Dynamic Variational Study of Chaos: Spin Glasses in Three Dimensions

A. Billoire[1], L. A. Fernandez[2,3], A. Maiorano[4,3], E. Marinari[4,5,6], V. Martin-Mayor[2,3], J. Moreno-Gordo[3,7], G. Parisi[4,5,6], F. Ricci-Tersenghi [4,5,6], J. J. Ruiz-Lorenzo[8,3]

[1] Institut de Physique Théorique, CEA Saclay and CNRS, 91191 Gif-sur-Yvette, France.
[2] Depto. Física Teórica I. Facultad de Ciencias Físicas. Universidad Complutense de Madrid. Madrid 28040. Spain.
[3] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), 50018 Zaragoza, Spain.
[4] Dipartimento di Fisica, Sapienza Università di Roma, I-00185 Rome, Italy.
[5] Nanotec, Consiglio Nazionale delle Ricerche, I-00185 Rome, Italy.
[6] Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1, I-00185 Rome, Italy.
[7] Departamento de Física Teórica, Universidad de Zaragoza, 50009 Zaragoza, Spain.
[8] Departamento de Física and Instituto de Computación Científica Avanzada (ICCAEx), Universidad de Extremadura, 06071 Badajoz, Spain.

Abstract.

We have introduced a variational method to improve the computation of integrated correlation times in the 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 spin by means of the study of correlations between different 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 distribution and it obeys finite-size scaling.

PACS numbers: 75.10.Nr,71.55.Jv,05.70.Fh

1. Introduction

Markov-chain Monte Carlo methods are a crucial tool to study non-perturbative problems in Statistical Mechanics and Quantum Field Theory [1, 2, 3, 4]. A major problem arises however when studying systems with rugged free-energy landscapes (think, for instance, of spin-glasses [5, 6], or glass-forming liquids [7]). The presence of many free-energy local minima often causes the simulation to get trapped and, as a consequence, a correct sampling of the phase space is impeded.

The Parallel-Tempering method (also known as exchange Monte Carlo) was devised to overcome these difficulties [8, 9]. One considers N_T system copies, each at a different temperature $T_1 < T_2 < \cdots < T_{N_T}$. The system copies are aimed to be statistically independent: the target probability distribution for the N_T systems is then the product of the Boltzmann distribution at the different temperatures. A Parallel Tempering simulation consists of two alternating steps. First, each system copy undergoes independently a standard Monte Carlo dynamics at its own temperature (e.g. Metropolis). Second, pairs of system copies attempt to exchange their temperatures ‡.

The rationale behind Parallel Tempering is simple. Each system copy performs a temperature random walk. When a system copy is at a low temperature, it explores the nearby free-energy minima only. Instead, when the temperature is high, free-energy barriers are easily overcome and distant minima can be reached. Therefore, for Parallel Tempering to thermalize, it is crucial that any system copy spend its time evenly at every temperature.

In fact, Parallel Tempering is so widespread that considerable efforts have been devoted to improve it. Different temperature-exchange rules have been developed and tested [10, 11, 12, 13, 14, 15]. Furthermore, it has been suggested that a significant gain can be achieved by optimizing the choice of the N_T temperatures [16, 17].

In order to assess the relative merits of the above suggestions, one needs a quantitative method. The theory of Markov chains suggest considering the *exponential* autocorrelation time $\tau_{\rm exp}$ of the Monte Carlo dynamics as the relevant figure of merit [2] (for the sake of completeness, we provide some definitions below). Indeed, $\tau_{\rm exp}$ tells us how long we should wait before equilibrium is reached. Unfortunately, $\tau_{\rm exp}$ is an elusive quantity. In the context of a Parallel Tempering simulation, it has been suggested that $\tau_{\rm exp}$ is best computed by studying the temperature-flow of the system copies [18, 19].

We shall specialize in the determination of $\tau_{\rm exp}$ 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 [20]. Furthermore, spin-glasses expose in a very direct way the major problems that a Parallel-Tempering simulation faces.

To be specific, we shall be considering the 3d Edwards-Anderson model [21, 22]. Our Ising variables ($s_i = \pm 1$) occupy the nodes of a cubic lattice of size L with periodic

‡ The temperature-exchange rule is designed to have the target probability distribution as the unique equilibrium measure.

boundary conditions. Spins interact with their nearest lattice-neighbors through the Hamiltonian

$$H = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j \,, \tag{1}$$

where quenched couplings are drawn (once for ever) from a bimodal probability distribution (so $J_{ij} = \pm 1$ 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 (the larger, the better) needs to be simulated due to non self averaging §. Besides, below the critical temperature T_c, the value of $\tau_{\rm exp}$ (i.e. the computational difficulty) presents huge fluctuations from sample to sample [18, 19] (see also Fig. B1, below). These fluctuations make the problem of computing $\tau_{\rm exp}$ a practical one: knowing the value of $\tau_{\rm exp}$ for each sample allows to save huge amount of computer time by setting the chain length proportional to its own $\tau_{\rm exp}$.

Actually, there is a fairly general physical mechanism behind the dramatic dispersion of $\tau_{\rm exp}$. We are referring to the temperature chaos effect [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40]. 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 [37, 41, 40]. Indeed, for some samples one encounters chaotic-events taking place at well defined temperatures, in the form of major changes of the encountered spin configurations. In fact, chaotic events are remindful of first-order phase transitions, of course rounded in our finite systems. 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). What we do know is that, 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 [37]. Lowering the temperature T_A, while keeping the size fixed, increases as well the probability of encountering a chaotic event.

As it is intuitively obvious, temperature chaos turns out to be a major obstacle for the Parallel Tempering temperature flow [19, 37, 41, 40]. Furthermore, temperature chaos is relevant in the analysis of spectacular experiments [42, 43, 44, 45, 46, 47, 48, 49] and in the performance analysis of commercial quantum annealers [41, 50, 51].

Here, we revisit the problem of estimating $\tau_{\rm exp}$ and present a variational method that can potentially save much computation time. Indeed, very often the simulation needs to be extended just because of the difficulties encountered in the computation of $\tau_{\rm exp}$. Having in our hands a safe mechanisms to estimate $\tau_{\rm exp}$ in an automated way (the number of samples go by the thousands) will avoid many unnecessary extensions of the simulation length. Besides, we investigate further the relationship between temperature chaos, which is a *static* equilibrium feature, and $\tau_{\rm exp}$ which characterizes a Markov-chain *dynamics*.

§ Stricitly speaking, non-selfaveraging occurs only when the correlation-length becomes similar to the system size (which is usually the case at the temperatures of interest).

The remaining part of this paper is organized as follow. In section 2 we introduce the different time scales that characterize a Monte Carlo Markov chain. Our simulations are described in Sect. 3. We present our characterization of temperature chaos in Sect. 4. The variational method for the computation of the autocorrelation time $\tau_{\rm exp}$ is in Sect. 4.1. Section 4.2 is devoted to the study of the scaling properties of the Parallel Tempering method $\tau_{\rm exp}$. We deepen the study of the thermodynamic equilibrium features that characterize temperature chaos [37] in Sect. 4.3. The relationship between static and dynamic chaos indicators is studied in Sect. 4.4. Our discussion of results concludes the paper in Sect. 5. We provide in appendix Appendix A a description of the parameters of our simulations; in appendix Appendix B we discuss the particular choice of samples we use [52]; in appendix Appendix C, we describe in detail the geometry used in our implementation of the synchronous multispin coding; finally, in appendix Appendix D we discuss some quantities which are surprisingly unrelated to chaos.

2. Time scales in a Markov chain

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

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 ($\tau_{\rm exp}$). In addition to this time scale, for any physical quantity f we can define a second time scale, the integrated autocorrelation time ($\tau_{{\rm int},f}$), which controls statistical errors in measuring f: two already equilibrated configurations whose time difference is $2\tau_{{\rm int},f}$ 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:

$$\tau_{{\rm int},f} \le \tau_{\rm exp} \,. \tag{2}$$

The crucial point is that $\tau_{{\rm int},f}$ is relatively easy to compute. Instead $\tau_{\rm exp}$ is rather elusive. Hence, we shall use Eq. (2) for a variational method analogous to the Rayleigh-Ritz variational principle in Quantum Mechanics. In Sect. 4.1, we shall try different quantities f and compute $\tau_{{\rm int},f}$ for each of them. The largest value of $\tau_{{\rm int},f}$ will be our variational estimate for $\tau_{\rm exp}$.

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

$$C_f(t) = E\left[f(t_1) f(t_2)\right] - E\left[f(t_1)\right]^2 \,, \quad t = t_1 - t_2 \,, \tag{3}$$

where $E[\ldots]$ stands for the expectation value and the two times t_1 and t_2 are large enough to reach equilibrium [hence $E\left[f(t_1)\right] = E\left[f(t_2)\right]$ and $C_f(t) = C_f(-t)$]. The integrated

autocorrelation time is defined from the normalized correlation function $\hat{C}_f(t)$:

$$\hat{C}_f(t) \equiv \frac{C_f(t)}{C_f(0)}\,, \quad \tau_{\mathrm{int},f} = \frac{1}{2} + \sum_{t=1}^{\infty} \hat{C}_f(t)\,. \tag{4}$$

The normalized autocorrelation function can be expressed in terms of the eigenvalues λ_n of the transition probability matrix projected on the subspace orthogonal to its eigenvalue 1 ($1 > |\lambda_1| \geq |\lambda_2| \geq \ldots$), see Ref. [2],

$$\hat{C}_f(t) = \sum_n A_{n,f} \lambda_n^{|t|}\,, \quad \sum_n A_{n,f} = 1\,, \tag{5}$$

where the index n runs from 1 to $N_T! 2^{N_T L^D} - 1$, in our case.

The amplitudes $A_{n,f}$ depend on f, while the λ_n are f-independent. In terms of the $A_{n,f}$'s and λ_n's one has

$$\tau_{\mathrm{int},f} = \frac{1}{2} \; + \; \sum_n A_{n,f} \, \frac{\lambda_n}{1-\lambda_n}\,. \tag{6}$$

Now, in practical applications the (leading) $A_{n,f}$'s and λ_n's are real positive. Hence, $\lambda_n = \mathrm{e}^{-1/\tau_n}$ defines the characteristic times τ_n. The exponential autocorrelation time of the Markov chain τ_{exp} is just τ_1, the largest of the τ_n. Now, for $\tau_n \gg 1$ one has $\lambda_n/(1-\lambda_n) = \tau_n + \mathcal{O}(1/\tau_n)$ and Eqs. (5,6) become

$$\hat{C}_f(t) = \sum_n A_{n,f} \mathrm{e}^{-|t|/\tau_n}\,, \quad \tau_{\mathrm{int},f} = \frac{1}{2} + \sum_n A_{n,f} \tau_n\,. \tag{7}$$

The variational method in Eq. (2) follows immediately from Eq. (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 Ref. [52], in which the metastate was studied. For this reason, our realizations of disorder $\{J_{ij}\}$ (samples) are particular. In appendices Appendix A and Appendix 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 Appendix A and Appendix 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) [53] techniques, that 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 [54, 55, 56, 57, 58, 59, 3] and it is denominated multisample

multispin coding (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 a sluggish dynamic 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 Sect. 4.2). For those misbehaving instances we turned to multisite multispin coding (MUSI-MSC): the 256 bits in a computer word now code 256 distinct spins of a single replica of a single sample [53]. In this way, we execute the Metropolis algorithm in $L^3/256$ steps. Our implementation for $L = 24$ use a geometric arrangement differing from Ref. [53], as explained in Appendix Appendix C.

The simulations were carried out using either Intel Xeon E5-2680 or AMD Opteron Processor 6272 processors. 12800 samples were simulated (and 4 replicas per sample). More details of the simulations are given in Appendix 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 Sect. 4.1. The finite-size scaling of the Parallel Tempering dynamics is studied in Sect. 4.2. The static viewpoint is considered in Sect. 4.3. Finally, in Sect. 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 Eq. (2) in a variational method to estimate the exponential autocorrelation time. Consider the eigenmode expansion in Eq. (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 ‖. We shall use our physical intuition to approach this ideal.

As explained in the Introduction, the temperature chaos effect suggests to focus our attention on the temperature flow along the parallel tempering dynamics [19, 41, 40]. 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

$$E(f) = \frac{1}{N_T} \sum_{i=0}^{N_T} f(i) \,. \tag{8}$$

‖ The reader is probably used to apply this formalism to observables f that can be directly computed over a spin configuration. Here we shall need to enlarge this viewpoint.

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

$$E(f) = \frac{1}{N_T(N_T-1)} \sum_{i=0}^{N_T} \sum_{j\neq i}^{N_T} f(i,j)\,. \tag{9}$$

We will optimize three parameters: the type of function f, the temperature T^* where f is zero, and a Wilson-Kadanoff renormalization block length, $l_{\rm blo}$. We will describe these three parameters in the next paragraphs.

We consider variational test-functions f belonging to eight different classes, see Table 1. One of these classes contains the linear functions studied in Ref. [19]. All our test-function have a vanishing expectation value $E(f) = 0$. We also request $f(T^*) = 0$ for some $T^* \in \{T_1, T_2, \ldots T_{N_T}\}$. The location of T^* is our second variational parameter. Specifically, our linear test-functions are

$$T > T^*\colon\ f_{T^*}(T) = a_+\,(T - T^*)\,, \tag{10}$$

$$T < T^*\colon\ f_{T^*}(T) = a_-\,(T - T^*)\,. \tag{11}$$

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, we consider quadratic ($p = 2$) and cubic ($p = 3$) test-functions:

$$T > T^*\colon\ f_{T^*}(T) = a_+\,(T - T^*)^p (2T_{N_T} - T^* - T), \tag{12}$$

$$T < T^*\colon\ f_{T^*}(T) = a_-\,(T^* - T)^p (2T_1 - T^* - T)\,. \tag{13}$$

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 also differentiable).

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

$$C_{f,T^*}(t) = \frac{n_{\rm Met}}{N_s - t_0 - t} \sum_{t'=t_0}^{N_s - t} \tilde{f}_{T^*}(i_{t'})\tilde{f}_{T^*}(i_{t'+t})\,. \tag{14}$$

Here, N_S is the number of times we stored the state of the PT indices i_t in the hard drive. Note that t_0 must be much greater than $\tau_{\rm int}$, in order to be safely in the equilibrium regime. The parameter $n_{\rm Met}$ is the periodicity with which we record the time indices i_t (in most of this work, $n_{\rm Met} = 25000$ 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 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]

$$\tau_{\mathrm{int},f,T^*} \approx n_{\mathrm{Met}} \left[\frac{1}{2} + \sum_{t=0}^{W} \hat{C}_f(t) \right] , \tag{15}$$

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

We have found it advantageous to consider a third variational parameter l_{blo} that 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_{blo} consecutive data (see e.g. Ref. [60]). We take the average of the $f_{T^*}(i_t)$ within a single block. This operation defines a new sequence of N_S/l_{blo} renormalized times, over which the integrated autocorrelation time can be estimated just as we did for the original data $l_{\mathrm{blo}} = 1$. The estimated autocorrelation time should be rescaled by l_{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 $\tau_{\mathrm{int},f,T^*}$ is much larger that the time step in the rhs. In fact, l_{blo} can be made much greater than the τ_{exp} that we aim to estimate. As a consequence, the renormalized correlation function will vanish for times $t \neq 0$. This means than the integrated autocorrelation time will be $1/2$ (over the renormalized time-mesh). When turning back to physical time units we shall find $\tau_{\mathrm{int}} = n_{\mathrm{Met}}\, l_{\mathrm{blo}}/2$ which diverges for large l_{blo}. Hence, we need a practical way to ensure that l_{blo} is not so large that all the physical information has been erased. Our solution imposes

$$\tau_{\mathrm{int},f,T^*,l_{\mathrm{blo}}} < \frac{5}{2} n_{\mathrm{Met}}\, l_{\mathrm{blo}} , \tag{16}$$

in order to consider the results of a given l_{blo}.

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

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

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

An example of the improvement obtained in the computation of the autocorrelation function is shown in Fig. 1. As it can be inferred from Eq. (7), a major difficulty is that the amplitude for τ_{exp}, namely $A_{1,f}$, can be very small. Indeed, the correlation function considered in a previous work [19] (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.

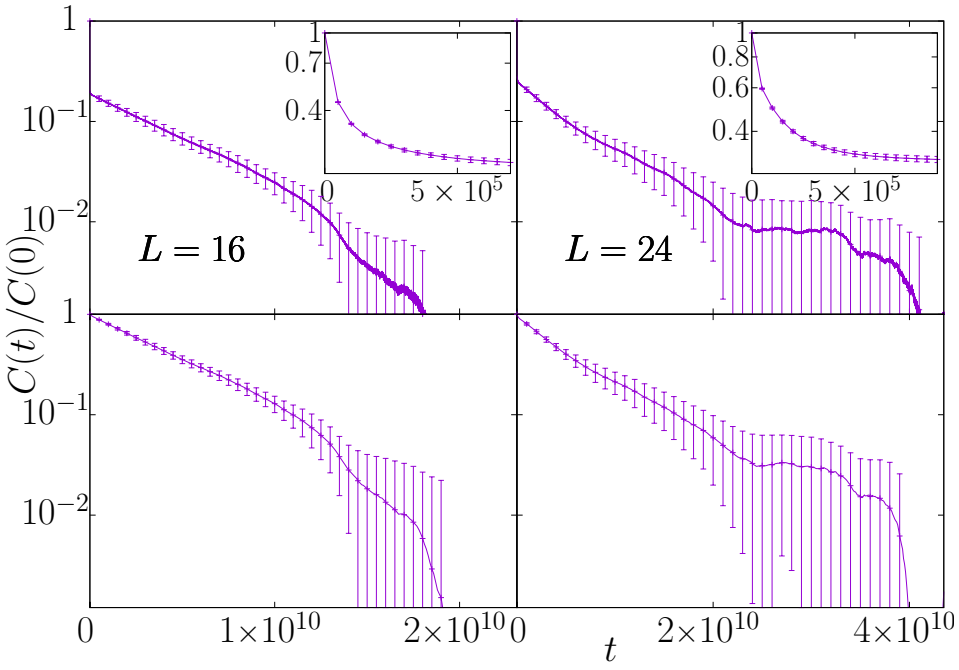

Figure 1. Auto-correlation function for the most chaotic sample for $L = 16$ (left) and $L = 24$ (right): (Top) Auto-correlation function computed using the method of [19] and (Bottom) using the variational method presented here. **Inset:** Linear-log plot showing the small t behavior of the autocorrelation function.

Identifier	Function
0	piecewise constant
1	piecewise linear
2	piecewise quadratic
3	piecewise cubic
|	OR in couples
&	AND in couples
$\wedge$	XOR in couples
$*$	Multiplication in couples

Table 1. Different choices of the function f used in the Variational Method.

L	0	1	2	3	|	&	$\wedge$	$*$	Total
16	2032	5320	3875	1374	4	115	74	6	12800
24	1556	7196	3089	820	0	127	11	1	12800

Table 2. 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.

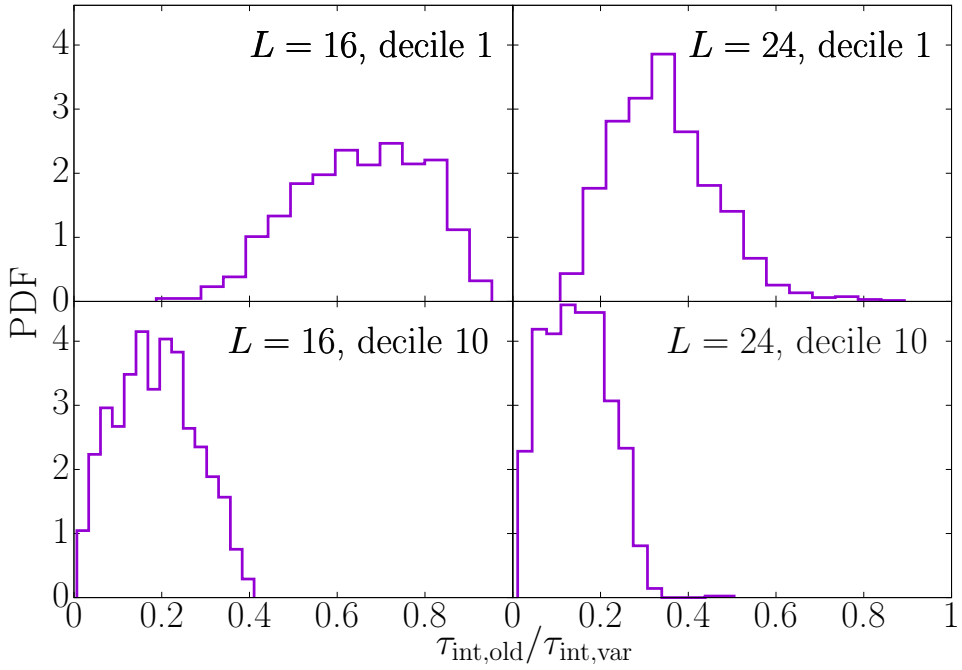

Figure 2. Conditional probability density function of the ratio $\tau_{\rm int,old}/\tau_{\rm int,var}$, given that $\tau_{\rm int,var}$ belongs to a given decile. We show the data for the first decile (left) and the 10th-decile (right) for $L = 16$ (top) and $L = 24$ (bottom).

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 to a great extent improves substantially the results obtained previously with a linear function and a parameter T^* chosen at the critical temperature [19].

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

4.2. The Finite Size Scaling of the Parallel Tempering Dynamics

In this section, we study comparatively the Parallel Tempering dynamics for $L = 8, 12, 16, 24$ and 32. Besides, we investigate the temperature chaos phenomenon from a dynamical point of view.

For the sake of clarity, from now on we shall denote the mimimal temperature reached on the Parallel Tempering simulation as $T_{\rm min}$ ($T_{\rm min}$ was denoted as T_1 in the

¶ Deciles are similar to percentiles. First, samples are ordered according to their τ. Then, we divide the samples in 10 sets (deciles) of equal size. Those samples with the lowest τ belong to decil 1, and so on.

previous section). We shall be focussing on the variational estimate $\tau_{\rm int,var}$. To lighten the notation, we shall be writting τ from now on.

An implicit assumption of our study, somewhat corroborated by the results in Sect. 4.4, is that the scaling of τ is mostly ruled by $T_{\rm min}$. Other details, such as the number of temperatures in the Parallel Tempering mesh, are expected to play a minor role (if kept within reasonable bounds).

For the comparative study of the dynamics we use the simulations at $T_{\rm min} \approx 0.7$ shown in the Table A1. 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 Ref. [19] and were obtained with the dedicated Janus computer [61]. The Janus simulation used heat-bath dynamics, rather than Metropolis, and the Parallel Tempering there had $N_T = 34$ and $T_{\rm min} = 0.703$. In order to convert the Heat-Bath autocorrelation times to Metropolis times we simulated with Janus ten randomly selected samples with both algorithms, finding consistently that $\tau_{\rm Metropolis} \approx \tau_{\rm heat-bath}/3$.

The probability distribution function of τ, $F(\tau)$, is shown in Fig. 3 (recall that we refer here to the *cumulative* distribution function, $F(\tau = 0) = 0$ and $F(\tau \to \infty) = 1$). Whereas the maximum slope of this function decreases with L for the small systems, this maximum slope seems to stabilize between $L = 24$ and $L = 32$, indeed the two later distributions can be approximately superposed by a simple translation. This observation suggest a behavior akin to critical slowing-down [62]

$$\tau \sim L^{z^{\rm PT}(T_{\rm min})} \,. \tag{17}$$

It is not obvious *a priori* that such a simple scaling should hold deep in the spin-glass phase. As an ansatz, we assume that the exponent $z^{\rm PT}$ depends mostly on the lowest temperature in the Parallel Tempering grid, $T_{\rm min}$ (and not on the number of temperatures).

As a first test of Eq. (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

$$z^{\rm PT}(L_1, L_2, p) = \frac{\log(\tau(L_1, p)/\tau(L_2, p))}{\log(L_1/L_2)} \,, \tag{18}$$

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 $z^{\rm PT}$ for three pairs of lattice sizes, (12,24), (16,24) and (24,32) and show the results in figure 4 as a function of the rank. We find that the values for the largest pair, namely $(24, 32)$, are independent of the rank, within statistical errors, in agreement with the ansatz. We also note an interesting coincidence with non-equilibrium simulations [63, 64, 53, 65], which may have some deep meaning. Indeed, in non-equilibrium conditions, it is found that the spin glass correlation length ξ, in a lattice $L \gg \xi$, grows at temperature $T = 0.7$ with the simulation time $t_{\rm w}$ as

$$\xi(t_{\rm w}) \propto t_w^{1/z(T)} \,, \quad z(T = 0.7) = 11.64(15) \,, \tag{19}$$

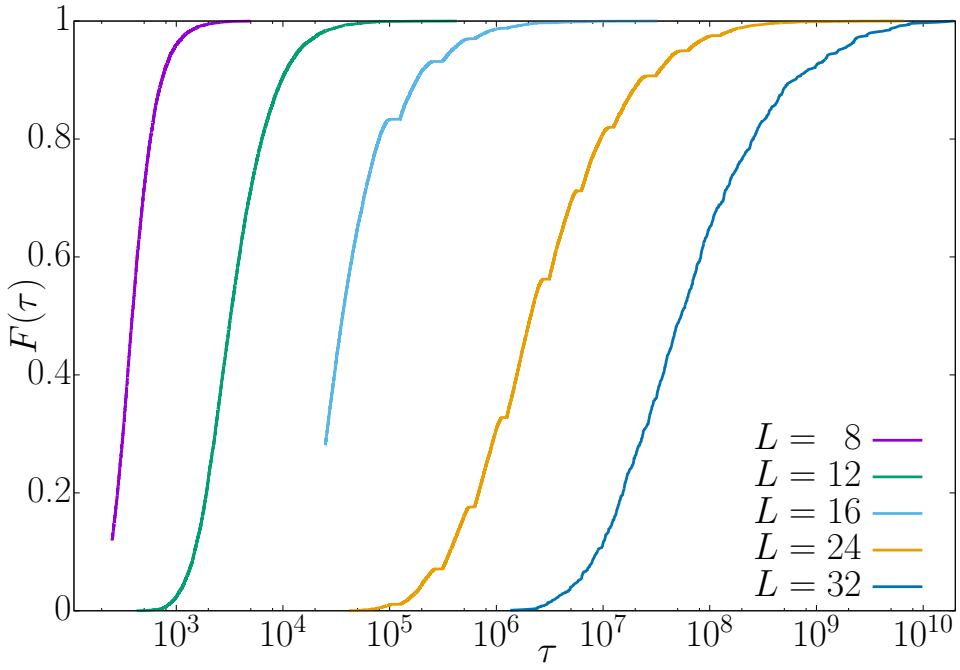

Figure 3. Empirical probability distribution of τ for $L = 8, 12, 16, 24$ and 32. Note that, for $L = 8$ and $L = 12$, some of the samples have τ smaller than our minimal resolution (indeed, if $\tau < n_{\rm Met}$ we cannot compute it safely). We show only the part of the distribution function that can be safely computed.

where $z(T)$ is the so-called dynamic critical exponent ($z(T)$ turns out to be temperature dependent in the spin-glass phase). Namely, our results for the lattice pair $(24, 32)$ suggest that

$$z(T = 0.7) \approx z^{\rm PT}(T_{\rm min} = 0.7)\,. \tag{20}$$

The estimates $z^{\rm PT}(L_1, L_2, p)$ for the lattice pairs $(16, 24)$ and $(12, 24)$ are compatible with this value for the largest percentiles ranks (i.e. the hardest samples) only.

As a further test, we can rescale the whole probability distribution by using Eqs. (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 for percentile ranks higher than 80 only and the curve corresponding to $L = 12$ seems to collapse for percentile ranks higher than 90. In figure 5 (inset), we show a log-log plot of $1-F(\tau)$ as a function of τ/L^z that emphasizes the large τ tail of the distribution. The fit presented shows that the probability density function of τ behaves, asymptotically for large y, like a fat-tailed distribution:

$$\rho(y \equiv \tau/L^z) \sim y^{-1-a_1}\,, \quad a_1 \approx 1.38\,. \tag{21}$$

We note that 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 turns out to be also asymptotic for $L \geq 24$ [37].

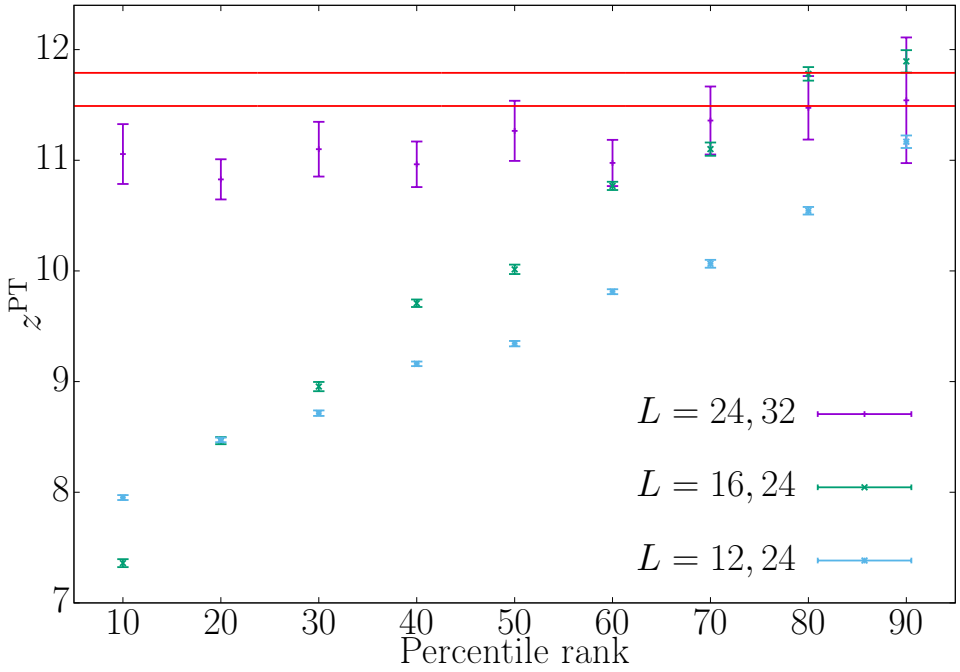

Figure 4. The effective exponent $z^{\rm PT}(L_1, L_2, p)$ for three different pairs of lattice sizes (12,24), (16,24) and (24,32) as a function of the percentile rank p. The two horizontal lines show the off-equilibrium value $z = 11.64(15)$ (see Eq. (19)). Note that the numerical values of $z^{\rm PT}$ for the largest pair are fully compatible with the off-equilibrium value.

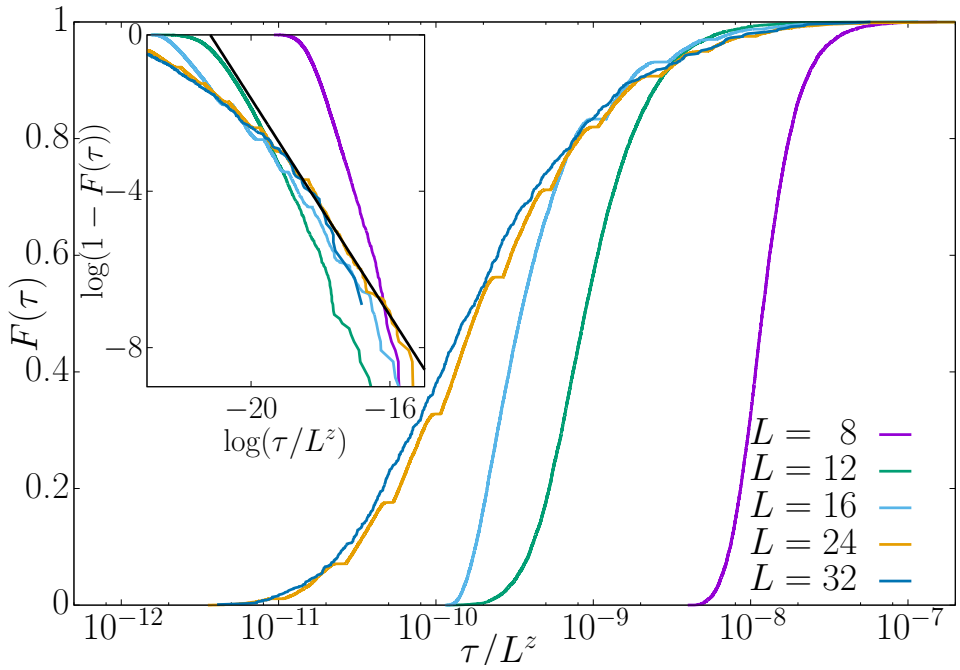

Figure 5. Probability distribution function of the rescaled variable $y = \tau/L^z$, [z is the dynamic exponent corresponding to $T_{\rm min} = 0.7$, namely $z(T = 0.7) = 11.64(15)$]. **(Inset)** Plot of $\log(1 - F(\tau))$ versus $\log(\tau/L^z)$; the straight black line is a fit to the form $a_0 -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

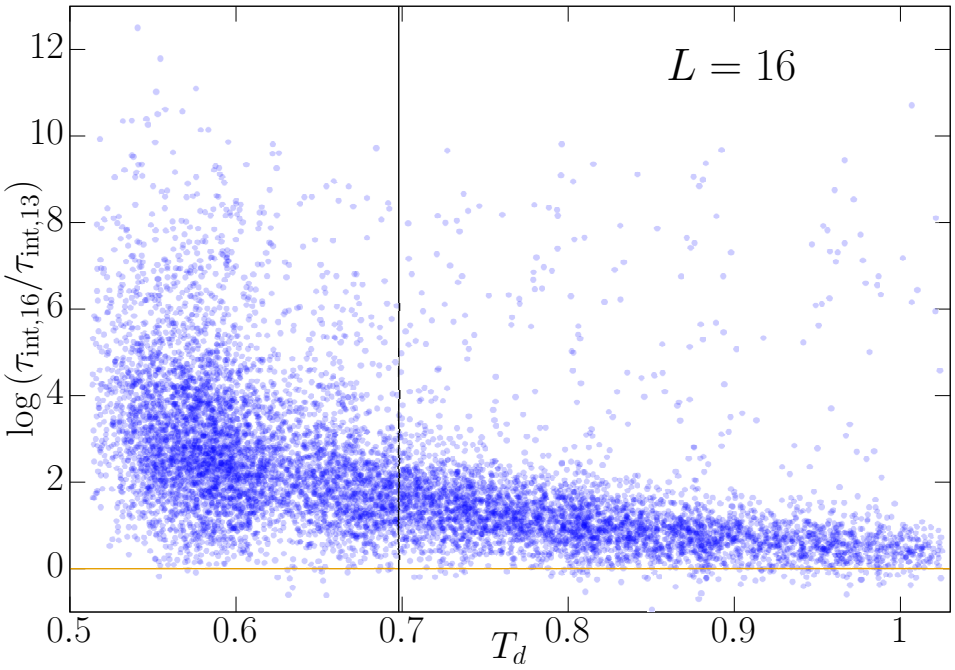

Figure 6. Scatter plot of $\log(\tau_{\rm int,16}/\tau_{\rm int,13})$ versus T_d. The lattice size is $L = 16$, $\tau_{\rm int,16}$ is the relaxation time for $N_T = 16$ ($T_{\rm min} = 0.479$), $\tau_{\rm int,13}$ is the relaxation time for $N_T = 13$ ($T_{\rm min} = 0.698$), T_d is the temperature of chaos from a dynamical point of view (defined in the variational method) of the simulation with $N_T = 16$. Of course the samples in both simulations are the same. The vertical black line represents the minimum temperature simulated in the $N_T = 13$ simulation. (We addded a small Gaussian white noise to T_d, which is a discrete variable, to avoid the cluttering of data in vertical lines).

In order to study how the range of temperatures in the Parallel Tempering affects the dynamics, we have performed an extra simulation for $L = 16$, simulating of course the very same samples. In the new simulation we take a lower minimum temperature ($T_{\rm min} = 0.479$ instead of $T_{\rm min} = 0.698$) 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 Parallel-Tempering temperature flow) that the simulation with $N_T = 13$ cannot see. This point is shown in the figure 6 where a scatter plot of $\log(\tau_{\rm int,16}/\tau_{\rm int,13})$ versus T_d for the 12800 samples is displayed ($\tau_{\rm int,16}$ and $\tau_{\rm int,24}$ are the autocorrelation times for $N_T = 13$ and 24 respectively; recall that T_d is the temperature T^* where the variational estimate $\tau_{{\rm int},f}$ reaches its maximum).

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

The same idea can be shown from a different point of view. Imagine that we have studied with great care a given sample down to some temperature $T_{\rm min}$. Can we say something about possible chaotic effects at lower temperatures? The question is

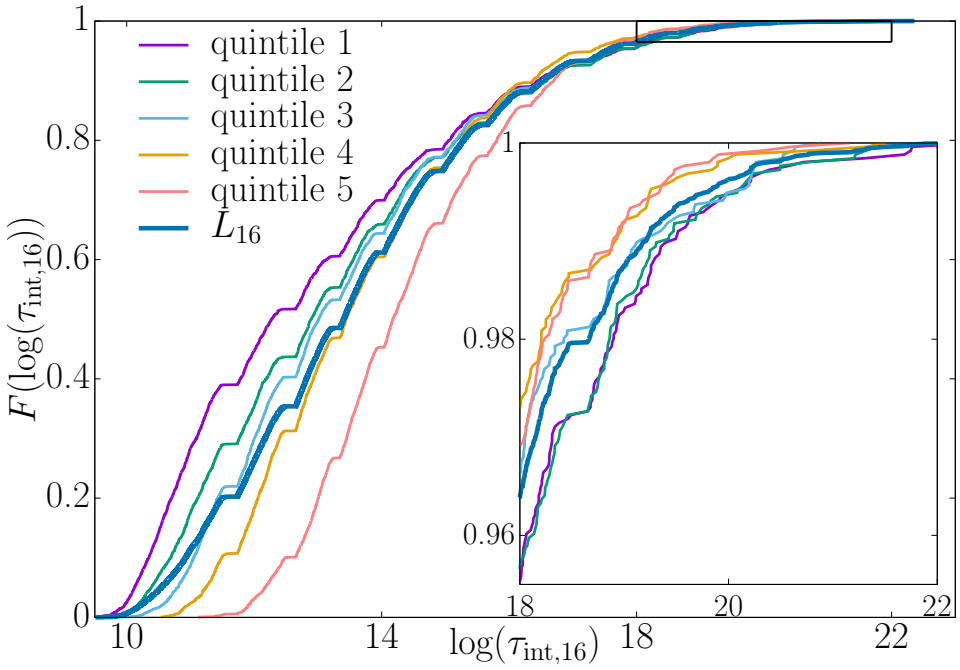

Figure 7. This figure explores how much information about the difficulty of our *hard* $L = 16$ simulation is contained in our *easy* $L = 16$ simulation (recall that the easy simulation had $N_T = 13$ temperatures and $T_{\mathrm{min}} = 0.698$, while the very same samples were studied in the hard simulation with $N_T = 16$ and $T_{\mathrm{min}} = 0.479$). To answer the question, we show the empirical probability distribution function of τ for the $N_T = 16$ simulation, conditional to the τ from $N_T = 13$ simulation belonging to a given quintile. The non-conditional probability distribution function is also shown (L_{16} curve). **Inset:** Top right blowup: For the truly hard samples, the simulation with $T_{\mathrm{min}} = 0.698$ conveys little or no information on the difficulty of the $T_{\mathrm{min}} = 0.479$ simulation.

answered by the negative in Fig. 7: The probability that a sample has a large τ_{int} for the simulation with a lower T_{min} is not correlated to the value of τ_{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

$$q_{T_1,T_2} = \frac{1}{V}\sum_x s_x^{T_1} s_x^{T_2}\,, \tag{22}$$

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 size effects. It has been suggested that static temperature chaos is a rare events driven phenomena, that should be studied via the distribution of the sample-dependent chaotic

parameter [37, 38]:

$$X^J_{T_1,T_2} = \frac{\langle q^2_{T_1,T_2}\rangle_J}{\sqrt{\langle q^2_{T_1,T_1}\rangle_J \langle q^2_{T_2,T_2}\rangle_J}}\,, \tag{23}$$

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.

It is convenient to introduce [37] the chaotic integral I

$$I = \int_{T_{\min}}^{T_{\max}} X^J_{T_{\min},T_2} dT_2\,. \tag{24}$$

A study of the temperature behavior of the chaotic parameter leads to the conclusion that chaos events happen at low temperatures only, therefore 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 the half-lower temperatures only.

Nevertheless, there exists certain samples that exhibit a huge $\tau_{\rm int}$ and have a relative large chaotic integral, so the correlation between static 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:

$$dX^J_{T_1,T_2}/dT_2\big|_{T_2=T_1} = 0\,. \tag{25}$$

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 Eq. (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 on knowing how these static and dynamics estimators are correlated.

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

$$K_i = 1 - X^J_{T_i,T_{i+1}} \tag{26}$$

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

$$I_X = aI_2 - b\min_i\left(-\log\left(K_i^2\right)\right) - c\sum_i\left(-\log\left(K_i^2\right)\right)\,, \tag{27}$$

where the coefficients a, b and c, that depends on the lattice size, are obtained through a minimization of the correlation coefficient r (recall that 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

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 Eq. 27, that maximize the correlation between I_X and $\log(\tau_{\rm int})$.

This finding is supported by Fig. 8. We see that the most chaotic samples in terms of the integrated autocorrelation time (Fig. 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 (Fig. 8, bottom), have a much smooth falling.

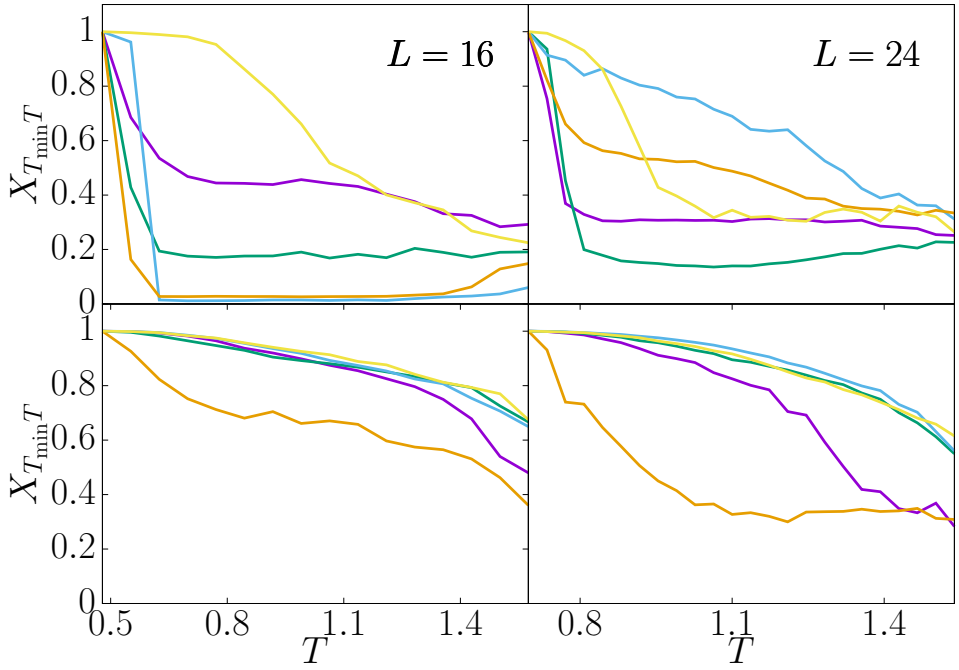

Figure 8. Plot of $X^J_{T_{\rm min},T}$ versus T for the five most chaotic samples (top) and the five less chaotic ones (bottom): $L = 16$ case (left) and $L = 24$ case (right).

In Fig. 9, we confront the most representative estimator for the dynamical chaos, namely the largest integrated autocorrelation time $\tau_{\rm int}$ found in our variational study, with the static chaotic integrals I, I_2 and I_X. We can observe how spurious values of the original parameter I (i.e. large values of I associated to large $\tau_{\rm int}$) are displaced towards lower values when we use improved parameters I_2 and I_X.

We perform the computation of the correlation coefficients [+] reported in Table 4. We observe a strong anti-correlation in I_X, that improves over the previous indicator of correlation I. [37] The improvement is less clear for I_2.

We can try to define other magnitudes (wither static or dynamical) that capture the

[+] By using bootstrap method for statistical error estimates.

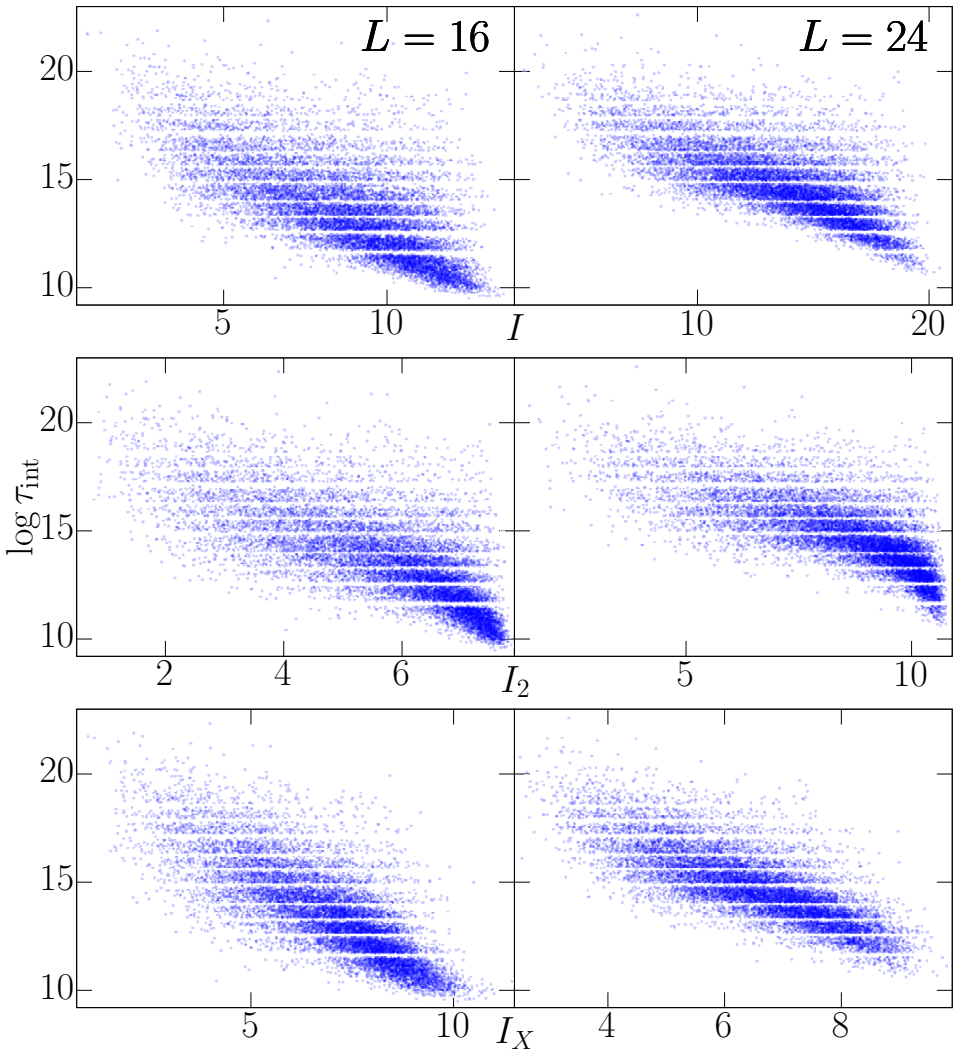

Figure 9. Scatter plot of $\log(\tau_{\rm int,var})$ versus the integrated chaotic parameter I. We present data for two lattice sizes and for the three definitions of the integrated chaotic parameter defined in the text (I, I_2 and I_X). The pattern of depleted horizontal bands is due to our choice of a few $l_{\rm blo}$.

chaos phenomenon. One possible choice is the temperature, T_s, in which $X^J_{T_{\rm min},T}$ presents the maximum (negative) slope. Unfortunately, we observe a weaker correlation between both estimators, τ and T_s, (see Fig. 10) and we can check it quantitatively through Table 5. Some further attempts along these lines are explored in Appendix Appendix D.

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 for 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).

L	Integral	r
16	I	-0.714 ± 0.005
16	I_2	-0.751 ± 0.005
16	I_X	-0.795 ± 0.004
24	I	-0.725 ± 0.005
24	I_2	-0.746 ± 0.005
24	I_X	-0.786 ± 0.004

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

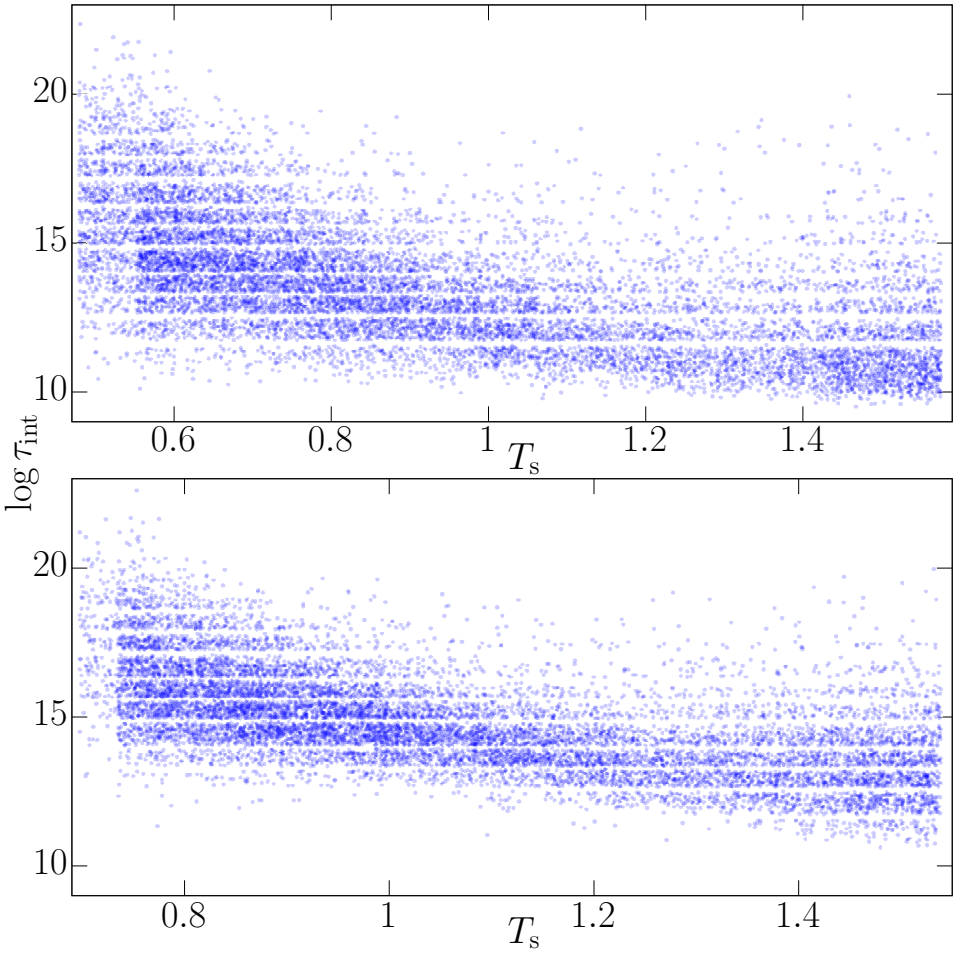

Figure 10. Scatter plot of $\log(\tau_{\rm int})$ against $T_{\rm s}$. We show $L = 16$ (top) and $L = 24$ (bottom). T_s is the temperature where $X^J_{T_{\rm min},T}$ presents the maximum (negative) slope.

We have checked that this procedure is very robust and can easily be implemented in an automatic way.

L	r
16	-0.621 ± 0.006
24	-0.621 ± 0.006

Table 5. Correlation coefficients for the scatter plot of $\log(\tau_{\rm int})$ versus $T_{\rm s}$ for the two simulated lattice sizes.

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 among these static chaotic indicators and the dynamical correlation times.

6. Acknowledgments

We thank the Janus collaboration for allowing us to analyze the $L = 32$ autocorrelation times from Ref. [19]. We also are 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 Unions 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 (*grinfishpc* 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 Ref. [52]. 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 replica (independent spin systems) for every inner and outer sample. Hence we have simulated 12800 disorder realizations (samples) with a

total of 12800×4 real spin systems. The parameters of the simulation can be found in Table A1.

The thermalization criteria that have been used is as follows (as explained above these criteria applied to every sample, individually). First of all, the number of iterations in $\tau_{\rm exp}$ units ($l_{\rm blo} = 1$) must be greater than 20; as a double-check to avoid failures in the automated fitting procedure, we recomputed $\tau_{\rm exp}$ with $l_{\rm blo} = 10$ (the total simulation length is also required to be longer than $20\tau_{\rm exp}^{l_{\rm blo}=10}$).

However, we had some additional safety checks to ensure that the computation of $\tau_{\rm exp}$ could be trusted. For those samples where any of the following two requirements was not met, we doubled the total simulation length and, only afterwards, we recomputed $\tau_{\rm exp}$. 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 $\tau_{\rm int}$, as computed for each of the four independent replica, must be less than two (for either $l_{\rm blo} = 1, 10, 100$). 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 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 Fig B1 that this is not the case for the probability distribution of τ: the probability distributions of τ for the samples sharing the same 10 inner parts are plotted separately. they are nearly indistinguishable. The average over the outer disorder (that we can call the metastate average in analogy with Ref. [52]) reduces dramatically the fluctuations due to the inner disorder. The same conclusion holds for the chaos integral (see Fig. B2)

On the other hand, the selection of the minimal tempeature in the Parallel Tempering could seem arbitrary, however the selection of $T_{\rm min}^{L=16}$ and $T_{\rm min}^{L=24}$ have been made carefully to assure that the most difficult samples had similar τ. This is shown in the figure B3.

Appendix C. The geometry of MUSI-MSC

The geometric construction explained in Ref. [53] for $L = 256$ turns out to be satisfactory as well for $L = 16$, but not for $L = 24$. Hence, we shall first recall the geometry that we employ for $L = 16$. Afterwards, we 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 Ref. [53].

MUSA-MSC

L	$L_{\rm int}$	N_T	$T_{\rm min}$	$T_{\rm max}$	$N_{\rm Met}$ ($\times 10^6$)	ps/s
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

MUSI-MSC

L	$L_{\rm int}$	N_T	$N_{\rm samp}$	$N_{\rm Met,min}$ $\times 10^6$	$N_{\rm Met,mean}$ $\times 10^6$	$N_{\rm Met,max}$ $\times 10^6$	ps/s
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

Table A1. Parameters of the simulations. L is the lattice size; $L_{\rm int}$ the size of the inner part of the lattice; N_T, $T_{\rm min}$ and $T_{\rm max}$ are the number of temperatures, the minimum and the maximum temperatures used in the Parallel Tempering method; $N_{\rm Met}$ 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 processors; $N_{\rm samp}$ denotes the number of bad samples whose simulations had to be extended in order to thermalize and finally $N_{\rm Met,min}$, $N_{\rm Met,mean}$ and $N_{\rm Met,max}$ 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 $N_{\rm MpPT} = 10$. 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.

For $L = 16$, the physical lattice of Cartesian coordinates $0 \le x, y, z < L$ is mapped to a *superspin* lattice. Each superspin is coded in a 256-bits 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-neighbors superspins. In particular, our superspins 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 superspin lattice (i_x, i_y, i_z) is

$$\begin{aligned} x &= b_x L_x + i_x\,,\ 0 \le i_x < L_x\,,\ 0 \le b_x < 8\,,\\ y &= b_y L_y + i_y\,,\ 0 \le i_y < L_y\,,\ 0 \le b_y < 8\,, \end{aligned} \tag{C.1}$$

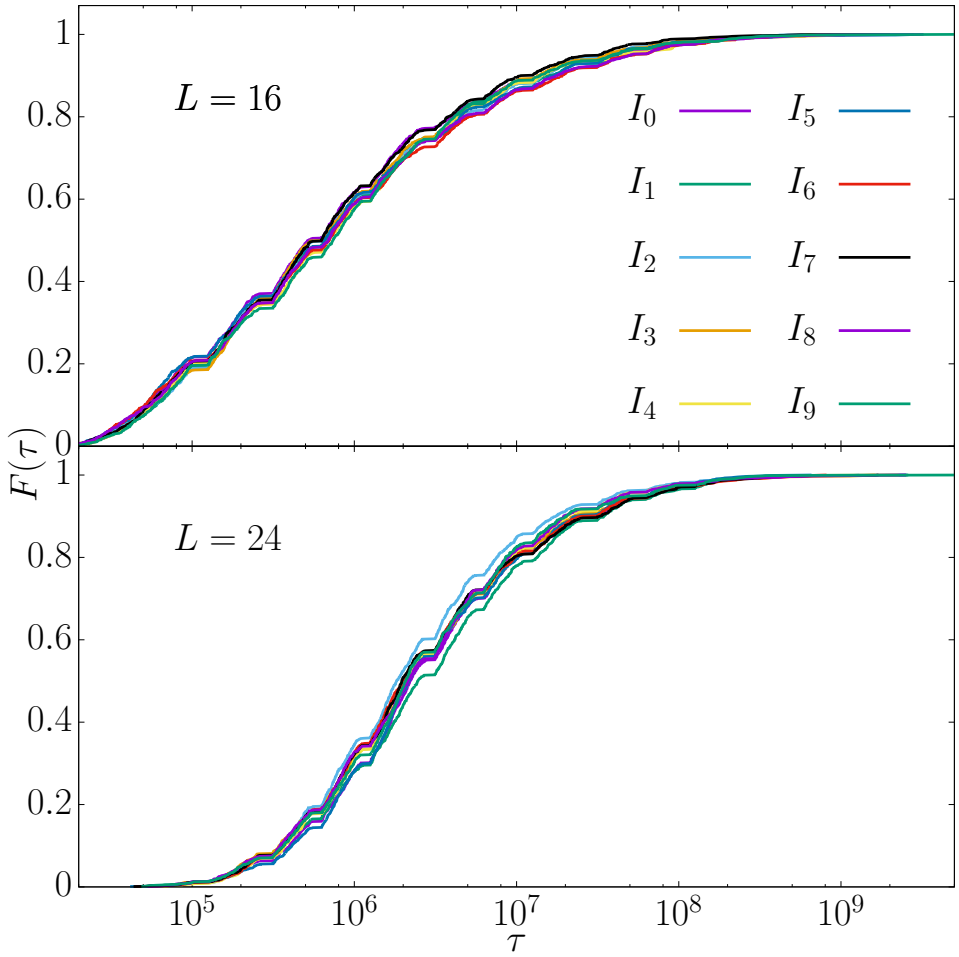

Figure B1. Empirical probability distribution function of τ represented for the 10 inner samples separately. $L = 16$ case (top) and $L = 24$ case (bottom). Averaging over the metastate (i.e. the outer samples) with fixed inner couplings reduces strongly the fluctuations between the inner samples.

$$z = b_z L_z + i_z\,,\ 0 \le i_z < L_z\,,\ 0 \le b_z < 4\,.$$

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

$$i_b = 64 b_z + 8 b_y + b_x\,,\ 0 \le i_b \le 255\,. \tag{C.2}$$

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 *even* and *odd* sublattices according to the parity of $x + y + z$. The two sublattices contain $L^3/2$ sites. Furthermore, odd spins interact only with even spins and vice versa. It follows that the update ordering is irrelevant, provide that our full-lattice sweep first updates all the (say) odd sites and next all the even sites. Now,

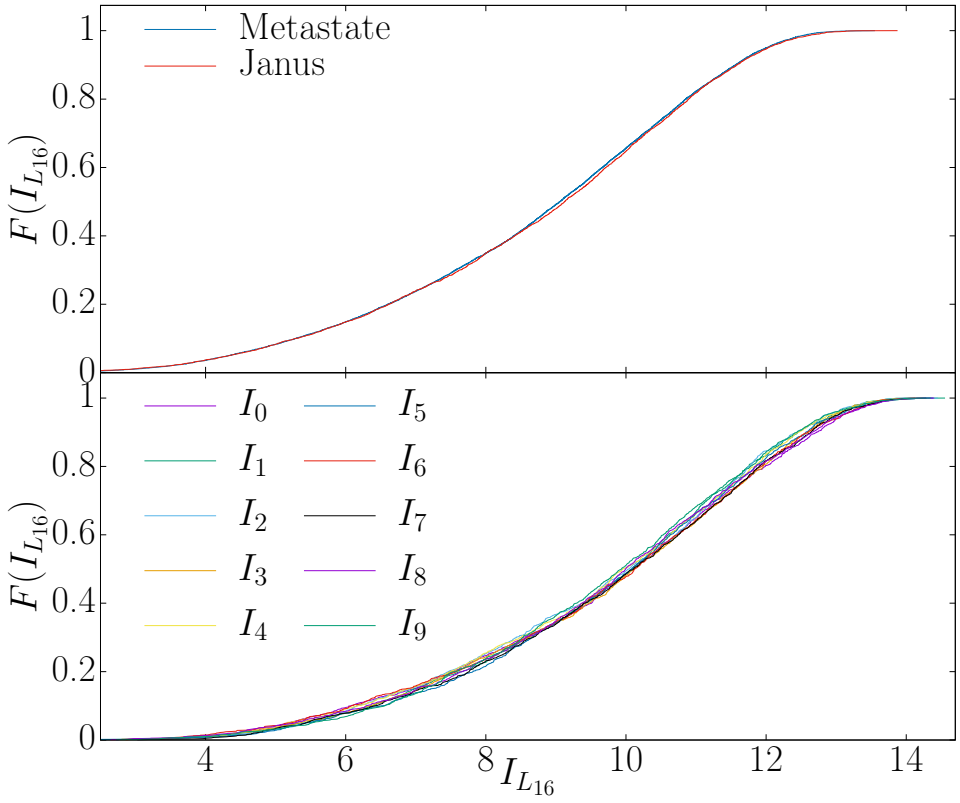

Figure B2. Empirical probability distribution function of the integrated chaotic parameter. **Top** We compare the distribution (labeled as "Metastate") obtained with our particulat choice of samples with the distribution obtained from 4000 fully independent samples (data from Janus). **Bottom:** Distributions obtained for the 10 inner samples plotted separately. Averaging over the metastate (over the outer couplings) reduces strongly the fluctuations between the innser samples.

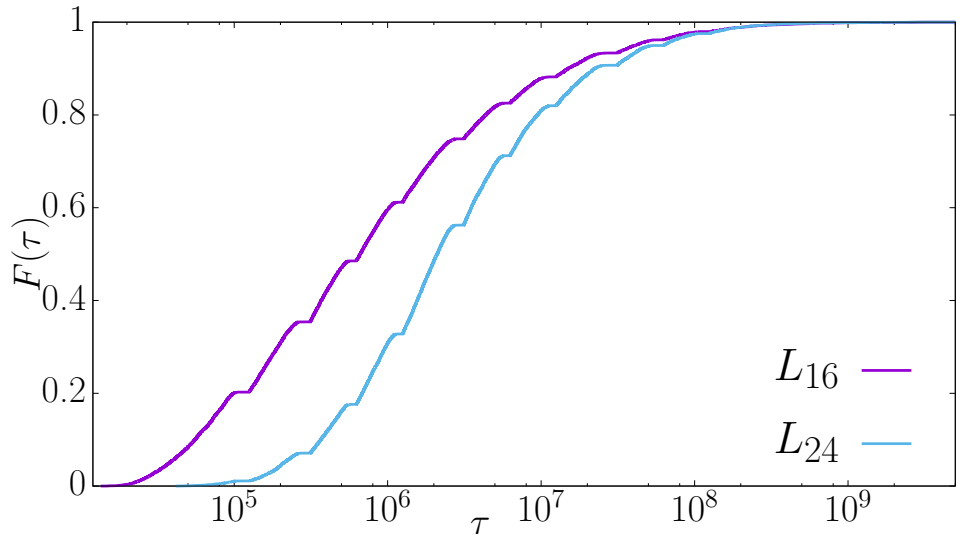

Figure B3. Empirical probability distribution function of τ. Comparison of results for the simulations ($L = 24$,$T_{\rm min} = 0.698$) and ($L = 16$,$T_{\rm min} = 0.479$). Note that at the high-end of very difficult samples, these two simulations are similarly challenging.

provided that L_x, L_y and L_z are all *even*, the parity of $x+y+z$ and $i_x+i_y+i_z$ coincide. This implies that all the spins coded in a single superspin share the same parity, making irrelevant the superspin lattice asymmetry. For $L = 16$ one finds that $L_x = L_y = 2$ and $L_z = 4$, the three of them 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 superspin lattice cannot be split into even and odd sublattices. Our solution consisted in introducing *logical* superspins of 512 physical spins, that were later on coded into two computer words of 256 bits each. The geometrical correspondence was ($L_x = L_y = L_z = L/8$)

$$\begin{aligned} x &= \tilde{b}_x L_x + j_x\,,\ 0 \le j_x < L_x\,,\ 0 \le \tilde{b}_x < 8\,,\\ y &= \tilde{b}_y L_y + j_y\,,\ 0 \le j_y < L_y\,,\ 0 \le \tilde{b}_y < 8\,,\\ z &= \tilde{b}_z L_z + j_z\,,\ 0 \le j_z < L_z\,,\ 0 \le \tilde{b}_z < 8\,. \end{aligned} \tag{C.3}$$

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

$$j_b = 64\tilde{b}_z + 8\tilde{b}_y + \tilde{b}_x\,,\ 0 \le i_b \le 511\,. \tag{C.4}$$

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 superspin coordinates (j_x, j_y, j_z) the 512 spins coded in the superspin 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 worse $L = 24$ 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 superspin lattice we found it preferable not to use concurrent threads in the simulation of a single system copies (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 along 12 CPU cores, achieving an average speed of 57 picoseconds per spin-flip.

Appendix D. Quantities not related to chaos

L	r
16	0.348 ± 0.008
24	0.342 ± 0.007

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

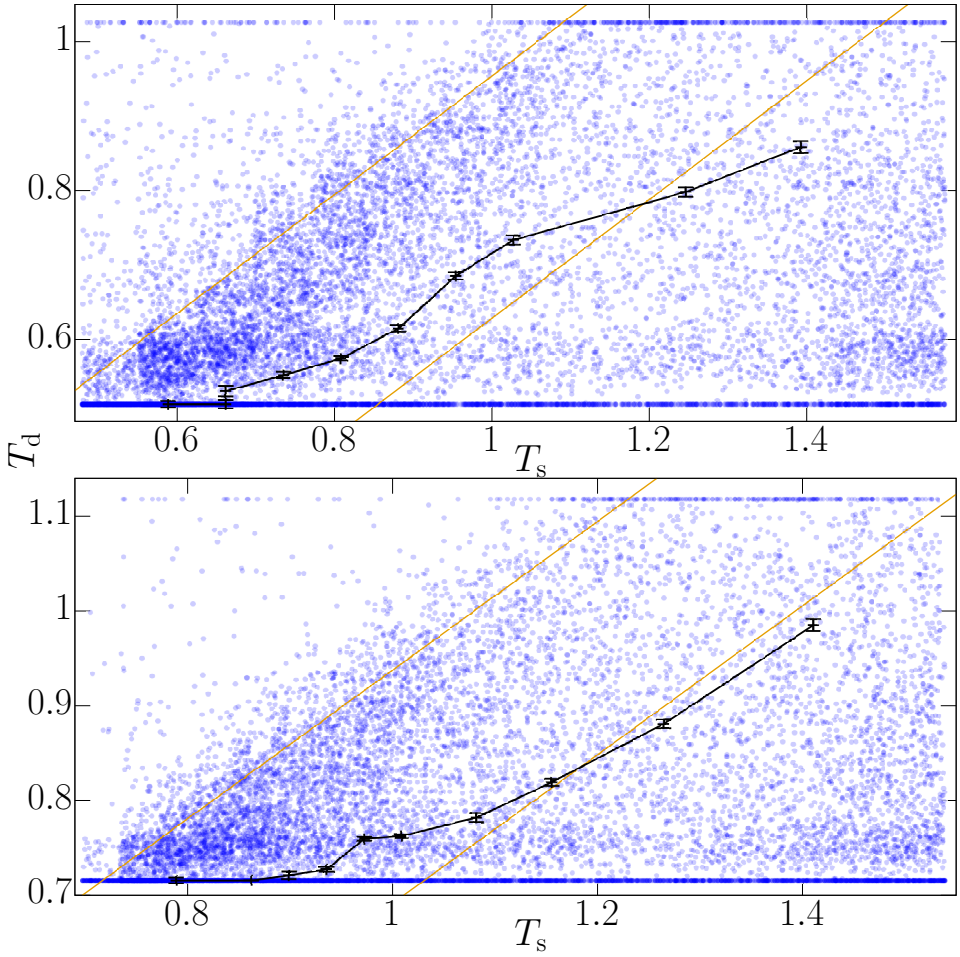

Figure D1. Scatter plot of T_d versus T_s. We present the $L = 16$-data (top) and the $L = 24$-ones (bottom). Points are calculated with an special procedure. First, samples are classified on deciles according to $\log(\tau_{\rm int})$. The points' coordinates were obtained by computing the median T_d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

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 $T_{\rm d} = \{T^* : \tau_{\rm int} = \tau_{\rm int,var}\}$ is correlated with the static temperature of chaos $T_{\rm s}$ (see Fig. D1).

In this case, Fig. D1 shows an over-density, however the points out 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.

References

[1] Sokal A 1992 *Quantum Fields on the Computer* (Singapore: Ed. Michael Creutz, 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 DeWitt-Morette C, Cartier P and Folacci A (N. Y.: Plenum)

[3] Newman M E J and Barkema G T 1999 *Monte Carlo Methods in Statistical Physics* (Oxford: Clarendon Press)

[4] Landau D P and Binder K 2005 *A Guide to Monte Carlo Simulations in Statistical Physics* 2nd ed (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 *Physics Reports* **476** 51–124

[8] Hukushima K and Nemoto K 1996 *J. Phys. Soc. Japan* **65** 1604

[9] Marinari E 1998 Optimized Monte Carlo methods *Advances in Computer Simulation* ed Kerstész J and Kondor I (Springer-Verlag)

[10] Sugita Y and Okamoto Y 1999 *Chemical Physics Letters* **314** 141 – 151 ISSN 0009-2614

[11] Calvo F 2005 *The Journal of Chemical Physics* **123** 124106

[12] Earl D J and Deem M W 2005 *Phys. Chem. Chem. Phys.* **7**(23) 3910–3916 URL `http://dx.doi.org/10.1039/B509983H`

[13] Brenner P, Sweet C R, VonHandorf D and Izaguirre J A 2007 *The Journal of Chemical Physics* **126** 074103

[14] Bittner E, Nußbaumer A and Janke W 2008 *Phys. Rev. Lett.* **101**(13) 130603

[15] Malakis A and Papakonstantinou T 2013 *Phys. Rev. E* **88**(1) 013312

[16] Sabo D, Meuwly M, Freeman D L and Doll J D 2008 *The Journal of Chemical Physics* **128** 174109

[17] Katzgraber H G, Trebst S, Huse D A and Troyer M 2006 *Journal of Statistical Mechanics: Theory and Experiment* **2006** P03018

[18] Fernandez L A, Martín-Mayor V, Perez-Gaviro S, Tarancon A and Young A P 2009 *Phys. Rev.* B **80** 024422

[19] Alvarez Baños R, Cruz A, Fernandez L A, Gil-Narvion J M, Gordillo-Guerrero A, Guidetti M, Maiorano A, Mantovani F, Marinari E, Martín-Mayor V, Monforte-Garcia J, Muñoz Sudupe A, Navarro D, Parisi G, Perez-Gaviro S, Ruiz-Lorenzo J J, Schifano S F, Seoane B, Tarancon A, Tripiccione R and Yllanes D (Janus Collaboration) 2010 *J. Stat. Mech.* **2010** P06026

[20] Barahona F 1982 *Journal of Physics A: Mathematical and General* **15** 3241

[21] Edwards S F and Anderson P W 1975 *Journal of Physics F: Metal Physics* F **5** 965

[22] Edwards S F and Anderson P W 1976 *J. Phys.* F **6** 1927

[23] McKay S R, Berker A N and Kirkpatrick S 1982 *Phys. Rev. Lett.* **48**(11) 767–770

[24] Bray A J and Moore M A 1987 *Phys. Rev. Lett.* **58**(1) 57–60

[25] Banavar J R and Bray A J 1987 *Phys. Rev. B* **35** 8888

[26] Kondor I 1989 *J. Phys. A* **22** L163

[27] Kondor I and Végsö 1993 *J. Phys. A* **26** L641

[28] Billoire A and Marinari E 2000 *J. Phys.* A **33** L265

[29] Rizzo T 2001 *J. Phys.* A **34** 5531

[30] Mulet R, Pagnani A and Parisi G 2001 *Phys. Rev.* B **63** 184438

[31] Billoire A and Marinari E 2002 *Europhys. Lett.* **60** 775

[32] Krzakala F and Martin O C 2002 *Eur. Phys. J.* B **28** 199

[33] Rizzo T and Crisanti A 2003 *Phys. Rev. Lett.* **90**(13) 137201

[34] Sasaki M, Hukushima K, Yoshino H and Takayama H 2005 *Phys. Rev. Lett.* **95** 267203

[35] Katzgraber H G and Krzakala F 2007 *Phys. Rev. Lett.* **98** 017201

[36] Parisi G and Rizzo T 2010 *Journal of Physics A: Mathematical and Theoretical* **43** 235003

[37] Fernandez L A, Martín-Mayor V, Parisi G and Seoane B 2013 *EPL* **103** 67003
[38] Billoire A 2014 *J. Stat. Mech.* **2014** P04016
[39] Wang W, Machta J and Katzgraber H G 2015 *Phys. Rev. B* **92**(9) 094410
[40] Fernández L A, Marinari E, Martín-Mayor V, Parisi G and Yllanes D 2016 *Journal of Statistical Mechanics: Theory and Experiment* **2016** 123301
[41] Martín-Mayor V and Hen I 2015 *Scientific Reports* **5** 15324
[42] Jonason K, Vincent E, Hammann J, Bouchaud J P and Nordblad P 1998 *Phys. Rev. Lett.* **81**(15) 3243–3246
[43] Bellon L, Ciliberto S and Laroche C 2000 *Europhys. Lett.* **51** 551
[44] Vincent E, Depuis V, Alba M, Hammann J and Bouchaud J P 2000 *Europhys. Lett.* **50** 674
[45] Bouchaud J P, Doussineau P, de Lacerda-Arôso T and Levelut A 2001 *Eur. Phys. J. B* **21** 335
[46] Ozon F, Narita T, Knaebel A, Debrégeas Hébraud P and Munch J P 2003 *Phys. Rev. E* **68** 032401
[47] Yardimci H and Leheny R L 2003 *Europhys. Lett.* **62** 203
[48] Mueller V and Shchur Y 2004 *Europhys. Lett.* **65** 137
[49] Guchhait S and Orbach R L 2015 *Phys. Rev. B* **92**(21) 214418
[50] Katzgraber H G, Hamze F, Zhu Z, Ochoa A J and Munoz-Bauza H 2015 *Phys. Rev. X* **5**(3) 031026
[51] Marshall J, Martin-Mayor V and Hen I 2016 *Phys. Rev. A* **94**(1) 012320
[52] 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**(3) 037203
[53] Fernández L A and Martín-Mayor V 2015 *Phys. Rev. B* **91**(17) 174202 URL `http://link.aps.org/doi/10.1103/PhysRevB.91.174202`
[54] Manssen M and Hartmann A K 2015 *Phys. Rev. B* **91**(17) 174433
[55] Baity-Jesi M, Baños R A, Cruz A, Fernandez L A, Gil-Narvion J M, Gordillo-Guerrero A, Iniguez D, Maiorano A, Mantovani F, Marinari E, Martín-Mayor V, Monforte-Garcia J, Muñoz Sudupe A, Navarro D, Parisi G, Perez-Gaviro S, Pivanti M, Ricci-Tersenghi F, Ruiz-Lorenzo J J, Schifano S F, Seoane B, Tarancon A, Tripiccione R and Yllanes D (Janus Collaboration) 2013 *Phys. Rev. B* **88** 224416
[56] Lulli M, Bernaschi M and Parisi G 2015 *Computer Physics Communications* **196** 290–303
[57] Leuzzi L, Parisi G, Ricci-Tersenghi F and Ruiz-Lorenzo J J 2008 *Phys. Rev. Lett.* **101**(10) 107203
[58] Baños R A, Fernandez L A, Martin-Mayor V and Young A P 2012 *Phys. Rev. B* **86**(13) 134416
[59] Fernández L A, Martin-Mayor V, Parisi G and Seoane B 2010 *Phys. Rev. B* **81**(13) 134403
[60] Amit D J and Martín-Mayor V 2005 *Field Theory, the Renormalization Group and Critical Phenomena* 3rd ed (Singapore: World Scientific)
[61] Belletti F, Guidetti M, Maiorano A, Mantovani F, Schifano S F, Tripiccione R, Cotallo M, Perez-Gaviro S, Sciretti D, Velasco J L, Cruz A, Navarro D, Tarancon A, Fernandez L A, Martín-Mayor V, Muñoz-Sudupe A, Yllanes D, Gordillo-Guerrero A, Ruiz-Lorenzo J J, Marinari E, Parisi G, Rossi M and Zanier G (Janus Collaboration) 2009 *Computing in Science and Engineering* **11** 48
[62] Zinn-Justin J 2005 *Quantum Field Theory and Critical Phenomena* 4th ed (Oxford: Clarendon Press)
[63] Belletti F, Cotallo M, Cruz A, Fernandez L A, Gordillo-Guerrero A, Guidetti M, Maiorano A, Mantovani F, Marinari E, Martín-Mayor V, Sudupe A M, Navarro D, Parisi G, Perez-Gaviro S, Ruiz-Lorenzo J J, Schifano S F, Sciretti D, Tarancon A, Tripiccione R, Velasco J L and Yllanes D (Janus Collaboration) 2008 *Phys. Rev. Lett.* **101** 157201
[64] Belletti F, Cruz A, Fernandez L A, Gordillo-Guerrero A, Guidetti M, Maiorano A, Mantovani F, Marinari E, Martín-Mayor V, Monforte J, Muñoz Sudupe A, Navarro D, Parisi G, Perez-Gaviro S, Ruiz-Lorenzo J J, Schifano S F, Sciretti D, Tarancon A, Tripiccione R and Yllanes D (Janus Collaboration) 2009 *J. Stat. Phys.* **135** 1121
[65] Baity-Jesi M, Calore E, Cruz A, Fernandez L A, Gil-Narvión J M, Gordillo-Guerrero A, Iiguez D, Maiorano A, Marinari E, Martin-Mayor V, Monforte-Garcia J, Muoz Sudupe A, Navarro D, Parisi G, Perez-Gaviro S, Ricci-Tersenghi F, Ruiz-Lorenzo J J, Schifano S F, Seoane B,

Tarancón A, Tripiccione R and Yllanes D 2017 *Proceedings of the National Academy of Sciences* **114** 1838–1843