Fig. 1.

IV. METHODS

We apply two methods in order to sample local minima. While the first is a more traditional approach using standard Monte Carlo techniques, the second method is an algorithm that has been derived from the recently introduced dynamic greedy algorithm [9]. Its efficiency relies on the specifically local nature of single spin flips, i.e., low connectivity, and in this study we only apply it to the Edwards-Anderson model.

A. Method I

This method is a standard Monte Carlo technique which employs flips of single spins and samples in principle all possible states of the spin glass. The ensemble is designed to include all local minima with a sufficiently high probability, such that their distribution can be inferred. Since the goal is to find local minima, i.e., states where all spins have negative energy and which, therefore, are stable against single spin flips, it is intuitive to use the number of spins with positive energy as a control parameter:

$$n^{\mathrm{p}}(\mathbf{S}) = \sum_{i=1}^{N} \Theta(e_i), \qquad (17)$$

where Θ is the Heaviside step function. However, simply minimizing this parameter would yield only local minima around the maximum of the minima distribution, but not in its tails. In order to sample the rare minima at high and at low energy, we also incorporate the Boltzmann weight. In the ensemble k of our simulation a state $\mathbf{S}$ is occupied with a probability

$$P_k(\mathbf{S}) \propto \omega_k(n^{\mathrm{p}}(\mathbf{S})) e^{-\beta_k \mathcal{H}(\mathbf{S})}, \qquad (18)$$

where β_k drives the energy of the system similarly to the inverse temperature in a canonical ensemble and $\omega_k(m)$ is a weight function that for a given number m equals the inverse sum of the Boltzmann weights of all spin configurations that have m spins with positive energy:

$$\omega_k(m) = \left(\sum_{\mathbf{S}} \delta_{n^{\mathrm{p}}(\mathbf{S}),m} e^{-\beta_k \mathcal{H}(\mathbf{S})} \right)^{-1}, \qquad (19)$$

where the sum goes over all possible states. The weight function ω_k causes a flat distribution over $n^{\mathrm{p}}(\mathbf{S})$ similar to the weights of a multicanonical simulation [12,13] or the inverse density of states from the Wang-Landau method [14] leading to a flat histogram over energy. Since ω_k is *a priori* not known we determine it before the actual simulation using an iterative procedure [15]. During the simulation a proposed step $\mathbf{S}_{\mathrm{old}} \to \mathbf{S}_{\mathrm{new}}$ is accepted with a probability according to the well-known Metropolis criterion:

$$p_{\mathrm{flip}}^k(\mathbf{S}_{\mathrm{old}}, \mathbf{S}_{\mathrm{new}}) = \min\left(1, \frac{P_k(\mathbf{S}_{\mathrm{new}})}{P_k(\mathbf{S}_{\mathrm{old}})}\right). \qquad (20)$$

Multiple such ensembles with different β are combined via the replica exchange method [16] and two ensembles k and l exchange configurations with the probability

$$p_{\mathrm{exch}}^{kl}(\mathbf{S}_k, \mathbf{S}_l) = \min\left(1, \frac{P_k(\mathbf{S}_l) P_l(\mathbf{S}_k)}{P_k(\mathbf{S}_k) P_l(\mathbf{S}_l)}\right), \qquad (21)$$

where $\mathbf{S}_k$ is the configuration belonging to ensemble k before the attempted exchange.

In order to estimate the distribution of local minima we apply the weighted-histogram analysis method (WHAM) [17,18]. It is possible to apply this algorithm directly to the various distributions of local minima measured at different β; however, since the data obtained at low n^{p} are carrying a large statistical error, we decided to determine the reweighting factors using all available data.

We first reweight in order to obtain the unnormalized canonical distributions

$$\tilde{\Pi}_k(E_i) = \sum_{\substack{t, \\ E_i - \epsilon < E(\mathbf{S}_{k,t}) < E_i + \epsilon}} \omega_k(n^{\mathrm{p}}(\mathbf{S}_{k,t}))^{-1}, \qquad (22)$$

where $\mathbf{S}_{k,t}$ are the configurations generated by the simulation in ensemble k and 2ϵ is the binning width $E_{i+1} = E_i + 2\epsilon$, and normalize

$$\Pi_k(E_i) = \frac{\tilde{\Pi}_k(E_i)}{\sum_j \tilde{\Pi}_k(E_j)} \approx \frac{G(E_i)e^{-\beta_k E_i}}{z_k}. \qquad (23)$$

Here, $G(E)$ denotes the density of states and z_k are the partition sums

$$z_k = \sum_i G(E_i)e^{-\beta_k E_i}, \qquad (24)$$

which can be self-consistently determined by iterating

$$z_k = \sum_i e^{-\beta_k E_i} \frac{\sum_l \Pi_l(E_i)}{\sum_l z_l^{-1} e^{-\beta_l E_i}}. \qquad (25)$$

One could now calculate the density of states

$$G(E_i) = \frac{\sum_l \Pi_l(E_i)}{\sum_l z_l^{-1} e^{-\beta_l E_i}}, \qquad (26)$$

or using the β-dependent distributions of the local minima,

$$\Pi_k^0(E_i) = \frac{\sum_{t, E_i - \epsilon < E(\mathbf{S}_{k,t}) < E_i + \epsilon} \delta_{n^{\mathrm{p}}(\mathbf{S}_{k,t}),0} \omega_k(0)^{-1}}{\sum_j \tilde{\Pi}_k(E_j)}, \qquad (27)$$

derive the overall distribution of local minima,

$$\Omega(E_i) = \frac{\sum_l \Pi_l^0(E_i)}{\sum_l z_l^{-1} e^{-\beta_l E_i}}. \qquad (28)$$

B. Method II

Basic concept. The aim of this method is to create an ensemble that contains all metastable states—and only those—with equal probability and to enable transitions between them, such that in a second step standard Monte Carlo techniques can be employed in order to investigate their properties.

As a first step, we set up a composite state that contains an unspecified spin configuration $\mathbf{S}$ and jN random numbers $\{\xi\} \in (0,1]^{jN}$ [19]. Here j specifies how many random numbers per lattice site are used. If basic Monte Carlo steps like spin flips and randomizations of elements of $\{\xi\}$ are applied to this state in an unbiased fashion, the system will perform simple sampling of the state space of spin configurations and simultaneously of $(0,1]^{jN} \subset \mathbb{R}^{jN}$.

Now we interpret the composite state as a random quench; i.e., the random numbers $\{\xi\}$ are used to create a sequence of spin configurations that starts at $\mathbf{S}$ and is guaranteed to end in a metastable state ρ. Applying the same Monte Carlo steps as before, changes in $\mathbf{S}$ and $\{\xi\}$ will, therefore, often cause changes of ρ such that a random walk in the space of local minima is performed. However, it cannot be expected that all metastable states are visited with equal probability.

We bias the ensemble such that a composite state is represented with a probability proportional to a weight $P_{\mathrm{goal}}(\mathbf{S},\{\xi\})$.

The function P_{goal} is chosen such that the resulting random walk in the space of local minima performs simple sampling; i.e., all metastable states ρ are occupied with uniform probability.

The principal ensemble. The initial spin configuration $\mathbf{S}$ and the set of random numbers $\{\xi\}$ is mapped onto a sequence of spin configurations by the primitive Monte Carlo method $\mathcal{M}$:

$$\mathcal{M} : (\mathbf{S},\{\xi\}) \mapsto (\sigma_0,\sigma_1,\sigma_2,\ldots,\sigma_f), \qquad (29)$$

where σ_i are spin configurations of the spin glass with $\sigma_0 \equiv \mathbf{S}$. In our case $\mathcal{M}$ stands for an energy minimization. The random numbers $\{\xi\}$ are used to randomly pick a spin with positive energy in σ_i, which is then flipped, thus creating σ_{i+1}. This is repeated until all spins have negative energy, such that the final state $\sigma_f \equiv \rho$ is a local energy minimum.

As usual, the primitive methods that are used to modify $\mathbf{S}$ and $\{\xi\}$ are unbiased; i.e., if simple sampling (ss) were used all $\mathbf{S}$ would be visited with equal probability and the random numbers $\{\xi\}$ would be uniformly distributed. In such a process the probability to obtain a given sequence $(\sigma_0,\ldots,\sigma_f)$ can easily be determined. Since $\mathcal{M}$ does an unbiased selection from all spins with positive energy, the probability of each individual draw equals the inverse numbers of spins with positive energy and the total probability is proportional to their product:

$$P_{\mathrm{ss}}((\sigma_0,\ldots,\sigma_f)) \equiv P_{\mathrm{ss}}(\mathcal{M}(\mathbf{S},\{\xi\})) = P(\mathbf{S}) \prod_{i=0}^{f-1} \frac{1}{n_i^{\mathrm{p}}}, \qquad (30)$$

where n_i^{p} is the number of spins with positive energy in the configuration σ_i. Using the inverse of this probability as the statistical weight of a state in a biased ensemble,

$$P_{\mathrm{is}}(\mathbf{S},\{\xi\}) \propto P_{\mathrm{ss}}(\mathcal{M}(\mathbf{S},\{\xi\}))^{-1}, \qquad (31)$$

during an importance-sampling (is) simulation will cause this simulation to create all sequences $(\sigma_0,\ldots,\sigma_f)$ with equal probability.

Of course, our goal is not to sample all sequences with equal probability, but all final states σ_f, which are the local energy minima. We have to assign an additional weight W to each sequence such that for all local minima ρ

$$p(\rho) = \sum_{\substack{(\sigma_0,\ldots,\sigma_f), \\ \sigma_f = \rho}} W((\sigma_0,\ldots,\sigma_f)) = \mathrm{const}. \qquad (32)$$

This is in a sense the inverse process to the first reweighting, where we introduced a weight function in order to move from a uniform distribution over the starting configurations $\mathbf{S}$ or σ_0 to a uniform distribution over all sequences. Now we wish to abandon the latter in favor of an ensemble with a uniform distribution over the final states σ_f.

Consider the partial tree depicted in Fig. 2. The full tree contains all those sequences that end in one particular final state, the local minimum σ_f. Each circle represents a spin configuration and any two connected states differ by exactly one spin value with the energy decreasing towards the root. The same configuration can appear multiple times in the tree. One possibility is that spins can be flipped in a different order which will cause states to appear more than once at the same level. We reconstruct the sequences in reverse order, i.e., starting
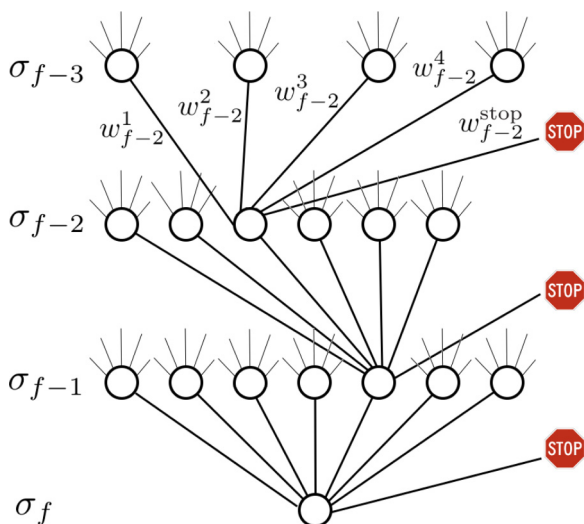
FIG. 2. Branches of the tree of sequences for one particular final state σ_f. Circles represent spin configurations and STOP nodes indicate that the initial configuration σ_0 has been identified with the node below. See text for a detailed description.

with the final state σ_f, and proceeding upwards to previous states. The length of the sequences is variable. Therefore, if all possible paths from the root σ_f should be considered, we have to accommodate the possibility of a premature stop while a continuation towards configurations of higher energy is still possible. This is symbolized by the STOP nodes. They are not configurations themselves, but identify their parent node with σ_0. If we assign weights w to all branches such that the sum over the weights of a node's outgoing (upward) branches equals unity

$$w_i^{\text{stop}} + \sum_{j=1}^{n_i^{\text{n}}} w_i^j = 1, \qquad (33)$$

then the product of these weights along each path has the desired property of W. Here, $n_i^{\text{n}} = N - n_i^{\text{p}}$ is the number of spins with negative energy in the configuration σ_i and, therefore, the number of spin configurations with higher energy to which σ_i can be connected. The most intuitive solution is to assign equal weights to all true continuations:

$$w_i^j = \frac{1 - w_i^{\text{stop}}}{n_i^{\text{n}}}, \qquad j = 1, 2, \dots, n_i^{\text{n}}. \qquad (34)$$

The remaining weights w^{stop} determine the length $f + 1$ of the sequence. We chose

$$w_i^{\text{stop}} = \begin{cases} 0, & \text{if } f + 1 - i < l_{\min}, \\ \frac{1}{l_{\max} - f + i}, & \text{else}, \end{cases} \qquad (35)$$

in order to obtain sequences of any length between (and including) $l_{\min}$ and $l_{\max}$ with equal frequency. If, for instance, $l_{\min} = 3$ and $l_{\max} = 7$ the STOP weights w_i^{stop} for the levels in Fig. 2 from the bottom to the top would read $0, 0, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1$. We find

$$W((\sigma_0, \dots, \sigma_f)) = w_0^{\text{stop}} \prod_{i=1}^{f} \frac{1 - w_i^{\text{stop}}}{n_i^{\text{n}}}, \qquad (36)$$

which with our choice of w_i^{stop} simplifies to

$$W((\sigma_0, \dots, \sigma_f)) = \frac{1}{l_{\max} - l_{\min} + 1} \prod_{i=1}^{f} \frac{1}{n_i^{\text{n}}} \qquad (37)$$

$$\propto \prod_{i=1}^{f} \frac{1}{n_i^{\text{n}}} \qquad (38)$$

if $f + 1 \in \{l_{\min}, \dots, l_{\max}\}$, otherwise $W((\sigma_0, \dots, \sigma_f)) = 0$.

Eventually, multiplying $P_{\text{is}}(\mathbf{S}, \{\xi\})$ and $W(\mathbf{S}, \{\xi\})$, we are left with an ensemble which includes all final states σ_f with equal probability

$$P_{\text{goal}}(\mathbf{S}, \{\xi\}) := \frac{\prod_{i=0}^{f-1} n_i^{\text{p}}}{\prod_{i=1}^{f} n_i^{\text{n}}}. \qquad (39)$$

The freedom to restrict the length of the sequence from above is important since during the construction of W we implicitly assumed that we can always move to states with higher energy and could, therefore, always chose $w^{\text{stop}} \neq 1$. To ensure that this assumption is justified during a simulation the sequences must not be too long. In our simulations we chose $l_{\max} = N/3$ and $l_{\min} = 1$.

The minimization. As stated above the minimization method $\mathcal{M}$ randomly chooses spins with positive energy and flips these until a stable configuration is reached. Normally, in such a procedure a spin would be selected by considering all candidates and using a single random number uniformly distributed between zero and the number of spins with positive energy, which rounded up will determine which spin to flip. However, this kind of global selection is unsuited in our case. It would cause a modification of the initial state σ_0 to potentially affect every single selection which would, therefore, require a complete reconstruction of the σ_i. The changes to σ_f would be considerable, which is undesirable in a Monte Carlo simulation, since the resulting acceptance rates would be very small. Instead, we implement strictly local conditions which collectively effect a global selection.

In the following as we construct the sequence $\sigma_0, \sigma_1, \dots, \sigma_f$ we will record the evolution of single spins by means of "spin states" ζ which describe a spin's value and its energy as determined by the environment, i.e., the adjacent spins. The initial state $\zeta_{0,k}$ of spin s_k as given by the spin configuration σ_0 will change to a new state $\zeta_{1,k}$ as soon as its energy changes, which happens if either spin s_k or one of its neighboring spins get flipped. Since consecutive σ_i only differ in exactly one spin, the first index of the ζ will in general not agree with the spin configurations to which they belong. For instance, most states $\zeta_{0,k}$ will be shared by σ_0 and σ_1.

If we assign a uniformly distributed random number $\eta_{0,i} \in (0, 1]$ to each initial spin state $\zeta_{0,i}$ with positive energy and sort the spins based on the magnitude of $\eta_{0,i}$ it is clear that we will obtain a completely random sequence. We identify the largest η_0 and flip the associated spin, thus creating σ_1. We then proceed with the new largest η. The following rules apply:

(1) If by the flip of a spin with random number $\eta_{\mu,k}$ an adjacent previously stable spin acquires positive energy in its new state $\zeta_{\nu,l}$, it can easily be inserted into the ordered set of spins with positive energy by choosing the random number $\eta_{\nu,l} \in (0, \eta_{\mu,k}]$.

(2) Albeit not strictly necessary, a new random number is also assigned in the same way if the energy of an unstable spin changes but remains positive.

(3) If a spin with positive energy changes to a stable state, i.e., with negative energy, its η is removed from the ordered set.

(4) The nonadjacent spins that retain their energy during a spin flip and whose state therefore does not change keep their random number.

For reasons of efficiency, in our simulation we reserve one random number $\xi \in (0,1]$ for each spin state regardless of the sign of its energy and calculate η from it when required. Random numbers of spins with negative energy have no impact and can be considered as temporarily decoupled degrees of freedom.

We can reformulate the algorithm by introducing a flipping "time"

$$t_{\mu,k} := -\ln \eta_{\mu,k} \tag{40}$$

and consequently [19]

$$t_{\mu,k} = t_{\mu,k}^{\text{orig}} - \ln \xi_{\mu,k}, \tag{41}$$

where $\xi_{\mu,k}$ is a uniformly distributed random number $\in (0,1]$ and $t_{\mu,k}^{\text{orig}}$ is the time at which the particular state of spin s_k was created, i.e., the time when a flip of one of its neighbors last changed its energy. A flip of a spin will not lead to a new flipping time for itself since after a flip its energy is per definition negative. In the beginning, i.e., for the state σ_0, all $t_{0,k}^{\text{orig}} = 0$. Although not used in our work, it is instructive to consider a biased selection. If the spin s_k in state $\zeta_{\mu,k}$ shall be selected with the relative weight $v_{\mu,k}$ it is relatively easy to show that

$$t_{\mu,k} = t_{\mu,k}^{\text{orig}} - \frac{\ln \xi_{\mu,k}}{v_{\mu,k}} \tag{42}$$

generates the desired behavior. If weights proportional to the Boltzmann weight of the respective energy change were chosen and if energy increases were allowed $\mathcal{M}$ would become the waiting-time Monte Carlo method [20].

We will now briefly discuss the different methods that we use in order to modify the sequence $\{\sigma\}$ during our Monte Carlo simulation.

Top-down update. If instead of the spin with the largest random number η (or the smallest flipping time t) the spin with the highest energy were flipped, the thus derived deterministic method would constitute a so-called greedy algorithm. Both methods are structurally very similar; each flip only affects adjacent spins and each spin state is characterized by a quantity (energy or η) whose maximum determines the next step. There is one difference that makes the random minimization easier to handle. During the greedy algorithm within the sequence of flipped spins the energy sometimes increases, while in the method used here η will always decrease (t will always increase). In a recent article [9] about "dynamical greedy algorithms" for the Edwards-Anderson model we have discussed algorithms that will propagate changes in the initial configuration σ_0 and determine the new σ_f with little computational effort. Since the same ideas are used here with

very little modification to implement a method that allows for changes to **S**, i.e., to σ_0, we refer to this publication for details.

Since random changes to the initial configuration will in general change the weight W_{goal} the acceptance probability must contain the ratio

$$R = \frac{P_{\text{goal}}(\mathbf{S}_{\text{new}}, \{\xi_{\text{new}}\})}{P_{\text{goal}}(\mathbf{S}_{\text{old}}, \{\xi_{\text{old}}\})} \tag{43}$$

as a bias correction.

Single random number update. Besides the starting configuration we can and should also modify the random numbers ξ during the simulation. All random numbers that belong to spin states with negative energy can be updated at leisure and those which belong to spin states $\zeta_{\mu,k}$ with positive energy, but which do not lead to a spin flip since the state is replaced before the flipping time,

$$t_{\mu,k} > t_{\mu+1,k}^{\text{orig}}, \tag{44}$$

can be assigned a new uniformly distributed random number

$$\xi_{\mu,k} \in \left(0, \exp\left(-t_{\mu+1,k}^{\text{orig}}\right)\right]. \tag{45}$$

In both cases the sequence of spin flips and therefore the weight W_{goal} remains unchanged. Hence, these updates can always be accepted with probability one. It is, of course, possible to modify any ξ without constraint and proceed to determine the resulting possibly altered sequence $\sigma_0, \ldots, \sigma_j, \sigma'_{j+1}, \ldots, \sigma'_f$. Then a nontrivial acceptance probability would ensue. In our simulations, however, we do not use such a method.

Bottom-up update. The distribution of sequences as defined by (34) and (35) enables us to introduce another update. Exploiting the fact that all sequence lengths between l_{min} and l_{max} are equally likely just as any upward continuation in Fig. 2, it is possible to create a new sequence that ends with the old local minima σ_f:

(1) Chose a new length $f' + 1$ from the allowed values randomly.

(2) Starting from $\sigma'_{f'} = \sigma_f$ create $\sigma'_{f'-1}, \sigma'_{f'-2}, \ldots, \sigma'_0$ by randomly flipping spins with negative energy.

(3) Assign new random numbers $\{\xi\}$ that are consistent with this sequence.

The third point warrants a more thorough discussion. Naturally, we start with σ'_0 since the random numbers η'_0 do not depend on the latter ones. If σ'_0 possesses n_0^{p} spins with positive energy and if σ'_1 is reached by flipping s_k we have to assign $\eta_{0,k}$ according to a distribution that equals the distribution of $\max(\chi_1, \chi_2, \ldots, \chi_{n_0^{\text{p}}})$, where the random numbers $\chi_i \in (0,1]$ are uniformly distributed. We find the distribution

$$p(\eta_{0,k}) = n_0^{\text{p}}(\eta_{0,k})^{n_0^{\text{p}}-1}, \quad \eta_{0,k} \in (0,1], \tag{46}$$

which means that we can set

$$\eta_{0,k} = \chi^{1/n_0^{\text{p}}}, \tag{47}$$

where $\chi \in (0,1]$ is uniformly distributed. Similarly, if the spin flip from σ'_{i-1} to σ'_i occurred at time τ_i and if the spin s_l has to be flipped in order to reach the new state $\zeta_{\mu+1,l}$ and σ'_{i+1},

$\eta_{\mu,l}$ is distributed the same way as $\max(\chi'_1, \chi'_2, \ldots, \chi'_{n_i^p})$, with $\chi'_i \in (0, e^{-\tau_i})$ uniformly distributed. Hence

$$p(\eta_{\mu,l}) = n_i^p (\eta_{\mu,l})^{n_i^p - 1} (e^{\tau_i})^{n^p}, \quad \eta_{\mu,l} \in (0, e^{-\tau_i}), \quad (48)$$

and therefore we can calculate $\eta_{\mu,l}$ from a uniformly distributed random number:

$$\eta_{\mu,l} = \chi^{1/n_i^p} e^{-\tau_i}. \quad (49)$$

Once all times and respective random numbers η of the performed flips are defined, their basic random numbers ξ can be calculated for the known times by inverting (41) and for the remaining spin states with positive energy according to (45). All other ξ are decoupled and may be kept or chosen at random. Since the update is designed to create sequences with the desired distribution there is no bias to correct by the acceptance probability:

$$R = 1. \quad (50)$$

It is worth noting that in principle this update allows for a true Markovian chain in the space of local minima. Without it, the selection of a new state ρ' does not exclusively depend on the current state ρ, but on the hidden degrees of freedom in **S** and $\{\xi\}$. A Markovian process is performed in their state space. Now, we can completely randomize these hidden degrees of freedom after each step using the bottom-up update, thus removing the surplus "memory" from the system. However, since the procedure involves the entire system and is, therefore, computationally expensive, it is not advisable to apply it that often.

In our simulations we randomly select a spin in **S** and attempt a spin flip in a top-down update. Then, the random variables at this lattice site are updated if possible. After N such combinations we perform a single bottom-up update.

Simulation. Once the framework introduced above is in place we can ignore all its inner workings and treat it as a normal system which ergodically changes from one single-flip stable configuration $\rho \equiv \sigma_f$ to another. Of course, the configurations of this particular system are a subset of the states of another system, but this concerns us no longer. In order to obtain statistics for a large energy range we apply a flat-histogram method. We introduce another weight function

$$W_{\text{flat}}(E) \approx \Omega(E)^{-1} \quad (51)$$

and require that in our simulation the probability to visit a certain metastable state ρ is

$$P_{\text{flat}}(\rho) \propto W_{\text{flat}}(E(\rho)), \quad (52)$$

which means that new states are accepted with the probability

$$P_{\text{flat}}^{\text{acc}}(\rho_{\text{old}} \to \rho_{\text{new}}) = \min\left(1, \frac{W_{\text{flat}}(E(\rho_{\text{new}}))}{W_{\text{flat}}(E(\rho_{\text{old}}))} R\right). \quad (53)$$

We initially approximate W_{flat} using a variant of the well-known Wang-Landau algorithm [14] with an additional restriction. The algorithm has difficulties in converging and in sampling the distribution in the extreme tails, i.e., at low and at high energy because only very few states exist there. The high-energy minima are much harder to find than the low-energy ones. We suspect the reason is that the latter are embedded in large basins and large metabasins which help to

TABLE I. Upper bounds for the weight function $W_{\text{flat}}(E)$.

L	Δ_{L}	Δ_{R}
4	∞	∞
6	∞	$0.18L^3$
8	$0.205L^3$	$0.17L^3$
10	$0.205L^3$	$0.13L^3$

guide the simulation. The problematic regions can be excluded from the simulation by restricting W_{flat}:

$$W_{\text{flat}}(E) < \begin{cases} \min(W_{\text{flat}}(E)) + \Delta_{\text{L}}, & \text{if } E < E^*, \\ \min(W_{\text{flat}}(E)) + \Delta_{\text{R}}, & \text{if } E > E^*, \end{cases} \quad (54)$$

where E^* is the position of the minimum of W_{flat}

$$W_{\text{flat}}(E^*) = \min(W_{\text{flat}}(E)). \quad (55)$$

This is more convenient than restricting the energy range directly because it can be applied in the same way to all samples. The values that we use are listed in Table I. Choosing $\Delta_{\text{L}} = 0.205N$ still allows for the sampling of the ground state, but prevents the algorithm from spending too much time at low energies during the weight determination.

Once $W_{\text{flat}}(E)$ is known with sufficiently high precision, we perform the main simulation and record a histogram $H(E)$ of the local minima from which their distribution can be calculated:

$$\Omega(E) = \frac{H(E)}{W_{\text{flat}}(E)}. \quad (56)$$

V. RESULTS

As a first goal we test the validity of our methods. In Fig. 3 we show $\Omega(E)$ for an $L = 4$ sample of the Edwards-Anderson
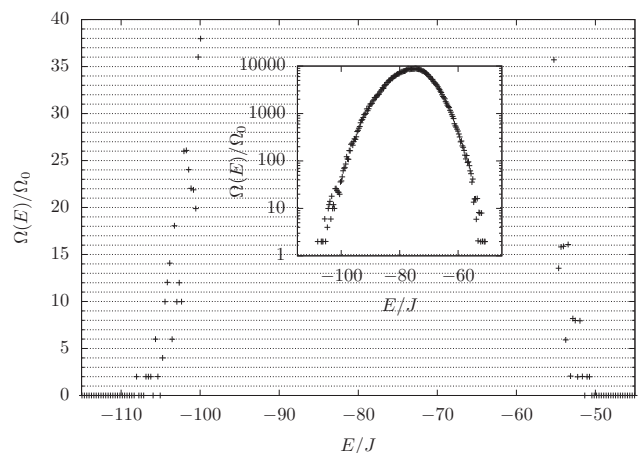

FIG. 3. The tails of the distribution of local minima for a single $4 \times 4 \times 4$ sample of the Edwards-Anderson model as measured with method II. The complete distribution is shown in the inset. The Ω values are integer multiples of the lowest nonzero value. The appropriate normalization Ω_0 is not obtained from the simulation, but is deliberately chosen such that the lowest occupied level equals 2 indicating that the respective energy intervals each contain a single twofold-degenerate metastable state.
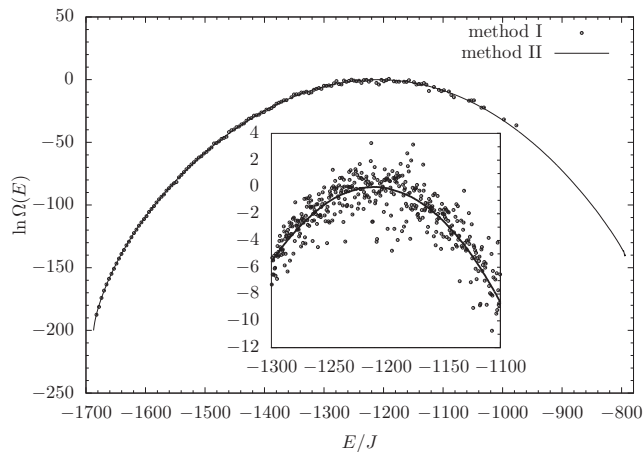
FIG. 4. The distribution of local minima for a single $10 \times 10 \times 10$ sample of the Edwards-Anderson model measured with both methods. In the main plot only every tenth data point for method I is displayed in order to ensure visibility. Notwithstanding, both data sets have a similar resolution. While the results are consistent, which suggests that both methods are accurate, the statistical error of method II is substantially lower (see inset). Besides, at high energy method I failed to find minima in a number of energy intervals leading to an apparent thinning out of data points.

model measured with method II. We use a binning method; i.e., every data point represents the aggregated statistics from a small energy interval. Each interval contains an integer number of local minima; hence a measurement of Ω should produce values that are integer multiples of the lowest nonzero value. This becomes clearly apparent in the tails of the distribution. We interpret the larger statistical fluctuations at high energy as evidence that sampling the high-energy minima is more demanding.

In order to compare the two methods we show results for a $10 \times 10 \times 10$ system in Fig. 4. We find that the results are in agreement. However, method I suffers from a much larger statistical error and for larger systems method II is able to cover a much larger energy range. Consequently, we proceed to apply method II if possible, i.e., for the Edwards-Anderson model.

A. Total number of minima

We note that the raw data from the simulations are *a priori* not normalized. Only for $L = 4$ and method II it is in principal possible to obtain a normalization constant since very precise measurements in the tails of the distribution are required. This was done "manually" for the distribution in Fig. 3. To automatize the normalization for small systems we use the lower values of the distribution of minima Ω to define a fitness function

$$\mathcal{F}(s) = \sum_{\substack{i, \\ 0 < \Omega(E_i) < S}} \frac{\cos\left[2\pi\,\Omega(E_i)/s\right]}{\Omega(E_i)}, \qquad (57)$$

which is intended to quantify how well the data match a supposed level distance s. Here, the parameter S is a suitably chosen upper threshold that limits the computational effort and increases precision while the division by $\Omega(E)$ is introduced

since we presume that $\ln \Omega(E)$ has a constant statistical error which implies that $\Omega(E)$ has a statistical error proportional to its absolute value and that larger values should, therefore, contribute less. The true level distance, i.e., the statistical contribution of a single (twofold degenerate) local minimum, is then estimated by the position of the maximum of $\mathcal{F}(s)$ and dividing by 2 accounts for the degeneracy:

$$\Omega_0 = \arg\max_{s > s_{\min}} \mathcal{F}(s)/2, \qquad (58)$$

where we considered

$$s_{\min} = \frac{\min[\Omega(E_i)]}{10} \qquad (59)$$

sufficient. It is highly unlikely that for $L = 4$ and the interval width we used ($\Delta E = E_{i+1} - E_i = 0.3J$) the interval with the lowest population already contains more than 10 local minima. We can now properly normalize the distributions of local minima and determine the total number of minima

$$N_S = \sum_i^\infty \Omega(E_i)/\Omega_0, \qquad (60)$$

take the disorder average, and obtain an estimate for the number [21] of metastable states for the three-dimensional EA model with $L = 4$:

$$\ln\langle N_S \rangle_J / N = 0.2111(3). \qquad (61)$$

This result matches the value 0.21125(1) that was calculated in Ref. [6] using an analytic approximation.

B. Averaging

Unfortunately, for larger system this normalization method cannot be applied and all other distributions presented henceforth are only determined up to unknown factors. Since we calculate the average of the logarithm of the distributions, these become unknown additive constants which do not have a direct effect besides creating an unknown additive constant for the average as well. However, the second moments $\langle [\ln \Omega(E)]^2 \rangle_J$ and hence the estimators of the statistical errors of $\langle \ln \Omega(E) \rangle_J$ depend on these constants. In our analysis we chose them such that the maximum of the canonical distributions $\Omega(E)e^{-\beta_{\text{sync}}E}$ is identical for all samples. This is equivalent to the (not entirely valid) assumption that all samples have about the same number of local minima and leads to an underestimation of the statistical error. With increasing system size this effect will vanish. Since we obtain very precise data for the Edwards-Anderson model we can use the natural choice $\beta_{\text{sync}} = 0$. For the SK model, however, the data are very noisy around the maximum of $\Omega(E)$ and we use $\beta_{\text{sync}} = 0.05/J$ for $N = 96$ and $\beta_{\text{sync}} = 0.2/J$ for $N = 128$. The first averaging procedure is given by

$$[\ln \Omega(E)]_1 := \langle \ln \Omega(E) - \ln \Omega(E^*) \rangle_J, \qquad (62)$$

where E^* is the position of the maximum of $\Omega(E)e^{-\beta_{\text{sync}}E}$.

Due to the variability of the interactions J_{ij} the energy interval at which local minima exist shifts especially for small systems. This means that we can obtain data from all samples only from a relatively small energy region. Outside this interval no average can be computed since the logarithm of the missing

TABLE II. Parameters used for the simulation of the SK model.

N	β_{min}	β_{max}
48	-1.44	1.4
64	-1.4	1.4
96	-0.6	1.4
128	-0.4	1.4

distributions is not defined. To obtain an averaged function over a larger interval we introduce a second averaging procedure for which we shift all distributions along the energy axis such that their maxima coincide with the average maximum position:

$$[\ln \Omega(E)]_2 := \langle \ln \Omega(E + E^* - \langle E^* \rangle_J) - \ln \Omega(E^*) \rangle_J. \quad (63)$$

C. Sherrington-Kirkpatrick model

We simulate systems of size $N = 48, 64, 96, 128$ with method I using parameters according to Table II. For each size we investigated 200 samples.

Figure 5 shows the average logarithmic distribution of local minima for the first averaging method and indicates that our results are basically in agreement with the analytical prediction. Details are discernible in Fig. 6 where we show the deviation of the averages from the finite-system approximation $g_0(\varepsilon) + \frac{1}{z} g_1(\varepsilon)$ with $z = N - 1$. Since we do not have a valid normalization and cannot determine the correct vertical position of the curves in Fig. 5 the absolute differences $\frac{1}{N}[\ln \Omega(\varepsilon N J \sqrt{N-1})] - [g_0(\varepsilon) + \frac{1}{N-1} g_1(\varepsilon)]$ have no meaning. Only the relation between the differences is relevant; i.e., a horizontal curve means agreement with the analytical prediction while a large slope indicates deviation. Consequently, the vertical positions of the curves have been adjusted for convenience and are not the result of a physically motivated normalization. Error bars result from the disorder average and represent two standard deviations.

With increasing system size the range in energy ε where local minima exist expands. Regardless of the averaging tech-

FIG. 6. Deviation of the averaged distribution from the analytical prediction for the SK model using (a) averaging method 1 and (b) method 2. The vertical positions $y(N)$ correspond to the unknown normalization constants and have here been chosen to avoid overlapping curves.

nique used, the distribution $\Omega(\varepsilon N J \sqrt{N-1})$ reaches smaller and greater ε for $N = 64$ than for $N = 48$. However, for even larger systems the shortcomings of the Monte Carlo method and the increasing complexity of the energy landscape make it impossible to find minima of high energy. In fact the downward curve at high ε for $N = 96$ and $N = 128$ suggests that even for the energies where we can find minima large populations are not accessed. The alternative explanation, that the analytical prediction is not accurate, seems less likely, especially since we obtain nice horizontal curves for $N = 48$ and $N = 64$. At low energies the measured distributions divert from the analytical solution as predicted by Bray and Moore.
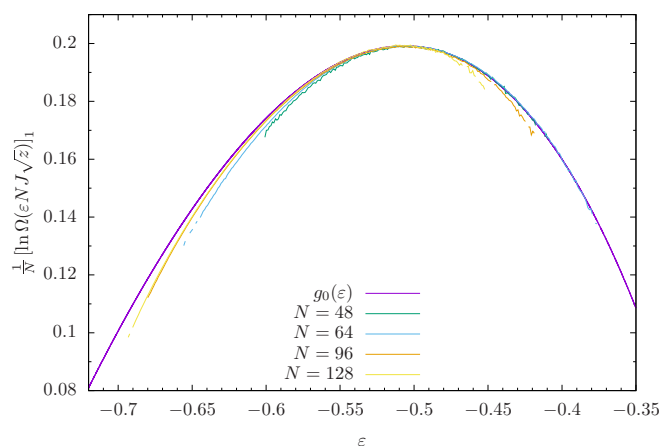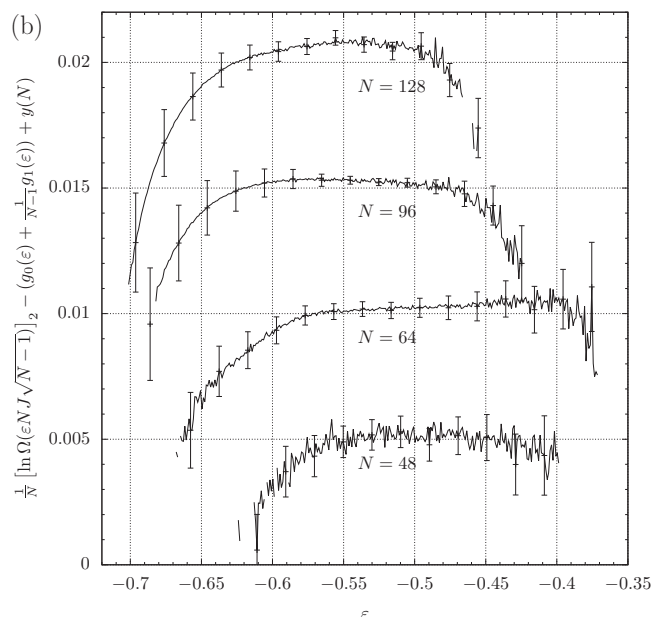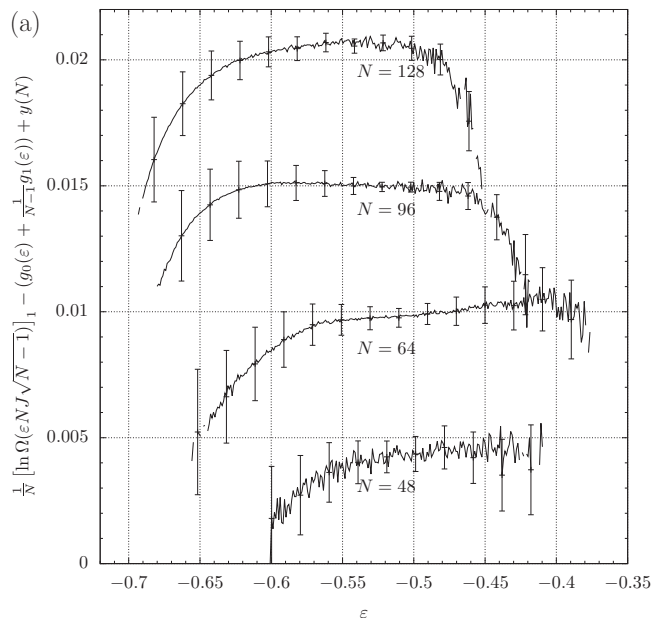

FIG. 5. The average 1 logarithmic density of local minima for the SK model. The vertical position of the curves, i.e., the normalization of Ω, has been chosen such that the maxima of the curves coincide.
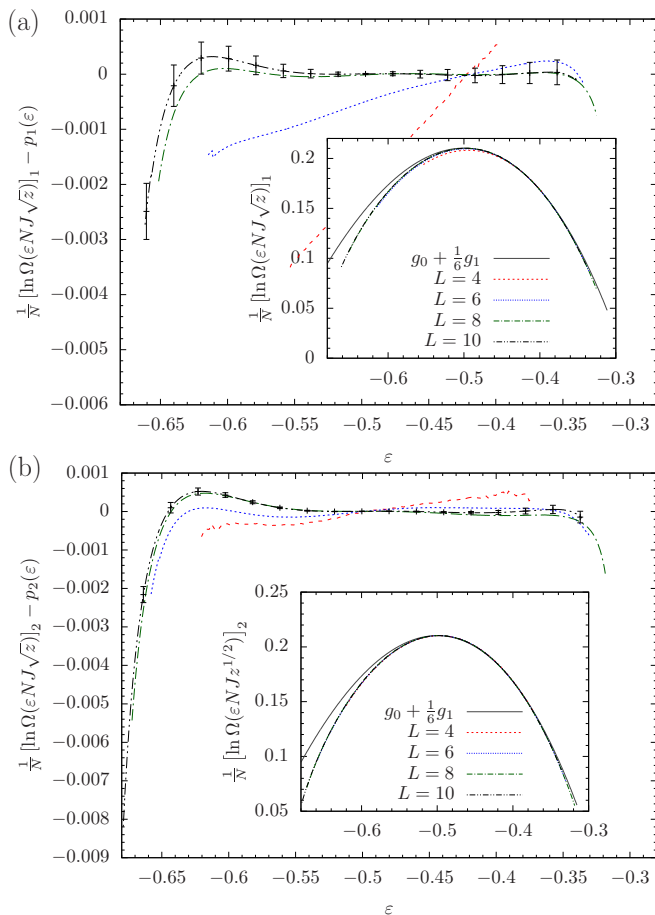
FIG. 7. The logarithmic density of local minima for the EA model with the sample average taken with (a) method 1 and (b) method 2. The main plots show the deviations from the polynomial approximations p_1 and p_2 (see text) in order to highlight size-dependent behavior. In the insets the distributions are plotted together with the analytic approximation by Bray and Moore.

The deviation is clearly visible for $\varepsilon < -0.6$. However, the errors are relatively large and it is difficult to judge whether the calculated $\varepsilon_c \approx -0.672$ will be realized for larger systems.

D. Edwards-Anderson model

For the investigation of the Edwards-Anderson model we are able to use method II and we obtain very precise results. We consider lattices of linear extension $L = 4, 6, 8, 10$ and investigated 1000 samples for each size. In the insets of Fig. 7 the distributions calculated with both averaging methods are plotted together with the approximation $g_0(\varepsilon) + \frac{1}{6} g_1(\varepsilon)$. We observe almost no dependence on the system size, except for the fact that the support becomes broader. Again, for large systems the sampling of local minima with high energies becomes very difficult, which led us to use the restrictions listed in Table I. Consequently, the rescaled energies reached

for $L = 10$ are not as high as for $L = 8$. While the data agree reasonably well with the analytical solution on the right flank of the distribution and the maximum is in a similar position, we see considerable deviations for energies below the peak. We find that the data are much better described by polynomials

$$p_1(\tilde{\varepsilon}) = -13.39\,\tilde{\varepsilon}^4 - 1.10\,\tilde{\varepsilon}^3 - 4.143\,\tilde{\varepsilon}^2 + \text{const.} \quad (64)$$

and

$$p_2(\tilde{\varepsilon}) = -13.73\,\tilde{\varepsilon}^4 - 1.07\,\tilde{\varepsilon}^3 - 4.142\,\tilde{\varepsilon}^2 + \text{const.}, \quad (65)$$

where $\tilde{\varepsilon}$ is a shifted energy such that $\tilde{\varepsilon} = 0$ at the maximum position for $L = 10$:

$$\tilde{\varepsilon} = \varepsilon - (-0.4978). \quad (66)$$

Both polynomials were obtained by fitting to the $L = 10$ averages for $\varepsilon \in [-0.55, \infty)$. Note that the contributions from the third- and fourth-order term are small and for $\varepsilon \in [-0.6, -0.4]$ the quadratic term alone provides a very good approximation. The deviations from these polynomials depicted in the main plots in Fig. 7 are very small and while with the first averaging method a clear dependence on the system size emerges, the curves for $L > 4$ for the second averaging method are much closer together. From the statistical errors which are shown for the $L = 10$ curves it also becomes clear that the second averaging method produces more precise results. We expect that both techniques would deliver the same curves for very large systems, hence we conclude that $p_2(\varepsilon)$ provides a better approximation of the true distribution of minima for large systems.

VI. CONCLUSION

In order to measure the distribution of metastable states of spin glasses we have introduced two different algorithms. The first method employs traditional Monte Carlo techniques such as flat-histogram, replica-exchange, and weighted-histogram analysis. This method is able to find local minima over a wide range in energy; however, it has great difficulties at high energy and the statistical errors are comparatively large. For the second approach we designed an ensemble that allows a direct uniform sampling of metastable states. This method can efficiently be applied for the Edwards-Anderson model and yields very precise results.

We find that for the Sherrington-Kirkpatrick model our results are consistent with the analytical predictions. Unfortunately, statistical errors are substantial and our simulations were only able to access the whole energy range for small system sizes.

Investigating the Edwards-Anderson model by means of our second method we were able to measure the distribution of metastable states with great precision. We found that the results—suitably rescaled—show very little dependence on the system size and we are, therefore, confident that also for much larger systems distributions very similar to the ones we describe would be obtained.

[1] A. J. Bray and M. A. Moore, J. Phys. C: Solid State Phys. **14**, 1313 (1981).

[2] T. Li, Phys. Rev. B **24**, 6579 (1981).

[3] C. M. Newman and D. L. Stein, Phys. Rev. E **60**, 5244 (1999).

[4] Z. Burda, A. Krzywicki, O. C. Martin, and Z. Tabor, Phys. Rev. E **73**, 036110 (2006).

[5] Z. Burda, A. Krzywicki, and O. C. Martin, Phys. Rev. E **76**, 051107 (2007).

[6] B. Waclaw and Z. Burda, Phys. Rev. E **77**, 041114 (2008).

[7] M. Baity-Jesi and G. Parisi, Phys. Rev. B **91**, 134203 (2015).

[8] A. Heuer, J. Phys.: Condens. Matter **20**, 373101 (2008).

[9] S. Schnabel and W. Janke, Comput. Phys. Commun. **220**, 74 (2017).

[10] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).

[11] S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).

[12] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).

[13] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).

[14] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[15] W. Janke, in *Computational Many-Particle Physics*, Lecture Notes in Physics, edited by H. Fehske, R. Schneider, and A. Weiße (Springer, Berlin, 2008), Vol. 739, pp. 79–140.

[16] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[17] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1195 (1989).

[18] S. Kumar, J. Rosenberg, D. Bouzida, R. H. Swendsen, and P. Kollman, J. Comput. Chem. **13**, 1011 (1992).

[19] Since we need to take the logarithm of the random numbers ξ, we exclude zero from the interval of possible values.

[20] J. Dall and P. Sibani, Comput. Phys. Commun. **141**, 260 (2001).

[21] Here, we have used the logarithm of the average of N_S in order to be able to compare our result. However, since both N_S and $\Omega(E)$ are expected to be log-normal distributed, in the rest of the paper we consider averages of logarithmic quantities.