State Hierarchy Induced by Correlated Spin Domains in short range spin glasses.

Eytan Domany[1,2*], Guy Hed[1], Matteo Palassini[3], and A. P. Young[4]

[1] *Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel*
[2] *Institute of Theoretical Physics, University of California Santa Barbara, CA 93106*
[3] *Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94118*
[4] *Physics Department, University of California Santa Cruz, Santa Cruz, CA 95064*

(August 3, 2001)

We generate equilibrium configurations for the three and four dimensional Ising spin glass with Gaussian distributed couplings at temperatures well below the transition temperature T_c. These states are analyzed by a recently proposed method using clustering. The analysis reveals a hierarchical state space structure. At each level of the hierarchy states are labeled by the orientations of a set of correlated macroscopic spin domains. Our picture of the low temperature phase of short range spin glasses is that of a State Hierarchy Induced by Correlated Spin domains (SHICS). The complexity of the low temperature phase is manifest in the fact that the composition of such a spin domain (i.e. its constituent spins), as well as its identifying label, are defined and determined by the "location" in the state hierarchy at which it appears. Mapping out the phase space structure by means of the orientations assumed by these domains enhances our ability to investigate the overlap distribution, which we find to be non-trivial. Evidence is also presented that these states may have a non-ultrametric structure.

I. INTRODUCTION

Whereas equilibrium properties of infinite range [1] spin glasses are completely understood within the framework of replica symmetry breaking (RSB) [2–5], spin glasses with short range interactions are the subject of considerable current debate and controversy. Open questions address the nature of the low temperature phases [2–8] and their theoretical description. Resolution of these issues by experiments or simulations is hindered by the extremely long relaxation time required for equilibration.

The most widely studied model of a short-range spin glass is the Edwards-Anderson model of an Ising spin glass

$$\mathcal{H} = \sum_{\langle ij \rangle} J_{ij} S_i S_j \ , \tag{1}$$

where $\langle ij \rangle$ denotes nearest neighbor sites of a simple (hyper) cubic lattice in D dimensions (we will consider $D = 3$ and $D = 4$) with periodic boundary conditions, $S_i = \pm 1$, and the couplings, J_{ij}, are independent random variables taken from a given distribution. The most commonly studied distribution, and the one we study here, is a Gaussian distribution with zero average and standard deviation $J = 1$.

The high temperature phase of the model is a disordered paramagnet. As the temperature decreases below a critical temperature T_c, the system (in three or more dimensions) undergoes a transition into a frozen spin-glass phase. In the spin glass phase, phase space is divided into "valleys" which we define as as an ergodic subset of the phase space, i.e. a maximal subspace that the system can span (or visit) as the time tends to infinity. For a finite system the definition is less clear, but a valley is usually referred to as a part of the phase space surrounded by free energy barriers, whose height diverges as the system size $L \to \infty$.

This definition of "valley" may not be identical to the notion of a "pure state" which has been used extensively in the literature [7–11]. and which is defined in terms of the set of correlation functions in a fixed *finite* region inside the system as $L \to \infty$ with some specified boundary conditions. In particular, it was recently emphasized [12–15] that a spin glass can in principle have many thermodynamically important valleys but just two pure states. This is realized when there are many valleys with free energies which differ by an amount of order unity, and configurations taken from different valleys have a vanishing density of (relative) domain walls as $L \to \infty$ (a *domain wall* is a surface separating a region where the two configurations are identical from a region where they are opposite). In contrast, if the density of domain walls is finite (*i.e.* the domain walls are space-filling), there is a non-vanishing probability to have a domain wall in any finite region of the system, and thus to have more than two pure states. In this paper we

*Electronic Address: fedomany@wicc.weizmann.ac.il

will be mainly concerned with the number and organization of valleys, and we will not investigate whether multiple valleys correspond to multiple pure states as defined above. In the following, by *state* we will always mean a microstate or spin configuration.

A. RSB, droplet and TNT scenarios

There are two traditional pictures of the spin glass phase; the droplet picture and RSB. According to the droplet picture of Fisher and Huse [6–8], the low energy excitations are in the form of *droplets* - compact regions with low surface tension that flip collectively. For a droplet of size L the typical (e.g. median) free energy F_L scales as L^θ, where θ is a dimension dependent exponent. Furthermore, the surface of these excitations has a vanishing density for large L. Therefore, thermodynamically important configurations have a vanishing density of relative domain walls, and hence a trivial overlap (defined below) over any finite region. It follows that in this approach within any finite region there are only two pure states, related by spin-flip symmetry.

A parameter commonly used to measure domain wall density is the link overlap and its distribution. Denote a configuration (or state) of an N-spin system by $\mathbf{S}^\mu = (S_1^\mu, S_2^\mu, ..., S_N^\mu)$. The link overlap $q_{\mu\nu}^{\mathrm{link}}$ between two configurations $\mathbf{S}^\mu$ and $\mathbf{S}^\nu$ is defined by

$$q_{\mu\nu}^{\mathrm{link}} = \frac{1}{\gamma N} \sum_{\langle ij \rangle} S_i^\mu S_j^\mu S_i^\nu S_j^\nu \ , \tag{2}$$

where the sum is over pairs of neighbor sites and γN is the number of bonds in the system. If the domain wall density vanishes, then the distribution $P(q^{\mathrm{link}})$ of the link overlap will be trivial: $P(q^{\mathrm{link}}) = \delta(q^{\mathrm{link}} - q_0)$. At $T = 0$ one has $q_0 = 1$, while q_0 decreases for $T > 0$ and becomes zero at T_c.

Another parameter commonly considered is the spin overlap $q_{\mu\nu}$ between configurations $\mathbf{S}^\mu$ and $\mathbf{S}^\nu$;

$$q_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} S_i^\mu S_i^\nu \ . \tag{3}$$

If there are only two pure states, as in the droplet model, the *local* overlap distribution, obtained *in a finite part of an infinite system*, would be trivial for all $T < T_c$, i.e. $P(q) = 0.5[\delta(q - q_{EA}) + \delta(q + q_{EA})]$, where q_{EA} is the average overlap inside a pure state. In addition, most conventional interpretations of the droplet picture [16,17] argue that the *global* $P(q)$, obtained from overlaps over the *whole* system, would also be trivial. This is realized if the droplets (with positive θ) are the only relevant excitations over all length scales. However the work of Huse and Fisher [7], and also Newman and Stein [9,18], is formulated in a sufficiently general fashion to accommodate a non-trivial global $P(q)$ if this arises from multiple valleys with non space-filling domain walls. In this situation, one would have a trivial link overlap distribution $P(q^{\mathrm{link}})$ in the infinite system size limit. Even though the global $P(q)$ would be non-trivial, the *local* $P(q)$, would be trivial because a vanishing density of domain walls means that the probability that a domain wall goes through a fixed finite part of the infinite sample also vanishes.

Numerical work has, so far, indicated a non-trivial global $P(q)$ [19,20]. For example, Marinari et al. [20] have used parallel tempering [21,22] to sample 3D Ising spin glasses of sizes up to $L = 16$ and for temperatures down to $T = 0.7 \simeq 0.74T_c$. They have found that $P(q)$ is non-trivial, and $P(0)$ does not vanish.

In the RSB picture, the Parisi [2–4] theory, which is exact for the infinite range model [1], is assumed to also apply to short range systems. Within the RSB solution, both $P(q)$ and $P(q^{\mathrm{link}})$ are non-trivial for $0 < T < T_c$, which implies that the system has many valleys and also many pure states. RSB suggests a tree-like hierarchical structure for the pure states. At every level of the hierarchy the states are divided into sets, so that the states in a given set are closer to each other than to states in other sets. At the next level down the hierarchy these sets are divided into subsets, and so on. Furthermore, according to the RSB solution the distances between the pure states exhibit *ultrametricity* [5]: the overlap between any two states is determined only by the lowest level in the hierarchy, at which they still belong to the same set. This means that for any triplet of pure states μ, ν and ρ the following relation always holds:

$$q_{\mu\nu} \geq \min(q_{\mu\rho},\, q_{\nu\rho}) \ . \tag{4}$$

Recently, a mixed picture has been proposed on the basis of numerical results of ground state computations [12–14], in which $P(q)$ is non-trivial but $P(q^{\mathrm{link}})$ is trivial (hence referred to as TNT; for **T**rivial and **N**on-**T**rivial). Houdayer, Krzakala and Martin [12,13] demonstrated the existence of macroscopic excitations with low energy cost in 3D Ising spin glasses of sizes up to $L = 11$. This suggests that the spin overlap distribution, $P(q)$ is non-trivial at finite temperature. Their results also indicate that the surface of these excitations is not space-filling, which suggests that the link overlap distribution, $P(q^{\mathrm{link}})$, is trivial.

Palassini and Young [14] studied changes to the ground state of a spin glass when a weak perturbation is applied to the bulk of the system. They considered short range models in three and four dimensions as well as the infinite range SK model and the Viana-Bray model. The results for the SK and Viana-Bray models agreed with the replica symmetry breaking picture as expected, but the data for the short range models agreed with the TNT picture. Effects of the type of perturbation considered in Ref. [14] on RSB have been investigated by Franz and Parisi [23].

Katzgraber et al. [24] measured directly the distributions $P(q)$ and $P(q^{\mathrm{link}})$ at finite temperature using parallel tempering [21,22] Monte Carlo, for 3D systems of linear size $L \leq 8$ at temperature $T \geq 0.2$, and 4D systems with $L \leq 5$ and the same temperature range. Extrapolating their results to large sizes they found that the variance of $P(q^{\mathrm{link}})$ vanishes as $L \to \infty$, and the distribution converges to delta function. They also found the distribution $P(q)$ to be non-trivial, as in [19,20], so their results also agree with the TNT picture. In the TNT scenario there are many valleys separated by free energy barriers, but only two pure states [25,15].

Although several pieces of work [12–14,24,26] supported a vanishing density of domain walls (and hence a fractal dimension of the domain walls, d_s, less than the space dimension), a large extrapolation is involved in deducing this result, and Marinari and Parisi [27–29] have argued, based on their own data and a somewhat different analysis, that actually $d_s = D$, which corresponds to RSB.

B. SHICS: State Hierarchy Induced by Correlated Spin Domains

Very recently a new method of analysis of the structure of the low temperature phase of short range spin glasses has been introduced [30,31]. Evidence for a novel picture of this phase, which is consistent with the TNT scenario, but inconsistent with RSB (since there is no ultrametricity), has been presented [30] on the basis of a "clustering analysis" of the degenerate ground states of the model (1) with $J_{ij} = \pm 1$ couplings. We denote this by "State Hierarchy Induced by Correlated Spin Domains" (SHICS).

In this picture there is a hierarchical tree-like structure of the states as in the RSB solution. The highest levels of the state hierarchy, are schematically illustrated in Fig. 1. At the first level of hierarchy the states divide into sets $\mathcal{C}$ and $\bar{\mathcal{C}}$, such that a state in $\mathcal{C}$ has a counterpart with the same energy in $\bar{\mathcal{C}}$, obtained by flipping all the spins. This equality of the energies follows, of course, from the symmetry of the Hamiltonian in zero field. However, this symmetry information is not imposed on the analysis; the method finds it by itself. In fact, it is not trivial, for a spin glass, to divide the states into two clusters such that every state in $\mathcal{C}$ has its reversed state in $\bar{\mathcal{C}}$. Suppose, for example, that one has two states μ and ν, and states $\bar{\mu}$ and $\bar{\nu}$ with reversed spins, such that the spin overlap $q_{\mu\nu}$ is close to zero. Should one put ν or $\bar{\nu}$ in the same cluster as μ? The analysis, used in Ref. [30] and here determines which one it is.

Many of the spins stay, with high probability, in the same relative orientation in most of the states $\mathcal{C}$. Most of these form a contiguous cluster $\mathcal{G}_1$, see Fig. 1. Among the remaining spins, an apparently macroscopic fraction form a contiguous domain, $\mathcal{G}_2$, such that the spins in it maintain, with high probability, their relative orientation in nearly all the states of $\mathcal{C}$. Hence $\mathcal{C}$ divides into two sub-clusters of states, $\mathcal{C}_1$ and $\mathcal{C}_2$, depending on the orientation of $\mathcal{G}_2$, see Fig. 1. In general, the domains $\mathcal{G}_1$ and $\mathcal{G}_2$ are distinct. In many samples, further levels of the hierarchy, with successively smaller domains $\mathcal{G}_3, \cdots$ can be clearly resolved, as discussed later. The excitations obtained by flipping the domains $\mathcal{G}_2, \mathcal{G}_3, \cdots$ appear to correspond to the large scale, low energy excitations investigated by Krzakala and Martin [12] and Palassini and Young [14]. Note that the local (or link) overlap was not investigated in Ref. [30].

By contrast, in the conventional interpretations of the droplet picture [16,17], the only substantial division of the states would be into $\mathcal{C}$ and $\bar{\mathcal{C}}$, and any further divisions emerging from the analysis would only correspond to microscopic spin domains. In the RSB scenario there would be a hierarchical structure to the states, similar to what we find here, but the nature of the spin domains would appear to be different, see e.g. [32]. We will discuss these differences further in Secs. IV and VII.

The purpose of the present paper is to use use the methodology of Ref. [30] to investigate whether the same picture of the spin glass phase found there also occurs for a spin glass with Gaussian couplings (which has a unique ground state apart from spin reversal) at *finite temperatures*. Both three and four dimensions are studied. We find that our data do fit this picture quite well. We also present here full details of the method.

Readers who like to skip ahead will find the picture of state clusters and spin domains that were obtained at $T = 0.2$ for a particular bond realization, conveniently summarized in Fig. 12. The corresponding overlap distribution $P(q)$ is presented in Fig. 14 (a).

The numerical procedure and parameters that were used in our simulations are described in Sec. II. In Sec. III we present the clustering methodology which we use to identify the states hierarchy, as described in Sec. IV. In Sec. V

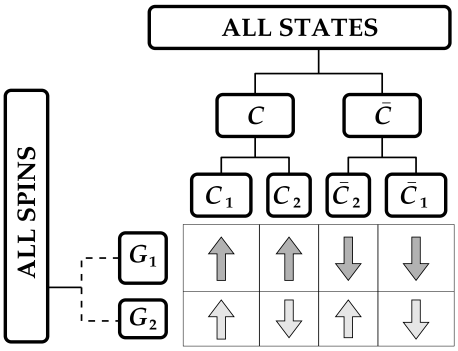

FIG. 1. Schematic representation of the SHICS picture; the two largest spin domains and the first two levels in the hierarchical organization of the states are shown. The structure of the states is explained by the spin domains' orientations; e.g. in the states of the two sets $\mathcal{C}_1, \mathcal{C}_2$, the spins of the larger domain, $\mathcal{G}_1$, have the same orientation, whereas the spins of the smaller domain, $\mathcal{G}_2$, have flipped. Spins not in $\mathcal{G}_1$ or $\mathcal{G}_2$ are in smaller domains which are not resolved at this level of the hierarchy.

we use the hierarchical partition of the state space to obtain the spin domains, show that their sizes scale with the system size and their correlation does not approach unity as $L \to \infty$. We also show that these spin domains, that were identified on physical grounds, can also be obtained by a cluster analysis of the N spins. Those domains yield a non-trivial overlap distribution $P(q)$ with peaks corresponding to the different domain sizes, as we show in Sec. VI. Since we find that the average correlation between spins in different domains does not approach unity with increasing system size, $P(q)$ will remain non-trivial as $L \to \infty$. The nature of our picture appears to yield a non-ultrametric structure, as indicated at the end of Sec. IV and demonstrated in Sec. VII, in which we present a parameter for ultrametricity and measure its distribution. Finally, our method and findings are summarized in Sec. VIII.

After this work was completed we received a preprint from Marinari et al [33] who have adopted and adapted the methodology of [30,31] to study the $J_{ij} = \pm 1$ model in $d = 3$ at a single temperature ($T = 0.5$ - whereas here we considered $T = 0.2$ and $T = 0.5$ for the Gaussian model). They also confirmed that the previously observed SHICS scenario [30] of a tree-like structure of the states, governed by correlated spin domains, remains valid at a non-zero temperature.

II. NUMERICAL METHOD

We simulate the Hamiltonian in Eq. (1) using the parallel tempering Monte Carlo method [21,22]. In this technique, one simulates several identical replicas of the system at different temperatures, and, in addition to the usual local moves, one performs global moves in which the temperatures of two replicas (with adjacent temperatures) are exchanged. This greatly speeds up equilibration at low temperatures. The detailed balance condition for temperature exchanges is satisfied by accepting these moves with probability min $[\exp(\Delta E \Delta\beta),\ 1]$, where $\Delta E = E^\mu - E^\nu$, E^μ and E^ν are the (total) energies of replicas μ and ν, and $\Delta\beta = \beta^\mu - \beta^\nu$ is the difference in inverse temperatures.

We choose a set of temperatures $T_i, i = 1, 2, \cdots, N_T$, in order that the acceptance ratio for the global moves is satisfactory, typically greater than about 0.3. We use the test for equilibration discussed in Ref. [24], which involves measurements of q_{link}. For that, we need, at each temperature, two copies of the system, so we actually run 2 sets of N_T replicas and perform the global moves independently in each of these two sets.

For the three-dimensional model we stored configurations for sizes $L = 4, 5, 6$ and 8 at $T = 0.20, 0.50$ and 2.0, which are to be compared with [20] $T_c \simeq 0.95$. We also stored size $L = 12$ configurations at $T = 0.50$. The parameters of the simulations are shown in Table I. The highest temperature was 2.0 and lowest 0.2 except for $L = 12$ where the lowest temperature was 0.5.

We generated randomly chosen interactions, J_{ij}, with a Gaussian distribution with zero mean and standard deviation unity. For each size, temperature and bond configuration (sample) we saved 500 spin configurations. These, together with the 500 obtained from them by spin reversal, constitute our ensemble of $M = 1000$ spin configurations, generated for each sample.

For the four-dimensional model we stored configurations for sizes $L = 3, 4$ and 5 at $T = 0.2, 0.8$ and 2.6, compared with [34] $T_c \simeq 1.80$. The highest temperature was 2.6 and the lowest 0.2. 500 spin configurations were saved for each sample. The other parameters of the simulations are also shown in Table I.

We are confident, based on the equilibration test used [24], that the spin configurations we generate are in thermal equilibrium. However, it is interesting to ask whether there are significant correlations between them. Our results

D	L	$n_{\rm equil}$	$n_{\rm meas}$	$N_{\rm samp}$	N_T
3	4	10^4	10^5	500	11
	5	5×10^4	5×10^5	500	15
	6	3×10^5	3×10^6	500	15
	8	10^6	10^7	335	18
	12	2×10^5	2×10^6	254	20
4	3	10^4	10^5	500	13
	4	4×10^4	4×10^5	500	13
	5	8×10^5	8×10^6	200	25

TABLE I. Parameters of the simulations in $D = 3$ and 4 dimensions. $N_{\rm samp}$ is the number of samples (i.e. sets of bonds), $n_{\rm equil}$ is the number of sweeps for equilibration and $n_{\rm meas}$ is the number of sweeps for measurements for each of the $2N_T$ replicas for a single sample. N_T is the number of temperatures used in the parallel tempering method.

do not require that correlations be absent, but the clustering method does require that a substantial number of independent configurations are generated for each sample.

For each set of bonds (and temperature) we store 500 spin configurations, 250 for each replica, so the number of sweeps between measurements, $t_{\rm meas}$, is given by $t_{\rm meas} = n_{\rm meas}/250$ where $n_{\rm meas}$ is given in Table I We will denote by "time", t, the number of Monte Carlo sweeps. A quantity which tests for correlations is the time-dependent Edwards-Anderson order parameter $[q(t)]_J \equiv [\langle S_i(t_0)S_i(t_0+t)\rangle]_J$, where $\langle \cdots \rangle$ indicates a thermal average. This is estimated from our spin configurations according to

$$[q(t)]_J = \left[\frac{1}{N_{t_0}} \sum_{t_o}^{N_{t_0}} \frac{1}{N} \sum_{i=1}^{N} S_i(t_0)S_i(t_0+t) \right]_J , \tag{5}$$

where we have averaged over N_{t_0} values for the initial time t_0 as well as over spins and bond configurations. Clearly $[q(0)]_J = 1$ and $[q(t)]_J \to 0$ for times sufficiently long that there are no correlations.

In Fig. 2 we show data for $[q(t)]_J$ in $D = 3$ for $L = 8, T = 0.2$. We see that the correlation is very small even for $t/t_{\rm meas} = 1$ (i.e. between the configurations of neighboring measurements). The same is true for smaller sizes and higher temperatures. For $L = 12, T = 0.5$, shown in the inset to Fig. 2, the correlations are larger, about 0.24 for $t/t_{\rm meas} = 1$, and then decrease, though less fast than exponentially. Thus, for $L = 12$, correlations will decrease, somewhat, the effective number of independent spin configurations. However, we feel that this is not crucial since we do not use the $D = 3, L = 12$ data for the clustering analysis, and only present it in one place, Figure 9.

In $D = 4$, for $L = 3$ and 5, the strength of the correlations at $T = 0.2$ is small, comparable to, or less than that for $D = 3, L = 8, T = 0.2$. For $L = 4$, the correlation is intermediate between the results shown in $D = 3$ for $L = 8$ and 12.

III. CLUSTERING METHODOLOGY

Clustering is an important technique to perform exploratory data analysis. The aim is to partition data according to natural classes present in it. By "natural classes" we mean groups of points that are close to one another and relatively far from other points, so that it is natural to assign them together, without using any preconceived information on the features according to which the set should be classified.

The standard definition of the clustering problem [35] is as follows. Partition N given data points (or objects) into K groups (i.e. clusters) so that two points that belong to the same group are, in some sense, more similar than two that belong to different groups. The $i = 1, 2, ...N$ data points are specified either in terms of their coordinates $\vec{X}_i$ in a D-dimensional space (representing the measured values of D attributes or features) or, alternatively, by means of an $N \times N$ "distance matrix", whose elements d_{ij} measure the dissimilarity of data points i and j. The traditional tasks of clustering algorithms are to determine K and to assign each data point to a cluster.

In the context of the present work we can think of our sample of M spin configurations as the objects to be clustered. Each object is represented by an $N-$component vector $\mathbf{S}^\mu = (S_1^\mu, S_2^\mu, ..., S_N^\mu)$, where $S_i^\mu = \pm 1$ is the value taken by spin i in state μ. An alternative view, which we also use, is to consider the N spins as the objects to be clustered.

Our first aim in this work was to look for a hierarchical structure of the states of a spin glass. Hence we wanted to find a hierarchy of partitions, where each partition is a refinement of the previous partition. This purpose calls for

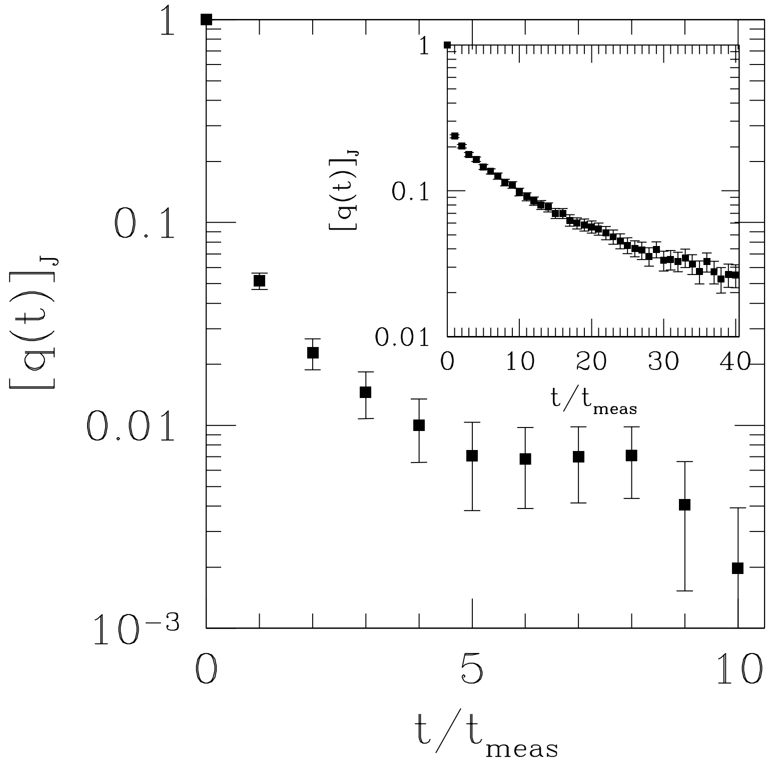

FIG. 2. The main part of the figure shows the correlation between spin configurations, $[q(t)]_J$, defined in Eq. (5) of the text, in $D = 3$ for $L = 8, T = 0.2$. The horizontal axis represents the number of Monte Carlo sweeps between the two configurations in units of the number of sweeps between individual measurements, $t_{\rm meas}$. For comparison, for each set of spins ("replica"), a total of 250 configurations are generated. The inset shows results for $D = 3, L = 12, T = 0.5$, which indicate that correlations between spin configurations are significantly larger than for $L = 8$.

using a hierarchical clustering algorithm. The output of such an algorithm is a tree of clusters, called a *dendrogram*. Each node in the tree corresponds to a cluster. The splitting of a cluster represents its partition into sub-clusters. The trunk is the single "cluster" that contains *all* the objects, representing the crudest partition; at the other extreme each leaf is a cluster of a single object, representing the finest partition.

There are many clustering algorithms that produce such a hierarchical partitioning of any data set. We tried two algorithms; a recently introduced one, SPC [36], which uses the physics of granular ferromagnets to identify clusters, and a graph-based algorithm proposed by Ward. In the present problem the state clusters are nearly always compact (i.e. consist of a high density of points concentrated in a relatively small volume), and the same holds for spin clusters. Therefore an algorithm that identifies compact clusters easily is most suitable for our needs and Ward's algorithm is designed to find such clusters. Furthermore, SPC is a "short-range" algorithm [37], in the sense that it couples directly only points within a characteristic length scale. If this scale is tuned by the distances inside valleys, which are much smaller than the distance between them, SPC identifies the valleys as different clusters, but may miss their relative hierarchical structure.

Ward's algorithm [35] is *agglomerative*, works its way up from the leaves to the trunk, by fusing two clusters at each step. It begins with an initial partition to $i = 1, 2, ..., N$ clusters, with a single data point in each. One calculates the distance D_{ij} between every pair of points i, j; one may use, for example, the Euclidean definition of distance, or (for binary valued coordinates) the Hamming distance.

At each step that pair of clusters, α, β, which are separated by the shortest effective distance $\rho_{\alpha\beta}$ from each other, are identified and fused to form a new cluster $\alpha' = \alpha \cup \beta$. The process stops when there is only one cluster, that contains all points.

Initially each data point $i = 1, 2, ...N$ constitutes a cluster and hence the distance ρ_{ij} between two such "clusters" is the original distance D_{ij} between points i and j. For subsequent steps, however, one must define an effective distance $\rho_{\alpha\beta}$, between any two clusters α and β. This distance is defined by the following update rule: if at a particular step we fuse two clusters, α and β, to form a new cluster α', we calculate the effective distances $\rho'_{\gamma\alpha'}$, between every

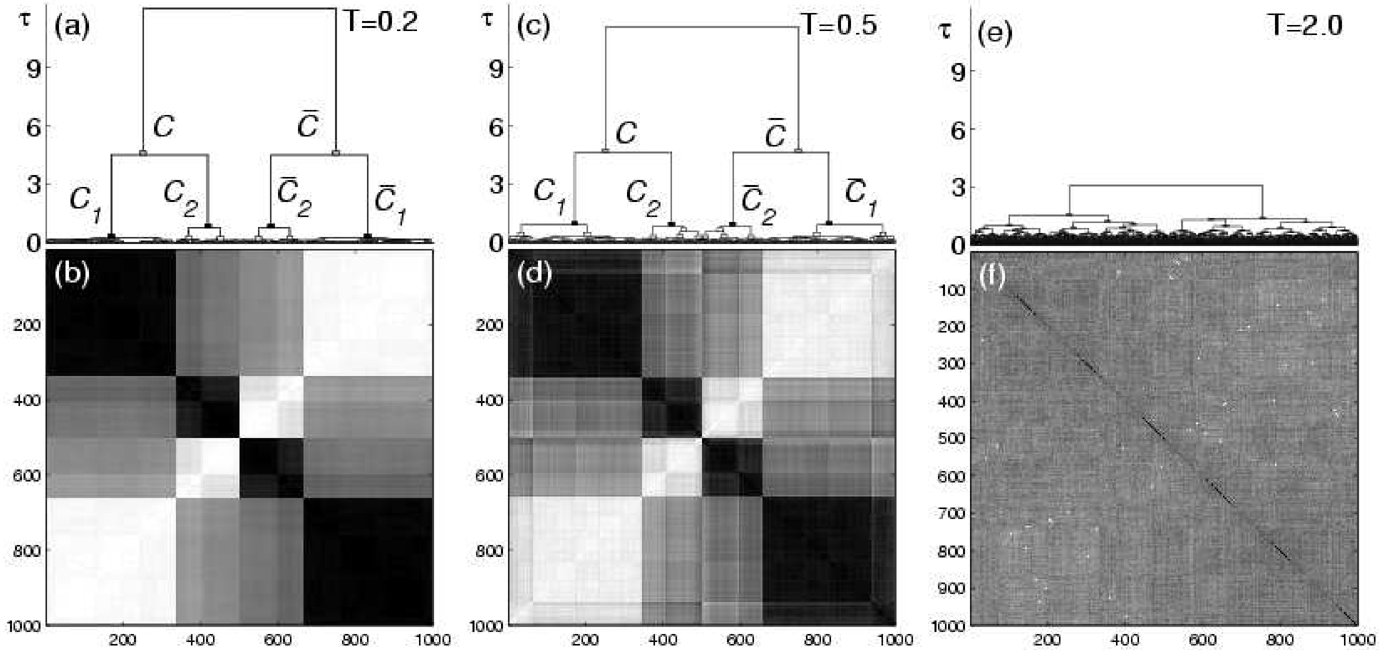

FIG. 3. **(a)** The dendrogram obtained by clustering the $M = 500 \times 2$ states of a specific realization in $D = 3$ of size $N = 8^3$ at $T = 0.2$. The vertical axis describes the value of τ, defined in Section III. **(b)** The distance matrix of the states used as an input to Ward's algorithm. Darker shades correspond to smaller distances. The states are ordered according to their position on the dendrogram (a). **(c), (d)** The same as in (a), (b), for the same realization $\{J\}$, but for an ensemble of states obtained at at $T = 0.5$. **(e), (f)** The same as in (a), (b), for the same realization, at $T = 2.0$, which is greater than $T_c \simeq 0.95$. Note that this dendrogram is not symmetric; almost all the distances are close to 0.5, so at each stage of the algorithm there were several possible partitions that gave minimal value to S. In the implementation we used, the algorithm chose a non-symmetric partition.

unchanged cluster, $\gamma \neq \alpha, \beta$, and the new α', according to

$$\rho'_{\alpha'\gamma} = \frac{n_\alpha + n_\gamma}{n_\alpha + n_\beta + n_\gamma}\rho_{\alpha\gamma} + \frac{n_\beta + n_\gamma}{n_\alpha + n_\beta + n_\gamma}\rho_{\beta\gamma} - \frac{n_\gamma}{n_\alpha + n_\beta + n_\gamma}\rho_{\alpha\beta} \;, \tag{6}$$

where n_x is the number of data points in cluster x. Distances between unfused clusters remain the same. Note that $\rho'_{\alpha'\gamma} > \rho_{\alpha\beta}$ and $\rho'_{\gamma\delta} > \rho_{\alpha\beta}$ for every two clusters γ, δ. Hence after every fusion step the minimal distance between clusters increases.

Whenever two clusters are fused, the quantity

$$S = \sum_\alpha \sigma_\alpha \tag{7}$$

where σ_α is the sum of squared distances over all pairs of points in cluster α,

$$\sigma_\alpha = \sum_{i,j\in\alpha} {D_{ij}}^2 . \tag{8}$$

increases. It can be shown [35] that Ward's fusion and distance update rules ensure that at each fusion step this increase is minimal.

We associate a value τ with each cluster α', where $\tau(\alpha') = \rho_{\alpha\beta}$ is the effective distance between the two clusters that were fused to form α'. For the initial single-point clusters we set $\tau = 0$. $\tau(\alpha)$ is related to σ_α, the sum of squared distances within cluster α. Clusters formed earlier have lower τ values, and their σ_α is smaller.

The result of the algorithm is a dendrogram, or tree, as in Fig. 3(a). The leaves at the bottom represent the individual data points; they are ordered on the horizontal axis in a way that reflects their proximity and hierarchical assignment [38]. The small boxes at the nodes represent clusters. The vertical location of cluster α is its τ value, and is thus related to its σ. When two relatively tight and well-separated clusters are fused, the τ value of the resulting cluster is much higher than those of the two constituents. Hence the length of the branch *above* cluster α provides a measure of its relative σ_α; long branches identify clear, tight clusters.

Every clustering algorithm is designed to work well for data that satisfy some (usually implicit) assumptions. When the actual distribution of the data points deviates from these assumptions, the algorithm may produce some "unnatural" partitions. For Ward's algorithm one has to look out for two potential problems.

The first problem arises from the implicit assumption that minimizing S, the variance within clusters, leads to "natural" partitions. This is not the case when, for example, the data consists of a set of points C whose natural partition is into two clusters C_1 and C_2 with very different sizes. We encountered this problem only for the classification of very small groups of states, and therefore it has very little statistical effect on our results.

The second, and seemingly more serious concern is the fact that like every agglomerative algorithm, Ward's algorithm will generate a tree-like structure when applied to *any* set of data. In fact, it is fairly easy to identify when the dendrogram and the corresponding partitions do correspond to real hierarchical structure, and when is it an artifact of the clustering algorithm used. We used three indicators for the "naturalness" of our state clusters: direct observations of (1) the dendrograms and (2) the distance matrices, as well as (3) a quantitative measurement of the sizes of our clusters, which are significantly smaller than the distance between clusters. These points are demonstrated in Sec. IV; for a detailed discussion see [37].

IV. STATE SPACE STRUCTURE

For a particular (randomly chosen) set of bonds $\{J\}$ of the system we generate, as discussed in Sec. II, a sample of 500 states, which constitute an equilibrium ensemble at a temperature T. Next, we add to this ensemble the set of 500 states obtained from the original set by spin reversal. Clearly the new ensemble of $M = 1000$ states also corresponds to thermal equilibrium [39] at T. We now address the following question:

> Do the M states of the equilibrium ensemble cover the 2^N points of state space or a part of it uniformly, or is there some underlying hierarchical organization?

As it turns out, the answer depends on T; whereas above T_c the M states do not exhibit any apparent structure, below T_c a very pronounced hierarchical organization is seen. To uncover this organization we use the clustering methodology of the previous Section, treating the M states of our ensemble as the data points to be clustered.

We describe here analysis of a single realization of the randomness, in order to help the reader perceive the qualitative nature of the results (see Figs. 4 and 5), and to define the observables that we measure. These observables were measured for each of the different realizations, and the distributions of their values were determined; the average and width of these distributions are also presented. This data demonstrate that the results described in this section for a single sample are typical and seen in many samples.

In order to cluster the states, each state μ is represented as an N-component vector $\mathbf{S}^\mu = (S_1^\mu, ..., S_N^\mu)$, where $S_i^\mu = \pm 1$ is the value taken by spin i in state μ. The complete data set can be represented as an $N \times M$ *data matrix*, whose columns are the vectors $\mathbf{S}^\mu$. For the set of $M = 1000$ states, obtained at $T = 0.2$ for a particular bond realization of an $N = 8^3$ spin system, the data matrix is presented in Fig 4(a). Pixel (i, μ) of this figure represents the sign of spin i in state μ; a black entry corresponds to $+1$ and white to -1. The spins appear in lexicographic order and the states in the random order generated by the simulation. As can be seen, the matrix appears fairly random, with no easily discernible structure; nevertheless, there is a clear organization of these M states into tight clusters. For the particular realization and ensemble of states presented here, these clusters of states can be seen by direct observation of the $M = 1000$ data-points $\mathbf{S}^\mu$, once one overcomes the hurdle of directly viewing a cloud of 1000 points in a $N = 512$ dimensional space.

A trivial way of visualizing points that lie in a high dimensional space is to project them onto a low (i.e. two or three) dimensional subspace. In order to reveal the underlying structure, it is important to choose with care the subspace onto which one projects. A widely used method to choose this subspace is that of *principal component analysis* (PCA) [40]. One constructs the $N \times N$ covariance matrix of the M points,

$$r_{ij} = \frac{1}{M} \sum_{\mu=1}^{M} \delta S_i^\mu \delta S_j^\mu \tag{9}$$

where

$$\delta S_i^\mu = (S_i^\mu - m_i)/\sigma_i \tag{10}$$

with m_i the average of the M variables S_i^μ and σ_i^2 their variance. For our case $m_i = 0$ and $\sigma_i = 1$ for all i, and hence the covariance matrix is the spin correlation matrix, *i.e.*

$$r_{ij} = c_{ij} = \frac{1}{M} \sum_{\mu=1}^{M} S_i^\mu S_j^\mu \ . \tag{11}$$

The eigenvectors $\mathbf{e}_i$ of this matrix are the principal directions or components of the variation in the data. They are ordered according to the size of the corresponding eigenvalues, with the largest coming first.

In Fig. 5 we present the projections of our $T = 0.2$ ensemble of $M = 1000$ states on the first two and three principal components. Even though projection of $N = 8^3$ dimensional data onto three and two dimensions involves a major loss of information, the cluster structure of the states is still clearly evident. In Fig. 5 (a) projection onto the largest eigenvector, $\mathbf{e}_1$, is represented by the horizontal axis, and on the second largest, $\mathbf{e}_2$, by the vertical. It is interesting to note that the two largest state clusters, $\mathcal{C}_1$ and $\bar{\mathcal{C}}_1$, project mostly onto $\mathbf{e}_1$ and the second largest pair, $\mathcal{C}_2$ and $\bar{\mathcal{C}}_2$ onto $\mathbf{e}_2$. Fig. 5 (b) indicates that the next sized variation, due to splitting of $\mathcal{C}_2$ into two subgroups, is captured by $\mathbf{e}_3$. The scale of the projections can be understood by the following argument: if the (normalized) eigenvector $\mathbf{e}_1$ is parallel to a typical vector from $\mathcal{C}_1$, then, since normalization of $\mathbf{e}_1$ involves a factor of $1/\sqrt{N}$, the maximum possible projection is $\sqrt{N} \approx 22.6$. Hence the projections shown in Fig. 5 are quite large, i.e. close to the maximum possible value.

Next we obtain a systematic quantitative measure of the hierarchical structure of state space by performing a cluster analysis of the M points. The choice of the particular clustering algorithm used was dictated by our idea of the state space structure, obtained from PCA and from the picture described in the Introduction and summarized in Fig. 1. Our aim is to find a hierarchy of partitions into compact clusters. That is, we would like states that belong to the same cluster to be closer to each other than to states in different clusters. Ward's algorithm, described in Section III, is tailored to perform this task for the kind of data distribution that we have in state space.

To start, we defined the $M \times M$ distance matrix D between the states μ, ν by

$$D_{\mu\nu} = \frac{1 - q_{\mu\nu}}{2} \ , \tag{12}$$

where $q_{\mu\nu}$ is the state overlap defined by Eq. (3). Next, we clustered the states using the distance matrix $D_{\mu\nu}$ as input to Ward's algorithm (see Eq. (6)). The algorithm results in a dendrogram, as shown in Figs. 3(a,c,e), for a sample at $T = 0.2, 0.5$ and 2.0, in three dimensions. The leaves, which represent the states, are ordered on the horizontal axis according to the order imposed by the dendrogram [38]. The nodes represents the clusters. The vertical location of each cluster corresponds to its τ value, and is thus related to the variance within it.

For $T = 0.2$ and 0.5, which are below $T_c \approx 0.95$ [20], we found clear partitions in the two highest levels of the dendrogram, as presented in Figs. 3(a,c). At the highest level the states are partitioned into $\mathcal{C}$ and $\bar{\mathcal{C}}$. At the next level, $\mathcal{C}$ is broken into two sub-clusters, which we denote as $\mathcal{C}_1$ and $\mathcal{C}_2$. For this specific sample the cluster $\mathcal{C}_2$ breaks further into two sub-clusters, which are clearly seen in Fig. 5 as well.

To gain insight into the manner in which similar states are grouped together, and to actually "look into the spin-glass" at the microscopic level, we present in Fig. 4(b) the same data matrix as shown in Fig. 4(a), but with the states again reordered according to the dendrogram of Fig. 3 (a). That is, to get Fig. 4 (b), the columns of Fig. 4(a) have been permuted according to their position in the dendrogram. The clear central vertical dividing line separates $\mathcal{C}$ from $\bar{\mathcal{C}}$. In addition to the central dividing line, another vertical line is also clearly visible - it separates the states that belong to the larger cluster $\mathcal{C}_1$ from the smaller one, $\mathcal{C}_2$.

We now demonstrate that the state clusters we found are indeed "correct" and "natural". First, we checked that the situation of merging two clusters of very different sizes occurs very rarely.

We showed that our partitions are "natural" and not an artefact of the algorithm (which produces a tree for any data), in three ways:

1. Note that direct observation of the dendrograms clearly distinguishes between the different situations above and below T_c. At $T = 0.2, 0.5$ ($< T_c$) the relative τ values of the state clusters $\mathcal{C}, \mathcal{C}_1$ and $\mathcal{C}_2$ – measured by the length of the branch above each cluster – are high. A long branch indicates that the size of the cluster is much smaller than the distance between it and its "brother", which indicates that the partition into these two groups is natural. In comparison, in the dendrogram obtained at $T = 2.0$ ($> T_c$), the relative τ values are much smaller than at $T = 0.2, 0.5$.

2. The genuinely hierarchical structure at $T = 0.2, 0.5$ is also evident from the states' distance matrix, as shown in Figs. 3(b,d). This distance matrix was obtained by reordering the states according to the results of the

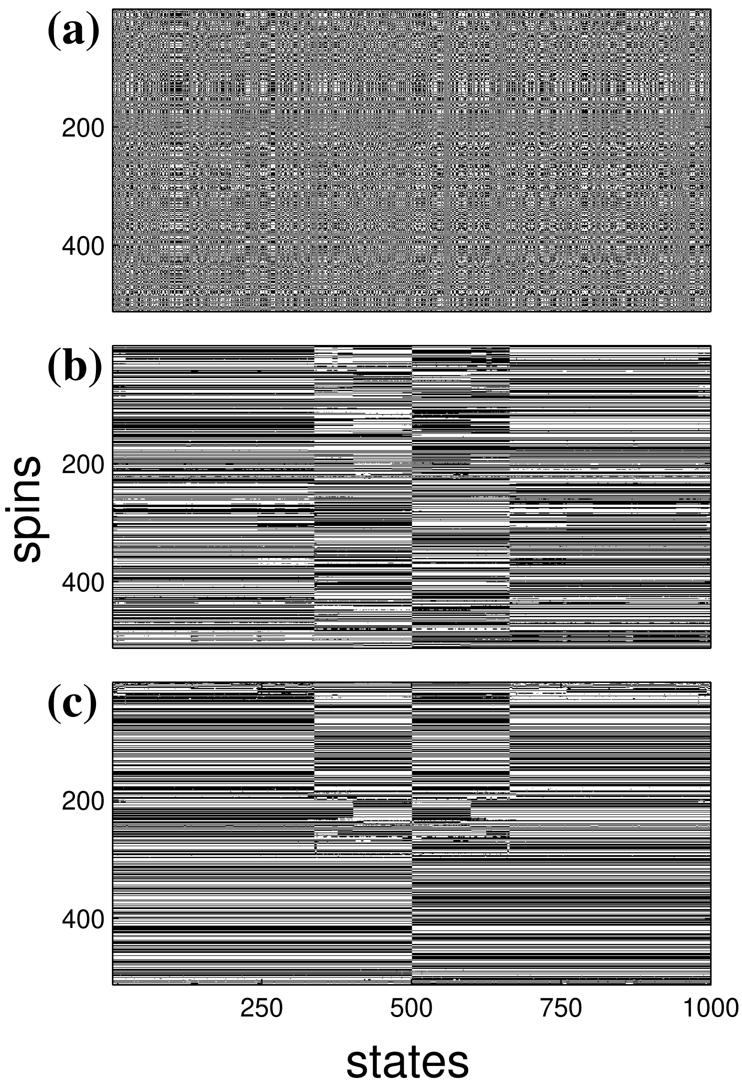

FIG. 4. **(a)** The original data matrix of 500×2 states $\mathbf{S}^{\mu}$, $S_i^{\mu} = \pm 1$, with black/white representing $+/-$. This $3D$ sample was generated for a realization of size 8^3 at $T = 0.2$ (the same one as in Fig. 3). The spins are in lexicographic order. **(b)** The same matrix, with the states ordered according to the dendrogram in Fig. 3. **(c)** The matrix in (b) with, in addition, the spins ordered according to the spin dendrogram $\mathcal{D}$ in Fig. 11.

cluster analysis, i.e. according to the order of the leaves of the corresponding dendrogram. When the states are randomly ordered (like in Fig. 4(a)), the resulting distance matrix is a homogeneous greyish square, like that of Fig. 3(f). The difference between this and Figs. 3(b,d) is striking: the distance matrices *within* clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ appear as dark squares (representing shorter distances) along the diagonal. The distances between clusters are represented by fairly uniform, lighter colored rectangles. In comparison, for $T = 2.0$ there is no real hierarchical organization of the states, and reordering them according to the dendrogram does not generate any ordered appearance of the distance matrix.

3. We measured the average distance between pairs of states that belong to each of the clusters $\mathcal{C}$, $\mathcal{C}_1$ and $\mathcal{C}_2$. The average $D(\mathcal{C})$ and the width $w(\mathcal{C})$ of the distribution of distances within $\mathcal{C}$ are

$$D(\mathcal{C}) = \tfrac{1}{|\mathcal{C}|^2} \sum_{\mu,\nu \in \mathcal{C}} D_{\mu\nu} \; ; \tag{13}$$

$$w(\mathcal{C}) = \left(\tfrac{1}{|\mathcal{C}|^2} \sum_{\mu,\nu \in \mathcal{C}} {D_{\mu\nu}}^2 - D(\mathcal{C})^2 \right)^{1/2} \quad , \tag{14}$$

where μ and ν refer to individual configurations. The average $D(\mathcal{C}_\alpha)$ and the width $w(\mathcal{C}_\alpha)$ for $\alpha = 1, 2$ are

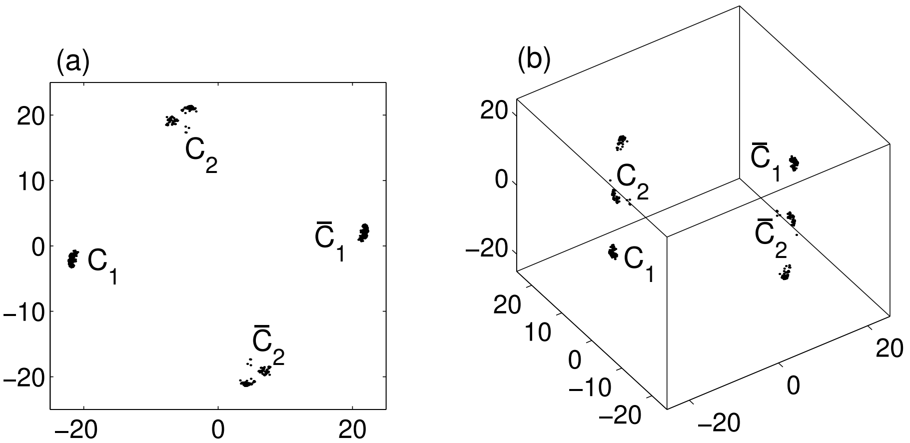

FIG. 5. Principal Component Analysis of a sample of $M = 500 \times 2$ states of a specific realization of $\{J\}$ in $3D$ with $N = 8^3$ spins at $T = 0.2$. Each point represents a state $\mathbf{S}^\mu$. The coordinates are projections on to eigenvectors $\mathbf{e}_i$, corresponding to the largest eigenvalues of the correlation matrix in Eq. (11). We show in (a) projections onto two eigenvectors, corresponding to the largest and next-largest eigenvalues of the correlation matrix, shown, respectively, on the horizontal and vertical axes. In (b) the three largest eigenvectors are used. The first and second level partitions of the hierarchy are clearly visible and, to some extent, the third level also.

defined in a similar way. The distribution of distances within clusters is to be compared with the distribution of distances between points that belong to different clusters. The average $D(\mathcal{C}_1, \mathcal{C}_2)$ and width $w(\mathcal{C}_1, \mathcal{C}_2)$ of the inter-cluster distance distribution are defined as

$$D(\mathcal{C}_1, \mathcal{C}_2) = \tfrac{1}{|\mathcal{C}_1||\mathcal{C}_2|} \textstyle\sum_{\mu \in \mathcal{C}_1} \sum_{\nu \in \mathcal{C}_2} D_{\mu\nu} \; ; \tag{15}$$

$$w(\mathcal{C}_1, \mathcal{C}_2) = \left(\tfrac{1}{|\mathcal{C}_1||\mathcal{C}_2|} \textstyle\sum_{\mu \in \mathcal{C}_1} \sum_{\nu \in \mathcal{C}_2} {D_{\mu\nu}}^2 - D(\mathcal{C}_1, \mathcal{C}_2)^2 \right)^{1/2} \; . \tag{16}$$

The clusters $\mathcal{C}, \bar{\mathcal{C}}$ are special in that each state $\mu \in \mathcal{C}$ has an inverted state $\bar{\mu} \in \bar{\mathcal{C}}$, so that $\mathbf{S}^\mu = -\mathbf{S}^{\bar{\mu}}$. Therefore $D(\mathcal{C}, \bar{\mathcal{C}}) = 1 - D(\mathcal{C})$ and $w(\mathcal{C}, \bar{\mathcal{C}}) = w(\mathcal{C})$.

A subset of the results is presented in Table II; for all temperatures, system sizes and both dimensions see [37]. We present for each variable x its mean $[x]_J$ (averaged over the disorder $\{J\}$) and its standard deviation $\Delta x = ([x^2]_J - {[x]_J}^2)^{1/2}$. For $T = 0.2$ and 0.5, which are below T_c, the average distances within the clusters are of the order of 0.1. $D(\mathcal{C}, \bar{\mathcal{C}})$ is around 0.9, which shows that there is a clear separation between these two clusters. $D(\mathcal{C}_1, \mathcal{C}_2)$ is much lower, but is still about two or three times larger than either $D(\mathcal{C}_1)$ or $D(\mathcal{C}_2)$. Note that the width of the distance distribution within a cluster is of the same order of the mean distance, so in general distances will not be much larger than twice the mean distance. At $T = 2.0$ ($> T_c$) the distances within and between clusters are almost equal and the differences are only due to statistical fluctuations, again indicating absence of natural structure, as we claimed on the basis of direct observation.

Measurement of some of the quantities listed above allows us to investigate the extent to which the state space structure of short-range spin glasses, as reflected by the data in Table II, is compatible with RSB. In the RSB [2–5] framework, the overlap between any pair of valleys (which correspond to pure states in the usual interpretation of RSB) from two different clusters that appear at the same level of the hierarchy is constant. It seems natural to associate the pure state clusters of RSB to our state clusters, e.g. $\mathcal{C}_1$ and $\mathcal{C}_2$. In this association, each state cluster contains states that belong to different "pure states". If the overlap between pure states of the two clusters is constant as in RSB, this should hold also for the overlap between each pair of states $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$, since the width of the overlap distribution inside a pure state approaches zero. In this case, all entries of the sub-matrix $\tilde{D}_{\mu\nu}$ for $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$ would be equal, so the width $w_{12} = [w(\mathcal{C}_1, \mathcal{C}_2)]_J$ should vanish as $L \to \infty$. To test whether this is the case, we present in Fig. 6 the values of $w_{12} = [w(\mathcal{C}_1, \mathcal{C}_2)]_J$ *vs* the system size L for $T = 0.2$ and $D = 3$. The error bars represent the statistical error (obtained by dividing the standard deviations, given in Table II, by $\sqrt{N_{\rm samp} - 1}$). We tried fits of the form

$$w_{12} = w_\infty + BL^{-y} \; , \tag{17}$$

with B and y as fit parameters. The overall best fit was for $w_\infty = 0.0205, B = 0.58$ and $y = 3.36$, which gives a very small χ^2 of 0.036. This is shown by the solid line in Fig. 6. We also tried the best fit assuming that $w_\infty = 0$, which

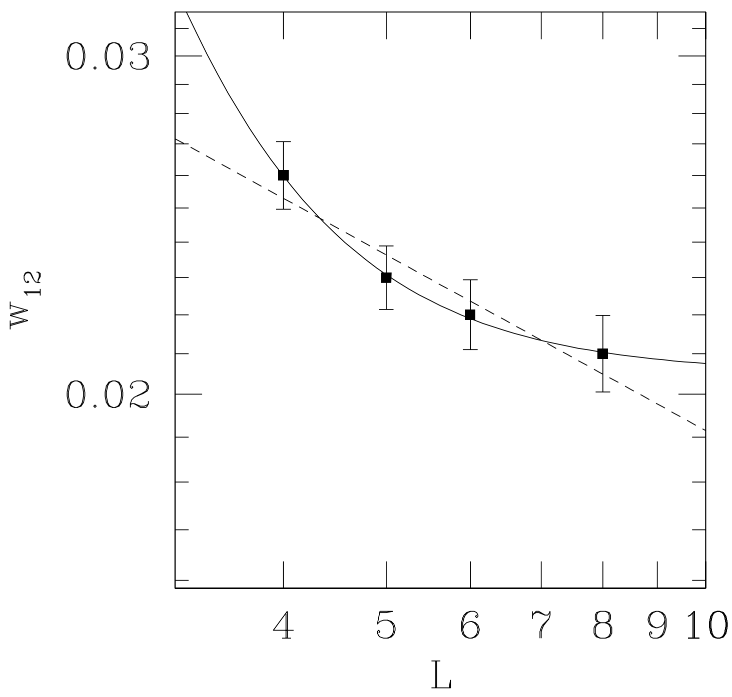

FIG. 6. A log-log plot of w_{12} against L for $T = 0.2$ and $D = 3$. The solid line is the best least squares fit to Eq. (17), while the dashed line is the best fit with the additional assumption that $w_\infty = 0$. .

has fit parameters $B = 0.039$ and $y = 0.30$, and is shown by the dashed line in the figure. This has a χ^2 of 1.41 which is much larger than the best fit with $w_\infty \neq 0$, but still acceptable. Hence even though our data suggests that $w_\infty \neq 0$, the possibility that $w_\infty = 0$, which corresponds to RSB, cannot be ruled out.

V. CORRELATED DOMAINS IN SPIN SPACE

A. Identifying the spin domains

According to our picture, splitting of a cluster at level a in the states hierarchy is induced by a macroscopic contiguous [41] spin domain $\mathcal{G}_a$. The size and shape of this domain determines the energy barrier separating two state clusters that were "born" at this level. In this subsection we describe how we identify from our data the two correlated domains $\mathcal{G}_1$ and $\mathcal{G}_2$, which determine the two highest levels of the states hierarchy, and also discuss whether they remain macroscopic at large L. Domains that emerge at the next level, $\mathcal{G}_3$ and $\mathcal{G}'_3$, are also discussed briefly.

Since the spins in such domains flip "collectively", they are highly correlated. The standard definition of the correlation c_{ij} of spins i and j is

$$c_{ij} = \langle S_i S_j \rangle = \frac{1}{\mathcal{Z}} \sum_{\mathbf{S}} S_i S_j \exp[-\beta \mathcal{H}(\mathbf{S})] \ , \tag{18}$$

where $\langle ... \rangle$ stands for the thermodynamic average for a particular realization of the disorder, and $\mathcal{Z}$ is the partition function at T. Using our equilibrium ensemble of states $\{\mathbf{S}^\mu\}$, we evaluate

$$c_{ij} = \frac{1}{M} \sum_{\mu} S_i^\mu S_j^\mu \ . \tag{19}$$

The correlation in itself is unimportant for spin glasses since it is gauge dependent and its average $[c_{ij}]_J$ over all the realizations of the disorder $\{J\}$ vanishes. The relevant measure of correlations in a spin glass is the square, ${c_{ij}}^2$. If two spins are independent of each other over the equilibrium ensemble of states, we have ${c_{ij}}^2 = 0$. On the other hand, for a pair of fully correlated spins we have ${c_{ij}}^2 = 1$; the two spins are either aligned or anti-aligned in all states.

To proceed, it is convenient to define, quite generally, $\mathcal{G}_{\mu\nu}$ as the set of spins whose sign is different in μ and ν, *i.e.*

$$\mathcal{G}_{\mu\nu} = \{ \, i \mid S_i^\mu \neq S_i^\nu \, \} \ . \tag{20}$$

We expect the largest domain, $\mathcal{G}_1$, to be in one orientation in the states of $\mathcal{C}$ and in the reversed one in the states of $\bar{\mathcal{C}}$. To identify the spins that indeed behave this way, we took all $(M/2)^2$ pairs of states $\mu \in \mathcal{C}$ and $\nu \in \bar{\mathcal{C}}$ and, for each pair, determined $\mathcal{G}_{\mu\nu}$. Ideally all the spins of $\mathcal{G}_1$ always flip together and maintain their relative orientation; if so, the set of spins $\mathcal{G}_{\mu\nu}$ for all pairs of states μ and ν would always include $\mathcal{G}_1$. However, at finite T we must allow for excitations of the order of J. So, even if a spin is highly correlated with the other spins of $\mathcal{G}_1$, it might lose its relative orientation in a few of the M states of the sample. In order not to "miss" such spins, we use a soft criterion when we determine whether a spin is a member of $\mathcal{G}_1$. We define a threshold θ and define $\tilde{\mathcal{G}}_1(\theta)$ as the set of spins i which are members of $\mathcal{G}_{\mu\nu}$, i.e. for which $S_i^\mu S_i^\nu = -1$, for at least a fraction θ of the pairs of states $\mu \in \mathcal{C}$ and $\nu \in \bar{\mathcal{C}}$. This can be written as

$$\tilde{\mathcal{G}}_1(\theta) = \left\{ \, i \ \middle| \ \frac{1}{|\mathcal{C}||\bar{\mathcal{C}}|} \sum_{\mu\in\mathcal{C}} \sum_{\nu\in\bar{\mathcal{C}}} S_i^\mu S_i^\nu < 1 - 2\theta \right\} \ , \tag{21}$$

since the terms in the normalized sum where $S_i^\mu S_i^\nu = 1$ must, by definition, sum up to less than $1-\theta$ and the sum of the terms with $S_i^\mu S_i^\nu = -1$ must be less than $-\theta$. We define our spin domain $\mathcal{G}_1(\theta)$ as the largest contiguous part of $\tilde{\mathcal{G}}_1(\theta)$. For large enough θ we found that for most realizations $\{J\}$, below T_c the sites of $\tilde{\mathcal{G}}_1(\theta)$ are contiguous and hence it is identical to $\mathcal{G}_1(\theta)$ (for detailed values of the ratio $|\mathcal{G}_a|/|\tilde{\mathcal{G}}_1|$, its mean over realizations and its standard deviation, see [37]). The next spin domain $\mathcal{G}_2(\theta)$ is defined in the same manner, on the basis of pairs of states $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$.

The above definition sets a lower bound on the correlation of spins within the domain. Consider two spins $i, j \in \mathcal{G}_1(\theta)$. By definition,

$${c_{ij}}^2 = \frac{1}{M^2} \sum_{\mu,\nu} S_i^\mu S_j^\mu S_i^\nu S_j^\nu. \tag{22}$$

Now the number of states in $\mathcal{C}$ and $\bar{\mathcal{C}}$ are both equal to $M/2$. In addition, for a given ν, we can replace μ by its inverse $\bar{\mu}$ and the product of the four spins doesn't change. Hence we get the same contribution from $\mu \in \mathcal{C}$ as $\bar{\mu} \in \bar{\mathcal{C}}$. As result we have

$${c_{ij}}^2 = \frac{1}{|\mathcal{C}||\bar{\mathcal{C}}|} \sum_{\mu\in\mathcal{C}} \sum_{\nu\in\bar{\mathcal{C}}} S_i^\mu S_j^\mu S_i^\nu S_j^\nu \ . \tag{23}$$

Now $S_i^\mu S_i^\nu$ will be -1 for a fraction of the states μ and ν which is greater than θ and $+1$ for a fraction less than $1-\theta$, and similarly for $S_j^\mu S_j^\nu$. Hence $S_i^\mu S_i^\nu$ and $S_j^\mu S_j^\nu$ will have the same sign with probability greater than $1-2(1-\theta) = 2\theta - 1$. Consequently, for $i, j \in \mathcal{G}_1(\theta)$, we have

$${c_{ij}}^2 > 2\theta - 1 - [1 - (2\theta - 1)] = 4\theta - 3. \tag{24}$$

The same constraint holds also for $\mathcal{G}_2$, with the sums taken over the states in clusters $\mathcal{C}_1$ and $\mathcal{C}_2$.

Since we introduced an arbitrary parameter θ into the definition of our spin clusters, it is important to consider the extent to which the value of θ affects their identification. As seen in Fig. 7, the sizes of the domains and their average correlation, defined below in (25), do not change much for $0.6 \leq \theta \leq 0.95$. For both $a = 1, 2$ we define (arbitrarily) $\mathcal{G}_a = \mathcal{G}_a(0.95)$. We do not choose $\theta = 1$ since, as discussed above, we do not want our results to be affected by small thermal fluctuations. In Fig. 8 we plot the spatial structure of $\mathcal{G}_1$ and $\mathcal{G}_2$ for a specific realization. For $T > T_c$ the correlations between each pair of spins are much smaller, and hence this analysis is meaningless. The procedure described above results in $\mathcal{G}_1(\theta) = \mathcal{G}_2(\theta) = \emptyset$ for any $\theta > 0.5$.

According to our picture these correlated spin domains govern the hierarchical structure of state space. It is important to clarify whether these domains survive as the system size L increases. There are two mechanisms by which increasing the system size can invalidate our picture: either the domains do not remain macroscopic when L increases, or they do remain macroscopic but merge as $L \to \infty$, i.e. the fraction of states in which $\mathcal{G}_2$ flips tends to zero. We now discuss each of these possibilities in turn. In addition, a simple figurative description of these two mechanisms is given Sec. V B.

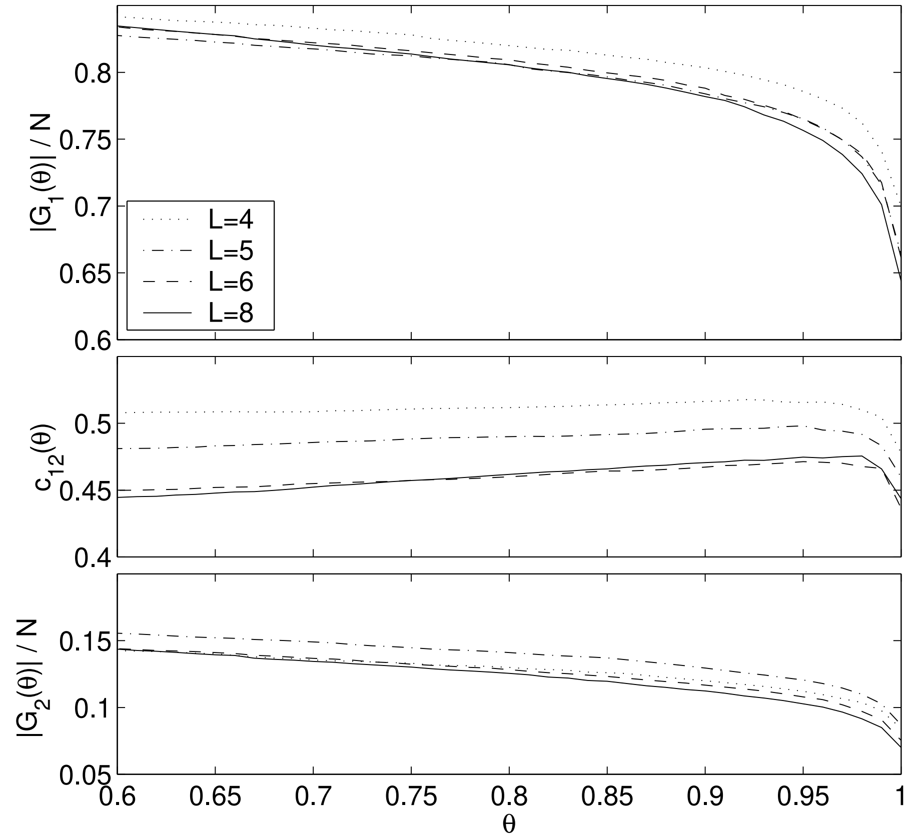

FIG. 7. The normalized sizes of the two largest spin domains, $\mathcal{G}_1(\theta)/N$ and $\mathcal{G}_2(\theta)/N$ and their correlation $\bar{c}_{12}$, defined in Eq. (25), as a function of the threshold θ for $D = 3$, $T = 0.2$.

1. The domains do not remain macroscopic when L increases. To study the finite size effects of our analysis we normalized the domain sizes by the number of spins and plotted the size distributions of the two domains for different system sizes, in $D = 3$ (see Fig. 9) and in $D = 4$ (Fig. 10), at two temperatures in both dimensions. The number of bond realizations, $N_{\rm samp}$, from which these distributions were obtained for various system sizes L, at both $D = 3, 4$, are given in Table I. For $T = 0.2$ in both dimensions, and at $T = 0.5$ for $D = 3$ the distributions seem to converge even for the small system sizes we use. We conclude with high certainty that at $T = 0.2$ for $D = 3, 4$ and at $T = 0.5$ for $D = 3$ the domain sizes $|\mathcal{G}_a|$ are proportional to L^D for both $a = 1, 2$. The mean and width of these distribution are presented in Table III. The width of the distributions does not vanish, so the sizes of the domains are non self-averaging quantities. On the other hand, for $T = 0.8$ in $D = 4$ we cannot determine conclusively whether the domain sizes do or do not remain proportional to $N = L^D$ as L increases.

2. $\mathcal{G}_1$ and $\mathcal{G}_2$ may remain macroscopic but merge as $L \to \infty$. If this occurs, we end up with a single domain and there will be no hierarchical structure in state space. To check that this does not happen we calculated the average correlation $\bar{c}_{12}$ between spins in $\mathcal{G}_1$ and $\mathcal{G}_2$,

$$\bar{c}_{12} = \frac{1}{|\mathcal{G}_1||\mathcal{G}_2|} \sum_{i \in \mathcal{G}_1} \sum_{j \in \mathcal{G}_2} {c_{ij}}^2 \ . \tag{25}$$

   If $\bar{c}_{12}$ approaches the value 1 as $L \to \infty$, the two domains indeed merge in the thermodynamic limit. In Table III we present, for systems of different sizes and dimensions, the average values of $\bar{c}_{12}$ (averaged over the disorder $\{J\}$) and the corresponding standard deviations. For $T = 0.2$, $D = 3, 4$ and for $T = 0.5$, $D = 3$ the average correlation decreases slightly as the system size increases, although, in $D = 3$ it seems to converge already for $L = 8$ to a fixed value of ~ 0.5. This means that the spins of $\mathcal{G}_1$ and $\mathcal{G}_2$ will not become fully correlated and the two domains will stay separate as L increases.

   Interestingly, in $D = 4$, the correlation for $L = 4, 5$ is higher at $T = 0.8$ than at $T = 0.2$. The reason for this is probably that as T increases, small pieces of $\mathcal{G}_1$ "fall of". Since $\mathcal{G}_2$ at $T = 0.2$ is small, one of these pieces,

T	L	$[D(\mathcal{C})]_J$	$\Delta D(\mathcal{C})$	$[w(\mathcal{C})]_J$	$\Delta w(\mathcal{C})$	$[D(\mathcal{C}_1)]_J$	$\Delta D(\mathcal{C}_1)$	$[w(\mathcal{C}_1)]_J$	$\Delta w(\mathcal{C}_1)$
0.2	4	0.045	0.049	0.055	0.052	0.015	0.017	0.019	0.018
	5	0.050	0.054	0.056	0.054	0.018	0.018	0.019	0.019
	6	0.053	0.056	0.054	0.053	0.021	0.020	0.019	0.019
	8	0.055	0.054	0.052	0.051	0.025	0.020	0.020	0.020
0.5	8	0.139	0.065	0.084	0.046	0.093	0.038	0.045	0.026
	12	0.151	0.065	0.078	0.046	0.106	0.036	0.041	0.024
2.0	8	0.487	0.006	0.053	0.002	0.477	0.009	0.055	0.002
		$[D(\mathcal{C}_2)]_J$	$\Delta D(\mathcal{C}_2)$	$[w(\mathcal{C}_2)]_J$	$\Delta w(\mathcal{C}_2)$	$[D(\mathcal{C}_1,\mathcal{C}_2)]_J$	$\Delta D(\mathcal{C}_1,\mathcal{C}_2)$	$[w(\mathcal{C}_1,\mathcal{C}_2)]_J$	$\Delta w(\mathcal{C}_1,\mathcal{C}_2)$
0.2	4	0.025	0.036	0.027	0.034	0.160	0.135	0.026	0.024
	5	0.025	0.032	0.025	0.031	0.169	0.147	0.023	0.020
	6	0.028	0.033	0.026	0.033	0.161	0.141	0.022	0.021
	8	0.030	0.027	0.024	0.026	0.161	0.139	0.021	0.018
0.5	8	0.112	0.057	0.057	0.037	0.253	0.126	0.053	0.027
	12	0.121	0.048	0.054	0.033	0.263	0.125	0.044	0.023
2.0	8	0.472	0.009	0.057	0.002	0.499	0.005	0.048	0.003

TABLE II. The average distances within and between state clusters, and the relations between them, for a subset of the $D = 3$ dimensional systems. For each variable x we present the average over all realizations, $[x]_J$, followed by its standard deviation, i.e. $\Delta x = ([x^2]_J - [x]_J{}^2)^{1/2}$. The statistical error of each mean $[x]_J$ is $\Delta x/\sqrt{N_{\rm samp}}$; the number of samples for each L, D is given in Table I.

D	T	L	$[\lvert\mathcal{G}_1\rvert]_J/N$	$\Delta\lvert\mathcal{G}_1\rvert/N$	$[\lvert\mathcal{G}_2\rvert]_J/N$	$\Delta\lvert\mathcal{G}_2\rvert/N$	$[\bar{c}_{12}]_J$	$\Delta\bar{c}_{12}$	$P(\mathcal{G}_2 \neq \emptyset)$
3	0.2	4	0.70(1)	0.21	0.099(4)	0.087	0.56(1)	0.33	0.856(6)
		5	0.66(1)	0.21	0.105(5)	0.104	0.55(1)	0.33	0.832(6)
		6	0.66(1)	0.20	0.090(4)	0.090	0.52(2)	0.34	0.836(6)
		8	0.64(1)	0.20	0.084(5)	0.094	0.53(2)	0.34	0.833(8)
	0.5	4	0.31(1)	0.21	0.062(3)	0.056	0.49(1)	0.32	0.56(1)
		5	0.26(1)	0.18	0.052(2)	0.043	0.49(1)	0.33	0.57(1)
		6	0.25(1)	0.16	0.046(2)	0.046	0.47(1)	0.33	0.52(1)
		8	0.22(1)	0.15	0.035(2)	0.034	0.47(2)	0.31	0.55(1)
		12	0.24(1)	0.15	0.033(2)	0.035	0.54(2)	0.31	0.56(2)
4	0.2	3	0.74(1)	0.19	0.107(5)	0.105	0.62(2)	0.34	0.840(6)
		4	0.73(1)	0.19	0.083(4)	0.092	0.53(2)	0.34	0.830(6)
		5	0.73(1)	0.19	0.082(7)	0.098	0.51(2)	0.34	0.77(1)
	0.8	3	0.154(7)	0.15	0.036(1)	0.031	0.47(1)	0.31	0.298(9)
		4	0.142(6)	0.12	0.025(1)	0.029	0.54(1)	0.31	0.37(1)
		5	0.139(8)	0.11	0.020(2)	0.025	0.57(2)	0.29	0.38(2)

TABLE III. The normalized sizes of the domains $\mathcal{G}_1$ and $\mathcal{G}_2$, and the average correlation between spins that belong to the two domains. The last two parameters are taken for realizations $\{J\}$ where $\mathcal{G}_2$ does not vanish. The probability for $\mathcal{G}_2$ not to vanish is also presented. For each quantity x the table contains $[x]_J$, its average over $N_{\rm samp}$ realizations of the disorder $\{J\}$ and the width of the distribution $\Delta x = \sqrt{[x^2]_J - [x]_J{}^2}$. Next to each $[x]_J$ we show its statistical error (in parentheses).

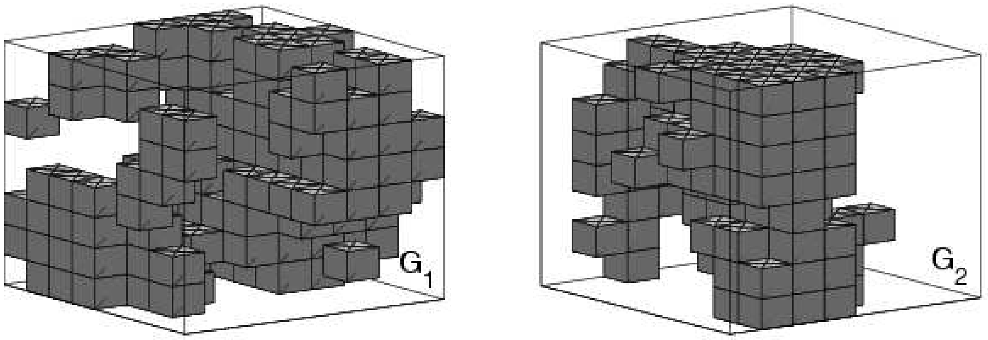

FIG. 8. The spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$, as found in the realization of Fig. 4. Note that we use periodic boundary conditions, so the domains are connected through the boundaries. No spin is shared by $\mathcal{G}_1$ and $\mathcal{G}_2$.

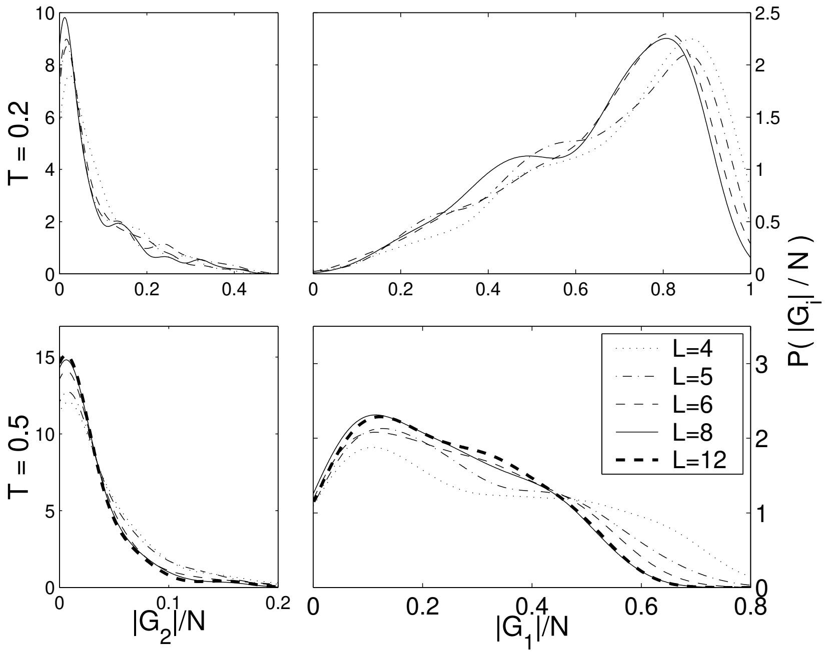

FIG. 9. Size distributions of the spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$ for $D = 3$ dimensions at $T = 0.2, 0.5$. The distributions seem to converge, despite the small system sizes.

which is larger than $\mathcal{G}_2$, assume the role of $\mathcal{G}_2$ at $T = 0.8$. Since this piece was part of $\mathcal{G}_1$ at $T = 0.2$, we expect its correlation, with what remains of $\mathcal{G}_1$ at $T = 0.8$, to be relatively high. Extrapolating from $L = 3, 4, 5$ is not useful, but we still believe that the correlation does not approach 1 as $L \to \infty$.

We also attempted to identify $\mathcal{G}_3$ and $\mathcal{G}'_3$, the spin domains associated with the third level of the state hierarchy (see below). $\mathcal{G}_3$ is the cluster which is associated with splitting $\mathcal{C}_1$ into its two descendents on the dendrogram, $\mathcal{C}_{1a}$ and $\mathcal{C}_{1b}$. The domain $\mathcal{G}'_3$ plays the same role in $\mathcal{C}_2$. Since by our notation $|\mathcal{C}_1| \geq |\mathcal{C}_2|$ we expected that in order to have a larger number of states, the spin correlations will be lower when measured over $\mathcal{C}_1$ than over $\mathcal{C}_2$. As a result we expect $|\mathcal{G}_3| \leq |\mathcal{G}'_3|$. Due to the small sizes of the systems we study, we cannot be sure if the sets of spins we identify as $\mathcal{G}_3$ and $\mathcal{G}'_3$ indeed play the role we attribute to them, or are just a microscopic noise and, therefore, only a finite size effect. The results are given in Table IV. We see that the normalized sizes of both domains decrease with the system size, perhaps due to finite size effects. We also measure the average correlation $\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2)$, of $\mathcal{G}_3$ with the largest domain correlated over $\mathcal{C}_1$, which includes $\mathcal{G}_1 \cup \mathcal{G}_2$ (this domain has a fixed orientation over the states of $\mathcal{C}_1$). This correlation is defined as

$$\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2) = \frac{1}{|\mathcal{G}_1 \cup \mathcal{G}_2||\mathcal{G}_3|} \sum_{i \in \mathcal{G}_1 \cup \mathcal{G}_2} \sum_{j \in \mathcal{G}_3} {c_{ij}}^2 \ . \tag{26}$$

In Table IV we see that the values of $\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2)$ decrease as L increases; hence if $\mathcal{G}_3$ survives as a macroscopic

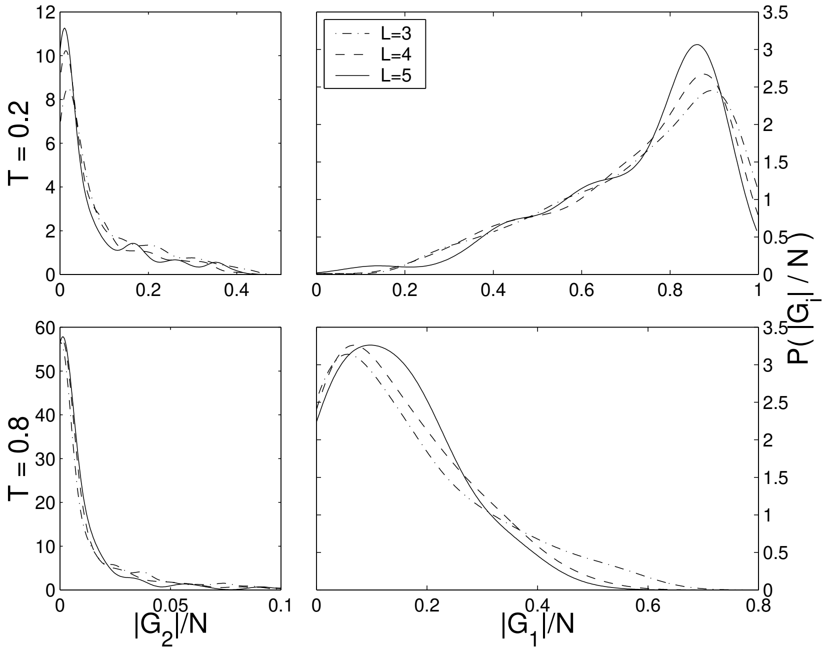

FIG. 10. Size distributions of the spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$ for $D = 4$ dimensions at $T = 0.2, 0.8$. For $T = 0.2$ the distributions seem to converge, despite the small system sizes. For $T = 0.8$, $|\mathcal{G}_2|/N$ converges to a narrow distribution around zero, and $|\mathcal{G}_1|/N$ does not show convergence yet.

cluster at large L, we expect it to remain distinct from the union of the two larger domains.

B. Spin space structure

So far we have obtained the spin domains using the results of the state space analysis. However, the existence of these domains can also be observed directly in spin space, i.e. without utilizing information about the previously identified hierarchical structure of state space, as we now demonstrate.

As described in Sec. IV, the equilibrium ensemble of states, obtained for each realization, is represented by an $N \times M$ data matrix $\{S_i^\mu\}$ (e.g. Fig 4(a)). In Sec. IV we treated each of the M *states*, represented by a column of this matrix, as a "data point" whose coordinates are the components of this N-dimensional vector. Now we view each of the N *spins* of the system as a data point, represented by a row of the same matrix. Each of these data points is a vector in an M-dimensional space.

The distance on the set of spins should be defined according to the nature of the clusters we are interested in. At this case, we expect highly correlated spins to be in the same cluster, and spins with low correlation to be in different clusters. Thus, we define the distance between a pair of data points i and j as

$$d_{ij} = 1 - {c_{ij}}^2 \ . \tag{27}$$

This $N \times N$ distance matrix serves as the input for clustering the spins, using Ward's algorithm. The dendrogram $\mathcal{D}$, obtained when the data of Fig. 4(a) are clustered, is presented in Fig. 11(a). The correlated spin clusters are represented by boxes in the dendrogram - let us denote them by $\tilde{g}_a$. When the spins are reordered according to the dendrogram, their distance matrix, shown in 11(b), clearly exhibits a non-trivial structure. There are large, highly correlated spin clusters on the lower levels of the dendrogram.

In order to "see" the manner in which the spins are ordered, we return to the data matrix of Fig. 4(a). We obtained Fig. 4(b) from (a) by reordering the columns according to the state dendrogram in Fig. 3. If we now reorder the rows of Fig. 4(b) according to the spin dendrogram $\mathcal{D}$ in Fig. 11, we get Fig. 4(c), which is redrawn as Fig. 12 with labeling of the largest state clusters and spin domains. The cluster structure of the spins can be clearly be seen in Fig. 12. Spins in $\mathcal{G}_1$ clearly have the same orientation in the states of $\mathcal{C}$ but are inverted in the corresponding states of $\bar{\mathcal{C}}$. Spins in $\mathcal{G}_2$ have opposite orientations in $\mathcal{C}_1$ and $\mathcal{C}_2$ and are inverted in the corresponding states of $\bar{\mathcal{C}}_1$ and $\bar{\mathcal{C}}_2$. One can also see that spins in domain $\mathcal{G}'_3$ separate $\mathcal{C}_2$ into two sub-clusters. As to $\mathcal{G}_3$, we point in Fig. 12 to a few (3 - 4) spins, which have the same sign in all states of $\mathcal{C}_2$ but change sign in $\mathcal{C}_1$.

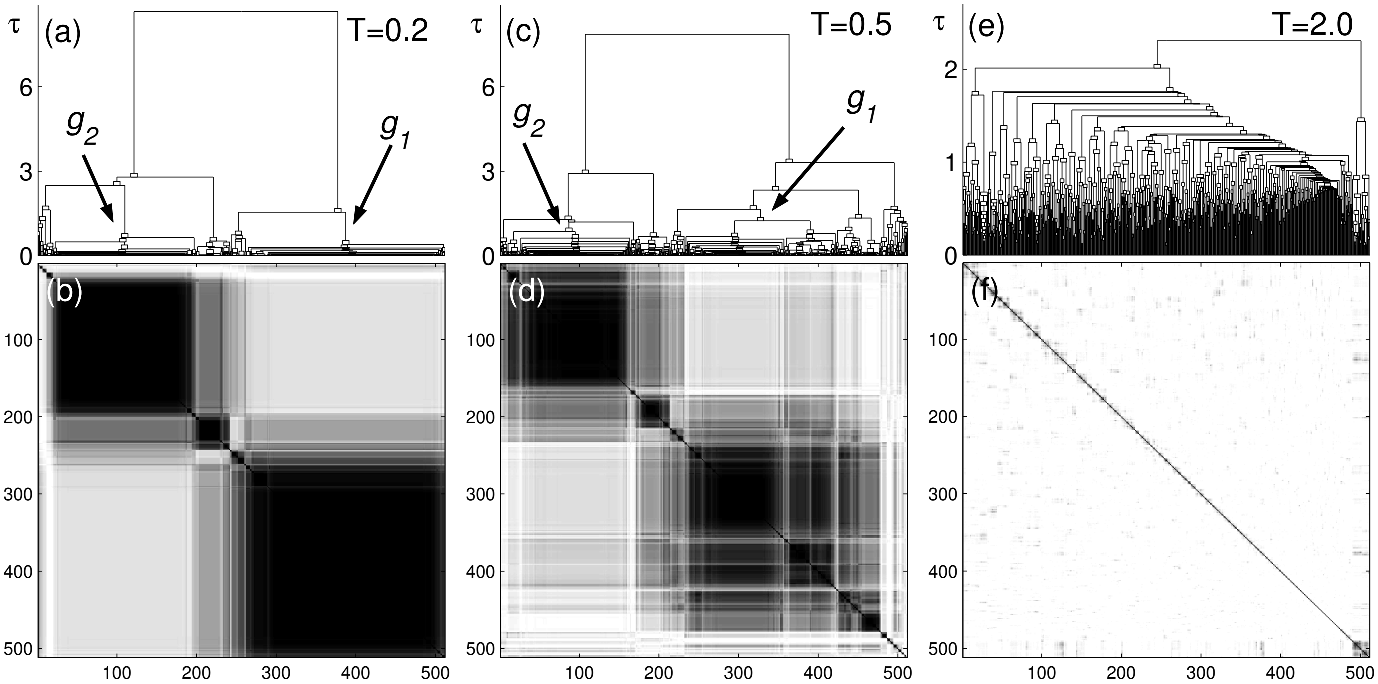

FIG. 11. **(a)** The spin dendrogram $\mathcal{D}$ for the data of Fig. 4(a) produced by Ward's algorithm. **(b)** The spin distance matrix d of this realization realization. The spins are ordered according to their clusters in $\mathcal{D}$. Darker shades correspond to smaller distances and higher correlations. **(c), (d)** The same as in (a), (b), for the same realization at $T = 0.5$. **(e), (f)** The same as in (a), (b), for the same realization at $T = 2.0$. The y-axis is rescaled to show the dendrogram, which clearly differ from the dendrograms in (a) and (c).

These data were obtained at $T = 0.2$ ($< T_c$). Above T_c the correlation between any two spins is low, and there is no cluster structure, as evident from Fig. 11(e,f). The relative τ values of this dendrogram are much smaller then those of the dendrograms in Figs. 11(a,c), and the reordered distance matrix is structure-less. If the domains $\mathcal{G}_a$ (that were identified in Sec. V A on the basis of the state hierarchy) are not an artifact of our analysis, they should be clearly identifiable in spin space, and appear as clusters in the spin dendrogram $\mathcal{D}$. To check this, for each realization we compared every spin cluster $\tilde{g}_a$, that appears in the corresponding spin dendrogram $\mathcal{D}$, to every spin domain $\mathcal{G}_a$ that was previously found for that realization. The spin cluster $\tilde{g}_a$ that was found to be most similar to $\mathcal{G}_a$ was identified and denoted by g_a. We used the similarity measure

$$\mathcal{S}(g_a, \mathcal{G}_a) = \frac{2|g_a \cap \mathcal{G}_a|}{|g_a| + |\mathcal{G}_a|} \tag{28}$$

which represents the fraction of shared spins by the "physical spin domain" $\mathcal{G}_a$ and the spin cluster g_a. For most realizations we have (at low T) $g_a = \mathcal{G}_a$ for both $a = 1, 2$; and when these groups are not precisely equal, they differ by only a few spins (see [37] for full details).

Fig. 12 also provides a convenient, simple "geometrical" interpretation of the two tests for the survival of our picture in the large L limit that we discussed in Sec. V A. Observe the rectangular region corresponding to spin domain $\mathcal{G}_2$ and state cluster $\mathcal{C}_2$. Validity of our picture relies on "survival" of this rectangle as we take the $L \to \infty$ limit. The first test we performed checked whether its vertical side, $|\mathcal{G}_2|$ stays finite. If this condition is not satisfied, the relative area of our rectangle goes to zero; a non-vanishing limiting $|\mathcal{G}_2|$ does not, however, guarantee that the rectangle stays finite; it may disappear if its horizontal dimension shrinks to zero when $L \to \infty$. The second test, showing that the correlation $\bar{c}_{12}$ does not approach 1, ensures that this does not happen either.

Overall, Fig. 12 summarizes in a convenient pictorial way our picture of the spin glass state in short range systems.

C. Spin domains and states hierarchy

Now that the spin domains have been well defined, we can examine the manner in which they govern the hierarchical partitioning of state space. Each state cluster at level a of the hierarchy can now be identified with one of two possible

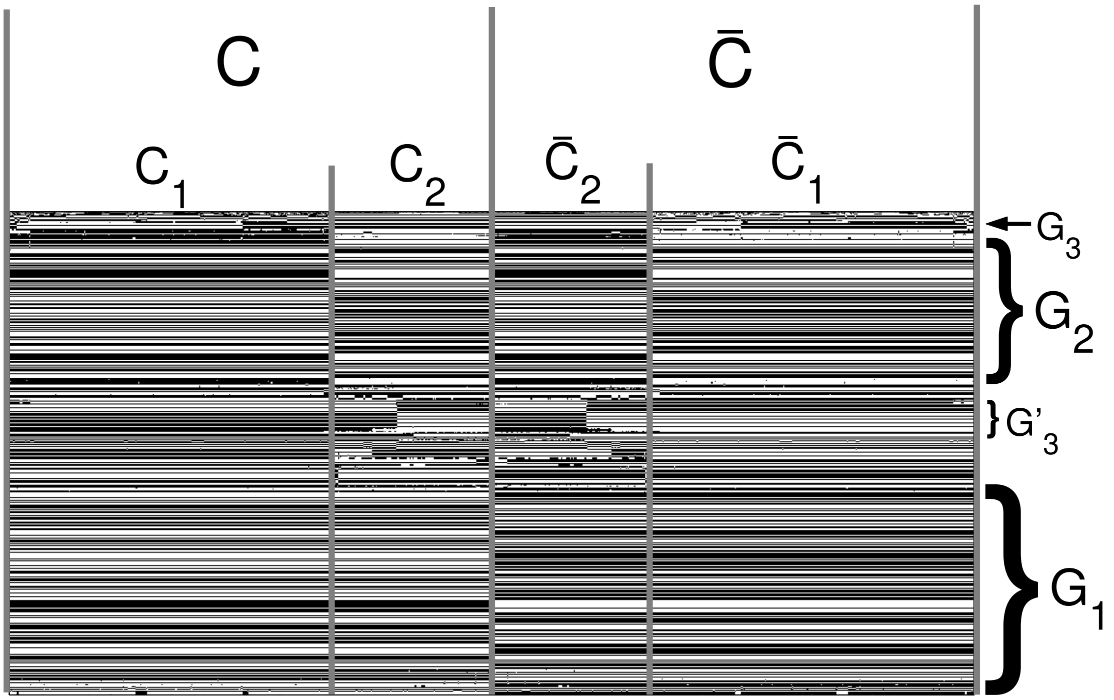

FIG. 12. A redrawing of the ordered data matrix of Fig. 4(c), in order to highlight the state clusters and spin domains discussed in the text. It is for a $3D$ realization of size $N = 8^3$ at $T = 0.2$. The columns represent the states μ and the rows represent the spins i, $S_i^\mu = \pm 1$, with black/white representing $+/-$. The states are ordered according to the dendrogram in Fig. 3, and the spins ordered according to the spin dendrogram $\mathcal{D}$ in Fig. 11. The state clusters and the spin domains are marked (see text).

configurations of the particular spin domain $\mathcal{G}_a$. We denote these two configurations as $\Uparrow_a$ and $\Downarrow_a$. Note that we have avoided the notation $+/-$ for the states of the spin domains, since in each state some of the spins have the $+$ sign and others $-$. For example, in the first level partition $\mathcal{G}_1$ has a certain characteristic configuration, $\Uparrow_1$, over all the states in $\mathcal{C}$, whereas over all the states of $\bar{\mathcal{C}}$ it is in the spin inverted configuration $\Downarrow_1$. The value $[\Uparrow_1]_i$, taken by spin $i \in \mathcal{G}_1$ in the configuration $\Uparrow_1$, is defined by

$$[\Uparrow_1]_i = \mathrm{sign}\left(\sum_{\mu \in \mathcal{C}} S_i^\mu\right) . \tag{29}$$

Our definition of $\mathcal{G}_1$, using Eq. (21) with $\theta = 0.95$, guarantees that the argument of the sign function in the above expression does not vanish. Hence, stating that $\mathcal{G}_1$ takes configuration $\Uparrow_1$ in a certain state μ implies that

$$\sum_{i \in \mathcal{G}_1} S_i^\mu [\Uparrow_1]_i > 0 . \tag{30}$$

The configuration assumed by $\mathcal{G}_1$ in any state μ determines that μ is assigned to $\mathcal{C}$ if $\mathcal{G}_1$ is in configuration $\Uparrow_1$, or to $\bar{\mathcal{C}}$ if $\mathcal{G}_1$ is in configuration $\Downarrow_1$.

The spin domain $\mathcal{G}_2$ determines, in a similar way, the partition of $\mathcal{C}$ into $\mathcal{C}_1$ and $\mathcal{C}_2$ (and the partition of $\bar{\mathcal{C}}$ into $\bar{\mathcal{C}}_1$ and $\bar{\mathcal{C}}_2$). $\mathcal{G}_2$ is in configuration $\Uparrow_2$ in states $\mathcal{C}_1$ and $\bar{\mathcal{C}}_2$, and in $\Downarrow_2$ in states $\mathcal{C}_2$ and $\bar{\mathcal{C}}_1$ (see Fig. 1 for a schematic illustration of this point).

Each spin domain $\mathcal{G}_a$ defines a partition of the states, at level a, into two groups - one in which $\mathcal{G}_a$ is in the $\Uparrow_a$ configuration and the other with $\Downarrow_a$. Picking a pair of states μ and ν, one from each group, the set of spins $\mathcal{G}_{\mu\nu}$, that are flipped in the transition between them, will always include $\mathcal{G}_a$ [42]. Thus, the distance $D_{\mu\nu} = |\mathcal{G}_{\mu\nu}|/N$ between two such states will almost always be larger then $|\mathcal{G}_a|/N$.

By our definition of $\mathcal{G}_a$, the probability that a large part of its spins will lose their relative orientation is small. Considering local dynamics, the time it will take $\mathcal{G}_a$ to flip is exponential in its size. If $\mathcal{G}_a$ is macroscopic (as we have shown for $a = 1, 2$) it may be associated with a macroscopic free energy barrier. In an infinite system it will take an infinite time to flip, thus inducing a separation of the phase space into two ergodic sub-spaces (or valleys).

The clear hierarchical organization of the state clusters suggest that the average distance (15) between state clusters formed at a high level of the hierarchy is significantly larger then the average distance between clusters formed at a lower level. Indeed, we show in Table II that in general $D(\mathcal{C},\bar{\mathcal{C}}) \gg D(\mathcal{C}_1,\mathcal{C}_2)$. We relate this characteristic of the state structure to the large variability of the spin domain sizes $|\mathcal{G}_a|$. Indeed, we have seen that typically $|\mathcal{G}_1| > 8|\mathcal{G}_2|$ for $T = 0.2$, $D = 3, 4$.

Now we have a complete picture, supported by our numerical findings, of a hierarchy of state clusters. The valleys are the leaves of this hierarchy [43]. At each level a of this hierarchy the partition of the states is refined according to the orientation of macroscopic spin domains $\mathcal{G}_a$. At different nodes of a certain level of the hierarchy there might be different correlated domains that determine their partition. Take, for example, the states in $\mathcal{C}_1$ (where $\mathcal{G}_1$ is in configuration $\Uparrow_1$ and $\mathcal{G}_2$ is in configuration $\Uparrow_2$). Over these states the largest *unlocked* [44] correlated domain is $\mathcal{G}_3 = \mathcal{G}_3(\Uparrow_1, \Uparrow_2)$. The two possible configurations of $\mathcal{G}_3$ inside $\mathcal{C}_1$ may be denoted as $\Uparrow_3(\Uparrow_1, \Uparrow_2)$ and $\Downarrow_3(\Uparrow_1, \Uparrow_2)$. Over the states of $\mathcal{C}_2$ we expect to find a different unlocked correlated domain $\mathcal{G}'_3 = \mathcal{G}'_3(\Uparrow_1, \Downarrow_2)$. We calculated the part of each domain which is included in the other. The results are given in Table IV. We see that $\mathcal{G}_3$ and $\mathcal{G}'_3$ share in general less than a fifth of their spins.

Note that in the ideal case (corresponding to $\theta = 1$), a spin domain $\mathcal{G}_a(\Uparrow_1, \Uparrow_2, \ldots \Uparrow_k)$, that appears at a particular level of the hierarchy, cannot share spins with the higher level domains $b = 1, 2, ..., k$, whose orientation is fixed while $\mathcal{G}_a$ flips. For $\theta = 0.95$ such sharing was also practically excluded. On the other hand, two domains such as $\mathcal{G}_3$ and $\mathcal{G}'_3$ *can* have shared spins, namely those that are free to flip in both the $(\Uparrow_1, \Uparrow_2)$ and $(\Uparrow_1, \Downarrow_2)$ situations.

Going all the way down the states hierarchy, we find that each valley can be characterized by a specific list of domain configurations, e.g. $\{\Uparrow_1,\ \Downarrow_2,\ \Downarrow_3(\Uparrow_1, \Downarrow_2),\ \Uparrow_4(\Uparrow_1, \Downarrow_2, \Downarrow_3),\ \ldots\}$.

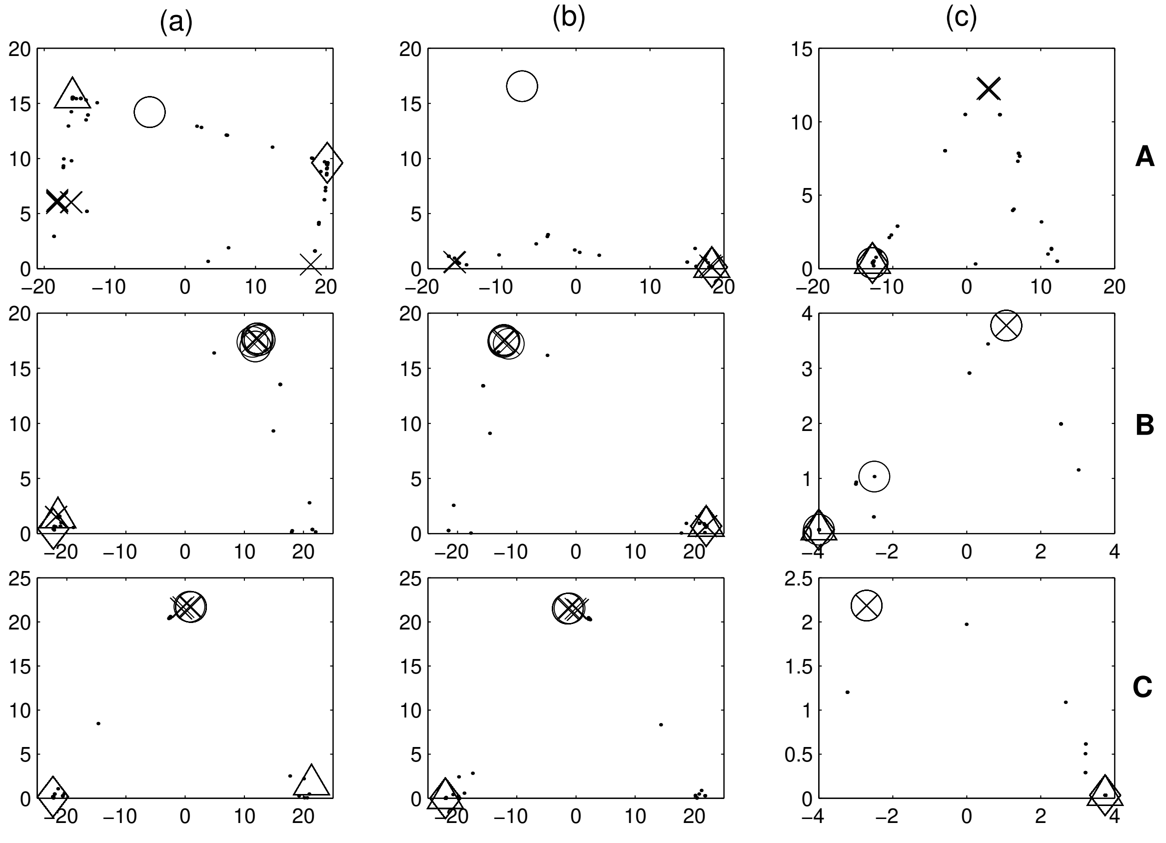

FIG. 13. The two principal components of the 512 spins of three realizations A, B and C (see text) in 3D. Each point represents a spin i and its coordinates are the projections of $\mathbf{S}_i = (S_i^1, S_i^2, \cdots, S_i^M)$ on to the two largest eigenvectors of the matrix R in Eq. (31). The analysis is carried over (a) all states; (b) the states of $\mathcal{C}_1$; and (c) the states of $\mathcal{C}_2$. The spins of $\mathcal{G}_1$ are marked by $\diamond$; of $\mathcal{G}_2$ by $\triangle$; of $\mathcal{G}_3$ by $\bigcirc$; and of $\mathcal{G}'_3$ by $\times$. Spins that belong to both $\mathcal{G}_3$ and $\mathcal{G}'_3$ are marked by $\otimes$. Spins that do not belong to any of these domains are marked with dots. The lower half of the plane is projected onto the upper half using $(x, y) \to (-x, -y)$. Spins in a correlated domain usually have the same values for the two principal components, and they fall on top of each other on the plot. Therefore, in most plots, a correlated domain seems to be represented by a single marker.

An additional insight is obtained from a PCA of the spins, which is to be distinguished from the PCA of the states in Fig. 5. To perform the PCA of the spins we form the covariance matrix

$$R_{\mu\nu} = \frac{1}{N}\sum_{i=1}^{N} \delta S_i^\mu \delta S_i^\nu \ , \tag{31}$$

which is analogous to Eq. (11), and project the two largest eigenvectors of R on to the spin configurations $\mathbf{S}_i$ for each site i.

The results for three realizations, labeled A, B and C, are shown in Fig. 13. Each data point represents one spin. Realization A is the one whose data matrix is shown in in Figs. 4(c) and Fig. 12.

In the upper left frame of Fig. 13 we see the results of the PCA analysis of the spins for realization A. We want highly correlated spins to be close on the plot. Since a spin $\mathbf{S}_i$ is fully correlated with its inverse $-\mathbf{S}_i$ each point (x, y) with $y < 0$ is projected on the plot to $(-x, -y)$. The spins of $\mathcal{G}_1$ are highly correlated with each other and all have the same values for the first two principal components of the spin space. Therefore they fall on top of each other, and we see only one $\diamond$ marker which represents all of them. The same is true for the spins of $\mathcal{G}_2$, marked by $\triangle$. As seen from Fig. 12 the spins of $\mathcal{G}_1$ are not correlated with the spins of $\mathcal{G}_2$ over the M states, and indeed the two domains are far from each other on the plot.

In column (b) of Fig. 13 we used only the states of $\mathcal{C}_1$ in the analysis. We can see in Fig. 12 that over $\mathcal{C}_1$ the spins of $\mathcal{G}_1$ and $\mathcal{G}_2$ are correlated, together with some of the spins of $\mathcal{G}'_3$, marked by $\times$. In the plot (the middle frame on the upper row of Fig. 13) we can see that indeed these spins are all plotted at the same coordinates. The spins of $\mathcal{G}_3$, marked as $\bigcirc$, are highly correlated, but are not correlated with $\mathcal{G}_1$ and $\mathcal{G}_2$. Note that the spins of $\mathcal{G}'_3$ are separated into two different sets, and are not correlated over $\mathcal{C}_1$.

When we perform the analysis using only the states of $\mathcal{C}_2$ we get the results presented in column (c) of Fig. 13. In the matrix of Fig. 12 we see that the spins of $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ are correlated together over $\mathcal{C}_2$, and indeed they all fall on top of each other in the plot. We also see $\mathcal{G}'_3$ as a separated correlated domain.

In the second row of Fig. 13 we give the results for realization B, in which $\mathcal{G}_3$ and $\mathcal{G}'_3$ share some of their spins. Those spins are marked by $\bigotimes$. In column (c) we see these spins inside $\mathcal{G}'_3$. The rest of the spins of $\mathcal{G}_3$ are not correlated with them. Some of them are correlated with $\mathcal{G}_1$ and $\mathcal{G}_2$, and others seem to be in another domain.

In the third row of Fig. 13 we present the results for realization C in which $\mathcal{G}_3 \subset \mathcal{G}'_3$. Here spins of $\mathcal{G}'_3$ seem to form a correlated set also over $\mathcal{C}_1$, though the correlations are not high enough for it to be considered as a domain by our definition.

VI. STATE OVERLAP

We have presented a description of the system in its low T phase, relating state space behavior to the microscopic structure in spin space. Most of the previous literature, however, did not directly measure the microscopic features of the system but examined their indirect implications on other parameters, such as the widely addressed overlap distribution $P(q)$. Beyond making contact with the literature, which concentrates on measuring $P(q)$, the aim of this Section is two fold: (i) we show how our methods allow a useful decomposition of this function into its physically relevant constituent parts, and (ii) we demonstrate that our picture provides a *microscopic interpretation* of the observed $P(q)$. To this end we focus here on $P_J(q)$, the overlap distribution for a specific realization $\{J\}$ of the bonds, whereas earlier works [19,20,24,45] presented results for the average over the disorder, $P(q) = [P_J(q)]_J$.

Two technical comments should be first made. First, because of overall spin-flip symmetry, the function $P_J(q)$ is symmetric and hence we can limit our attention to $q > 0$. Second, since for most realizations $|\mathcal{G}_1| > N/2$, we have

$$P_J^{\mathcal{CC}}(q) \simeq \begin{cases} P_J(q) & q \geq 0 \\ 0 & q < 0 \end{cases} \tag{32}$$

where by $P_J^{\mathcal{CC}}(q)$ we denote the distribution of overlaps between pairs of states $\mu, \nu \in \mathcal{C}$, so that we have to deal only with such pairs.

A. Decomposition of $P_J(q)$ and $P(q)$

The overlap distribution for a specific realizations of the randomness, $P_J(q)$, is expected to be the sum of two main parts

$$P_J(q) = P_J^i(q) + P_J^o(q), \tag{33}$$

where $P^i_J(q)$ is the overlap distribution *within* a valley (and between a valley and its spin reversed counterpart), and $P^o_J(q)$ is the overlap distribution between states that belong to *two different* valleys. $P^i_J(q)$ converges to $\delta(|q|-q_{\rm EA})/2$ in the thermodynamic limit, where $q_{\rm EA}$ is the Edwards-Anderson order parameter, which will also be denoted as the "self-overlap". $P^o_J(q)$ is the sum of several contributions, corresponding to different pairs of valleys.

In the thermodynamic limit this separation is unambiguous; if two microstates μ and ν are separated by a macroscopic energy barrier, they belong to two different valleys and their overlap $q_{\mu\nu}$ contributes to $P^o_J(q)$. For finite systems this separation is problematic; our picture and method, however, does allow us to estimate $P^o_J(q)$ or, to be more precise, to calculate a function $\tilde{P}^o_J(q)$ defined below, which is a *lower bound* to it. In our picture, the transition between such pairs of microstates (that belong to two different valleys) is associated with flipping a specific set of spin domains. Consequently, having identified the relevant spin domains, we can identify when μ and ν belong to different valleys and also the level in the states' hierarchy at which they differ.

A remaining apparent ambiguity concerns the level of the state hierarchy at which we "stop" and decide whether a particular pair of microstates belongs to different valleys or not. Suppose we stop the decomposition of $\mathcal{C}$ at some level n and denote by $\mathcal{C}^n_\alpha$ the clusters obtained at this level. The overlaps obtained from pairs of microstates that belong to different valleys *at this level* are assigned to the distribution $P^{o,n}_J(q)$, and pairs from the same valleys to $P^{i,n}_J(q)$:

$$
\begin{aligned}
P^{i,n}_J(q) &= \sum_\alpha P_J^{C^n_\alpha C^n_\alpha}(q) \\
P^{o,n}_J(q) &= \sum_{\alpha\neq\beta} P_J^{C^n_\alpha C^n_\beta}(q),
\end{aligned} \tag{34}
$$

where, from Eq. (33)

$$
P_J(q) = P^{i,n}_J(q) + P^{o,n}_J(q) \qquad \text{for } q \geq 0. \tag{35}
$$

Clearly, by going down a level further, to $n+1$, some pairs that were assigned to $P^{i,n}_J(q)$ will be reassigned to $P^{o,n+1}_J(q)$, but if a pair was in $P^{o,n}_J(q)$ it will stay in $P^{o,n+1}_J(q)$. This argument clearly shows that $P^{o,n}_J(q)$ obtained at any level is a *lower bound* to $P^o_J(q)$. This point is explained again below for the particular case of $n=2$.

To demonstrate how natural is the separation of eq. (33), we consider pairs of states $\mu\in\mathcal{C}_1$ and $\nu\in\mathcal{C}_2$, i.e. pairs taken from state clusters that appear at the second ($n=2$) level of the states' hierarchy. According to our picture such pairs contribute a non-vanishing part of $P_J(q)$, which we denote by $P^{C_1C_2}_J(q)$ ($\equiv P^{o,2}_J(q)$, since for $n=2$ $\mathcal{C}$ has only these two sub clusters) This function, as well as its complement $P_J(q)-P^{C_1C_2}_J(q)$ are presented, for $T=0.2$ and $L=8$ in Fig. 14, for four realizations of the randomness. The figure shows clearly that the separation is natural, and not just an artifact of our analysis.

For all these four realizations the spin domain $\mathcal{G}_2$ is clearly identifiable and is "macroscopic" (note that this holds for more than 80% of the realizations, see table III). In all these cases the states μ and ν belong to different valleys, and contribute to $P^o_J(q)$. There may be, however, pairs of states which also contribute to $P^o_J(q)$, but are *not* included in $P^{C_1C_2}_J(q)$. This happens when (at least) one of the state clusters $\mathcal{C}_1,\mathcal{C}_2$ has internal structure and decomposes into sub-clusters (i.e. higher level valleys). Say $\mathcal{C}_1$ contains two such sub-clusters, $\mathcal{C}_{1a},\mathcal{C}_{1b}$. The overlap of a pair of states $\mu\in\mathcal{C}_{1a}$ and $\nu\in\mathcal{C}_{1b}$ contributes to $P^o_J(q)$, and is not included in $P^{C_1C_2}_J(q)$; hence the latter function is a *lower bound* on the former. As discussed above in Sec. V, such internal structure of $\mathcal{C}_1$ (or $\mathcal{C}_2$) is associated with a spin domain $\mathcal{G}_3$ (or $\mathcal{G}'_3$). This structre is clearly present for the realizations in Fig. 14 (a) and (d), as evident from the multi-peaked structure of $P^{C_1C_2}_J(q)$s which is discussed further below.

We now generate a distribution $\widetilde{P}^{C_1C_2}(q)$ which is a lower bound on the contribution of $P^{C_1C_2}(q)\equiv[P^{C_1C_2}_J(q)]_J$ to the *average* distribution $P(q)$. In order to assure that $\widetilde{P}^{C_1C_2}(q)$ constitutes a lower bound to $P^{C_1C_2}(q)$, we included in $\widetilde{P}^{C_1C_2}(q)$ *only* contributions $P^{C_1C_2}_J(q)$ from those realizations J in which $\mathcal{G}_2$ was relatively large, namely $|\mathcal{G}_2|>0.05N$. For the other realizations we set the contribution to the average over J to zero; hence our $\widetilde{P}^{C_1C_2}(q)$ is a lower bound to the true $P^{C_1C_2}(q)$ (which, in turn, is a lower bound to $P^o(q)$). In Figs. 15 we show the distributions $P(q)$ and $\widetilde{P}^{C_1C_2}(q)$. The data indicates that the weight in the tail for small q stays finite with increasing L (at least for this range of sizes), in agreement with earlier studies [19,20,24,45] which just measured $P(q)$. For systems with Gaussian couplings $P^i(q)$ has a very small contribution at $|q|<0.7$ and $\widetilde{P}^o(q)$ is the dominant part of $P(q)$ in this range. For an Ising spin-glass with binary couplings, however, the difference between the distributions is significant and proper care must be taken when delicate issues, such as triviality of $P(q)$, are investigated [31].

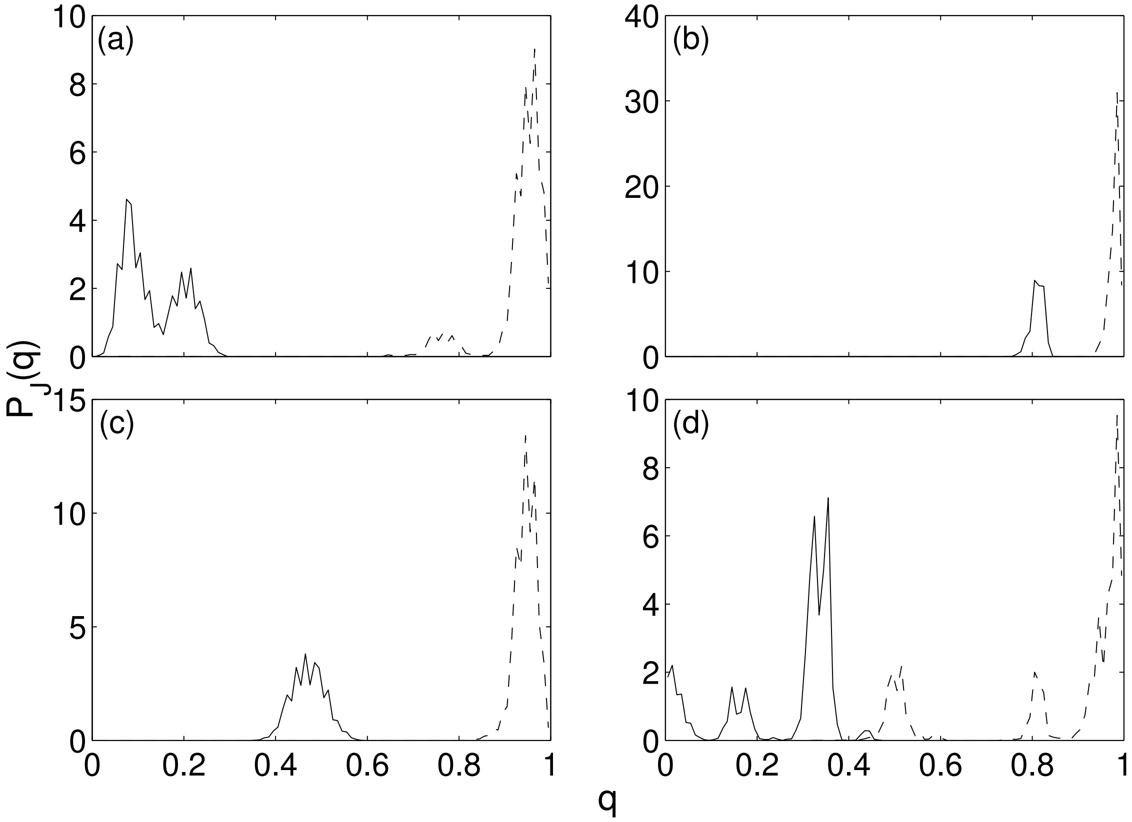

FIG. 14. The distribution $P_J(q)$ for four realizations of $\{J\}$ at $T = 0.2$ in $3D$. The distribution in (a) is the same as in the top frame of Fig. 16. The solid line describes $P_J^{C_1 C_2}(q)$ and the dashed line plots the rest of the distribution, $P_J(q) - P_J^{C_1 C_2}(q)$. The latter contains a large peak at $q \approx 1$ which is the distribution $P_J^i(q)$, of overlaps inside the valleys.

B. Interpretation of $P_J(q)$ in terms of spin domains

Our aim is to interpret the distribution $P_J(q)$, obtained for a particular realization, in terms of the state clusters $\mathcal{C}_i$ and spin domains $\mathcal{G}_a$ that were discussed in the previous sections. Before going into a detailed discussion and analysis, we state the interpretation that arises, for the four realizations whose $P_J(q)$ was shown in Fig. 14. The first of these, Fig. 14 (a), corresponds to a system in which $\mathcal{C}_2$ has internal structure, due to a sizeable domain $\mathcal{G}_3'$; its counterpart, $\mathcal{G}_3$ is too small to have a clear signature. The size of $\mathcal{G}_3'$ governs the splitting of the peak drawn with a solid line and also of the peak at high q (dashed line). In the systems of Fig. 14 (b) and (c) neither $\mathcal{C}_1$ nor $\mathcal{C}_2$ have noticeable internal structure; the domains $\mathcal{G}_3, \mathcal{G}_3'$ are microscopic. The system of 14 (d) has internal structure for both $\mathcal{C}_1$ and $\mathcal{C}_2$, induced by domains $\mathcal{G}_3$ and $\mathcal{G}_3'$, respectively. The sizes of these two domains govern the observed splitting of both the solid and dashed curves.

One can associate each peak of $P_J(q)$ with the overlaps of pairs of states that are related by flipping one or more of the previously identified spin domains. In this regard our interpretation resembles the RSB picture [5] which also relates the peaks of $P(q)$ [20] to overlaps between configurations in different valleys.

To substantiate these claims and make them more precise we consider in detail the realization whose (ordered) state and spin data matrix is given in Fig. 12, and whose $P_J(q)$ (shown in Fig. 14 (a)) is reproduced and magnified in Fig. 16. For this realization we clearly identified *three* spin domains; $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3'$. Disregarding the splitting induced by $\mathcal{G}_3'$ (and $\mathcal{G}_3$, if present) we identify two main peaks that dominate $P_J^{\mathcal{CC}}(q)$. We performed a fit of $P_J^{\mathcal{CC}}(q)$ to a sum of two Gaussians,

$$P_J^{\mathcal{CC}}(q) = b_1 \exp[(q - q_1)^2/{a_1}^2] + b_2 \exp[(q - q_2)^2/{a_2}^2] \ , \tag{36}$$

with a_i, b_i and q_i as fit parameters, yielding the dotted curves in the upper part of Fig. 16. The center of the (split) peak at low q is q_1 and the high-q data is centered at q_2.

To see how these q_i are related to our state clusters and spin domains, note that the overlap $q_{\mu\nu}$ between states μ and ν is related to the size of the set $\mathcal{G}_{\mu\nu}$ (defined in (20)), of spins that flip when passing from state μ to ν;

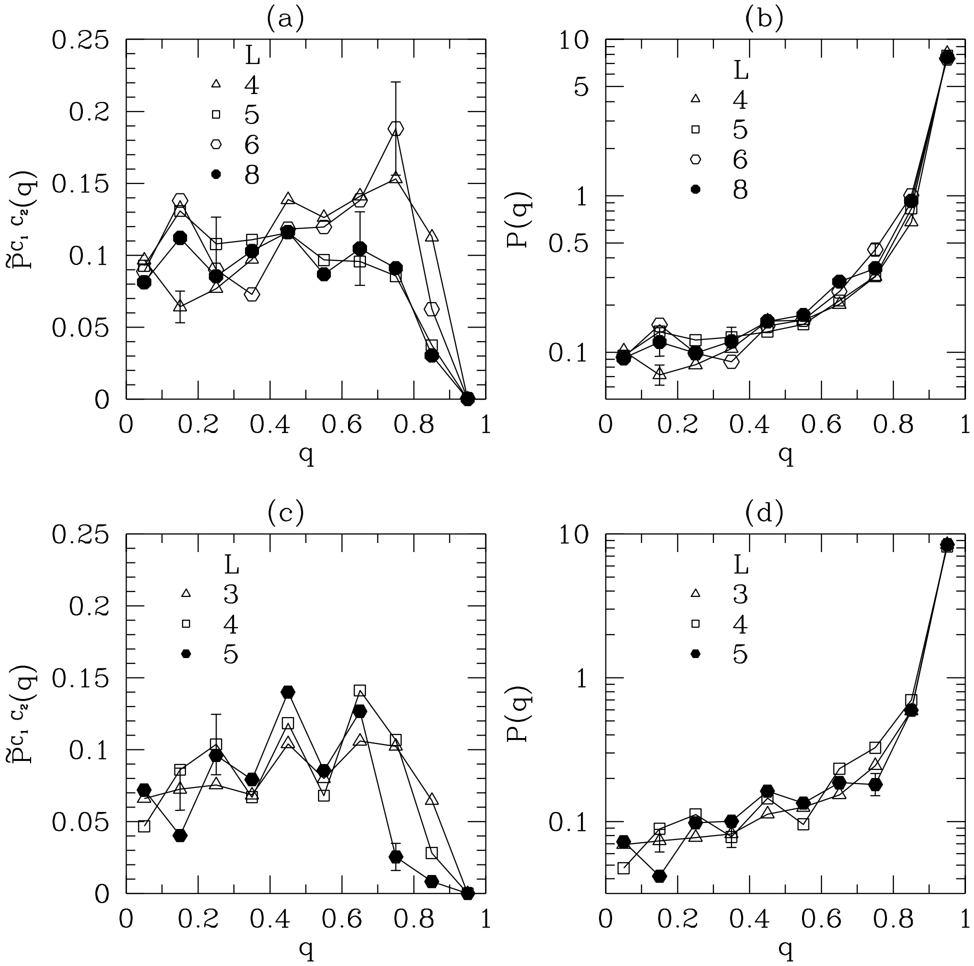

FIG. 15. **(a)** The partial distribution $\widetilde{P}^{C_1C_2}(q)$ for $D=3$ $L=4,5,6,8$. It is normalized so that $2\int_0^1 \widetilde{P}^{C_1C_2}(q)$ is its weight in the total $P(q)$. For clarity only a few representative error bars are shown. **(b)** The distribution $P(q)$ for the same systems as in (a). **(c)** $\widetilde{P}^{C_1C_2}(q)$ as in (a) but for $D=4$ $L=3,4,5$. **(d)** $P(q)$ for the same systems as in (c).

$$q_{\mu\nu} = 1 - 2|\mathcal{G}_{\mu\nu}|/N \; . \tag{37}$$

For nearly all state pairs $\mu,\nu \in \mathcal{C}$ the domain $\mathcal{G}_1$ is in the state $\Uparrow_1$; hence $|\mathcal{G}_{\mu\nu}| \le 1 - |\mathcal{G}_1|$, so that $q_{\mu\nu} \ge 2|\mathcal{G}_1|/N - 1$. The state pairs belong to one of two types:

1. Pairs in which $\mathcal{G}_2$ flips between $\Uparrow_2$ to $\Downarrow_2$ or vice versa. These pairs contribute to $P_J^{o,1}(q) = P_J^{\mathcal{C}_1\mathcal{C}_2}(q)$. The definition of $\mathcal{G}_2$ yields that in most such cases $\mathcal{G}_2 \in \mathcal{G}_{\mu\nu}$ and hence $2|\mathcal{G}_1|/N - 1 \le q_{\mu\nu} \le 1 - 2|\mathcal{G}_2|/N$.

2. Pairs in which neither $\mathcal{G}_1$ nor $\mathcal{G}_2$ flip contribute to $P_J^{i,1}(q) = P_J^{\mathcal{C}_1\mathcal{C}_1}(q) + P_J^{\mathcal{C}_2\mathcal{C}_2}(q)$. For these pairs in most cases $|\mathcal{G}_{\mu\nu}| \le N - |\mathcal{G}_1 \cup \mathcal{G}_2|$ and hence $q_{\mu\nu} \ge 2(|\mathcal{G}_1| + |\mathcal{G}_2|)/N - 1$.

The peak centered at q_1, is attributed to state pairs of the first type, and hence

$$2|\mathcal{G}_1|/N - 1 \le q_1 \le 1 - 2|\mathcal{G}_2|/N \tag{38}$$

The other peak, centered at q_2, is attributed to state pairs of the second type, and thus we expect

$$q_2 \ge 2(|\mathcal{G}_1| + |\mathcal{G}_2|)/N - 1 \tag{39}$$

These two inequalities yield $q_2 - q_1 \ge 2|\mathcal{G}_2|/N$. Evidently, this structure of $P_J(q)$ is completely consistent with our picture of spin domains that govern partition of state space into well defined clusters. By a detailed analysis [37] we have shown that the (at least) two-peaked structure of $P_J(q)$ survives for large L.

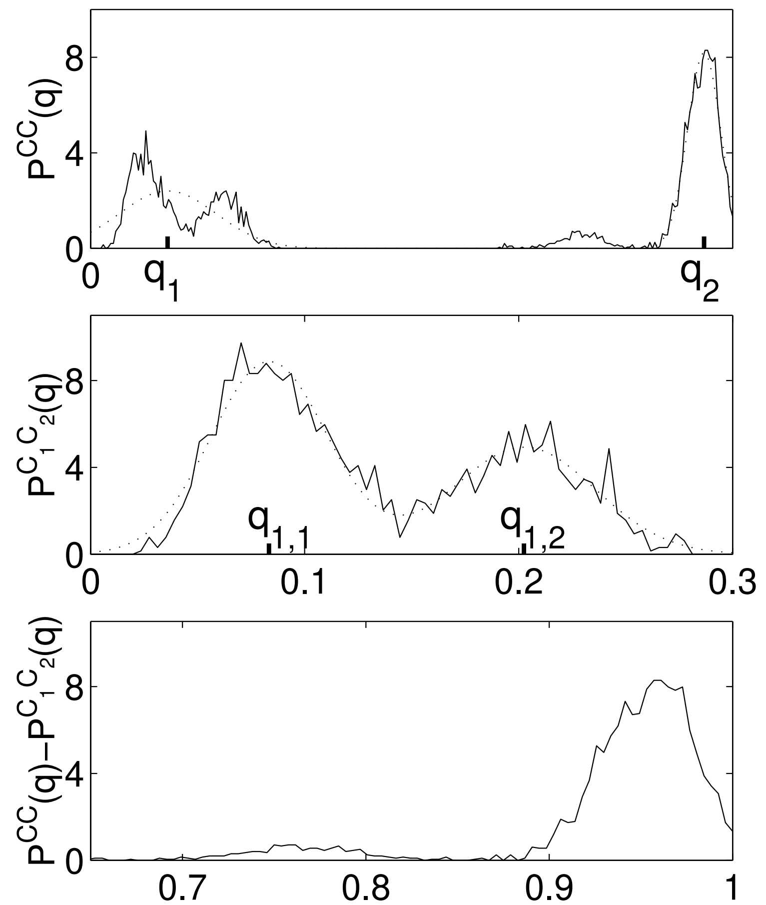

FIG. 16. (Top:) The distribution $P_J^{\mathcal{C}\mathcal{C}}(q)$ for the same $3D$ realization whose data were presented in Fig. 12. The dotted line is a fit to the sum of two Gaussians (see text). (Middle:) The partial distribution $P_J^{\mathcal{C}_1\mathcal{C}_2}(q)$ for the same realization. The dotted line represents a fit to the sum of two Gaussians. (Bottom:) The difference between the two previous distributions.

In some realizations, such as the ones that yield Figs. 14(a) and 14(d) $P_J(q)$ has more peaks, since $P_J^{\mathcal{C}_1\mathcal{C}_2}(q)(\equiv P_J^{o,2}(q))$ exhibits two or more peaks; this splitting is due, as mentioned above, to spin domains $\mathcal{G}_3$ and $\mathcal{G}_3'$. We analyzed $P_J^{\mathcal{C}_1\mathcal{C}_2}(q)$ in the same way as we did for $P_J^{\mathcal{C},\mathcal{C}}(q)$, using the same form of fit as in Eq. (36). For example, in the middle part of Fig. 16, $\tilde{q}_{1,1}$ and $\tilde{q}_{1,2}$ denote the centers of the two Gaussians, with $\mathcal{G}_2$ and $\mathcal{G}_3$ playing the previous roles of $\mathcal{G}_1$ and $\mathcal{G}_2$.

For much larger systems, for which the state hierarchy is expected to have more than two clear levels, we expect to find a finer structure in $P(q)$. It will exhibit multiple peaks, each related to different domain sizes. The heights and widths of the peaks are expected to be governed by the sizes of the state clusters that contribute to it which, in turn, are determined by the correlations between the spin domains that generating these clusters. Each of these peaks can be isolated and measured separately by observing the overlap of states of the corresponding clusters.

The shape of $P(q)$ we describe above resembles the one assumed by RSB. It is important to re-emphasize, however, that our $P(q)$ was obtained for finite systems; its resemblance to the form predicted by RSB does not necessarily mean that the latter picture is the correct one. In fact, previous studies [12,14,24,26] of the link overlap (defined in Eq. (2)) indicate that it is trivial, which contradicts the RSB scenario, though this conclusion has been disputed in Refs. [27–29]. In fact, our picture and results also do *not* appear to be consistent with RSB since we find a non-ultrametric state structure, as we show in Sec. VII.

VII. ULTRAMETRICITY

Ultrametricity is one of the main characteristics of the mean field RSB picture. Efforts to establish [46] or dismiss [47] the existence of ultrametricity in short range spin glasses did not yield conclusive results. We presented in Sec. IV indications that w_{12}, the width of the distance distribution between states from $\mathcal{C}_1$ and $\mathcal{C}_2$, does not vanish, implying a non-ultrametric structure of state space. Here we look for a more direct test of ultrametricity. The main problem is that we can equilibrate only small systems, where ultrametricity is hindered by finite size effects. Ultrametricity is a statement about the geometrical properties of *triangles* formed by three "pure states" (or by three micro states that belong to different pure states). All three have to belong [48] to $\mathcal{C}$, and for small systems only a small fraction of the realizations contain such triplets of states.

For $D = 3$ at $T = 0.2$ we measured $\tilde{p}$, the fraction of realizations for which $\mathcal{G}_3$ (or $\mathcal{G}_3'$) were large enough to induce two clearly separated peaks of $P_J^{\mathcal{C}_1,\mathcal{C}_2}(q)$ (see Sec. VI B). We found, for $L = 4, 5, 6, 8$ the values $\tilde{p} = 0.006, 0.026, 0.056, 0.090$, respectively. At $D = 4$ the similar fractions, at $T = 0.2$ and for $L = 3, 4, 5$ are $\tilde{p} = 0.02, 0.030, 0.080$. Note that for both $D = 3, 4$, $\tilde{p}$ increases with the size of the system.

Our method of analysis allows us to identify the realizations that do contain such triangles of states and use exclusively them to investigate whether ultrametricity does or does not hold. In this way we avoid many finite size effects that might obscure the results.

A set of objects with a distance measure D is ultrametric if any three objects α, β and γ form an isosceles triangle, with the base equal to or smaller than the two equal sides. This demand can be formulated as the requirement that the inequality

$$D_{\alpha\beta} \leq \max\{D_{\alpha\gamma}, D_{\beta\gamma}\} \ . \tag{40}$$

be satisfied for all three choices of the distance placed on its left side.

When the system is in the high T paramagnetic phase it will exhibit ultrametricity, since, as $L \to \infty$ the probability distribution of distances will be $P(D_{\mu\nu}) = \delta(D_{\mu\nu} - 1/2)$ and all triangles will be equilateral. Similar behavior occurs *inside* a specific valley at $T < T_c$, since for two states μ and ν inside the valley $P(D_{\mu\nu}) \to \delta(D_{\mu\nu} - (1 - q_{\rm EA})/2)$, where q_{EA} is the Edwards-Anderson order parameter.

The non-trivial result of RSB is that the valleys *themselves* are ultrametric. In order to investigate this claim, we have to focus on triplets of states, each chosen from a different valley. For large systems with many valleys this does not require special care, since almost all triplets of states will belong to three different valleys. For small systems, however, a large fraction of the possible triplets will have at least two states from the same valley. Such triplets should be disregarded.

Our way of analysis provides us with tools to examine ultrametricity for small systems. We utilize the state hierarchy obtained in Sec. IV to carefully choose triplets of states from different state clusters. We chose three clusters: $\mathcal{C}_2$, $\mathcal{C}_{1a}$ and $\mathcal{C}_{1b}$. The last two clusters are the "children" of $\mathcal{C}_1$ in the state dendrogram, i.e. $\mathcal{C}_1 = \mathcal{C}_{1a} \cup \mathcal{C}_{1b}$. According to our picture a triplet of states, one from each of these three clusters, belong to three different valleys, since we have to flip a correlated domain with a macroscopic number of spins in order to move from one cluster to another. To move from $\mathcal{C}_2$ to $\mathcal{C}_1$ we have to flip $\mathcal{G}_2$ from configuration $\Downarrow_2$ to configuration $\Uparrow_2$. Similarly, when moving from $\mathcal{C}_{1a}$ to $\mathcal{C}_{1b}$ we have to flip $\mathcal{G}_3$ from $\Uparrow_3$=$\Uparrow_3$ $(\Uparrow_1, \Uparrow_2)$ to $\Downarrow_3$=$\Downarrow_3$ $(\Uparrow_1, \Uparrow_2)$ (see Subsection V C). Due to the small sizes studied, in this paper we do not present any conclusive evidence that $\mathcal{G}_3$ is indeed macroscopic. However, if (in the $L \to \infty$ limit) it is not macroscopic, our method predicts that there are only four valleys (determined by $\mathcal{G}_1$ and $\mathcal{G}_2$) and hence the the RSB picture clearly does not hold.

In order to have a quantitative measure of ultrametricity we define an index K in the following manner. Let μ, ν and ρ be three states, so that $D_{\mu\nu} \geq D_{\mu\rho} \geq D_{\nu\rho}$. We define

$$K_{\mu\nu\rho} = \frac{D_{\mu\nu} - D_{\mu\rho}}{D_{\nu\rho}} \ . \tag{41}$$

The triangle inequality requires $D_{\nu\rho} \geq D_{\mu\nu} - D_{\mu\rho}$ so we have $0 \leq K_{\mu\nu\rho} \leq 1$. Ultrametricity demands $D_{\mu\nu} = D_{\mu\rho}$ so if there is ultrametricity we expect $P(K) \to \delta(K)$ as $L \to \infty$.

We measured $P(K_{\mu\nu\rho})$ for $\mu \in \mathcal{C}_2$, $\nu \in \mathcal{C}_{1a}$ and $\rho \in \mathcal{C}_{1b}$. We used our samples for $T = 0.2$; since as the temperature is lower and more distant from T_c, the state structure should be clearer and less blurred by finite size effects. We measured the distribution of K for each realization, and then obtained $P(K)$ by averaging over the disorder $\{J\}$. In all systems we found with high probability that $K_{\mu\nu\rho} = 1$ exactly (see Table V).This happens when $\mathcal{G}_{\mu\nu}$, the set of spins one has to flip when going from μ to ν, coincides precisely with $\mathcal{G}_{\mu\rho} \cup \mathcal{G}_{\nu\rho}$, the union of the two sets that are flipped when we go from ρ to μ and to ν. This is, however, clearly a finite size effect; as L increases the probability $P(K = 1)$ decreases dramatically. Therefore we do not include this part of the distribution in our estimation of $P(K)$. If this part of $P(K)$ broadens as L increases, its exclusion cannot be achieved by simply ignoring the triangles with $K = 1$. This, however, is clearly not the case: we present in Table V the probability $P(0.9 \leq K < 1)$, and show that its increase with L is much too small to compensate for the decrease in $P(K = 1)$.

In order to disregard this finite size effect we truncated $P(K = 1)$ from $P(K)$ and renormalized to get the distribution

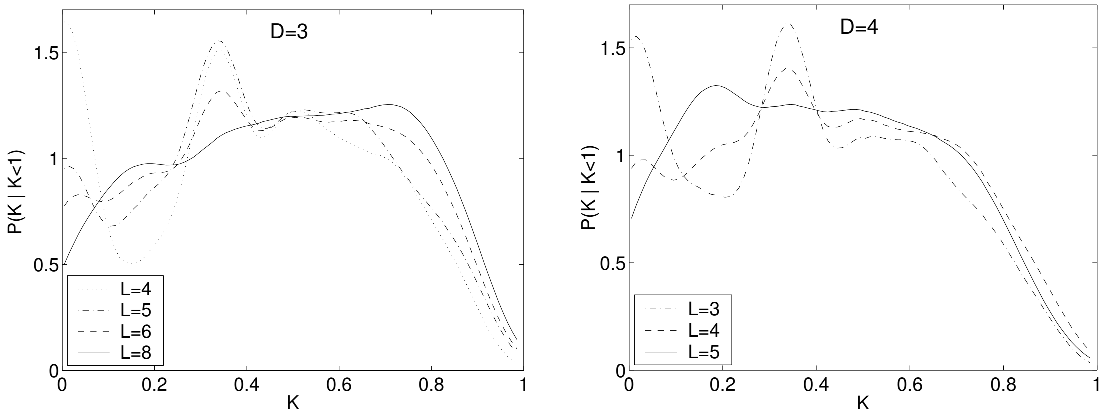

FIG. 17. The distribution $P(K|K<1)$ of $K_{\mu\nu\rho}$, for $\mu \in \mathcal{C}_2$, $\nu \in \mathcal{C}_{1a}$ and $\rho \in \mathcal{C}_{1b}$. All systems are sampled at $T = 0.2$.

$$P(K|K<1) = \begin{cases} P(K)\,/\,P(K<1) & K<1 \\ 0 & K=1 \end{cases} . \tag{42}$$

For large L we expect $P(K=1)$ to vanish, and $P(K)$ will approach $P(K|K<1)$. The results are plotted in Fig. 17. In Table V we give the mean and variance of $P(K|K<1)$. Though we deal with small systems, it seems that $P(K|K<1)$ converges to a distribution with non-vanishing mean and variance, indicating breakdown of ultrametricity for the three valleys studied.

Again one should address the question: do these results remain valid in the large L limit? We have to show that the state triplets we used, from $\mathcal{C}_2$, $\mathcal{C}_{1a}$ and $\mathcal{C}_{1b}$, have a finite statistical weight as $L \to \infty$. In Sec. VI we showed that $|\mathcal{C}_2|/N$ remains finite if the average correlation $\bar{c}_{12}$ between $\mathcal{G}_1$ and $\mathcal{G}_2$ does not approach one. From the same argument we conclude that if the correlation of $\mathcal{G}_3$ with $\mathcal{G}_1 \cup \mathcal{G}_2$ does not approach one then both $|\mathcal{C}_{1a}|/N$ and $|\mathcal{C}_{1b}|/N$ do not vanish and the weight of such state triplets remains finite, and the system does not exhibit ultrametricity. We do have evidence that the average correlation $\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2)$ of $\mathcal{G}_3$ with $\mathcal{G}_1 \cup \mathcal{G}_2$ in fact decreases as L increases, but it is not conclusive.

VIII. SUMMARY AND DISCUSSION

We have presented a new picture of the spin glass-phase in finite dimensional systems. This picture - State Hierarchy Induced by Correlated Spin domains (SHICS) - is consistent with numerical findings of a non-trivial overlap distribution [19,20,24] and macroscopic spin domains which cost only a *finite* energy to flip [12,14]. Our results differ from the conventional interpretations [16,17] of the droplet picture; nevertheless, the scenario presented in the original work of Fisher and Huse [6–8], and also the work of Newman and Stein [9,18], is of sufficient generality to allow consistency with our findings.

In the spin glass phase, the system consists of macroscopic spin domains of variable sizes. Each of these domains flips as a coherent entity, and the flipping costs only a finite free energy. The variability in size gives rise to a hierarchical structure in state space. At each level in the hierarchy some state clusters split; each such splitting is associated with a spin-domain. The first (highest) level splitting (to $\mathcal{C}, \bar{\mathcal{C}}$) is associated with the largest domain $\mathcal{G}_1$; at the next level the two observed splittings ($\mathcal{C} \to \mathcal{C}_1, \mathcal{C}_2$ and $\bar{\mathcal{C}} \to \bar{\mathcal{C}}_1, \bar{\mathcal{C}}_2$) are related by symmetry and hence governed by the same, second largest domain $\mathcal{G}_2$. At each level, the state clusters are labeled according to the orientation of the corresponding domains.

Below the second level, different spin domains are involved depending on which state cluster is being subdivided, e.g. $\mathcal{G}_3$ is the domain whose orientation splits the states in $\mathcal{C}_1$, while a different domain $\mathcal{G}_3'$ is involved in splitting $\mathcal{C}_2$. Although $\mathcal{G}_3 \neq \mathcal{G}_3'$, in general they may share some of their spins. The state space structure in the lower levels of the hierarchy has to be further investigated for larger systems. Specifically, one has to verify that $\mathcal{G}_3$ and $\mathcal{G}_3'$ do not vanish as $L \to \infty$.

Some details of our hierarchical picture do not appear to be consistent with RSB. According to the RSB scenario, the states have an ultrametric structure, which implies that for any two state clusters defined at a certain level of the hierarchy, e.g. $\mathcal{C}_1$ and $\mathcal{C}_2$, the distribution of overlaps q_{ij} between $i \in \mathcal{C}_1$ and $j \in \mathcal{C}_2$ should approach a delta-function for large L. We presented in Sec. IV indications that the width of the distribution $P(\tilde{D}_{ij})$, of values in $\tilde{D}$, may not vanish for $L \to \infty$, indicating absence of ultrametricity. We also presented direct evidence for lack of ultrametricity in Sec. VII. However, studies on larger sizes are needed to verify that the test which indicates lack of ultrametricity will still yield the same conclusion as $L \to \infty$.

In Sections VI and VII we demonstrated how, by separating the state space into its components, we can calculate various quantities using only a chosen part of this space, thus obtaining more reliable numerical results and reducing finite size effects.

Clustering analysis can be applied also to other systems with a non-trivial phase space structure, *i.e.* which have several valleys which are not related by any apparent symmetry, such as random field models, see *e.g.* the discussion in Ref. [7], or other models with random anisotropy [49]. It can help not only in the investigation of the macroscopic properties of a system, but also in understanding the micro-structure that give rise to its properties.

Acknowledgements The research of ED was partially supported by grants from the Germany-Israel Science Foundation (GIF); his stay at the ITP, UCSB was supported in part by the NSF under grant PHY99-07949. APY acknowledges support from the NSF through grant DMR 0086287. We thank A. Hartmann, I. Kanter, E. Marinari and M. Mézard for most helpful discussions.

---

[1] D. Sherrington and S. Kirkpatrick, Physical Review Letters **35**, 1792 (1975).
[2] G. Parisi, Physical Review Letters **43**, 1754 (1979).
[3] G. Parisi, J. Phys. A **13**, 1101 (1980).
[4] G. Parisi, Physical Review Letters **50**, 1946 (1983).
[5] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. A. Virasoro, J. Physique **45**, 843 (1984).
[6] D. S. Fisher and D. A. Huse, Physical Review B **38**, 386 (1988).
[7] D. A. Huse and D. S. Fisher, J. Phys A **20**, L997 (1987).
[8] D. S. Fisher and D. A. Huse, J. Phys A **20**, L1005 (1987).
[9] C. M. Newman and D. L. Stein, Physical Review E **57**, 1356 (1998).
[10] A. A. Middleton, Physical Review Letters **83**, 1672 (1999).
[11] M. Palassini and A. P. Young, Physical Review B **60**, R9919 (1999).
[12] F. Krzakala and O. C. Martin, Physical Review Letters **85**, 3013 (2000).
[13] J. Houdayer and O. C. Krzakala, F.and Martin, Eur. Phys. J. B **18**, 467 (2000).
[14] M. Palassini and A. P. Young, Physical Review Letters **85**, 3017 (2000).
[15] C. M. Newman and D. L. Stein, cond-mat/0103616 (unpublished).
[16] A. J. Bray and M. A. Moore, in *Heidelberg Colloquium on Glassy Dynamics and Optimization*, edited by L. Van Hemmen and I. Morgenstern (Springer-Verlag, Berlin, 1986), p. 121.
[17] M. A. Moore, H. Bokil, and B. Drossel, Physical Review Letters **81**, 4252 (1998).
[18] C. M. Newman and D. L. Stein, Physical Review E **63**, 16101 (2001).
[19] R. N. Bhatt and A. P. Young, Physical Review Letters **64**, 1859 (1990).
[20] E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, Physical Review B **58**, 852 (1998).
[21] K. Hukushima and K. Nemoto, Journal of the Physical Society of Japan **65**, 1604 (1996).
[22] E. Marinari, in *Advances in computer simulations*, edited by J. Kertész and I. Kondor (Springer-Verlag, Berlin, 1998), p. 50, (cond-mat/9612010).
[23] S. Franz and G. Parisi, cond-mat/0006188 (unpublished).
[24] H. G. Katzgraber, M. Palassini, and A. P. Young, cond-mat/0007113 (unpublished).
[25] C. M. Newman and D. L. Stein, cond-mat/0010033 (unpublished).
[26] M. Palassini and A. P. Young, Physical Review Letters **83**, 5126 (1999).
[27] E. Marinari and G. Parisi, Physical Review Letters **85**, 3332 (2000).
[28] E. Marinari and G. Parisi, Physical Review B **62**, 11677 (2000).
[29] E. Marinari and G. Parisi, cond-mat/0007493 (unpublished).
[30] G. Hed, A. K. Hartmann, D. Stauffer, and E. Domany, Physical Review Letters **86**, 3148 (2001).
[31] G. Hed, A. K. Hartmann, and E. Domany, cond-mat/0012451 (unpublished).
[32] M. Mézard and M. A. Virasoro, J. Physique **46**, 1293 (1985).
[33] E. Marinari, O. C. Martin, and F. Zuliani, cond-mat/0103534 (unpublished).
[34] G. Parisi, F. Ricci-Tersenghi, and J. J. Ruiz-Lorenzo, J. Phys. A **29**, 7943 (1996).
[35] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data* (Prentice–Hall, Englewood Cliffs, 1988).
[36] M. Blatt, S. Wiseman, and E. Domany, Neural Computation **9**, 1805 (1997).
[37] G. Hed, Master's thesis, Feinberg Graduate School of the Weizmann Institute of Science, January 2001.

[38] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, Proc. Natl. Acad. Sci. **96**, 6745 (1999), in fact we used a variant of this method, ensuring that the dendrogram is symmetric.
[39] Strictly, it is not an equilibrium canonical ensemble, since for each state in the ensemble we have also its exact inverse, but this does not affect our results.
[40] R. D. Reed and R. J. Marks, in *Neural Smithing* (The MIT Press, Cambridge, Massachusetts, 1999), p. 299.
[41] If say $\mathcal{G}_1$, the spin domain that flips, consists of two disjoint parts $\mathcal{G}_1^a$ and $\mathcal{G}_1^b$, one can flip each of these separately, one at a time. In such a case we identify the largest of the two sub-domains as our $\mathcal{G}_1$, whereas the smaller induces the finer level splitting in the state hierarchy.
[42] Except maybe a small part of $\mathcal{G}_a$, that may lose its relative orientation due to local thermal excitations.
[43] The leaves of the hierarchy discussed here are *not* the individual states $\mathbf{S}^\mu$ of the state dendrogram of Sec. IV.
[44] The largest *locked* correlated domain over the states of $\mathcal{C}_1$ includes $\mathcal{G}_1 \cup \mathcal{G}_2$.
[45] A. K. Hartmann, Eur. Phys. J. B **13**, 539 (2000).
[46] S. Franz and F. Ricci-Tersenghi, Physical Review E **61**, 1121 (2000).
[47] A. K. Hartmann, Europhysics Letters **44**, 249 (1998).
[48] Ultrametricity is trivially violated if we include in the same triangle states with negative and positive overlaps.
[49] R. Fisch, Physical Review B **51**, 11507 (1995).

D	L	$[\lvert\mathcal{G}_3\rvert]_J/N$	$[\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2)]_J$	$P(\mathcal{G}_3 \neq \emptyset)$	$[\lvert\mathcal{G}'_3\rvert]_J/N$	$P(\mathcal{G}'_3 \neq \emptyset)$	$[\lvert\mathcal{G}_3 \cap \mathcal{G}'_3\rvert/\lvert\mathcal{G}_3\rvert]_J$	$P(\mathcal{G}_3 \neq \emptyset$ and $\mathcal{G}'_3 \neq \emptyset)$
3	4	0.048±0.003	0.55±0.015	0.914(4)	0.087±0.008	0.834(6)	0.23±0.019	0.772(8)
	5	0.046±0.003	0.52±0.015	0.914(4)	0.085±0.009	0.882(5)	0.15±0.016	0.818(7)
	6	0.043±0.003	0.48±0.015	0.924(3)	0.081±0.009	0.896(4)	0.19±0.017	0.832(6)
	8	0.036±0.003	0.43±0.017	0.905(5)	0.076±0.010	0.905(5)	0.16±0.019	0.827(8)
4	3	0.045±0.003	0.56±0.015	0.928(3)	0.094±0.010	0.838(6)	0.25±0.020	0.782(8)
	4	0.037±0.003	0.48±0.015	0.908(4)	0.061±0.007	0.920(3)	0.16±0.016	0.844(6)
	5	0.034±0.005	0.43±0.024	0.84(1)	0.072±0.014	0.865(8)	0.19±0.027	0.73(1)

TABLE IV. The size of the spin domain $\mathcal{G}_3$ and $\mathcal{G}'_3$, the correlation of $\mathcal{G}_3$ with $\mathcal{G}_1 \cup \mathcal{G}_2$ and the relative part of $\mathcal{G}_3$ and of $\mathcal{G}'_3$, which is common to both these spin domains. All results are taken for realizations where the domains concerned do not vanish, and we give also the probability of this to happen. All data was taken for $T = 0.2$. We present the average over these realizations $\{J\}$ ± the statistical error, obtained by dividing the standard deviation by $\sqrt{N_s}$, where N_s is the number of realizations that contributed to each average.

D	L	$P(K=1)$	$P(0.9 \leq K < 1)$	mean(K)	var(K)
3	4	0.78	0.0007	0.385	0.073
	5	0.57	0.0082	0.426	0.066
	6	0.35	0.0126	0.447	0.068
	8	0.08	0.0269	0.476	0.066
4	3	0.74	0.0012	0.362	0.068
	4	0.38	0.0116	0.413	0.067
	5	0.10	0.0095	0.406	0.061

TABLE V. The third and fourth columns show the probability for $K_{\mu\nu\rho} = 1$ and $0.9 \leq K_{\mu\nu\rho} < 1$, for $\mu \in \mathcal{C}_2$, $\nu \in \mathcal{C}_{1a}$ and $\rho \in \mathcal{C}_{1b}$. The fifth and sixth columns give the mean and variance of the distribution of $P(K|K<1)$. All systems are sampled at $T = 0.2$.