Dynamic variational study of chaos: spin glasses in three dimensions

A Billoire\textsuperscript{1}, L A Fernandez\textsuperscript{2,3}, A Maiorano\textsuperscript{3,4}, E Marinari\textsuperscript{4,5,6}, V Martin-Mayor\textsuperscript{2,3}, J Moreno-Gordo\textsuperscript{3,7,9}, G Parisi\textsuperscript{4,5,6}, F Ricci-Tersenghi\textsuperscript{4,5,6} and J J Ruiz-Lorenzo\textsuperscript{3,8}

\textsuperscript{1} Institut de Physique Théorique, CEA Saclay and CNRS, 91191 Gif-sur-Yvette, France
\textsuperscript{2} Departamento de Física Teórica I. Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Madrid 28040, Spain
\textsuperscript{3} Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), 50018 Zaragoza, Spain
\textsuperscript{4} Dipartimento di Fisica, Sapienza Università di Roma, I-00185 Rome, Italy
\textsuperscript{5} Nanotec, Consiglio Nazionale delle Ricerche, I-00185 Rome, Italy
\textsuperscript{6} Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1, I-00185 Rome, Italy
\textsuperscript{7} Departamento de Física Teórica, Universidad de Zaragoza, 50009 Zaragoza, Spain
\textsuperscript{8} Departamento de Física and Instituto de Computación Científica Avanzada (ICCAEx), Universidad de Extremadura, 06006 Badajoz, Spain
E-mail: jmorenogordo@gmail.com

Received 27 September 2017
Accepted for publication 12 December 2017
Published 2 March 2018

Abstract. We have introduced a variational method to improve the computation of integrated correlation times in parallel tempering dynamics, obtaining a better estimate (a lower bound, at least) of the exponential correlation time. Using this determination of the correlation times, we revisited the problem of the characterization of the chaos in temperature in finite dimensional spin glasses, by way of the study of correlations between various chaos indicators computed in the static and the correlation times of the parallel tempering dynamics. The sample-distribution of the characteristic time for the parallel tempering dynamics turns out to be fat-tailed, and to obey finite-size scaling.

Keywords: spin glasses, slow relaxation, glassy dynamics, aging, extreme value statistics

\textsuperscript{9} Author to whom any correspondence should be addressed.
1. Introduction

Markov-chain Monte Carlo methods are a crucial tool to study non-perturbative problems in statistical mechanics and quantum field theory [1–4]. A major problem arises, however, when studying systems with rugged free-energy landscapes: we have in mind, for example, spin glasses [5, 6], or glass-forming liquids [7]). The presence of many free-energy local minima often causes the numerical simulation to get trapped and, as a consequence, does not allow a correct sampling of the phase space.

The parallel tempering method—connected to the original simulated tempering method [8], and also known as the replica exchange method—was devised to overcome these difficulties [9–11]. One considers $N_T$ copies (or clones) of the system, and uses for each of them a different temperature $T_i$, with $T_1 < T_2 < \cdots < T_{N_T}$. As explained in appendix E, the target probability distribution for the $N_T$ systems is the product of the Boltzmann distributions at the various temperatures. A parallel tempering numerical simulation is based on two alternating sets of steps. First, each system copy independently undergoes standard Monte Carlo dynamics (for example Metropolis) at its own temperature. Then, a swap between two copies is attempted. The swap is accepted with a probability that depends on the difference in energy and temperature between the two systems. This process allows the system to explore the phase space more efficiently and helps it escape from local minima.
temperature: one can use one or more Monte Carlo steps each time. Second, pairs of spin configurations attempt to exchange their temperatures\footnote{The temperature-exchange rule is designed to have the target probability distribution as the unique equilibrium measure. In other words, the restriction of the total measure to a single temperature is exactly the appropriate Boltzmann distribution at that temperature—see appendix E.}

The rationale behind parallel tempering is simple. Each system copy undergoes a random walk in temperature space. When a system copy is at a low temperature, it only explores the nearby free-energy local minima. When its temperature is high, however, free-energy barriers disappear: the copy can freely wander in phase space, and when it cools again it will typically fall in a different free energy valley, with different local minima. For parallel tempering to effectively thermalize, it is crucial that any copy of the system spends its time roughly evenly at every temperature: high temperatures are needed to ensure visiting all the phase space; low temperatures are needed to visit its low free energy regions. In fact, parallel tempering is currently used in a very large number of very different applications (for example in physics, biology, chemistry, engineering, statistics), and considerable efforts have been devoted to improving it, from various communities. Various temperature-exchange rules have been developed and tested\cite{12–17}. Furthermore, it has been suggested that a significant gain can be achieved by optimizing the choice of the $N_T$ temperatures\cite{18, 19}.

In order to assess the relative merits of the above suggestions, one needs a quantitative method. The theory of Markov chains suggests considering the \textit{exponential auto-correlation time} $\tau_{\text{exp}}$ of the Monte Carlo dynamics as a relevant figure of merit\cite{2}. $\tau_{\text{exp}}$ tells us how long we should wait before equilibrium is reached. Unfortunately, $\tau_{\text{exp}}$ is an elusive quantity. In the context of a parallel tempering simulation, it has been suggested that $\tau_{\text{exp}}$ is best computed by studying the temperature-flow of the system copies\cite{20, 21}: the exchange of temperatures is, indeed, the slow mode of the combined numerical simulation based on parallel tempering and Metropolis moves, and it is an interesting process to quantify. We will focus here on the determination of $\tau_{\text{exp}}$ for a parallel tempering simulation of a spin glass. Our choice entails no generality loss, because the problem of finding the ground state (or low-temperature configurations) in a spin glass is NP-complete\cite{22}: understanding it sheds light on a large class of very interesting phenomena. Furthermore, spin glasses show very clearly the major problems that a parallel tempering simulation faces.

To be specific, we shall be considering the three dimensional Edwards–Anderson model\cite{23, 24}. Ising variables ($s_i = \pm 1$) occupy the nodes of a cubic lattice of size $L$ with periodic boundary conditions. Spins interact with their nearest lattice-neighbors through the Hamiltonian

$$H = - \sum_{\langle i,j \rangle} J_{ij}s_is_j ,$$

where the quenched couplings $J_{ij}$ are drawn from a bimodal probability distribution (so that $J_{ij} = \pm 1$ with $1/2$ probability) at the beginning of the simulation. A choice of couplings $\{ J_{ij} \}$ will be called a (disorder) sample (or realization) hereafter.

A major complication in the numerical study of the Hamiltonian (1) is that a large number of samples of the system (the larger, the better) needs to be studied due to the
non-self-averaging property of the system\textsuperscript{11}. Besides, below the critical temperature $T_c$, the value of $\tau_{\text{exp}}$ (i.e. the computational difficulty that characterizes the physical system) presents huge sample to sample fluctuations [20, 21] (see also figure B1). The presence of these fluctuations makes the problem of computing $\tau_{\text{exp}}$ very relevant for saving CPU time (allowing, in this way, larger and more accurate simulations for a given cost): knowing the value of $\tau_{\text{exp}}$ for each individual sample makes it possible to save a huge amount of computer time by setting the chain length for a given sample proportional to its own $\tau_{\text{exp}}$.

There is a fairly general physical mechanism behind the dramatic dispersion of $\tau_{\text{exp}}$ (and behind its severe growth with the system size)—the so called temperature chaos [25–42]. Temperature chaos consists of a major reorganization of the typical equilibrium configurations upon tiny temperature changes. A detailed inspection shows how the effect arises on finite systems [39, 42, 43]. Indeed, for some samples, one encounters chaotic events taking place at well defined temperatures, in the form of major changes of the spin configurations as the temperature is lowered. Chaotic events are reminiscent of first-order phase transitions (rounded in a finite system). In a fixed temperature interval, $T_A < T < T_B$ with $T_B < T_c$, a given sample may undergo zero, one or even more chaotic events (the temperature location of the chaotic events is also random). Given $T_A < T < T_B$, the larger the system, the larger is the probability of finding samples displaying chaotic events in that temperature region [39]. Lowering $T_A$ while keeping the size fixed also increases the probability of encountering a chaotic event.

As is intuitively obvious, temperature chaos turns out to be a major obstacle for the parallel tempering temperature flow [21, 39, 42, 43]. The main point is that equilibrium in parallel tempering implies equilibrium at all temperatures. Now, let us assume that the typical equilibrium spin-configurations at two neighboring temperatures in the temperature grid are vastly different. Clearly, if one spin configuration of the low-temperature type is momentarily placed at the high temperature, it will have a hard time traveling to the highest temperatures in the temperature grid (because the clones at the higher temperatures are fitter, and the local spin-flip dynamics is obviously inefficient to remediate this problem). Furthermore, temperature chaos is relevant in the analysis of crucial experimental results [44–51], and in the performance analysis of commercial quantum annealers [43, 52, 53].

Here, we revisit the problem of estimating $\tau_{\text{exp}}$, and present a variational method that can potentially save a large amount of computation time. Very often, a numerical simulation needs to be extended just because of the difficulties encountered in the computation of $\tau_{\text{exp}}$. Having in our hands a safe mechanism to estimate $\tau_{\text{exp}}$ in an automated way (the number of samples needed in a state-of-the-art numerical simulation goes by the thousands) can avoid unnecessary extensions of the simulation length. We also investigate further the relationship between temperature chaos, which is a static equilibrium feature, and $\tau_{\text{exp}}$, which characterizes a Markov chain dynamics.

This paper is organized as follows. In section 2, we introduce two different time scales that characterize a Monte Carlo Markov chain. Our simulations are described in section 3. We present our characterization of temperature chaos in section 4. The variational method for the computation of the autocorrelation time $\tau_{\text{exp}}$ is discussed in

\textsuperscript{11}Strictly speaking, non-self-averaging occurs only when the correlation length reaches the order of magnitude of the system size (which is usually the case at the temperatures of interest).
section 4.1. Section 4.2 is devoted to the study of the scaling properties of the parallel tempering method \( \tau_{\text{exp}} \). We study in very precise detail the thermodynamic equilibrium features that characterize temperature chaos [39] in section 4.3. The relationship between static and dynamic chaos indicators is studied in section 4.4. Our discussion of results concludes the paper in section 5. We provide in appendix A a description of the parameters of our simulations. In appendix B, we discuss the particular choice of samples we use [54]. In appendix C, we describe in detail the geometry used in our implementation of the synchronous multispin coding. In appendix D, we discuss some quantities which are surprisingly unrelated to chaos. Finally, in appendix E, we discuss in some detail the relationship between time-correlations and system equilibration.

2. Time scales in a Markov chain

This section is a quick reminder of some basic concepts. The interested reader is referred to [2] for further details. Specific examples and computational recipes will be discussed in section 4.1 (see also appendix E).

Almost all the Monte Carlo methods used in statistical physics are based on the theory of Markov chains. A Markov chain starts from some initial configuration and we need to know how long the Markov dynamics must be run in order to reach equilibrium. This time scale is the exponential autocorrelation time (\( \tau_{\text{exp}} \)). In addition to this time scale, we can define for any physical quantity \( f \) a second time scale: the integrated autocorrelation time (\( \tau_{\text{int},f} \)). This controls statistical errors in measuring \( f \): two already equilibrated configurations whose time difference is \( 2\tau_{\text{int},f} \) are statistically independent in an effective sense (but only as far as the quantity \( f \) is concerned).

Under very mild assumptions (see below) it is possible to show that the following inequality holds for any \( f \):

\[
\tau_{\text{int},f} \leq \tau_{\text{exp}}.
\]

The crucial point is that \( \tau_{\text{int},f} \) is relatively easy to compute. \( \tau_{\text{exp}} \), on the other hand, is rather elusive. Hence, we shall use equation (2) for a variational method analogous to the Rayleigh–Ritz variational principle in quantum mechanics. In section 4.1, we shall try different quantities \( f \) and compute \( \tau_{\text{int},f} \) for each of them. The largest value of \( \tau_{\text{int},f} \) will be our variational estimate for \( \tau_{\text{exp}} \).

Let us recall that the equilibrium autocorrelation function for quantity \( f \) is

\[
C_f(t) = E[f(t_1)f(t_2)] - E[f(t_1)]^2, \quad t = t_1 - t_2,
\]

where \( E[\ldots] \) stands for the expectation value and the two times \( t_1 \) and \( t_2 \) are large enough to reach equilibrium (hence \( E[f(t_1)] = E[f(t_2)] \) and \( C_f(t) = C_f(-t) \)). The integrated autocorrelation time is defined from the normalized correlation function \( \hat{C}_f(t) \):

\[
\hat{C}_f(t) \equiv \frac{C_f(t)}{C_f(0)}, \quad \tau_{\text{int},f} = \frac{1}{2} + \sum_{t=1}^{\infty} \hat{C}_f(t).
\]
The normalized autocorrelation function can be expressed in terms of the eigenvalues \( \lambda_n \) of the transition probability matrix projected onto the subspace orthogonal to its eigenvector of eigenvalue 1 (\( 1 > |\lambda_1| \geq |\lambda_2| \geq \ldots \)) (see [2]):

\[
\hat{C}_f(t) = \sum_n A_{n,f} \lambda_n^{[t]}, \quad \sum_n A_{n,f} = 1,
\]

(5)

where the index \( n \) runs from 1 to \( N_T!2^{N_TLD} - 1 \), in our case.

The amplitudes \( A_{n,f} \) depend on \( f \), while the \( \lambda_n \) are \( f \)-independent. In terms of \( A_{n,f} \) and \( \lambda_n \), one has

\[
\tau_{\text{int},f} = \frac{1}{2} + \sum_n A_{n,f} \frac{\lambda_n}{1 - \lambda_n}.
\]

(6)

Now, in practical applications the (leading) \( A_{n,f} \) and \( \lambda_n \) values are real positive. Hence, \( \lambda_n = e^{-1/\tau_n} \) defines the characteristic time \( \tau_n \). The exponential autocorrelation time of the Markov chain \( \tau_{\text{exp}} \) is just \( \tau_n \), the largest of the \( \tau_n \). Now, for \( \tau_n \gg 1 \), one has \( \lambda_n/(1 - \lambda_n) = \tau_n + \mathcal{O}(1/\tau_n) \) and equations (5) and (6) become

\[
\hat{C}_f(t) = \sum_n A_{n,f} e^{-|t|/\tau_n}, \quad \tau_{\text{int},f} = \frac{1}{2} + \sum_n A_{n,f} \tau_n.
\]

(7)

The variational method in equation (2) follows immediately from equation (7). The optimal choice for the observable \( f \) would have \( A_{1,f} = 1 \) and \( A_{n>1,f} = 0 \) in its decomposition in characteristic times.

3. Numerical simulations

We develop our study in the context of [54], in which the metastate was studied. For this reason, our realizations of disorder \( \{ J_{ij} \} \) (samples) are particular. In appendices A and B, we explain how the samples have been chosen, and argue that this choice does not affect the results.

We have simulated this model using the parallel tempering method with Metropolis updates. See appendices A and B for the reasons behind our choice of the minimal temperature in the parallel tempering. Regarding the Metropolis updates, we have used either the multisample multispin coding (MUSA-MSC) [3] or the multisite multispin coding (MUSI-MSC) [55] techniques, which we will briefly describe.

Intel and AMD CPUs support 128 and 256-bit words in their streaming extensions. It is known that we can perform the Metropolis update of a single spin by using a sequence of Boolean operations [3], so we can take advantage of current CPU technology to simulate 128 or 256 systems simultaneously. This method is widely used in computational physics [3], [56–61], and it is denominated MUSA-MSC. The most efficient version of our MUSA-MSC code turned out to be the one with 128 bits.

However, there exist certain samples with such sluggish dynamics that MUSA-MSC ceases to be efficient. Indeed, if only a few of the 128 samples coded in a computer word are not yet thermalized, continuing the simulation of the already equilibrated samples is a waste of computer time. This problem is particularly acute for \( L = 16 \) and
24, because the width of the autocorrelation time distribution increases with $L$ (see section 4.2). For those misbehaving instances, we turn to MUSI-MSC: the 256 bits in a computer word now code 256 distinct spins of a single replica of a single sample [55]. In this way, we execute the Metropolis algorithm in $L^3/256$ steps. Our implementation for $L = 24$ use a geometric arrangement differing from [55], as explained in appendix C.

The simulations were carried out using either Intel Xeon E5-2680 or AMD Opteron 6272 processors. 12800 samples were simulated (and four replicas per sample). More details of the simulations are given in appendix A.

4. Characterizations of temperature chaos

Temperature chaos will be studied from two complementary viewpoints. The perspective offered by the parallel tempering dynamics is considered in section 4.1. The finite-size scaling of the parallel tempering dynamics is studied in section 4.2. The static viewpoint is considered in section 4.3. Finally, in section 4.4, we will study the correlation between the parallel tempering dynamics and temperature chaos.

4.1. Dynamics: the variational method

Our scope here is to use equation (2) in a variational method to estimate the exponential autocorrelation time. Consider the eigenmode expansion in equation (7). The optimal choice for the observable $f$ would have $A_{1,j} = 1$ (and $A_{n>1,j} = 0$) in its decomposition in characteristic times$^{12}$. We shall use our physical intuition to approach this ideal.

As explained in the Introduction, the temperature chaos effect suggests focusing our attention on the temperature flow along the parallel tempering dynamics [21, 42, 43]. Let us consider one of the $N_T$ system copies in the parallel tempering dynamics. We shall describe the temperature random-walk through the index $i_t$ that indicates that, at time $t$, our system copy is at temperature $T_{i_t}$. The equilibrium probability for $i_t$ is just the uniform probability over the set $\{1, 2, \ldots, N_T\}$. If we consider an arbitrary function of $i_t$ its equilibrium expectation value will be

$$E(f) = \frac{1}{N_T} \sum_{i=0}^{N_T} f(i).$$

We shall consider, as well, pairs of system copies. These pairs will be described by two integer indices, $i_t \neq j_t$. The equilibrium value of an arbitrary function of a pair of system copies is

$$E(f) = \frac{1}{N_T(N_T - 1)} \sum_{i=0}^{N_T} \sum_{j \neq i}^{N_T} f(i, j).$$

$^{12}$ The reader is probably used to applying this formalism to the evolution of a single spin configuration. Here, we shall need to enlarge this viewpoint to a parallel tempering simulation that involves several spin chains, and to a function $f$ that is related to the temperature of a given chain. More details can be found in appendix E.
We also request $f(T)$ to be continuous at $T^*$. Indeed, we need to fix only the ratio $E f(T)$ is our second variational parameter. Specifically, our linear test-functions are

$$T > T^*: \quad f_{T^*}(T) = a_+(T - T^*),$$

$$T < T^*: \quad f_{T^*}(T) = a_-(T - T^*).$$

We require the two amplitudes $a_+$ and $a_-$ to be positive. Their ratio is fixed by imposing $E(f_{T^*}) = 0$. Indeed, we need to fix only the ratio $a_+/a_-$, because the overall scale of the test function $f_{T^*}$ is irrelevant. Besides these, we consider quadratic ($p = 2$) and cubic ($p = 3$) test-functions:

$$T > T^*: \quad f_{T^*}(T) = a_+(T - T^*)^2(2T_{N_T} - T^* - T),$$

$$T < T^*: \quad f_{T^*}(T) = a_-(T^* - T)^2(2T_1 - T^* - T).$$

We choose again $a_+, a_- > 0$, and the ratio $a_+/a_-$ is fixed by imposing $E(f_{T^*}) = 0$. Note that all our test-functions are continuous at $T^*$ (the cubic $f_{T^*}$ are even differentiable at $T^*$).

Now, for each $f$ and $T^*$, we need to estimate the autocorrelation function $C_{f,T^*}(t)$, recall equation (3), and the related integrated autocorrelation time (4). Let $\hat{f}_{T^*} \equiv f_{T^*} - E(f_{T^*})$. $C_{f,T^*}(t)$ is estimated as

$$C_{f,T^*}(t) = \frac{n_{\text{Met}}}{N_s - t_0 - t} \sum_{t' = t_0}^{N_s - t} \hat{f}_{T^*}(i_{t'}) \hat{f}_{T^*}(i_{t' + t}).$$

Here, $N_S$ is the number of times we have stored the state of the PT indices $i_t$ in the hard drive. Note that $t_0$ must be much greater than $\tau_{\text{int}}$, in order to be safely in the equilibrium regime. The parameter $n_{\text{Met}}$ is the periodicity with which we record the time indices $i_t$ (in most of this work, $n_{\text{Met}} = 250 000$ Metropolis sweeps). Note that $C_{f,T^*}(t)$ is independent of the system copy. Therefore, we can average over the $N_T$ numerical estimations of $C_{f,T^*}(t)$ (as well as over the four independent replicas), which greatly

---

**Table 1.** Different choices of the function $f$ used in the variational method.

| Identifier | Function                  |
|------------|---------------------------|
| 0          | Piecewise constant        |
| 1          | Piecewise linear          |
| 2          | Piecewise quadratic       |
| 3          | Piecewise cubic           |
| $|$          | OR in couples             |
| &          | AND in couples            |
| ^          | XOR in couples            |
| *          | Multiplication in couples |

J. Stat. Mech. (2018) 033302

https://doi.org/10.1088/1742-5468/aaa387
enhances the statistics. The computation for functions \( f \) depending on a pair of system copies is analogous.

Once we have computed \( C_{f, T^*}(t) \), the normalized correlation function is just \( \hat{C}_{f, T^*}(t) \), and the integrated autocorrelation time can be computed in the standard way \([2]\)

\[
\tau_{\text{int}, f, T^*} \approx n_{\text{Met}} \left[ \frac{1}{2} + \sum_{t=0}^{W} \hat{C}_f(t) \right],
\]

(15)

where \( W \) is a self-consistent window \([2]\) that avoids the divergence of the variance of \( \tau_{\text{int}, f, T^*} \) (we impose \( \tau_{\text{int}, f, T^*} < 10W \)).

We have found it advantageous to consider a third variational parameter \( l_{\text{blo}} \), which we now describe. We build Wilson–Kadanoff blocks: the Monte Carlo sequence \( f_{T^*}(i_1), f_{T^*}(i_2), \ldots f_{T^*}(i_{N_S}) \) is divided into blocks of \( l_{\text{blo}} \) consecutive data (see e.g. \([62]\)). We take the average of the \( f_{T^*}(i_t) \) within a single block. This operation defines a new sequence of \( N_S/l_{\text{blo}} \) renormalized times, over which the integrated autocorrelation time can be estimated just as we did for the original data \( l_{\text{blo}} = 1 \). The estimated autocorrelation time should be rescaled by \( l_{\text{blo}} \) in order to recover the original time units. The purpose of the blocking is to reduce high-frequency fluctuations.

There is a danger in the use of Wilson–Kadanoff blocks, though. Formula (15) was obtained assuming that \( \tau_{\text{int}, f, T^*} \) is much larger than the time step in the right-hand side. In fact, \( l_{\text{blo}} \) can be made much greater than the \( \tau_{\text{exp}} \) that we aim to estimate. As a consequence, the renormalized correlation function will vanish for times \( t \neq 0 \). This means that the integrated autocorrelation time will be 1/2 (over the renormalized time-mesh). When turning back to physical time units, we shall find \( \tau_{\text{int}} = n_{\text{Met}} l_{\text{blo}}/2 \), which diverges for large \( l_{\text{blo}} \). Hence, we need a practical way to ensure that \( l_{\text{blo}} \) is not so large that all the physical information has been erased. Our solution imposes

\[
\tau_{\text{int}, f, T^*, l_{\text{blo}}} < \frac{5}{2} n_{\text{Met}} l_{\text{blo}},
\]

(16)
in order to consider the results of a given \( l_{\text{blo}} \).

We obtain, for each sample, a huge number of values of \( \tau_{\text{int}} \) corresponding to the eight different functions and the different choices of \( T^* \). We have tried for \( T^* \) all the temperatures \( T_i \) in the lower half of the set of temperatures in our parallel tempering simulation. The values of \( l_{\text{blo}} \) are taken from the list \{1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000\}.

Our variational estimate \( \tau_{\text{int}, \text{var}} \) is the largest of these numbers. This is a robust estimate (i.e. this methodology does not provide spurious values), and thus can be implemented in an automatic way in the analysis, and allows for a precise estimate of the thermalization time needed.

We shall also consider below the temperature \( T_d \) which is the \( T^* \) for which the variational maximum is attained.

An example of the improvement obtained in the computation of the autocorrelation function is shown in figure 1. As can be inferred from equation (7), a major difficulty is that the amplitude for \( \tau_{\text{exp}} \), namely \( A_{1,f} \) can be very small. Indeed, the correlation function considered in a previous work \([21]\) (which is our piece-wise linear \( f \), identifier \#1 in
table 1, and $T^*$ set to the critical temperature), has $A_{1,f} \approx 0.1$. Instead, the optimized autocorrelation function has an amplitude $A_{1,f}$ almost ten times larger.

We observe in table 2 that, for almost all samples, the variational method chooses a function $f$ depending on one system copy only. Moreover, this variational method in many cases substantially improves the results obtained previously with a linear function and a parameter $T^*$ chosen at the critical temperature [21].

We can do a quantitative comparison between the variational method proposed here and the old approach. Let us histogram the ratio $\tau_{\text{int,old}}/\tau_{\text{int,var}}$, conditioned to the value of $\tau_{\text{int,var}}$ (which is a good indicator of how chaotic a sample is). We represent the result of this study in figure 2, where $\tau_{\text{int,old}}/\tau_{\text{int,var}}$ is represented for the first and last deciles of $\tau_{\text{int,var}}$\textsuperscript{13}. The advantages of the variational estimator are evident when one focuses on decile 10 (i.e. for the most chaotic samples), where we observe a significant fraction of samples with $\tau_{\text{int,old}}/\tau_{\text{int,var}} < 0.1$.

\textbf{4.2. The finite size scaling behavior of the parallel tempering dynamics}

In this section, we study the parallel tempering dynamics for $L = 8, 12, 16, 24$ and 32, and we investigate temperature chaos from a dynamical point of view. In the following

\textsuperscript{13} Deciles are similar to percentiles. First, samples are ordered according to their $\tau$. Then, we divide the samples among 10 sets (deciles) of equal size. Those samples with the lowest $\tau$ belong to decile 1, and so on.

\begin{table}[h]
\centering
\begin{tabular}{lllllllll}
\hline
$L$ & 0 & 1 & 2 & 3 & $\&$ & $\wedge$ & $*$ & Total \\
\hline
16 & 2032 & 5320 & 3875 & 1374 & 4 & 115 & 74 & 6 & 12800 \\
24 & 1556 & 7196 & 3089 & 820 & 0 & 127 & 11 & 1 & 12800 \\
\hline
\end{tabular}
\caption{Number of times the variational method has picked one of the eight choices among the functions $f$ described in the text. $L$ denotes the lattice size.}
\end{table}
we will denote the minimal temperature allowed in a parallel tempering simulation as $T_{\text{min}}$ (it was $T_1$ in the previous section). We will focus on the variational estimate of $\tau_{\text{int,var}}$ (that we will call simply $\tau$ from now on).

An implicit assumption of our study, corroborated by the results in section 4.4, is that the scaling behavior of $\tau$ is mostly decided by the value $T_{\text{min}}$. Other details, such as the number of temperatures in the parallel tempering mesh, are expected to play a minor role (if kept in a reasonable range). For the comparative analysis of the dynamics we use the simulations at $T_{\text{min}} \approx 0.7$ shown in table A1. This is the lowest value of $T$ we have thermalized for all our lattice sizes. An important advantage of $T_{\text{min}} \approx 0.7$ is that temperature chaos has already been characterized at such temperatures, in the equilibrium setting [39]. Lowering $T_{\text{min}}$ would increase chaos effects, which would have been good in principle, but it would also have made it extremely difficult to reach thermal equilibrium. On the other hand, increasing $T_{\text{min}}$ to approach the critical point would make the results irrelevant, because samples displaying temperature chaos would be too scarce (besides, we want to study the spin glass phase, rather than critical effects).

For $L \leq 16$ we have $N_T = 13$. For $L = 24$ we needed to increase $N_T$ in order to keep constant the acceptance rate of the temperature exchange step of the parallel tempering simulation. The $L = 32$ data are from [21], and have been obtained with the dedicated Janus computer [63]. The Janus simulation used heat bath dynamics, rather than Metropolis, and the parallel tempering there had $N_T = 34$ and $T_{\text{min}} = 0.703$. In order to be sure that heat bath autocorrelation times are consistent with Metropolis times (as we would expect), we simulated with Janus ten randomly selected samples with both algorithms, finding that $\tau_{\text{Metropolis}} \approx \tau_{\text{heat-bath}}/3$.

We show in figure 3 the cumulative distribution function of $\tau$, $F(\tau)$. The maximum slope of $F$ decreases with $L$ for the small systems, and stabilizes between $L = 24$ and

![Figure 2. Conditional probability density function of the ratio $\tau_{\text{int,old}}/\tau_{\text{int,var}}$, given that $\tau_{\text{int,var}}$ belongs to a given decile. We show the data for the first decile (left) and the tenth decile (right) for $L = 16$ (top) and $L = 24$ (bottom).](image)
$L = 32$; indeed, these two distributions can be approximately superposed by a simple translation. This is reminiscent of a critical slowing-down\cite{64}:

$$\tau \sim L^{z_{PT}(T_{\text{min}})}.$$  \hfill (17)

It is not obvious \textit{a priori} that such a simple scaling should hold in the spin glass phase. As a working, simplifying hypothesis, we assume that the exponent $z_{PT}$ only depends on the value of the lowest temperature in the parallel tempering grid, $T_{\text{min}}$ (and not on the number of temperatures).

As a first test of equation (17), we compute an effective $z$ exponent by comparing the probability distributions for two lattice sizes $(L_1, L_2)$, by means of the definition

$$z_{PT}(L_1, L_2, p) = \frac{\log(\tau(L_1, p)/\tau(L_2, p))}{\log(L_1/L_2)},$$  \hfill (18)

where $\tau(L_i, p)$ is determined by the implicit equation $F(\tau(L_i, p)) = p/100$ where $p = 1, \ldots, 100$ is the so-called percentile rank (i.e. $\tau(L_i, p)$ is the $p$th percentile of the distribution for the size $L_i$). We have computed $z_{PT}$ for three pairs of lattice sizes, $(12, 24)$, $(16, 24)$ and $(24, 32)$; in figure 4 we show the results as a function of the rank. The values for the largest pair, $(24, 32)$, are independent of the rank, within statistical errors, in agreement with the ansatz. Smaller size couples give smaller estimates (in the same ball park) for low ranks, but converge to the $(24, 32)$ value for high ranks (i.e. for the harder samples), and the coincidence improves and extends to smaller ranks for larger lattices. We remark that this dynamic behavior is consistent with the static findings in this temperature range\cite{39}: for $L = 8$ it is almost impossible to find samples displaying strong temperature chaos. One needs to go to systems as large as $L = 24, 32$ to find chaotic samples with a significant probability.

An interesting coincidence with the results of non-equilibrium simulations\cite{55, 65–67} could have a deep meaning. Indeed in non-equilibrium conditions one finds that the

\textbf{Figure 3.} Empirical probability distribution of $\tau$ for $L = 8$, 12, 16, 24 and 32. For $L = 8$ and $L = 12$ some of the samples have $\tau$ smaller than our minimal resolution (if $\tau < n_{\text{Met}}$ we cannot compute it safely). We show only the part of the distribution function that can be safely computed.
spin glass correlation length $\xi$, in a lattice of size $L \gg \xi$, at temperature $T = 0.7$ grows with the simulation time $t_w$ as

$$\xi(t_w) \propto t_w^{1/z(T)}$$

where $z(T)$ is the so-called dynamic critical exponent, that turns out to be strongly temperature dependent in the spin glass phase $z(T) \propto T_c/T$. Our results for the lattice pair $(24,32)$ suggest that $z(T = 0.7) \approx z^{PT}(T_{\text{min}} = 0.7)$.

As a further test we can rescale the whole probability distribution by using equations (17) and (20). This is done in figure 5 (main) that shows $F(\tau)$ as a function of $y = \tau/L^z$. As expected, the data for $L = 24$ and $L = 32$ present a nice collapse. The curve corresponding to $L = 16$ collapses with them only for percentile ranks higher than 80, and the curve corresponding to $L = 12$ collapses for percentile ranks higher than 90. This is a nice, smooth behavior. On the larger lattice sizes, we reach a perfect scaling, but already on smaller lattices we see a partial scaling, which improves for increasing size. In figure 5 (inset), we show a log–log plot of $1 - F(\tau)$ as a function of $\tau/L^z$, that emphasizes the large $\tau$ tail of the distribution. The fit presented shows that the probability density function of $\tau$ behaves, asymptotically for large $y$, like a fat-tailed distribution:

$$\rho(y \equiv \tau/L^z) \sim y^{-1-a_1}, \quad a_1 \approx 1.38.$$  

The distribution seems to reach its asymptotic form for $L \geq 24$. Perhaps unsurprisingly, the thermodynamic (i.e. equilibrium) effective potential that characterizes temperature chaos also turns out to be asymptotic for $L \geq 24$ [39].

In order to study how the range of temperatures in the parallel tempering affects the dynamics, we have performed an extra simulation for $L = 16$. In the new simulation we

$$https://doi.org/10.1088/1742-5468/aaa387$$
take a lower minimum temperature ($T_{\text{min}} = 0.479$ instead of $T_{\text{min}} = 0.698$) increasing $N_T$ from 13 to 16 in order to keep the interval between adjacent temperatures fixed—see table A1. Since the simulation with $N_T = 16$ reaches a lower minimum temperature than the simulation with $N_T = 13$ we expect to find chaos events (i.e a jam in the

![Figure 5](https://example.com/figure5.png)

**Figure 5.** Probability distribution function of the rescaled variable $y = \tau/L^z$, ($z$ is the dynamic exponent corresponding to $T_{\text{min}} = 0.7$, namely $z(T = 0.7) = 11.64(15)$). (Inset) plot of $\log(1 - F(\tau))$ versus $\log(\tau/L^z)$; the straight black line is a fit to the form $a_0 - a_1 \log(\tau/L^z)$ yielding $a_0 = -29.33$ and $a_1 = -1.38$.

![Figure 6](https://example.com/figure6.png)

**Figure 6.** Scatter plot of $\log(\tau_{\text{int},16}/\tau_{\text{int},13})$ versus $T_d$. The lattice size is $L = 16$, $\tau_{\text{int},16}$ is the relaxation time for $N_T = 16$ ($T_{\text{min}} = 0.479$), $\tau_{\text{int},13}$ is the relaxation time for $N_T = 13$ ($T_{\text{min}} = 0.698$), $T_d$ is the temperature of chaos from a dynamical point of view (defined in the variational method) of the simulation with $N_T = 16$. Disorder samples are the same in the two simulations. The vertical black line represents the minimum temperature simulated in the $N_T = 13$ simulation. (We have added a small Gaussian white noise to $T_d$, which is a discrete variable, to avoid the cluttering of data in vertical lines).
parallel tempering temperature flow) that the simulation with $N_T = 13$ cannot see. In figure 6 we show a scatter plot of $\log(\tau_{\text{int},16}/\tau_{\text{int},13})$ versus $T_d$ for the 12800 samples ($\tau_{\text{int},16}$ and $\tau_{\text{int},13}$ are the autocorrelation times for $N_T = 16$ and 13 respectively. $T_d$ is the temperature $T^*$ where the variational estimate $\tau_{\text{int},f}$ reaches its maximum).

For $T_d > 0.698$, the ratio takes values of order one for most samples, while for $T_d < 0.698$ there is a huge number of samples with $\tau_{\text{int},16} \gg \tau_{\text{int},13}$, i.e. there are a lot of samples with a chaotic behavior in a temperature range below $T_{\text{min}} = 0.698$.

The same idea can be analyzed from a different point of view. Imagine that we have studied with great care a given sample down to some temperature $T_{\text{min}}$. Can we say something about possible chaotic effects at lower temperatures? The question is answered negatively in figure 7: the probability that a sample has a large $\tau_{\text{int}}$ for the simulation with a lower $T_{\text{min}}$ is not correlated to the value of $\tau_{\text{int}}$ for the first simulation.

4.3. Statics

In the infinite volume limit, static temperature chaos is the complete rearrangement of the equilibrium configuration under any change of temperature. It has been studied numerically mostly through the disorder average of the probability density function of the overlap between the spin configurations at temperatures $T_1$ and $T_2$,

$$q_{T_1,T_2} = \frac{1}{V} \sum_x s_x^{T_1} s_x^{T_2},$$

or through ratio of moments of this distribution. However, because of the size of the systems that can be currently simulated, the overlap is strongly influenced by finite

https://doi.org/10.1088/1742-5468/aaa387

15
size effects. It has been suggested that static temperature chaos is a rare-event driven phenomenon, that should be studied via the distribution of the sample-dependent chaotic parameter [39, 40]:

\[ X^J_{T_1, T_2} = \frac{\langle q^2_{T_1, T_2} \rangle^J}{\sqrt{\langle q^2_{T_1, T_1} \rangle^J \langle q^2_{T_2, T_2} \rangle^J}}, \]

(23)

where \( \langle \cdots \rangle^J \) is the thermal average within a given sample \( (J) \). Notice that \( 0 < X^J_{T_1, T_2} \leq 1 \); \( X^J_{T_1, T_2} = 1 \) means that equilibrium spin configuration of the \( J \) sample at temperature \( T_1 \) and temperature \( T_2 \) are indistinguishable, while \( X^J_{T_1, T_2} = 0 \) means that the equilibrium spin configurations are completely different.

The temperature evolution of \( X^J_{T_1, T_2} \) is shown in figure 8 for selected samples (in the figure, \( T_1 \) is kept fixed to \( T_1 = T_{\text{min}} \), while \( T_2 \) is made to vary). In some samples, we find chaotic events, namely sharp drops of \( X^J_{T_1, T_2} \) at very well defined temperatures, implying that the typical spin configurations significantly differ at the two sides of the chaotic event. It was empirically observed in [39] that chaotic events occurring at low temperatures are most harmful to the performance of parallel tempering. To quantify the effect, the chaotic integral \( I \) was introduced

\[ I = \int_{T_{\text{min}}}^{T_{\text{max}}} X^J_{T_{\text{min}}, T_2} \, dT_2. \]

(24)

Note that a sharp drop of \( X^J_{T_{\text{min}}, T_2} \) at a low \( T_2 \) will result in a very low value of the chaotic integral \( I \). Furthermore, a study of the temperature behavior of the chaotic

Figure 8. Plot of \( X^J_{T_{\text{min}}, T} \) versus \( T \) for the five most chaotic samples (top) and the five less chaotic ones (bottom): \( L = 16 \) case (left) and \( L = 24 \) case (right).
parameter leads to the conclusion that chaos events happen only at low temperatures, wherefore the high temperatures introduce only noise in the estimate of $I$. In order to eliminate this noise, we introduce a new integrated chaotic parameter $I_2$ that involves only the lower half of the temperature range.

Nevertheless, there exist certain samples that exhibit a huge $\tau_{\text{int}}$, and have a relatively large chaotic integral, so the correlation between statics and dynamics is more complicated than one could hope. Therefore, in order to improve our thermodynamic understanding of the parallel tempering dynamics, we need to look elsewhere. We have found it useful to consider the temperature derivative of the chaotic parameter. Indeed, it is easy to prove that

$$\left. \frac{dX^J_{T_1,T_2}}{dT_2} \right|_{T_2=T_1} = 0.$$  \hfill (25)

for any temperature $T_1$. However, if we focus on these outlier samples, we notice that these samples present a sharp drop in $X^J_{T_1,T_2}$ at two consecutive temperatures. This observation will motivate the definition in equation (26), below.

### 4.4. Correlations dynamics-static

Once we have characterized the chaos phenomena from both dynamical and static point of view, we are interested in knowing how these static and dynamic estimators are correlated.

Besides the chaos integrals $I$ and $I_2$, we introduce a new quantity for further use:

$$K_i = 1 - X^J_{T_i,T_{i+1}}.$$  \hfill (26)

After some trials, we have finally defined a last parameter:

$$I_X = a I_2 - b \min_i \left( -\log \left( K_i^2 \right) \right) - c \sum_i \left( -\log \left( K_i^2 \right) \right),$$  \hfill (27)

where the coefficients $a$, $b$ and $c$, that depend on the lattice size, are obtained through a minimization of the correlation coefficient $r$ between $I_X$ and $\log(\tau_{\text{int}})$ ($r$ is negative, and it would be $r = -1$ if we managed to achieve a perfect understanding of our dynamical data). The values of these coefficients are given in table 3.

This finding is supported by figure 8. We see that the most chaotic samples in terms of the integrated autocorrelation time (figure 8, top), present a sharp fall in the chaotic parameter. On the other hand, we can see that less chaotic samples in terms of the integrated time (figure 8, bottom), have a much smoother fall.

In figure 9, we confront the most representative estimator for the dynamical chaos, namely the largest integrated autocorrelation time $\tau_{\text{int}}$ found in our variational study, with the static chaotic integrals $I$, $I_2$ and $I_X$. We can observe how spurious values of the

| $L$ | $a$   | $b$   | $c$   |
|-----|-------|-------|-------|
| 16  | 0.6143| 0.2865| 0.1373|
| 24  | 0.2963| 0.3217| 0.0120|

Table 3. Value of the coefficients $a$, $b$ and $c$ in equation (27), that maximize the correlation between $I_X$ and $\log(\tau_{\text{int}})$. 
Figure 9. Scatter plot of $\log(\tau_{\text{int, var}})$ versus the integrated chaotic parameter $I$. We present data for two lattice sizes and for the three definitions of the integrated chaotic parameter defined in the text ($I$, $I_2$ and $I_X$). The pattern of depleted horizontal bands is due to our choice of a few $b_{\text{blo}}$.

Table 4. Correlation coefficients for the scatter plot of $\log(\tau_{\text{int}})$ versus the integrated chaotic parameter, for two lattice sizes and for the three definitions of the parameter ($I$, $I_2$ and $I_X$).

| $L$ | Integral | $r$       |
|-----|----------|-----------|
| 16  | $I$      | $-0.714 \pm 0.005$ |
| 16  | $I_2$    | $-0.751 \pm 0.005$ |
| 16  | $I_X$    | $-0.795 \pm 0.004$ |
| 24  | $I$      | $-0.725 \pm 0.005$ |
| 24  | $I_2$    | $-0.746 \pm 0.005$ |
| 24  | $I_X$    | $-0.786 \pm 0.004$ |
original parameter \( I \) (i.e. large values of \( I \) associated to large \( \tau \)) are displaced towards lower values when we use the improved parameters \( I_2 \) and \( I_X \).

The value of the correlation coefficients are reported in table 4\(^{14} \). We observe a strong anti-correlation in \( I_X \), that improves over the previous indicator of correlation \( I \). \(^{[39]} \) The improvement is less clear for \( I_2 \).

We can try to define other magnitudes (whether static or dynamic) that capture the chaos phenomenon. One possible choice is the temperature, \( T_s \), for which \( X^T_{\min,T} \) presents the maximum (negative) slope. Unfortunately, we observe a weaker correlation between both estimators, \( \tau \) and \( T_s \), (see figure 10) and we can check it quantitatively through table 5. Some further attempts along these lines are explored in appendix D.

\(^{14} \) Statistical-error estimates were computed using the bootstrap method.
5. Discussion and conclusions

We have proposed an efficient variational method to estimate the elusive exponential autocorrelation time of a Monte Carlo Markov chain, specific to the (arguably important) case of a parallel tempering simulation. In this variational method, we have introduced three parameters (a temperature $T^*$, a function $f$ and a block length). We have checked that this procedure is very robust, and can easily be implemented in an automatic way.

In addition, we have studied the scaling properties of the probability distribution of the autocorrelation time, obtained using the proposed variational approach. In particular, we have shown that scaling holds for lattices of sizes $L \geq 24$, consistently with previous studies using effective potentials.

Moreover, we have introduced additional static chaotic indicators, and finally we have checked the statistical correlations between these static chaotic indicators and the dynamical correlation times.

Acknowledgments

We thank the Janus collaboration for allowing us to analyze the $L = 32$ autocorrelation times from [21]. We are also grateful for allowing us to carry out a short simulation on the Janus computer, in order to establish the correspondence between the Metropolis and the Heat-bath autocorrelation times.

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 694925). We were partially supported by MINECO (Spain) through Grant Nos. FIS2015-65078-C2, FIS2016-76359-P and by the Junta de Extremadura (Spain) through Grant No. GRU10158 (these three contracts were partially funded by FEDER). Our simulations were carried out at the BIFI supercomputing center (using the Memento and Cierzo clusters), at the TGCC supercomputing center in Bruyères-le-Châtel (using the Curie computer, under the allocation 2015-056870 made by GENCI) and at ICCAEx supercomputer center in Badajoz (Grinfisphc and Iccaexhpc). We thank the staff at BIFI, TGCC and ICCAEx supercomputing centers for their assistance.

Appendix A. Parameters of the simulation

Whereas in numerical simulations of spin glasses the disorder samples are usually independent, the samples we use here are not fully independent. The motivations of our choice are explained in [54]. We consider cubes with $L^3$ spins and $3L^3$ couplings, divided into an inner part of $(L/2)^3$ spins and an outer part surrounding it. We simulate 10 independent inner samples, and, for each inner sample, 1280 independent outer samples. We simulate four replicas (independent spin systems) for every inner and outer sample. Hence, we have simulated 12 800 disorder realizations (samples) with a total of $12 800 \times 4$ real spin systems. The parameters of the simulation can be found in table A1.
The thermalization criteria that have been used are as follows (as explained above, these criteria applied to every sample, individually). First of all, the number of iterations in $\tau_{\text{exp}}$ units ($l_{\text{blo}} = 1$) must be greater than 20; as a double-check to avoid failures in the automated fitting procedure, we recomputed $\tau_{\text{exp}}$ with $l_{\text{blo}} = 10$ (the total simulation length is also required to be longer than 20 $\tau_{\text{blo}} = 10$).

However, we had some additional safety checks to ensure that the computation of $\tau_{\text{exp}}$ could be trusted. For those samples where either of the following two requirements was not met, we doubled the total simulation length, and only then recomputed $\tau_{\text{exp}}$. First, in order to make sure that every sample spends enough time at high temperatures, we require that each copy of the system in the parallel tempering method spends at least 35% of the time in the upper half temperature region. Second, the ratio between the larger and the smaller values of $\tau_{\text{int}}$, as computed for each of the four independent replicas, must be less than two (for either $l_{\text{blo}} = 1, 10, 100$). This last requirement can

---

**Table A1.** Parameters of the simulations. $L$ is the lattice size; $L_{\text{int}}$ the size of the inner part of the lattice; $N_T$, $T_{\text{min}}$ and $T_{\text{max}}$ are the number of temperatures, the minimum and the maximum temperatures used in the parallel tempering method; $N_{\text{Met}}$ is the number of Metropolis sweeps (at each temperature); ps/spin is the average CPU time per spin-flip in MUSI-MSC, using an Intel Xeon CPU E5-2680 processor; $N_{\text{samp}}$ denotes the number of bad samples whose simulations had to be extended in order to thermalize and finally $N_{\text{Met, min}}$, $N_{\text{Met, mean}}$ and $N_{\text{Met, max}}$ are the minimum, mean and maximum number of Metropolis sweeps per temperature needed to reach thermalization (bad samples). The set of temperatures used is clearly the same in the MUSI-MSC and MUSA-MSC parts of this table. The number of Metropolis sweeps between two consecutive parallel tempering sweeps is always $N_{\text{mp, PT}} = 10$. For the MUSI-MSC simulation of $L = 24$ we parallelized, using *Pthreads*, by distributing the $N_T = 24$ system copies among 12 CPU cores in the Intel Xeon CPU E5-2680.

| L  | $L_{\text{int}}$ | $N_T$ | $T_{\text{min}}$ | $T_{\text{max}}$ | $N_{\text{Met}}$ ($\times 10^6$) | ps/spin |
|----|----------------|-------|-----------------|-----------------|-------------------------------|---------|
| 24 | 12             | 24    | 0.698           | 1.538           | 500                           | 104     |
| 16 | 8              | 16    | 0.479           | 1.575           | 250                           | 107     |
| 16 | 8              | 13    | 0.698           | 1.575           | 250                           | 119     |
| 16 | 12             | 13    | 0.698           | 1.575           | 250                           | 119     |
| 14 | 12             | 13    | 0.698           | 1.575           | 500                           | 120     |
| 12 | 6              | 13    | 0.698           | 1.575           | 250                           | 119     |
| 8  | 4              | 13    | 0.698           | 1.575           | 250                           | 126     |

| L  | $L_{\text{int}}$ | $N_T$ | $N_{\text{samp}}$ | $N_{\text{Met, min}}$ ($\times 10^6$) | $N_{\text{Met, mean}}$ ($\times 10^6$) | $N_{\text{Met, max}}$ ($\times 10^6$) | ps/spin |
|----|----------------|-------|------------------|--------------------------------------|---------------------------------------|--------------------------------------|---------|
| 24 | 12             | 24    | 2441             | 1000                                 | 4262                                  | 326000                               | 57      |
| 16 | 8              | 16    | 2898             | 500                                  | 5096                                  | 355500                               | 304     |
| 16 | 8              | 13    | 338              | 500                                  | 543                                   | 4000                                  | 306     |
| 16 | 12             | 13    | 314              | 500                                  | 578                                   | 8000                                  | 306     |

The thermalization criteria that have been used are as follows (as explained above, these criteria applied to every sample, individually). First of all, the number of iterations in $\tau_{\text{exp}}$ units ($l_{\text{blo}} = 1$) must be greater than 20; as a double-check to avoid failures in the automated fitting procedure, we recomputed $\tau_{\text{exp}}$ with $l_{\text{blo}} = 10$ (the total simulation length is also required to be longer than 20 $\tau_{\text{blo}} = 10$). However, we had some additional safety checks to ensure that the computation of $\tau_{\text{exp}}$ could be trusted. For those samples where either of the following two requirements was not met, we doubled the total simulation length, and only then recomputed $\tau_{\text{exp}}$. First, in order to make sure that every sample spends enough time at high temperatures, we require that each copy of the system in the parallel tempering method spends at least 35% of the time in the upper half temperature region. Second, the ratio between the larger and the smaller values of $\tau_{\text{int}}$, as computed for each of the four independent replicas, must be less than two (for either $l_{\text{blo}} = 1, 10, 100$). This last requirement can
help us to identify a lack of thermalization for those samples whose leading term in the autocorrelation function has a very small amplitude.

Appendix B. On the selection of relevant parameters of the simulation

The natural question is whether our particular choice of samples (see appendix A) affects our results. One could imagine that the results obtained from configurations sharing the same inner part could be strongly correlated, and that with only 10 inner parts, our statistics would be insufficient. We show in figure B1 that this is not the case for the probability distribution of \( \tau \): the probability distributions of \( \tau \) for the samples sharing the same 10 inner parts are plotted separately. They are nearly indistinguishable. The average over the outer disorder (which we can call the metastate average, in analogy with [54]) dramatically reduces the fluctuations due to the inner disorder. The same conclusion holds for the chaos integral (see figure B2).

On the other hand, the selection of the minimal temperature in the parallel tempering could seem arbitrary; however, the selection of \( T_{\text{min}}^{L=16} \) and \( T_{\text{min}}^{L=24} \) has been made carefully, to ensure that the most difficult samples had similar \( \tau \). This is shown in the figure B3.

Figure B1. Empirical probability distribution function of \( \tau \) represented for the 10 inner samples separately. \( L = 16 \) case (top) and \( L = 24 \) case (bottom). Averaging over the metastate (i.e. the outer samples) with fixed inner couplings strongly reduces the fluctuations between the inner samples.
Appendix C. The geometry of MUSI-MSC

The geometric construction explained in [55] for $L = 256$ turns out to be satisfactory for $L = 16$ as well, but not for $L = 24$. Hence, we shall first recall the geometry that we employ for $L = 16$. After this, we shall explain the modifications that we introduced.
for \( L = 24 \). Note that multispin coding is not usually employed in single-sample simulations—because, in common schemes, one needs an independent random number per bit. Fortunately, this problem can be circumvented, as explained in [55].

For \( L = 16 \), the physical lattice of Cartesian coordinates \( 0 \leq x, y, z < L \) is mapped to a super-spin lattice. Each super-spin is coded in a 256-bit computer word (of course, the 256 bits correspond to 256 physical spins, which are updated in parallel). The crucial requirement is that spins which are nearest neighbors in the physical lattice are coded into nearest-neighbor super-spins. In particular, our super-spins are placed at the nodes of a cubic lattice with the geometry of a parallelepiped of dimensions \( L_x = L_y = L/8 \), and \( L_z = L/4 \). The relation between physical coordinates \((x, y, z)\) and the coordinates in the super-spin lattice \((i_x, i_y, i_z)\) is

\[
\begin{align*}
x &= b_x L_x + i_x, \quad 0 \leq i_x < L_x, \quad 0 \leq b_x < 8, \\
y &= b_y L_y + i_y, \quad 0 \leq i_y < L_y, \quad 0 \leq b_y < 8, \\
z &= b_z L_z + i_z, \quad 0 \leq i_z < L_z, \quad 0 \leq b_z < 4.
\end{align*}
\]  
(C.1)

In this way, exactly 256 sites in the physical lattice are given the same super-spin coordinates \((i_x, i_y, i_z)\). We differentiate between them by means of the bit index:

\[
i_b = 64b_x + 8b_y + b_z, \quad 0 \leq i_b \leq 255.
\]  
(C.2)

Since we have to simulate \( N_T \) independent system copies in our parallel tempering simulation, we simply carry out successively the simulation of the \( N_T \) systems.

The alert reader will note that the above geometric construction is very anisotropic (we start with a cube, but end-up with a parallelepiped). Fortunately, this unsightly feature can be easily fixed by noticing that the single-cubic lattice is bipartite. Indeed, the lattice splits into the odd and even sub-lattices according to the parity of \( x + y + z \). The two sub-lattices contain \( L^3/2 \) sites. Furthermore, odd spins interact only with even spins, and vice versa. It follows that the update ordering is irrelevant, provided that our full-lattice sweep updates (say) all the odd sites first, and then all the even sites. Now, provided that \( L_x, L_y \) and \( L_z \) are all even, the parities of \( x + y + z \) and \( i_x + i_y + i_z \) coincide. This implies that all the spins coded in a single super-spin share the same parity, making the super-spin lattice asymmetry irrelevant. For \( L = 16 \), one finds that \( L_x = L_y = 2 \) and \( L_z = 4 \), all three being even numbers, and hence the above geometric construction works smoothly.

Unfortunately, for \( L = 24 \) one has \( L_x = L_y = 3 \) and \( L_z = 6 \), which implies that the super-spin lattice cannot be split into even and odd sub-lattices. Our solution consisted of introducing logical super-spins of 512 physical spins, that were later coded into two computer words of 256 bits each. The geometrical correspondence was \((L_x = L_y = L_z = L/8)\)

\[
\begin{align*}
x &= b_x L_x + j_x, \quad 0 \leq j_x < L_x, \quad 0 \leq b_x < 8, \\
y &= b_y L_y + j_y, \quad 0 \leq j_y < L_y, \quad 0 \leq b_y < 8, \\
z &= b_z L_z + j_z, \quad 0 \leq j_z < L_z, \quad 0 \leq b_z < 8.
\end{align*}
\]  
(C.3)

In this way, exactly 512 sites in the physical lattice are given the same super-spin coordinates \((j_x, j_y, j_z)\). We differentiate between them by means of the bit index:

\[
j_b = 64b_x + 8b_y + b_z, \quad 0 \leq i_b \leq 511.
\]  
(C.4)

https://doi.org/10.1088/1742-5468/aaa387

24
Now, the crucial observation is that (because \( L_x = L_y = L_z = 3 \) for \( L = 24 \)), the parity of \( x + y + z \) coincides with that of \( j_x + j_y + j_z \) if (and only if) the parity of \( \tilde{b}_x + \tilde{b}_y + \tilde{b}_z \) is even. In other words, given super-spin coordinates \((j_x, j_y, j_z)\) the 512 spins coded in the super-spin split into 256 even spins and 256 odd spins. Because same-parity spins are guaranteed to be mutually non-interacting, we decided to code the 256 bits with the same parity in the same computer word, with the corresponding bit index being the integer part of \( j_b/2 \).

However, the acceleration obtained with the MUSI-MSC was not enough for some of the worst \( L = 24 \) samples. Hence, we decided to add an extra layer of parallelism by using Pthreads to simulate a single sample in multicore processors. Given the smallness of the super-spin lattice, we found it preferable not to use concurrent threads in the simulation of copies of a single system (recall that we have \( N_T = 24 \) system copies in the parallel tempering simulation of \( L = 24 \)). Rather, we distributed the \( N_T \) system copies among 12 CPU cores, achieving an average speed of 57 picoseconds per spin-flip.

Appendix D. Quantities not related to chaos

Some perfectly reasonable quantities turn out to have surprisingly little relation to temperature chaos. To illustrate this effect, we test whether or not the temperature obtained through the variational method \( T_d = \{ T^* : \tau_{\text{int}} = \tau_{\text{int, var}} \} \) is correlated with the static temperature of chaos \( T_s \) (see figure D1).

In this case, figure D1 shows an over-density; however, the points outside of the principal density are too dispersed. For \( L = 16 \) (top) the number of points within the lines are 8017 (62.63% of the total) while for \( L = 24 \) (bottom) the number of points within the lines are 7539 (58.90% of the total). If we calculate the correlation coefficients, we obtain the table D1.

Appendix E. Analyzing parallel tempering simulations

In the main text we have used theoretical tools to analyze the time series produced by a Markov chain [2] in a setting that might be unfamiliar in the context of Statistical Mechanics. In particular, in our parallel tempering simulations we have a number \( N_T \) of independent copies (or clones) of the spin system that we want to simulate. Each clone wanders along the temperature axis, and our analysis is focused solely on these temperature excursions. At first sight, the reader might be surprised by the fact that this temperature wandering may teach us something about how far the spins are from thermal equilibrium at each temperature. The purpose of this appendix is to briefly clarify the relationship between the two types of degree of freedom, namely the clone temperatures and the spins (see also [11, 19–21, 68]).

For the sake of clarity, this appendix is organized in three paragraphs. A Markov Chain Monte Carlo describes a random-walk process: in appendix E.1, we describe the phase space where our random-walk takes place. We also discuss in appendix E.1, the stationary probability distribution (i.e. the equilibrium distribution) that our random walk is targeted to reach. In appendix E.2, we analyze some basic facts about the dynamics of a Markov process (see for example [2] for a more detailed discussion).
Finally, in appendix E.3, we consider an example that will hopefully clarify the matter further.

**E.1. The phase space and the equilibrium distribution**

We consider a cubic lattice of linear size $L$ with periodic boundary conditions. We define a set of $N_T$ temperatures, with $T_1 < T_2 < \cdots < T_{N_T}$. Our random walk moves in a discrete, very large phase space. Each state point, denoted $X, Y, Z...$ hereafter, is composed of two elements.

**Table D1.** Correlation coefficients of the scatter plot of $T_d$ against $T_s$ for the simulated two lattice sizes.

| $L$ | $r$ |
|-----|-----|
| 16  | 0.348 ± 0.008 |
| 24  | 0.342 ± 0.007 |

Figure D1. Scatter plot of $T_d$ versus $T_s$. We present the $L = 16$-data (top) and the $L = 24$-ones (bottom). Points are calculated with a special procedure. First, samples are classified in deciles according to $\log(\tau_{\text{int}})$. The points coordinates were obtained by computing the median $T_d$ and the median $T_s$ within each decile (errors from bootstrap). The parallel red lines enclose the area of over-density that presents a higher correlation for later recount.
Dynamic variational study of chaos: spin glasses in three dimensions

- The spins: for each lattice site \( x \) we have \( N_T \) binary variables \( s_x^{(\alpha)} = \pm 1 \). Here, \( \alpha \) is the clone index, which takes values \( \alpha = 1, \ldots, N_T \).
- The clone permutation \( \pi: \pi \) is a permutation of \( N_T \) symbols (there are \( N_T! \) such permutations). The action of the permutation over the clone index \( \alpha, \pi(\alpha) \), has a simple interpretation: it means that clone \( \alpha \) is currently at temperature \( T_{\pi(\alpha)} \).

In order to emphasize the composite nature of our state-point, we use the notation \( X = \{ \pi, \{ s_x^{(\alpha)} \}_{\alpha=1}^{N_T} \} \). The state point can take \( N_T!\cdot 2^{N_T\cdot L^3} \) values. The position of the random walk in phase space depends on time: \( X_t = \{ \pi_t, \{ s_x^{(\alpha)} \}_{\alpha=1}^{N_T} \} \). The random walk has, by construction, the stationary distribution

\[
P_{\text{eq}}(X) = \frac{1}{N_T!} \prod_{\alpha=1}^{N_T} \frac{\exp[-H(\{ s_x^{(\alpha)} \})/T_{\pi(\alpha)}]}{Z_{T_{\alpha}}},
\]

(E.1)

where \( H \) is the Edwards–Anderson Hamiltonian defined in equation (1) and \( Z_{T_{\alpha}} \) is the partition function at temperature \( T_{\alpha} \). One can also write it as

\[
P_{\text{eq}}(X) = \frac{1}{N_T!} \prod_{\alpha=1}^{N_T} \frac{\exp[-H(\{ s_x^{\pi^{-1}(\alpha)} \})/T_{\alpha}]}{Z_{T_{\alpha}}},
\]

(E.2)

where \( \pi^{-1} \) is the inverse permutation of \( \pi \) (\( \pi(\pi^{-1}(\alpha)) = \alpha \) for any \( \alpha \)). Let us now consider the conditional probability conditioned to a given value of \( \pi \). Without loss of generality, we select \( \pi = \mathbb{1} \), the identity permutation, such that \( \mathbb{1}(\alpha) = \alpha \) for all \( \alpha \):

\[
P_{\text{eq}}(X|\pi = \mathbb{1}) = \frac{e^{-H(\{ s_1^{(1)} \})/T_1}}{Z_{T_1}} \frac{e^{-H(\{ s_2^{(2)} \})/T_2}}{Z_{T_2}} \cdots \frac{e^{-H(\{ s_{N_T}^{(N_T)} \})/T_{N_T}}}{Z_{T_{N_T}}},
\]

(E.3)

This conditional probability is a product of distributions (i.e. the spins for clones \( \alpha \neq \beta \) are statistically independent, provided that \( \pi \) is kept fixed), and the equilibrium probability distribution for the spins \( \{ s_x^{(\alpha)} \} \) is the Boltzmann distribution for temperature \( T_{\alpha} \).

Two marginal probabilities extracted from \( P_{\text{eq}}(X) \) are of interest:

- Tracing out the spin degrees of freedom in equation (E.1), one sees that the equilibrium probability for the clones permutation is uniform:

\[
P_{\text{eq,marginal}}(\pi) = \frac{1}{N_T!}.
\]

(E.4)

Specializing to clone \( \alpha \), we find \( P_{\text{eq}}(\pi(\alpha) = \beta) = 1/N_T \) for any \( \beta \). Checking that this has been achieved with good accuracy for all clones is one of the important tests of thermalization.

- The equilibrium probability for the spins of the clone currently at temperature \( T_{\beta} \), namely \( \alpha = \pi^{-1}(\beta) \), is

\[
P_{\text{eq,marginal}}(\{ s_x^{(\alpha)} \}|\pi(\alpha) = \beta) = \frac{\exp[-H(\{ s_x^{(\alpha)} \})/T_{\beta}]}{Z_{T_{\beta}}}.
\]

(E.5)

https://doi.org/10.1088/1742-5468/aaa387 27
In other words, when the random-walk equilibrates, Boltzmann equilibrium is reached at all $N_T$ temperatures: the spin configuration of the clone currently at temperature $T_\beta$ is a typical configuration of the Boltzmann distribution at such temperature.

### E.2. The random walk and its correlation functions

We consider a stationary Markov process [2]. When going from time $t$ to time $t+1$ the system is updated $X_t \to X_{t+1}$ with a time-independent rule, that only uses as input the current state $X_t$. Previous states $(X_{t-1}, X_{t-2}, \ldots)$ have no influence on the decision of where to move at time $t+1$.

In our case, the Markov dynamics is generated by a square matrix $G_{X,Y}$ of dimension $N_T!2^{N_T}L^3$ that meets two basic conditions, viz. $G_{X,Y} \geq 0$ and $\sum_X G_{X,Y} = 1$. In fact, $G_{X,Y}$ is a conditional probability: it is the probability for having $X_{t+1} = X$ when one knows that $X_t = Y$. It follows that the probability for having $X_t = k = X$, namely $P_{t=0}(X)$, obeys the master equation

$$P_{t=k}(X) = \sum_Y [G^k]_{X,Y} P_{t=0}(Y), \quad \text{(E.6)}$$

where $G^k$ is the $k$th power of the generating matrix $G$. Matrix $G$ is carefully crafted to fulfill the balance condition

$$\sum_Y G_{X,Y} P_{\text{eq}}(Y) = P_{\text{eq}}(X). \quad \text{(E.7)}$$

The balance condition states that the equilibrium distribution (E.1) is a right-eigenvector of matrix $G$, with eigenvalue 1. When combined with the master equation, the balance condition tells us that the equilibrium distribution is a stationary distribution for our random walk.

Let us consider the spectral decomposition of the initial distribution on the $N_T!2^{N_T}L^3$ right-eigenvectors of matrix $G$, $Gu_{n} = \lambda_{n} u_{n}$ (ordered in such a way that $1 > |\lambda_1| > |\lambda_2| > \ldots$):

$$P_{t=0} = P_{\text{eq}} + \sum_n c_n u_{n}. \quad \text{(E.8)}$$

The master equation implies that

$$P_{t=k} = P_{\text{eq}} + \sum_n c_n \lambda_{n}^k u_{n}. \quad \text{(E.9)}$$

15 [2] employs a reversed convention, where our $G_{X,Y}$ is named $T_{Y,X}$. As a consequence, [2] reverses the ordering of vector and matrices in matrix products—see e.g. equation (E.6).

16 Specifically, our $G$ is factorized as $G = G_{\text{Temperature Swap}}(G_{\text{Metropolis}})^{10}$. During the Metropolis part of the dynamics the spins of clone $\alpha$ evolve with a standard Metropolis dynamics at temperature $T_{\pi(\alpha)}$ (each factor $G_{\text{Metropolis}}$ corresponds to a full-lattice sweep). The permutation $\pi$ is changed by matrix $G_{\text{Temperature Swap}}$. We try to exchange sequentially $\pi^{-1}(\alpha)$ with $\pi^{-1}(\alpha+1)$, for $\alpha = 1, 2, \ldots N_T - 1$ (in this way, the clone at the lowest temperature has a theoretical chance to reach the highest temperature in a single parallel tempering iteration). Each temperature swap attempt is accepted or rejected according to a Metropolis test—see e.g. [68].
Hence, $P_{t=k}$ converges exponentially to $P_{eq}$ and the corresponding exponential auto-correlation time is $\tau_{exp} = -1/\log|\lambda_1|$.

However, the spectral analysis of the equilibrium correlation functions (see section 2) is carried in terms of the left eigenvectors of matrix $G$, $\tilde{u}_n G = \lambda_k \tilde{u}_n$. Fortunately, for any matrix, left-eigenvalues coincide with right-eigenvalues (instead, left and right eigenvectors typically differ). In fact, these are the eigenvalues appearing in equation (5), which we repeat here for the reader’s convenience:

$$\hat{C}_f(t) = \sum_n A_{n,f} \lambda_n^{[t]}, \quad \sum_n A_{n,f} = 1. \tag{E.10}$$

In particular, the constant vector $\tilde{u}_0$ ($\tilde{u}_0(X) = 1$ for all states $X$) is a left eigenvector with eigenvalue 1. The generic observable $f$ considered in equation (E.10) can be decomposed as

$$f(X) = E(f) \tilde{u}_0(X) + \sum_n B_{n,f} \tilde{u}_n(X), \tag{E.11}$$

where $E(f)$ is the equilibrium expectation value. The coefficients $A_{n,f}$ in equation (E.10) are $A_{n,f} = \hat{B}_{n,f}/(\sum_n \hat{B}_{n',f})$, where $\hat{B}_{n,f} = B_{n,f} E(\tilde{u}_n(X)[f(X) - E(f)])$.

**Figure E1.** Top: Monte Carlo history for the overlap $q_4(t)$—see equation (E.12), as computed for each of the four clones in the truncated simulation (see text). Note that our simulation time is much too short to expose the symmetry $q_4 \leftrightarrow -q_4$. As a consequence, we know for sure that thermal equilibrium has not been reached for the truncated simulation. Bottom: as in top panel, for the first four clones in one of our standard simulations with $N_T = 24$ temperatures (there were 10, completely independent, standard simulations). For each clone, the overlap $q_4(t)$ changes sign many times along the simulation (as it is to be expected for a well equilibrated simulation). Note that, with small probability, each clone reaches a state where $|q_4| \sim 0.8$. These events, which are not observed for the other three overlaps $q_a$ $a = 1,2,3$, make it particularly interesting to study the dynamics of $q_4$. 

https://doi.org/10.1088/1742-5468/aaa387
The crucial message from this analysis is that the characteristic time scales \( \tau_n \) that one identifies by studying the correlation functions, as we did in the main text, are exactly the timescales that govern the approach to equilibrium—see equation (E.9).

These characteristic times \( \tau_n \) can be obtained from any convenient observable \( f \). Whether \( f \) is a spin observable or something related to the clone permutation is immaterial. The only thing that really matters is that \( A_n = 1, f \) should be as large as possible.

\[ \lambda_n = e^{-1/\tau_n}. \]

**E.3. An example**

Just to show how deeply the spin and the temperature dynamics are intertwined, we consider an example, here, in detail. We shall consider a typical \( L = 24 \) sample instance (neither extremely easy, nor extremely hard: it roughly corresponds to percentile 90 of difficulty—see figure 3).

We consider the standard parallel tempering simulation protocol from the main text: \( N_T = 24, T_{\text{min}} = 0.698 \). For this particular sample one needs to run the simulation for \( 2 \times 10^9 \) Metropolis sweeps (for each clone) in order to meet our thermalization criteria. We also consider a truncated simulation where we only keep the lowest four temperatures: \( N_T = 4, T_{\text{min}} = 0.698, T_2 = 0.735, T_3 = 0.771 \) and \( T_4 = 0.808 \) (all four
deep in the spin glass phase, since $T_c = 1.102(3)$ [57]). The truncated simulation is also run for $2 \times 10^9$ Metropolis sweeps per clone.

Our expectation is that the standard simulation will equilibrate, while the truncated simulation will not. The rationale for this expectation is simple: in the standard simulation, each clone spends some $2 \times 10^9 / 24 \approx 8 \times 10^7$ Monte Carlo steps at the highest temperature. Yet, the exponential auto-correlation time for the Metropolis dynamics at $T = 1.6$ is about $10^4$ lattice sweeps [69]. Hence, the time spent by each clone at the highest temperature is long enough to effectively de-correlate the system. Instead, the highest temperature in the truncated simulation $T_{\text{max, truncated}} = 0.808$ lies well below $T_c$. At such a low temperature, the Metropolis dynamics is too inefficient to decorrelate the system in only $2 \times 10^9 / 4 = 5 \times 10^8$ Metropolis sweeps.

Besides the temperature dynamics already considered in the main text, we shall also study here the dynamics of spin observables. Using the fact that we have already equilibrated this sample, we have selected randomly four equilibrium spin configurations at our lowest temperature $T_{\text{min}} = 0.698$, $\{\tau_{x,a}\} a = 1, 2, 3, 4$. Then, for each clone, we compute the time-dependent overlap

$$q_{a,\alpha}(t) = \frac{1}{L^3} \sum_x \tau_{x,a} s_{x}^{(\alpha)}(t).$$

We always compute the overlap with a given clone $\alpha$, irrespective of its time-dependent temperature $T_{\pi_t(\alpha)}$.

We compute the overlaps $q_{a,\alpha}(t)$ from a set of ten new standard simulations ($N_T = 24$), with a random start, where we measure the overlaps very often (every $5 \times 10^4$ Metropolis sweeps). We also compute the overlaps $q_{a,\alpha}(t)$ from our new truncated simulation with $N_T = 4$ (the truncated simulation had a random start, as well). Recall that, as we said above, the spin masks $\{\tau_{x,a}\}$ are taken from the previous sets of simulations that were discussed in the main text.

The global spin flip symmetry of the Edwards–Anderson Hamiltonian implies that the equilibrium distribution for $q_{a,\alpha}$ is symmetric under $q_{a,\alpha} \leftrightarrow -q_{a,\alpha}$. It is important to
check this symmetry, since it is believed that the largest dynamical barriers are related to global spin-flips [70].

The Monte Carlo history of the time-dependent overlap with \( \tau_4 \), that we call \( q_4 \) in figure E1, shows very clearly that the truncated simulation is not able to reach thermal equilibrium within the time span of our simulations. The Monte Carlo histories for the other overlaps (not shown), \( q_{a=1,2,3} \) are qualitatively similar. Instead, the standard simulation displays the expected symmetry under \( q_{4,a} \leftrightarrow -q_{4,a} \). The Monte Carlo histories (in the standard simulation) for \( q_{a,a} \) with \( a = 1,2,3 \) (not shown) are symmetric as well. Only \( q_4 \) uncovers a state that arises with small probability, characterized by \( |q_4| \sim 0.8 \). This feature suggests that \( q_4 \) is the most interesting overlap to look at.

In order to make the above impressions quantitative, we show in figure E2 some equilibrium correlation functions, which can be computed, of course, only for the standard simulation. As could be expected from appendix E.2, the very same exponential auto-correlation time is computed from the temperature random walk, or from the \( q_4 \) correlation (specifically, and measuring time in Metropolis sweeps, we find \( 10^{-7} \tau_{\text{exp}} = 3.0(4) \) from \( q_4 \), while we find the fully compatible value \( 10^{-7} \tau_{\text{exp}} = 3.1(6) \) from the \( T \) random-walk). One could conclude from figure E2 that the computation of \( \tau_{\text{exp}} \) is simpler by considering \( q_4 \) than by studying the temperature random walk. This is a misleading conclusion, though: we had to equilibrate the system, in the first place, in order to find the spin mask \( \{\tau_{x,a}=4\} \) that defines the overlap \( q_4 \). Furthermore, the other spin masks, \( \{\tau_{x,a=1,2,3}\} \), turned out not to be particularly useful in the computation of the exponential auto-correlation time. It is in no way guaranteed that one can identify an interesting overlap by randomly picking a small number of equilibrated configurations.

Finally, one could consider a different question. Figure E1 shows beyond any question that the truncated simulation does not reach equilibrium. However, there are only four clones in that run, and one could believe that it should not be that difficult to equilibrate the clone permutation. The question is investigated in figure E3 by means of an occupation histogram (it is not possible to compute equilibrium correlation functions for a simulation that does equilibrate). The answer to our query is an unqualified no: the fact that the spins are out from equilibrium also makes it impossible to equilibrate the clone permutations.

References

[1] Sokal A 1992 Quantum Fields on the Computer ed M Creutz (Singapore: World Scientific)
[2] Sokal A D 1997 Monte Carlo methods in statistical mechanics: foundations and new algorithms Functional Integration: Basics and Applications (1996 Cargèse School) ed C DeWitt-Morette et al (New York: Plenum)
[3] Newman M E J and Barkema G T 1999 Monte Carlo Methods in Statistical Physics (Oxford: Clarendon)
[4] Landau D P and Binder K 2005 A Guide to Monte Carlo Simulations in Statistical Physics 2nd edn (Cambridge: Cambridge University Press)
[5] Mézard M, Parisi G and Virasoro M 1987 Spin-Glass Theory and Beyond (Singapore: World Scientific)
[6] Young A P 1998 Spin Glasses and Random Fields (Singapore: World Scientific)
[7] Cavagna A 2009 Phys. Rep. 476 51

18 The alert reader will point out that the eigenvectors of the dynamical matrix \( G \) can be classified according to their parity with respect to global spin-flip symmetry. However, because spin-flip symmetry is spontaneously broken in the low temperature-phase, spin-flip transitions are exponentially (in some power of \( L \) suppressed in local Monte Carlo at fixed temperature. The only efficient mechanism for producing a global spin-reversal is having the clone travel to the high-temperature end of the parallel tempering temperature grid.
Dynamic variational study of chaos: spin glasses in three dimensions

[8] Marinari E and Parisi G 1992 Europhys. Lett. 19 451
[9] Geyer C J 1991 Markov chain Monte Carlo maximum likelihood Computing Science and Statistics: Proc. of the 23rd Symp. on the Interface ed E M Keramigas (Fairfax: Interface Foundations)
[10] Hukushima K and Nemoto K 1996 J. Phys. Soc. Japan 65 1604
[11] Marinari E 1998 Optimized Monte Carlo methods Advances in Computer Simulation ed J Kerstész and I Kondor (Berlin: Springer)

[12] Sugita Y and Okamoto Y 1999 Chem. Phys. Lett. 314 141
[13] Calvo F 2005 J. Chem. Phys. 123 124106
[14] Earl D J and Deem M W 2005 Phys. Chem. Chem. Phys. 7 3910
[15] Brenner P, Sweet C R, Von Handorf D and Izaguirre J A 2007 J. Chem. Phys. 126 074103
[16] Bittner E, Nu A and Janke W 2008 Phys. Rev. Lett. 101 130603
[17] Malakis A and Papakonstantinou T 2013 Phys. Rev. E 88 013312
[18] Sabo D, Meuwly M, Freeman D L and Doll J D 2008 J. Chem. Phys. 128 174109
[19] Katzgraber H G, Trebst S, Huse D A and Troyer M 2006 J. Stat. Mech. P03018
[20] Fernandez L A, Martin-Mayor V, Perez-Gaviro S, Tarancon A and Young A P 2009 Phys. Rev. B 80 024422
[21] Alvarez Baños R et al and Janus Collaboration 2010 J. Stat. Mech. P06026
[22] Barahona F 1982 J. Phys. A: Math. Gen. 15 3241
[23] Edwards S F and Anderson P W 1975 J. Phys. F: Met. Phys. 5 965
[24] Edwards S F and Anderson P W 1976 J. Phys. F: Met. Phys. 6 1927
[25] McKay S R, Berker A N and Kirkpatrick S 1982 Phys. Rev. Lett. 48 767
[26] Bray A J and Moore M A 1987 Phys. Rev. Lett. 58 57
[27] Banavar J R and Bray A J 1987 Phys. Rev. B 35 8888
[28] Kondor I 1989 J. Phys. A: Math. Gen. 22 L163
[29] Kondor I and Véghő 1993 J. Phys. A: Math. Gen. 26 L641
[30] Billoire A and Marinari E 2000 J. Phys. A: Math. Gen. 33 L265
[31] Rizzo T 2001 J. Phys. A: Math. Gen. 34 5531
[32] Mulet R, Pagnani A and Parisi G 2001 Phys. Rev. B 63 184438
[33] Billoire A and Marinari E 2002 Europhys. Lett. 60 775
[34] Krzakala F and Martin O C 2002 Eur. Phys. J. B 28 199
[35] Rizzo T and Crisanti A 2003 Phys. Rev. Lett. 90 137201
[36] Sasaki M, Hukushima K, Yoshino H and Takayama H 2005 Phys. Rev. Lett. 95 267203
[37] Katzgraber H G and Krzakala F 2007 Phys. Rev. Lett. 98 017201
[38] Parisi G and Rizzo T 2010 J. Phys. A: Math. Theor. 43 235003
[39] Fernandez L A, Martin-Mayor V, Parisi G and Seoane B 2013 Europhys. Lett. 103 67003
[40] Billoire A 2014 J. Stat. Mech. P04016
[41] Wang W, Machta J and Katzgraber H G 2015 Phys. Rev. B 92 094410
[42] Fernández L A, Marinari E, Martin-Mayor V, Parisi G and Yllanes D 2016 J. Stat. Mech. 123301
[43] Martin-Mayor V and Hen I 2015 Sci. Rep. 5 15324
[44] Jonason K, Vincent E, Hammann J, Bouchaud J P and Nordblad P 1998 Phys. Rev. Lett. 81 3243
[45] Bellon L, Ciliberto S and Laroche C 2000 Europhys. Lett. 51 551
[46] Vincent E, Depuis V, Alba M, Hammann J and Bouchaud J P 2000 Europhys. Lett. 50 674
[47] Bouchaud J P, Doussineau P, de Lacerda-Árboz T and Levelut A 2001 Eur. Phys. J. B 21 335
[48] Ozon F, Narita T, Knaebel A, Debrégeas Hébrard P and Munch J P 2003 Phys. Rev. E 68 032401
[49] Yardımcı H and Leheny R L 2003 Europhys. Lett. 62 203
[50] Mueller V and Shchur Y 2004 Europhys. Lett. 65 137
[51] Guchhait S and Orbach R L 2015 Phys. Rev. B 92 214418
[52] Katzgraber H G, Hamze F, Zhu Z, Ochoa A J and Munoz-Bauza H 2015 Phys. Rev. X 5 031026
[53] Marshall J, Martin-Mayor V and Hen I 2016 Phys. Rev. A 94 012320
[54] Billoire A, Fernandez L A, Maiorano A, Marinari E, Martin-Mayor V, Moreno-Gordo J, Parisi G, Ricci-Tersenghi F and Ruiz-Lorenzo J J 2017 Phys. Rev. Lett. 119 037203
[55] Fernández L A and Martin-Mayor V 2015 Phys. Rev. B 91 174202
[56] Manssen M and Hartmann A K 2015 Phys. Rev. B 91 174433
[57] Baity-Jesi M et al and Janus Collaboration 2013 Phys. Rev. B 88 224416
[58] Lulli M, Bernaschi M and Parisi G 2015 Comput. Phys. Commun. 196 290
[59] Leuzzi L, Parisi G, Ricci-Tersenghi F and Ruiz-Lorenzo J J 2008 Phys. Rev. Lett. 101 170203
[60] Baños R A, Fernandez L A, Martin-Mayor V and Young A P 2012 Phys. Rev. B 86 134416
[61] Fernández L A, Martin-Mayor V, Parisi G and Seoane B 2010 Phys. Rev. B 81 134403
[62] Amit D J and Martin-Mayor V 2005 Field Theory, the Renormalization Group and Critical Phenomena 3rd edn (Singapore: World Scientific)

https://doi.org/10.1088/1742-5468/aaa387
[63] Belletti F et al and Janus Collaboration 2009 Comput. Sci. Eng. 11 48
[64] Zinn-Justin J 2005 Quantum Field Theory and Critical Phenomena 4th edn (Oxford: Clarendon)
[65] Belletti F et al and Janus Collaboration 2008 Phys. Rev. Lett. 101 157201
[66] Belletti F et al and Janus Collaboration 2009 J. Stat. Phys. 135 1121
[67] Baity-Jesi M et al 2017 Proc. Natl Acad. Sci. 114 1838
[68] Marinari E, Parisi G and Ruiz-Lorenzo J J 1998 Numerical simulations of spin glass systems Spin Glasses and Random Fields ed A P Young (Singapore: World Scientific)
[69] Ogielski A T 1985 Phys. Rev. B 32 7384
[70] Billoire A and Marinari E 2001 J. Phys. A: Math. Gen. 34 L727