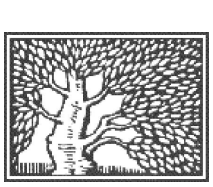

The nature of the spin-glass phase in finite dimension: Correlated spin domains generate hierarchical structure in state space

Guy Hed

Thesis for the M.Sc. degree
under the supervision of **Prof. Eytan Domany**

Submitted to the Feinberg Graduate School of
the Weizmann Institute of Science

January 2001

Acknowledgments

I would like to thank Eytan Domany for guidance, advise, and illuminating discussions. He taught me a great deal of physics, and contributed much to my understanding of science. His enthusiasm and enjoyment of research were inspiring and contagious.

The work presented here was done in collaboration with several people: Dietrich Stauffer was the one who came up with the idea to cluster the ground states of a binary spin-glass, and gave me the first motivation for this research; Alexander Hartmann has supplied me with the ground state sampless; A. Peter Young and Matteo Palassini generated the simulated tempering samples. I would like to thank them for their direct contribution to this research and for a most helpful correspondence.

To Gaddy Getz, Noam Shental and Erel Levine for interesting talks and ideas, and a good time.

To two people with whom I worked previous to this year: Ronnie Kosloff, who helped me to start my way in science; and Rob Bisseling, who gave me a chance to try a different field of research. Both contributed much to my experience, capabilities and my love of science.

To my mother, father, brother and sisters, for their love, support and encouragement.

To my wife, Dorit

Contents

1. Introduction

Whereas equilibrium properties of infinite range [1] spin glasses are completely understood within the framework of replica symmetry breaking (RSB) [2], spin glasses with short range interactions are the subject of considerable current debate and controversy [3–5]. Open questions address the nature of the low temperature phases [2, 6] and their theoretical description [3,6–9]. 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

$$\mathcal{H} = \sum_{\langle ij \rangle} J_{ij} S_i S_j \ , \tag{1.1}$$

where $\langle ij \rangle$ denotes nearest neighbor sites of a simple cubic lattice and the couplings, J_{ij}, are 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 variance $J = 1$.

The high temperature phase of the model is disordered, paramagnetic. As the temperature decreases below a critical temperature T_c, the system (in 3 or more dimensions) undergoes a transition into a frozen spin-glass phase. The main issue of controversy is the number of *pure states* in this phase. For an infinite system, a pure (or thermodynamic) state is defined as an ergodic subset of the phase space, i.e. a maximal subspace that the system can span (or visit) in a finite time. For a finite system the definition is less clear, but a pure state is usually referred to as a part of the phase space surrounded by free energy barriers, whose height diverges as the system size $L \to \infty$.

There are two main approaches to this problem; the droplet picture and RSB. According to the droplet picture of Fisher and Huse [6] locally, in a finite region of an infinite system, there are only two pure states. 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.

Within the droplet approach pure states must have a trivial overlap over any finite region, i.e. have a vanishing density of domain walls. Otherwise, there is a non-vanishing probability to have a domain wall in a finite region of the system, and thus to have more then two pure states in that region.

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

$q^{\rm link}_{\mu\nu}$ between two states $\mathbf{S}^\mu$ and $\mathbf{S}^\nu$ is defined by

$$q^{\rm link}_{\mu\nu} = \frac{1}{\gamma N} \sum_{\langle ij \rangle} S^\mu_i S^\mu_j S^\nu_i S^\nu_j \ , \tag{1.2}$$

where the sum is over pairs of neighbor sites and γ is the coordination number of the system. If the domain wall density vanishes, then the distribution $P(q^{\rm link})$ of the link overlap will be trivial: $P(q^{\rm link}) = \delta(q^{\rm link} - 1)$.

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

$$q_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} S^\mu_i S^\nu_i \ . \tag{1.3}$$

The distribution $P(q)$ of this parameter reflects the number of pure states for the N-spin system. If there are only two global pure states, the overlap distribution in the thermodynamic limit will be trivial: $P(q) = 0.5[\delta(q - q_{EA}) + \delta(q + q_{EA})]$, where q_{EA} is the average overlap inside a pure state. Hence a trivial $P(q)$ is consistent with the droplet picture; this, however, does *not* mean that finding a non-trivial $P(q)$ invalidates it! Huse and Fisher [10] clearly state that a non-trivial $P(q)$ *is* possible if there are domain walls with microscopic free energy cost, as long as the domain walls density is zero for an infinite system, so that the probability to find a domain wall at any finite region vanishes. This is the case, for example, for an Ising ferromagnet with anti-periodic boundary conditions. That is, an *infinite system* can have a trivial link overlap distribution and a non-trivial $P(q)$. On the other hand, if the link-overlap distribution $P(q^{\rm link})$ is trivial, then the overlap distribution $P(q)$, when observed *in a finite part* of an infinite system, *must* also be trivial. The reason is that a trivial $P(q^{\rm link})$ means vanishing density of domain walls, and hence the probability that a finite part will have a domain wall going through it also vanishes.

The statements made above concern infinite systems, for which measurements are made for either the entire system or on a finite part of it. Numerical work, however, is done on finite systems. In recent numerical work strong evidence for non-trivial $P(q)$ was found [7]. In the light of the statements made above, this does *not* disprove the droplet picture. On the contrary - the scenario referred to as TNT [11], of **T**rivial $P(q)$ and **N**on-**T**rivial $P(q^{\rm link})$, which was found to be most consistent with numerical data [11,12], does support the droplet picture. This scenario was also supported by recent analytic arguments [13].

Marinari et al. [7] have used parallel tempering [14,15] to sample 3D Ising spin glasses of sizes up to $L = 16$ and for temperatures down to $T = 0.7 \simeq 0.74 T_c$. They have found that $P(q)$ is non-trivial, and $P(0)$ does not vanish.

Krzakala and Martin [11] demonstrated the existence of macroscopic excitations with low energy cost in 3D Ising spin glasses of sizes up to $L = 11$. For a specific realization of $\{J\}$ they first identified the ground-state of the system $\mathbf{S}^0$. Then they added a constraint, forcing a randomly chosen pair of spins to change their relative orientation, i.e. $S^0_i S^0_j S_i S_j = -1$. A new ground state $\mathbf{S}^1$ was found under this constraint. In this new state a contiguous region

$\mathcal{G}_{01}$, which includes one of the spins i or j, was flipped (relative to $\mathbf{S}^0$). Having done this for three different pairs of spins per realization and more then 2000 different realizations, they found that in a finite fraction of the instances the size of the flipped domain was between $N/4$ to $N/2$, i.e. $q_{01} = [1-2|\mathcal{G}_{01}|/N]$ did not approach 1. On the other hand, the excitation energy $\mathcal{H}(\mathbf{S}^0) - \mathcal{H}(\mathbf{S}^1)$ - the surface energy of $\mathcal{G}_{01}$ - remained of order J. They also found that the dimension of the surface of $\mathcal{G}_{01}$ is smaller than the dimension of the system.

Katzgraber et al. [12] measured directly the distributions of $q^{\rm link}$ and q. They used parallel tempering [14, 15] to simulate and equilibrate different realizations $\{J\}$ of the bonds, for 3D systems of linear size $L \leq 8$ at temperature $T \geq 0.2$, and 4D systems with $L \leq 5$. Extrapolating their results they found that the variance of $P(q^{\rm link})$ vanishes as $L \to \infty$, and the distribution converges to $\delta(q^{\rm link}-1)$. This confirmed that the surface dimension of the spin clusters $\mathcal{G}_{01}$ found in [12] is indeed smaller than the system's dimension. They also found the the distribution $P(q)$ to be non-trivial, as in [7].

Newman and Stein [13] have recently proposed a spin-glass model with temperature dependent couplings J_{ij}. At $T_0 < T_c$ this model is identical to the EA model (1.1) with a coupling distribution with a finite width $J(T_0)$. As T decreases the couplings are changed so that $J(T)$ increases and yet the energy per spin remains constant and the system stays at the spin-glass phase. $J(T) \to \infty$ as $T \to 0$, so at $T = 0$ this model is is identical to the highly disordered model [16], whose set of ground states (in the infinite volume limit) is rigorously shown to have trivial $P(q^{\rm link})$ below eight dimensions. For any $0 < T \leq T_0$ the correspond to an EA model below T_c. As $T \to 0$ the Boltzmann weight of any state which is not a ground state of the disordered model will decrease arbitrarily. Under the assumptions that the (a) the properties of the EA model is independent of the specific coupling distribution; and (b) the number of pure states in the spin-glass phase is monotonically increasing with $1/T$; they deduce that the number of ground states of the disordered model is an upper bound on the number of pure states of the EA model at any finite T, so below eight dimensions there will be (at a finite region of an infinite system) only two pure states.

According to the alternative picture, the RSB framework can be applied also to short range systems [7]. Within the RSB solution, both $P(q)$ and $P(q^{\rm link})$ are non-trivial. 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* [2]: 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:

$$q_{\mu\nu} \geq \min(q_{\mu\rho},\, q_{\nu\rho}) \; . \tag{1.4}$$

Franz and Ricci-Tersenghi [9] found indirect evidence for a dynamic ultrametricity of 3D spin glasses, which implies static ultrametricity [17]. They forced (using a soft constraint) two replicas $\mathbf{S}^\mu$ and $\mathbf{S}^\nu$ (of the same system $\{J\}$) of an Ising spin glass of size $L = 20$

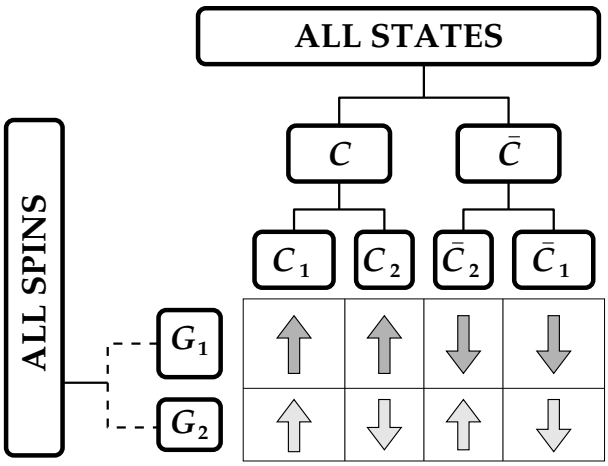

Figure 1.1: Schematic representation of our picture; the two largest spin domains and the first two levels in the hierarchical organization of the states are shown. The structure of the states is explained by the spin domains' orientations; e.g. in the states of the two sets $\mathcal{C}_1, \mathcal{C}_2$, the spins of $\mathcal{G}_1$ have the same orientation, whereas the spins of the smaller cluster, $\mathcal{G}_2$, have flipped.

to remain at a constant overlap q_0. Following the temporal evolution of the systems they found that the as the waiting time t_w increases the system came more close to fulfill the relation $q(\mathbf{S}^\mu(t_w)\mathbf{S}^\nu(t)) \geq \min(q(\mathbf{S}^\mu(t)\mathbf{S}^\nu(t)), q_0)$, for $t_w \leq 10^7$ and $t \leq 10^8$ MCS.

As far as we are aware, no published work gives direct evidence for a non-trivial $P(q^{\text{link}})$.

Very recently a new picture of the low temperature phase of short-range spin glasses has been proposed [18], on the basis of a numerical study of the ground states of the model (1.1) with $J_{ij} = \pm 1$ couplings. On the one hand, this picture is in agreement with the TNT scenario; on the other, a hierarchical tree-like structure of the pure states is proposed, just like in the RSB solution. This state hierarchy is, however, *non-ultrametric*; at each level of the hierarchy, the states split into two subsets, generated by flipping a macroscopic contiguous domain of strongly correlated spins [19]. As we go down in the state hierarchy, the size of the corresponding spin domain decreases. The highest levels of the state hierarchy, induced by the two largest spin domains, is schematically illustrated in Fig. 1.1. The spin domains have been identified as the cores of the excitations found by Krzakala and Martin. Once the state hierarchy has been identified and the assignment of states at each level completed, one could identify these spin domains or "cores" in a systematic precise manner. This, in turn, provides a statistically meaningful investigation of the way in which the overlap distribution $P(q)$ is governed and dictated by the underlying hierarchy of correlated spin domains.

In Chapter 2 we present the clustering methodology which we use in Chapter 3 to study the model (1.1) with Gaussian couplings at finite temperatures.

In Chapter 4 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 1 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 Chapter 5. Since the average correlation between domains does not increase with the system size, $P(q)$ will remain non-trivial as $L \to \infty$.

The nature of our picture yields non-ultrametric structure, as demonstrated in Chapter 6. We present a parameter for ultrametricity, and measure its distribution. The results suggest the system does *not* become ultrametric as L increases.

In Chapter 7 we apply our method to the model (1.1) with binary couplings, $J_{ij} = \pm 1$, at $T = 0$. We find there the same structure found for Gaussian spin glasses at finite T. We conclude that $P(q)$ is non-trivial, in contradiction with recent papers [20–22]. We show how to perform a correct extrapolation of $P(q)$, by isolating a non-trivial component of the distribution.

Finally, our method and findings are summarized in Chapter 8.

2. 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 [23] 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 some cases, as in the present work, there may be available some knowledge about the nature of the desired partition. For example, one may wish to have a preassigned number of clusters, with data assigned to them in such a way that minimizes the variance within each cluster[1]. If such information about the desired clusters is available, it makes sense to choose a clustering method accordingly, i.e. one that is designed to generate (or identify) clusters of the desired characteristics.

In the context of the present work we can think of our sample of M ground states as the objects to be clustered. Each object is represented by an $N-$component vector $\vec{S}^\mu = (S_1^\mu, S_2^\mu, ..., S_N^\mu)$, where $S_i^\mu = \pm 1$ is the value taken by spin i in ground 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 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

[1] Such is the case, for example, in image compression algorithms.

any data set. We tried two algorithms; a recently introduced one, SPC [24], which uses the physics of granular ferromagnets to identify clusters, and a graph-based algorithm proposed by Ward. In the present problem the ground state clusters are compact (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 (see Appendix A), in the sense it regard only nearest neighbour distances with a particular characteristic length scale. Since the distances inside pure states are much smaller than the distance between them, SPC disregards the latter. It identifies the pure states as different clusters, but misses the hierarchical structure.

Ward's algorithm [23] 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 N clusters, with a single data point in each. 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.

To implement the algorithm one must define an effective distance $\rho_{\alpha\beta}$, between any two clusters α and β. 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. 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 unchanged cluster, $\gamma \neq \alpha, \beta$, and the new α', according to the rule

$$\rho'_{\alpha'\gamma} = \frac{n_\alpha + n_\gamma}{n_\alpha + n_\beta + n_\gamma}\rho_{\alpha\gamma} + \frac{n_\beta + n_\gamma}{n_\alpha + n_\beta + n_\gamma}\rho_{\beta\gamma} - \frac{n_\gamma}{n_\alpha + n_\beta + n_\gamma}\rho_{\alpha\beta} \ , \tag{2.1}$$

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.

Ward's algorithm tries, in effect, to minimize the quantity

$$S = \sum_\alpha \sigma_\alpha \tag{2.2}$$

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

$$\sigma_\alpha = \sum_{i,j \in \alpha} {D_{ij}}^2. \tag{2.3}$$

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. 3.3(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 [25]. The small boxes at the nodes represent clusters. The vertical location of cluster α is its τ value, and is thus related to it's 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.

Like every clustering algorithm, Ward's also has various problems which have to be kept in mind. One such problem we encountered has to do with the manner in which it deals with a set of points C whose natural partition is into two clusters C_1 and C_2 with very different sizes. Let the typical distance between points inside C_1 is D_1 and the typical distance between points in C_1 and C_2 is D_2. If we have

$$|C_1| > (D_1/D_2)^2 |C_2| \tag{2.4}$$

then S will gain the lowest value for a partition of C into two sub clusters, namely $\tilde{C}_1$ and $\tilde{C}_2$, such that $\tilde{C}_1$ contains about half the points of C_1 and $\tilde{C}_2$ contains the other half *and* the points of C_2. Even though this is not the natural partition of C, it is preferred by Ward's algorithm. 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.

A 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. Therefore we have to verify that the results we obtain do correspond to a natural partition of the 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 to test our clustering.

First, compare the dendrograms obtained in two cases with genuine structure, shown in Fig. 3.3(a,c) with that of structureless data, Fig. 3.3 (e). In the first two dendrograms the relative τ values of the clusters are much higher then in the third dendrogram.

Second, compare the distance matrices obtained when the states (datapoints) are reordered according to the results of the clustering process. For the first two datasets, that have genuine structure, the reordered distance matrices (Fig. 3.3(b,d)) have clear dark submatrices along the diagonal; these indicate that distances *within* a cluster are significantly shorter than *between* clusters. On the other hand, when the states of the structureless dataset are reordered according the results of its clustering, the resulting distance matrix, Fig. 3.3(f), is homogenous and greyish, indicating that the distances within and between the "clusters" of Fig. 3.3(e) are similar.

Third, as will be described below, we measured directly the distribution of distances within clusters and between them. We did this for each realization seperately, and then we measured the distribution of the results over the disorder (see Chapter 3). The quantitative measures obtained this way are in full agreement with the first two, more qualitative indices of "naturalness" of our partitions. In the cases with real structure the distribution of distances within a cluster is narrow, concentrated about a low distance, whereas the

distibution of distances between points of two different clusters is also narrow, but about a considerably higher value, also shown in detail in Chapter 3). These tests and indicators clearly support our claim that we are able to use Ward's algorithm in a discriminating way, identifying beyond doubt the clear, natural partitions and those that are mere artifacts of our method.

3. State space structure

For a particular (randomly chosen) set of bonds $\{J\}$ of the system we generate, by Parallel Tempering [14, 15], a sample of $M/2$ states, which constitute an equlibrium ensemble at a temperature T. Next, we add to this ensemble the set of $M/2$ states obtained from the original set by spin reversal. Clearly the new ensemble of M states also corresponds to thermal equilibrium[1] 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 Chapter, treating the M states of our ensemble as the data points to be clustered.

To this end, 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 $N \times M$ 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, this matrix is presented in Fig 3.1(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.

The trivial way to visualize points that lie in a high dimensional space is one of projecting 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) [26]. One constructs the $N \times N$ covariance matrix of the M points; the eigenvectors of this matrix are the principal directions or components of the variation in the data, which are ordered according to the size of the corresponding eigenvalues, with the largest coming first.

[1] It does not an equilibrium cannonical ensemble, since for each state in the ensemble we have also its exact inverse, but that does not effect our results

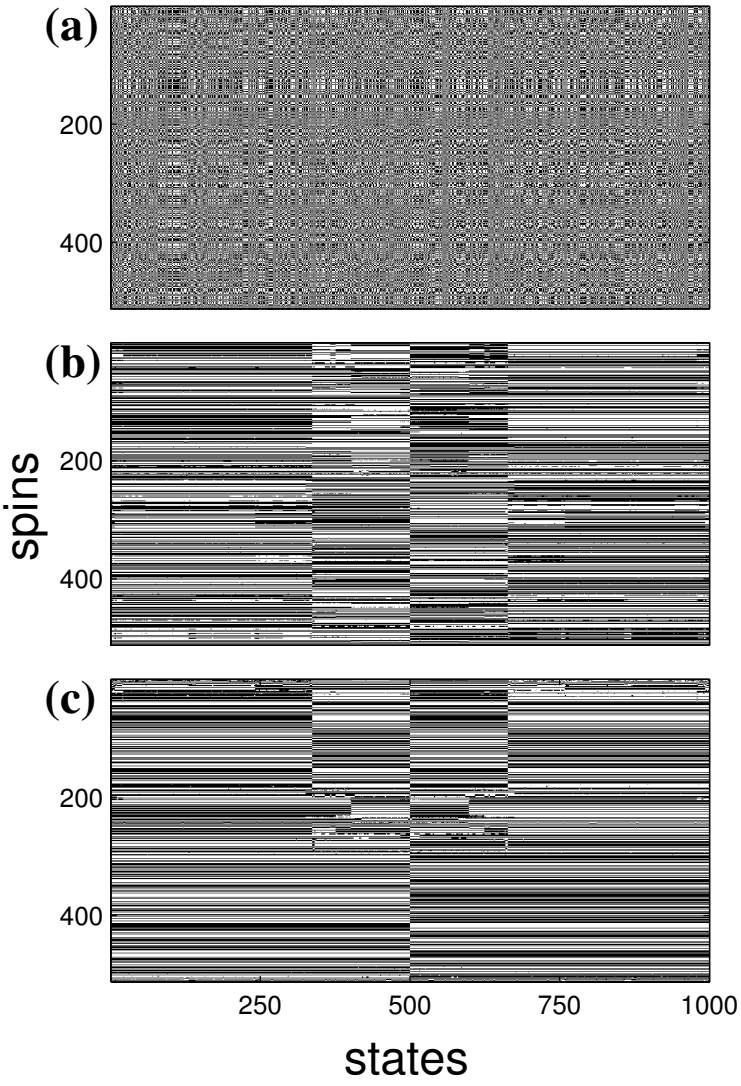

Figure 3.1: **(a)** The original data matrix of 500×2 states $\mathbf{S}^\mu$, $S_i^\mu = \pm 1$, with black/white representing $+/-$. This sample was generated for a realization of size 8^3 at $T = 0.2$ (same as for Fig. 3.3). The spins are in lexicographic order. **(b)** The same matrix, with the states ordered according to the dendrogram in Fig. 3.3. **(c)** The matrix in (b), with the spins ordered according to the dendrogram $\mathcal{D}$ in Fig. 4.7.

In Fig. 3.2 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.

To obtain a systematic quantitative measure of the hierarchical structure of state-
! space we perform cluster analysis of the M points. The choice of the particular clustering

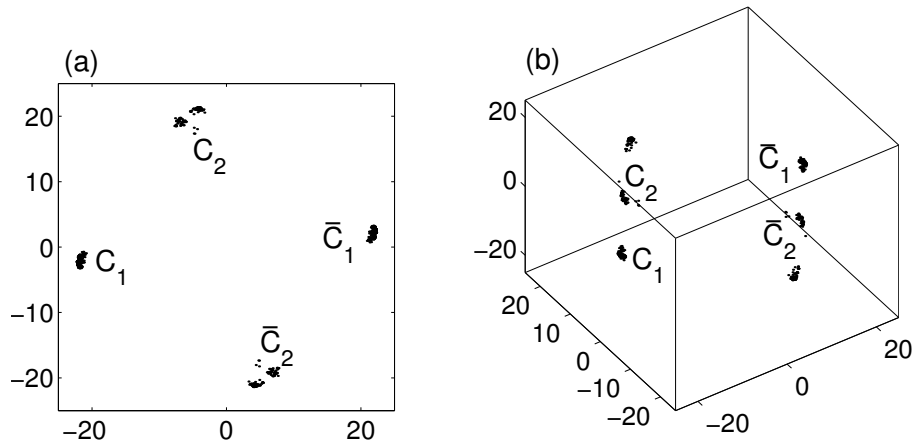

Figure 3.2: Principal Component Analysis of a sample of $M = 500 \times 2$ states of a specific realization of $\{J\}$ with $N = 8^3$ spins at $T = 0.2$. Each point represents a state $\mathbf{S}^\mu$. The three coordinates are the (a) two and (b) three largest principal components of the vactors in the sample. The first and second level partitions of the hierarchy are clearly visible.

algorithm used was dictated by our idea of the state space structure, obtained from PCA and from our picture (as described in the Introduction and summarized in Fig. 1.1).

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 Chapter 2, is tailored to preform 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

$$D_{\mu\nu} = \frac{1 - q_{\mu\nu}}{2} \,, \tag{3.1}$$

where $q_{\mu\nu}$ is the state overlap defined by Eq. 1.3.

Next, we clustered the spins using the distance matrix $D_{\mu\nu}$ as input to Ward's algorithm (see Eq. (2.1)). The algorithm results in a dendrogram, as shown in Figs. 3.3(a,c,e), for $T = 0.2, 0.5$ and 2.0, respectively (in three and four dimensions $T_c \approx 1.0$ [7, 27]). The leaves, which represent the states, are ordered on the horizontal axis according to the order imposed by the dendrogram [25]. 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 $T = 0.2, 0.5$ we found clear partitions in the two highest levels of the dendrogram, as presented in Figs. 3.3(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 set of states the cluster $\mathcal{C}_2$ breaks further into two sub-clusters, which are clearly seen in Fig. 3.2 as well.

We now address the issue mentioned in Chapter 2; namely, are the state clusters "natural", or an artifact of our clustering technique? As mentioned above, the first pitfall one

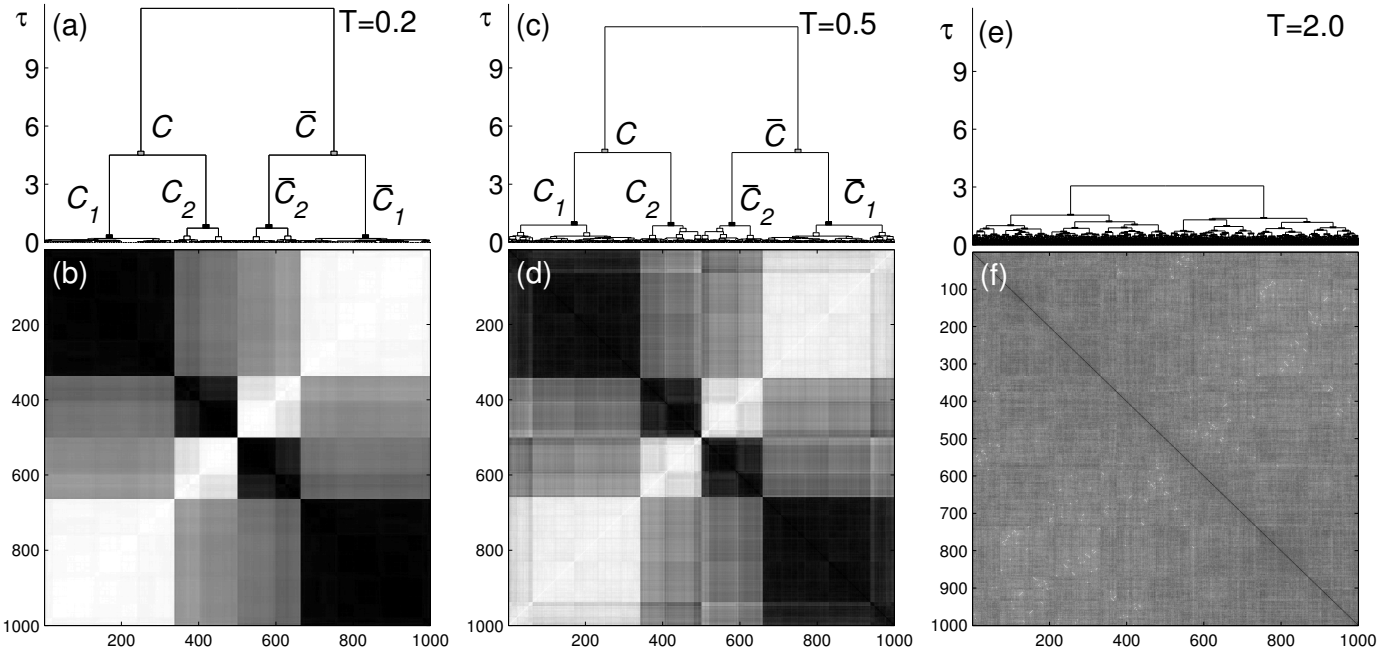

Figure 3.3: **(a)** The dendrogram obtained by clustering a sample of the $M = 500 \times 2$ states of a specific realization of size $N = 8^3$ at $T = 0.2$. The vertical axis describes the value of τ, defined in Chapter 2. **(b)** The ground states distance matrix used as an input to Ward's algorithm. 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 $T = 2.0$. 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.

should worry about is due to the fact that Ward's algorithm might miss the partition into two "natural" classes $\mathcal{C}_2, \mathcal{C}_1$ if their sizes satisfy the condition (2.4). As we show below, this is usually not the case. The other possible source of error is the fact (also discussed in Chapter 2) that Ward's algorithm will always generate a tree-like structure, even when there are no natural partitions in the data. This is precisely the situation above T_c, for example at $T = 2.0$, for which cluster analysis yields the dendrogram presented in Fig. 3.3(e).

We addressed this issue in three ways. First, note that direct observation of the dendrograms clearly differentiates between the two situations. At $T = 0.2, 0.5$ 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 - is relatively 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 $T = 2.0$, the relative τ values are much smaller than at $T = 0.2, 0.5$.

The genuinely hierarchical structure at $T = 0.2, 0.5$ is also evident from the states' distance matrix, as shown in Figs. 3.3(b,d). This distance matrix was obtained by reordering the states according to the results of the cluster analysis, i.e. according to the order of the leaves of the corresponding dendrogram. When the states are randomly ordered (like in Fig. 3.1(a)), the resulting distance matrix is a homogenous greyish square, like that of Fig. 3.3(f). The difference between this and Figs. 3.3(b,d) is striking: the distance matrices *within* clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ appear as dark (representing shorter distances) squares along the diagonal. When a cluster has internal structure, its distance matrix also has darker squares along the diagonal. The distances between clusters are represented by fairly uniform, lighter colored rectangles. In comparison, for $T = 2.0$, when there is no real hierarchical organization of the states, reordering them according to the dendrogram does not generate any ordered appearance of the distance matrix.

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. 3.1(b) the same data matrix as shown in Fig. 3.1(a), but with the states again reordered according to the dendrogram of Fig. 3.3 (a). That is, to get Fig. 3.1 (b), the colums of Fig. 3.1(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 C_1 from the smaller one, $\mathcal{C}_2$.

To obtain a quantitative assessment of the extent to which our state clusters are real, 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

$$D(\mathcal{C}) = \tfrac{1}{|\mathcal{C}|^2} \textstyle\sum_{\mu,\nu\in\mathcal{C}} D_{\mu\nu} \; ; \tag{3.2}$$

$$w(\mathcal{C}) = \left(\tfrac{1}{|\mathcal{C}|^2} \textstyle\sum_{\mu,\nu\in\mathcal{C}} {D_{\mu\nu}}^2 - D(\mathcal{C})^2 \right)^{1/2} \; . \tag{3.3}$$

The average $D(\mathcal{C}_\alpha)$ and the width $w(\mathcal{C}_\alpha)$ for $\alpha = 1, 2$ are 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

$$D(\mathcal{C}_1, \mathcal{C}_2) = \tfrac{1}{|\mathcal{C}_1||\mathcal{C}_2|} \textstyle\sum_{\mu\in\mathcal{C}_1} \sum_{\nu\in\mathcal{C}_2} D_{\mu\nu} \; ; \tag{3.4}$$

$$w(\mathcal{C}_1, \mathcal{C}_2) = \left(\tfrac{1}{|\mathcal{C}_1||\mathcal{C}_2|} \textstyle\sum_{\mu\in\mathcal{C}_1} \sum_{\nu\in\mathcal{C}_2} {D_{\mu\nu}}^2 - D(\mathcal{C}_1, \mathcal{C}_2)^2 \right)^{1/2} \; . \tag{3.5}$$

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 $w(\mathcal{C}, \bar{\mathcal{C}}) = w(\mathcal{C})$.

The results, presented in Table 3.1, confirm our claims that our state clusters are natural. We present for each variable x its mean $[x]_J$ (averaged over the disorder $\{J\}$) and standard deviation $([x^2]_J - [x]_J{}^2)^{1/2}$. For $T < T_c$ 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 of three fold 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 two-fold 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.

In the RSB [2] framework the overlap between any pair of pure states from two different clusters in 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 $\mathcal{C}_2$. Each of them contains states that belong to different pure states. If the overlap between pure states of the two clusters is constant, this should hold also for the overlap between each pair of states $i \in \mathcal{C}_1$ and $j \in \mathcal{C}_2$, since the width of the overlap distribution inside a pure state approaches zero. In this case the sub-matrix $\tilde{D}_{ij}$ for $i \in \mathcal{C}_1$ and $j \in \mathcal{C}_2$ is uniform, and the width $w(\mathcal{C}_1, \mathcal{C}_2)$ of the distribution $P(\tilde{D}_{ij})$ should vanish as $L \to \infty$. We used the data for $T = 0.2,\ D = 3$ to perform a fit of the form

$$w(\mathcal{C}_1, \mathcal{C}_2, L) = w_\infty + BL^y \ , \tag{3.6}$$

with B and y as fit parameters. The minimum of $\chi^2 = 1.9 \times 10^{-8}$ was found for $w_\infty = 0.021$ (with $B = 0.58(10),\ y = 3.37(12)$). Setting $w_\infty = 0$ we get $\chi^2 = 1.3 \times 10^{-6}$ with $B = 0.039(4),\ y = 0.31(7)$). Since for $w_\infty = 0$ we get a much higher value for χ^2, our data supports a non-vanishing value of w_∞.

D	T	L	$D(\mathcal{C})$	$w(\mathcal{C})$	$D(\mathcal{C}_1)$	$w(\mathcal{C}_1)$
3	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	4	0.118±0.074	0.095±0.052	0.067±0.044	0.052±0.030
		5	0.130±0.074	0.094±0.051	0.078±0.044	0.050±0.028
		6	0.130±0.068	0.088±0.048	0.084±0.041	0.049±0.028
		8	0.139±0.065	0.084±0.046	0.093±0.038	0.045±0.026
	2.0	4	0.435±0.030	0.114±0.009	0.394±0.039	0.103±0.007
		5	0.458±0.020	0.090±0.006	0.430±0.025	0.085±0.005
		6	0.473±0.011	0.074±0.004	0.453±0.018	0.072±0.003
		8	0.487±0.006	0.053±0.002	0.477±0.009	0.055±0.002
4	0.2	3	0.036±0.046	0.047±0.048	0.012±0.018	0.016±0.020
		4	0.042±0.047	0.046±0.050	0.016±0.016	0.015±0.017
		5	0.042±0.045	0.044±0.051	0.017±0.013	0.013±0.014
	0.8	3	0.162±0.074	0.105±0.046	0.105±0.049	0.063±0.028
		4	0.161±0.065	0.089±0.042	0.116±0.042	0.052±0.025
		5	0.167±0.065	0.081±0.042	0.124±0.042	0.045±0.024
	2.6	3	0.456±0.019	0.099±0.007	0.426±0.029	0.093±0.005
		4	0.480±0.009	0.066±0.003	0.465±0.014	0.065±0.003
		5	0.490±0.004	0.048±0.002	0.483±0.006	0.049±0.002
			$D(\mathcal{C}_2)$	$w(\mathcal{C}_2)$	$D(\mathcal{C}_1,\mathcal{C}_2)$	$w(\mathcal{C}_1,\mathcal{C}_2)$
3	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	4	0.089±0.059	0.066±0.042	0.248±0.135	0.064±0.033
		5	0.096±0.056	0.064±0.038	0.262±0.133	0.061±0.030
		6	0.100±0.053	0.061±0.038	0.250±0.131	0.056±0.027
		8	0.112±0.057	0.057±0.037	0.253±0.126	0.053±0.027
	2.0	4	0.400±0.032	0.103±0.007	0.484±0.027	0.104±0.008
		5	0.429±0.022	0.086±0.005	0.492±0.017	0.083±0.006
		6	0.449±0.016	0.073±0.004	0.496±0.011	0.068±0.004
		8	0.472±0.009	0.057±0.002	0.499±0.005	0.048±0.003
4	0.2	3	0.016±0.023	0.018±0.025	0.148±0.135	0.018±0.019
		4	0.020±0.026	0.020±0.029	0.134±0.131	0.017±0.016
		5	0.024±0.030	0.021±0.031	0.131±0.135	0.014±0.014
	0.8	3	0.128±0.061	0.076±0.036	0.282±0.123	0.074±0.030
		4	0.138±0.057	0.068±0.035	0.279±0.121	0.061±0.025
		5	0.143±0.053	0.061±0.035	0.277±0.120	0.049±0.022
	2.6	3	0.425±0.025	0.093±0.006	0.492±0.017	0.091±0.006
		4	0.461±0.014	0.067±0.003	0.498±0.008	0.061±0.003
		5	0.477±0.008	0.053±0.003	0.500±0.004	0.043±0.002

Table 3.1: The average distances within and between state clusters, and the relations between them. The numbers are the averages over all realizations ± the standard deviation, i.e. $[x]_j \pm ([x^2]_J - [x]_j^{\,2})^{1/2}$.

4. Correlated domains in spin space

4.1. 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 spin domain $\mathcal{G}_a$. The size and shape of this domain determines the free energy barrier separating two state clusters that were "born" at this level. We describe here 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. These domains are the cores of sets of macroscopic contiguous spin clusters with microscopic surface tension, that are discussed in Chapter 5.

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

$$c_{ij} = \langle S_i S_j \rangle = \frac{1}{\mathcal{Z}} \sum_{\mathbf{S}} S_i S_j \exp[-\mathcal{H}(\mathbf{S})] \ , \tag{4.1}$$

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

$$c_{ij} = \frac{1}{M} \sum_{\mu} S_i^\mu S_j^\mu \ . \tag{4.2}$$

The correlation in itself is irrelevant for spin glasses. It is gauge dependent and its average $[c_{ij}]_J$ over all the realizations of the disorder $\{J\}$ vanishes. The quantity ${c_{ij}}^2$ is the relevant measure of correlations in a spin glass. 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 in all states or always have opposite signs.

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 identified $\mathcal{G}_{\mu\nu}$, the set of spins whose sign is different in μ and ν,

$$\mathcal{G}_{\mu\nu} = \{ \ i \mid S_i^\mu \neq S_i^\nu \ \} \ . \tag{4.3}$$

Ideally all the spins of $\mathcal{G}_1$ always flip together and maintain their relative orientation. At finite T, however, 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 $\nu \in \bar{\mathcal{C}}$:

$$\tilde{\mathcal{G}}_1(\theta) = \left\{ i \;\middle|\; \frac{1}{|\mathcal{C}||\bar{\mathcal{C}}|} \sum_{\mu\in\mathcal{C}} \sum_{\nu\in\bar{\mathcal{C}}} S_i^\mu S_i^\nu < 1 - 2\theta \right\} . \tag{4.4}$$

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 $\{J\}$ the sites of $\tilde{\mathcal{G}}_1(\theta)$ are contiguous and hence it is identical to $\mathcal{G}_1(\theta)$, as seen in Tab. 4.1. 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. Two spins $i, j \in \mathcal{G}_1(\theta)$ must flip together at least for a fraction $2\theta - 1$ of the state pairs $\mu \in \mathcal{C}$ and $\nu \in \bar{\mathcal{C}}$. Thus for any two spins $i, j \in \mathcal{G}_1$

$$\begin{aligned} {c_{ij}}^2 &= \frac{1}{M^2} \sum_{\mu,\nu} S_i^\mu S_j^\mu S_i^\nu S_j^\nu \\ &= \frac{1}{|\mathcal{C}||\bar{\mathcal{C}}|} \sum_{\mu\in\mathcal{C}} \sum_{\nu\in\bar{\mathcal{C}}} S_i^\mu S_j^\mu S_i^\nu S_j^\nu > 4\theta - 3 . \end{aligned} \tag{4.5}$$

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. 4.1, the sizes of the domains and their average correlation, defined below in (4.6), 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. 4.2 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. *First* - if 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. These distributions were obtained from 500 (and 335 for $L = 8$) bond realizations in $D = 3$ (see Fig. 4.4(a)) and from 500 (and 200 for $L = 5$) realizations for $D = 4$ (Fig. 4.4(b)), at two temperatures in both dimensions. For $T = 0.2$, in both dimensions the distributions

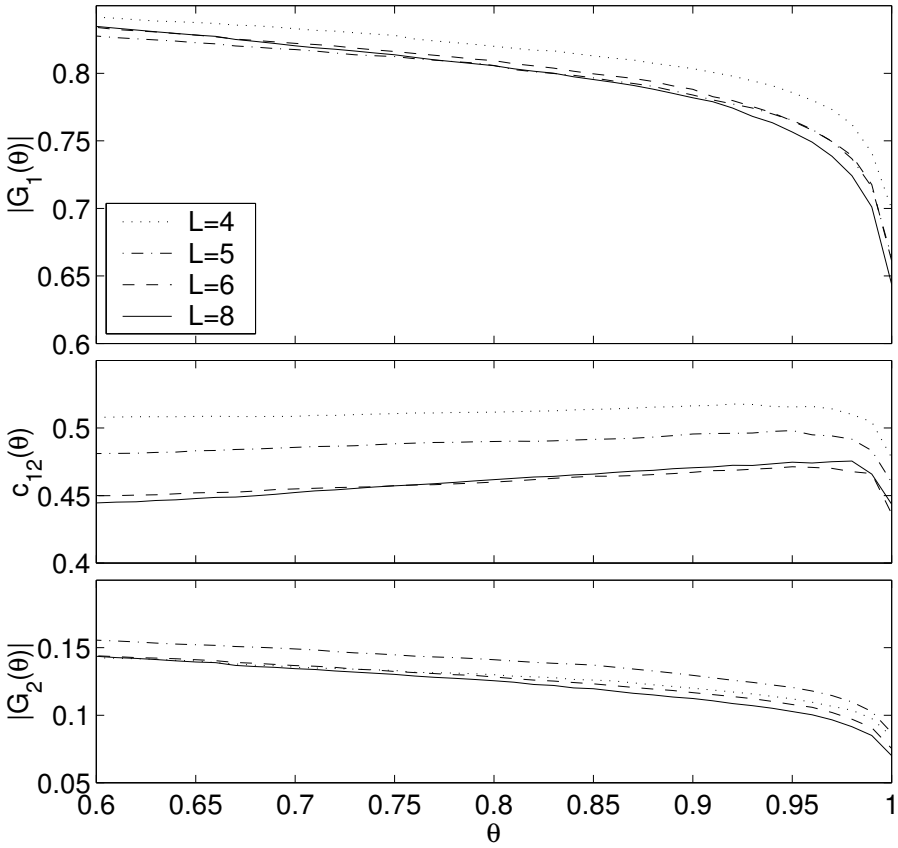

Figure 4.1: The sizes of $\mathcal{G}_1(\theta)$ and $\mathcal{G}_2(\theta)$ and their correlation $\bar{c}_{12}$ as a function of the threshold θ for $D = 3$, $T = 0.2$.

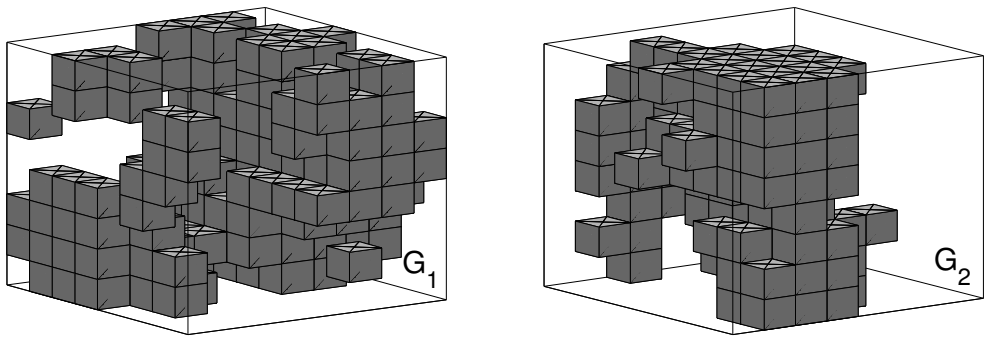

Figure 4.2: The spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$, as found in the realization of Fig. 3.1. Note that we use periodic boundary conditions, so the domains are connected through the boundaries.

seem to converge even for the small system sizes we use. We conclude with high certainty that at $T = 0.2$ the domain sizes $|\mathcal{G}_a| \propto L^D$ for both $a = 1, 2$. The mean and width of

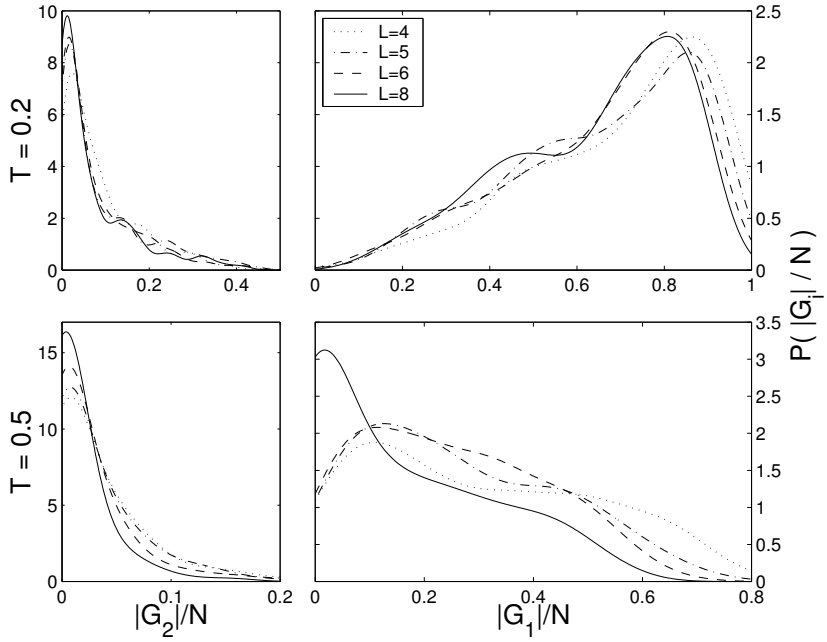

Figure 4.3: Size distributions of the spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$ for $D = 3$ dimensions at $T = 0.2, 0.5$. For $T = 0.2$ the distributions seem to converge, despite the small system sizes.

these distribution are presented in Tab. 4.2. On the other hand, for $T = 0.5$ in $D = 3$ and $T = 0.8$ in $D = 4$ we cannot determine conclusively whether the domain sizes do or do not remain proportional to $N = L^D$ as L increases.

There is, however, a *second* mechanism by which our picture can become invalid. Even if $\mathcal{G}_1$ and $\mathcal{G}_2$ do remain macroscopic for large L, we must show that they do not merge as $L \to \infty$. If they do, 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 $\mathcal{G}_2$,

$$\bar{c}_{12} = \frac{1}{|\mathcal{G}_1||\mathcal{G}_2|} \sum_{i \in \mathcal{G}_1} \sum_{j \in \mathcal{G}_2} {c_{ij}}^2 \ . \tag{4.6}$$

If $\bar{c}_{12}$ approaches the value 1 as $L \to \infty$, the two domains indeed merge in the thermodynamic limit.

In Table 4.2 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 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, the correlation for $L = 4, 5$ is higher at $T = 0.8$ than at $T = 0.2$. The

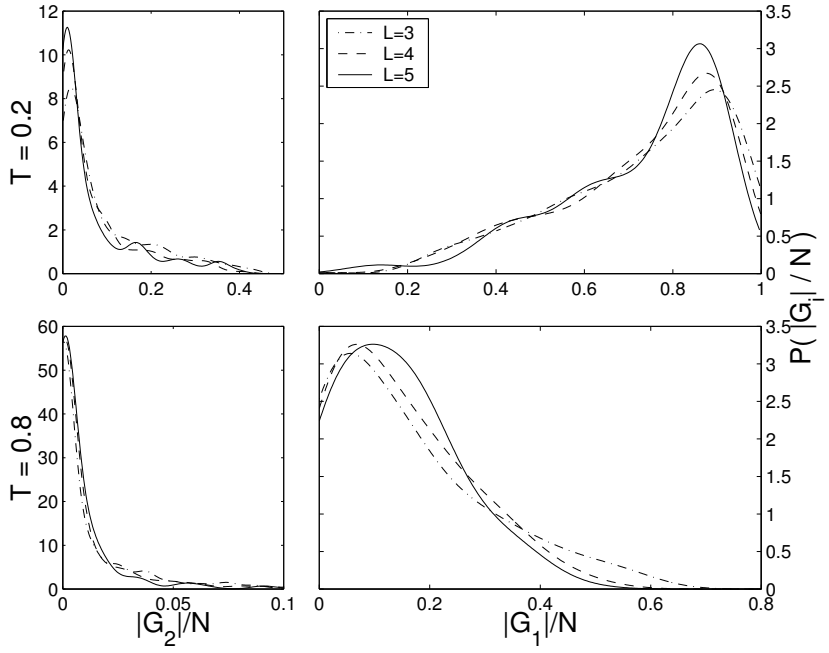

Figure 4.4: Size distributions of the spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$ for $D = 4$ dimensions at $T = 0.2, 0.8$. For $T = 0.2$ the distributions seem to converge, despite the small system sizes. For $T = 0.8$, $|\mathcal{G}_2|/N$ converges to a narrow distribution around zero, and $|\mathcal{G}_1|/N$ does not show convergence yet.

reason for this is probably that as T increases, small pieces of $\mathcal{G}_1$ "fall of" at its edges. Since $\mathcal{G}_2$ at $T = 0.2$ is small, one of these pieces, 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 attempted to identify also $\mathcal{G}_3$ 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 $\mathcal{C}_2$. We expect $|\mathcal{G}_3| \leq |\mathcal{G}_3'|$ since by our notation $|\mathcal{C}_1| \geq |\mathcal{C}_2|$, so in general there are more unlocked spins in $\mathcal{C}_1$ then in $\mathcal{C}_2$. 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 Tab. 4.3. 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

$$\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2) = \frac{1}{|\mathcal{G}_1 \cup \mathcal{G}_2||\mathcal{G}_3|} \sum_{i \in \mathcal{G}_1 \cup \mathcal{G}_2} \sum_{j \in \mathcal{G}_3} {c_{ij}}^2 \,. \tag{4.7}$$

In Table 4.3 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 cluster at large L, we expect it to be well separated from the union of the two larger domains.

4.2. 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 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

$$[\Uparrow_1]_i = \text{sign}\left(\sum_{\mu \in \mathcal{C}} S_i^\mu\right) \ . \tag{4.8}$$

Our definition of $\mathcal{G}_1$ guarantees that the argument of the sign function in the above expression does not vanish. By stating that $\mathcal{G}_1$ takes configuration $\Uparrow_1$ in a certain state μ we mean that

$$\sum_{i \in \mathcal{G}_1} S_i^\mu [\Uparrow_1]_i > 0 \ . \tag{4.9}$$

Hence the configuration assumed by $\mathcal{G}_1$ in any state μ determines the assignment of μ to either $\mathcal{C}$ or $\bar{\mathcal{C}}$.

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 $\Uparrow_2$ over $\mathcal{C}_1$ and $\bar{\mathcal{C}}_2$, and in $\Downarrow_2$ over $\mathcal{C}_2$ and $\bar{\mathcal{C}}_1$ (see Fig. 1.1 for a schematic illustration of this point).

Each spin domain $\mathcal{G}_a$ defines a partition of the states, at level a, into two sets - one in which $\mathcal{G}_a$ is in the $\Uparrow_a$ configuration and the other with $\Downarrow_a$. The distance $D_{\mu\nu}$, defined by (3.1), between two states μ and ν, one of each set, is almost always larger than $|\mathcal{G}_a|/N$, as we will show in Chapter 5. In order to move from one set to the other we have to flip most of the spins of $\mathcal{G}_a$. Thus, if $\mathcal{G}_a$ is macroscopic (as we have seen for $a = 1, 2$) it can be associated with a macroscopic free energy barrier.

The clear hierarchical organization of the state clusters suggest that the average distance (3.4) between state clusters formed at a high level of the hierarchy is larger then the average distance between clusters formed at a lower level. Indeed, we show in Table 3.1 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. 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*[1] 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 4.5. We see that $\mathcal{G}_3$ and $\mathcal{G}'_3$ share in general less than fifth of their spins.

An additional insight is obtained from a PCA of the spins. Each spin i, $i = 1 \ldots N$, is represented by a vector $\mathbf{S}_i = (S_i^1, S_i^2, \ldots, S_i^M)$ of its values over the M states in the sample. By PCA we project these N vectors onto the two dimensional surface, which contains, in a sense, maximum information on the data. In Fig. 4.5 we present the results of a PCA analysis for three realizations, labeled A, B and C. The data matrix for realization A, with the states and spins ordered according to the clustering results[2] is presented in Fig. 3.1(c) and in Fig. 4.6, where we also marked the different state clusters and spin domains.

In the upper left frame of Fig. 4.5 we see the results of the PCA analysis of the spins, whose vectors are the rows of the matrix in Fig. 4.6. We want highly correlated spins to be close on the plot. Since a spin $\mathbf{S}_i$ is fully correlated with its inverse $-\mathbf{S}_i$ we project each point (x, y) on the plot with $y < 0$ 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. Therefor 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. 4.6 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. 4.5 we used only the states of $\mathcal{C}_1$ in the analysis. We can see in Fig. 4.6 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. 4.5) 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. 4.5. In the matrix of Fig. 4.6 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. 4.5 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

[1] The largest *locked* correlated domain over the states of $\mathcal{C}_1$ includes $\mathcal{G}_1 \cup \mathcal{G}_2$.

[2] for clustering of the spins see Section 4.3.

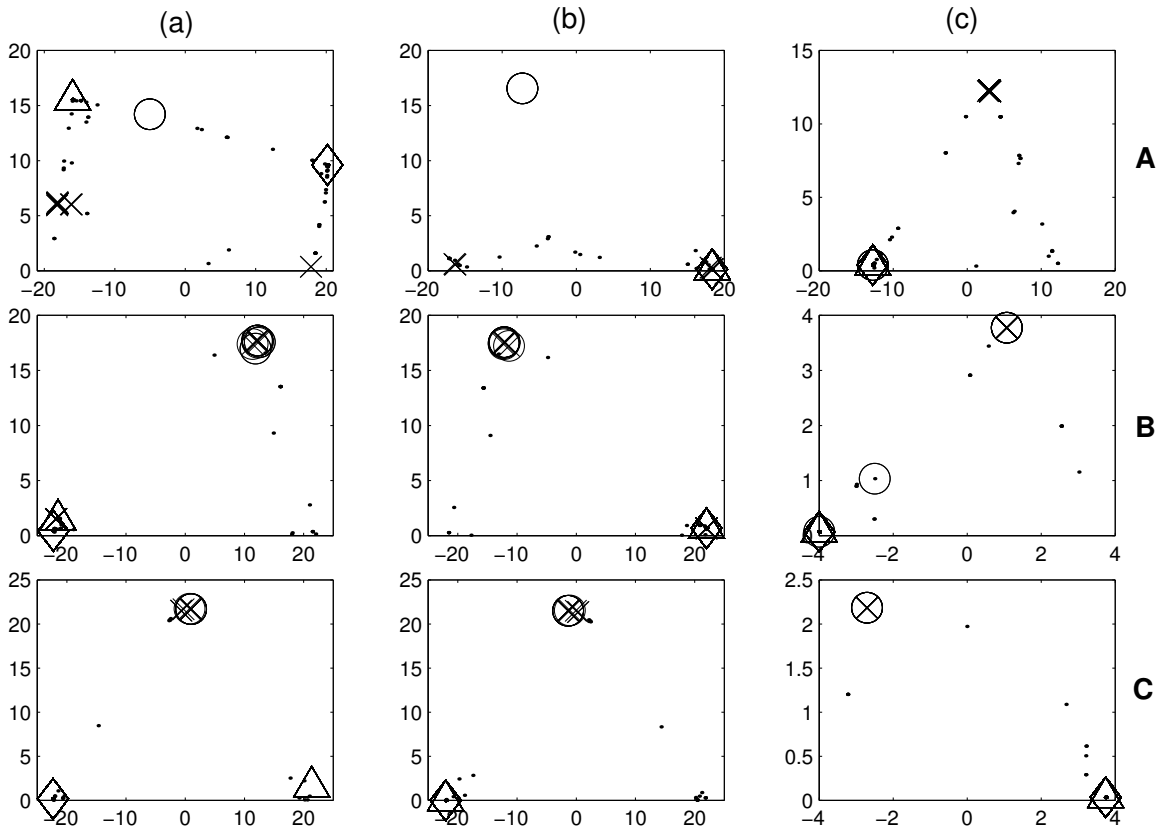

Figure 4.5: The two principal components of the 512 spins of three realizations A, B and C (see text). The analysis is carried over (a) all states; (b) the states of $\mathcal{C}_1$; and (c) the states of $\mathcal{C}_2$. The spins of $\mathcal{G}_1$ are marked by $\Diamond$; of $\mathcal{G}_2$ by $\triangle$; 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 $\bigotimes$. Spins that do not belong to any of these domains are marked with dots. The lower half of the plain is projected onto the upper half using $(x, y) \to (-x, -y)$. Spins in a correlated domain usually has the same values for the two principal components, and they fall on top of each other on the plot. Therefor in most plots a correlated domain is seemed to be represented by a single marker.

correlated with $\mathcal{G}_1$ and $\mathcal{G}_2$, and others seem to be in another domain.

In the third row of Fig. 4.5 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.

In the case of $\mathcal{C}$ and $\bar{\mathcal{C}}$ the largest unlocked correlated domain over each of them is identical, $\mathcal{G}_2(\Uparrow_1) = \mathcal{G}_2(\Downarrow_1) = \mathcal{G}_2$, since these state clusters are related by spin-flip symmetry. There is no symmetry that relates $\mathcal{C}_1$ and $\mathcal{C}_2$.

The pure states are the leaves of the hierarchy[3], as in the ultrametric RSB picture [2].

[3]The leaves of the hierarchy discussed here are *not* the individual states $\mathbf{S}^\mu$ of the state dendrogram of

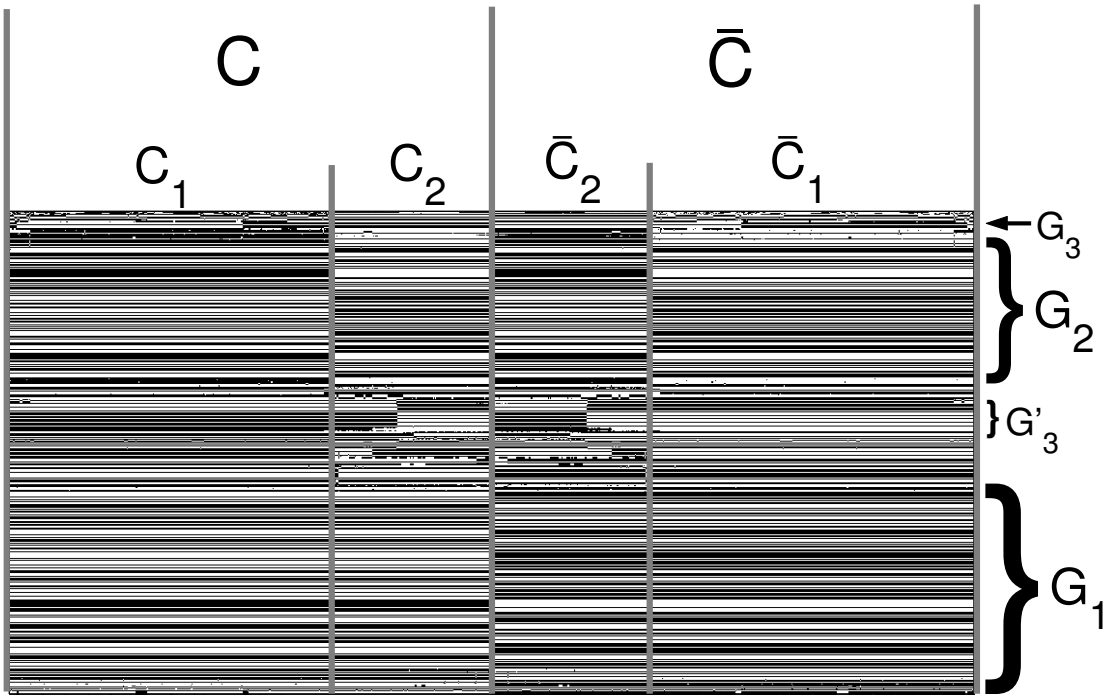

Figure 4.6: The ordered data matrix of Fig. 3.1(c), for a realization of size $N = 8^3$ at $T = 0.2$. The columns represent the states μ and the rows represent the spins i, $S_i^\mu = \pm 1$, with black/white representing $+/-$. The states are ordered according to the dendrogram in Fig. 3.3, and the spins ordered according to the dendrogram $\mathcal{D}$ in Fig. 4.7. The state clusters and the spin domains are marked (see text).

In a pure state each domain $\mathcal{G}_a$ is in a particular configuration $\Uparrow_a$ or $\Downarrow_a$. These domains are associated with free energy barriers separating the pure states. Thus a certain pure state can be characterized by a specific list of domain configurations, e.g. $\{\Uparrow_1, \Downarrow_2, \Downarrow_3(\Uparrow_1, \Downarrow_2), \Uparrow_4(\Uparrow_1, \Downarrow_2, \Downarrow_3), \ldots\}$.

4.3. Spin space structure

So far we have obtained the spin domains using the results of the state space analysis, and have demonstrated the way in which they induce the observed hierarchical structure of state space. The existence of these domains can be observed directly in spin space, i.e. without utilizing information about state space, as we now demonstrate.

As described in Chapter 3, the equilibrium ensemble of states, obtained for each realization, is represented by an $N \times M$ matrix $\{S_i^\mu\}$ (e.g. Fig 3.1(a)). In Chapter 3 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

Chapter 3.

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 defined the distance between a pair of data points i and j as

$$d_{ij} = 1 - {c_{ij}}^2 \ . \tag{4.10}$$

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. 3.1(a) are clustered, is presented in Fig. 4.7(a). When the spins are reordered according to the dendrogram, their distance matrix, shown in 4.7(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. 3.1(a). We obtained Fig. 3.1(b) from (a) by reordering the columns according to the state dendrogram. If we now reorder the rows of Fig. 3.1(b) according to the spin dendrogram $\mathcal{D}$, we get Fig. 3.1(c). The cluster structure of the spins can be seen also here. One can see clearly that spins with indices from 250 up belong to $\mathcal{G}_1$ and distinguish between $\mathcal{C}$ and $\bar{\mathcal{C}}$; the spins with indices up to 200 belong to $\mathcal{G}_2$, and differentiate $\mathcal{C}_1$ from $\mathcal{C}_2$; finally, the spins 210-240 belong to a third, less clear, domain $\mathcal{G}_3'$, which separates $\mathcal{C}_2$ into two sub-clusters. We note that the correlation between two spins that belong to different spin clusters does not vanish, but is much lower than the correlation within a cluster.

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. 4.7(e,f). The relative τ values of this dendrogram are much smaller then those of the dendrograms in Figs. 4.7(a,c), and the reordered distance matrix is structure-less.

If the domains $\mathcal{G}_a$, identified above 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, we located in the spin dendrogram $\mathcal{D}$ of each realization those clusters g_a, which are most similar to $\mathcal{G}_a$. We used the similarity measure

$$\mathcal{S}(g_a, \mathcal{G}_a) = \frac{2|g_a \cap \mathcal{G}_a|}{|g_a| + |\mathcal{G}_a|} \tag{4.11}$$

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, as seen in Tab. 4.5.

D	T	L	$\lvert\mathcal{G}_1\rvert/\lvert\tilde{\mathcal{G}}_1\rvert$	$\lvert\mathcal{G}_2\rvert/\lvert\tilde{\mathcal{G}}_2\rvert$
3	0.2	4	1.00±0.00	1.00±0.01
		5	1.00±0.03	1.00±0.02
		6	1.00±0.02	0.99±0.04
		8	1.00±0.02	1.00±0.02
	0.5	4	0.99±0.05	0.99±0.05
		5	0.99±0.06	0.98±0.08
		6	0.98±0.08	0.98±0.09
		8	0.98±0.08	0.98±0.07
4	0.2	3	1.00±0.00	1.00±0.03
		4	1.00±0.02	0.99±0.04
		5	1.00±0.01	0.99±0.04
	0.8	3	0.97±0.09	0.98±0.08
		4	0.94±0.13	0.93±0.16
		5	0.92±0.15	0.88±0.18

Table 4.1: The part of the largest contiguous set of spins $\mathcal{G}_a$ in the set $\tilde{\mathcal{G}}_a$ defined in (4.4), for $a = 1, 2$. We present the average over $\{J\}$ $\pm$ the standard deviation, for $\theta = 0.95$

D	T	L	$\lvert\mathcal{G}_1\rvert/N$	$\lvert\mathcal{G}_2\rvert/N$	$\bar{c}_{12}$	$P(\mathcal{G}_2 \neq \emptyset)$
3	0.2	4	0.70±0.21	0.099±0.087	0.56±0.33	0.856(6)
		5	0.66±0.21	0.105±0.104	0.55±0.33	0.832(6)
		6	0.66±0.20	0.090±0.090	0.52±0.34	0.836(6)
		8	0.64±0.20	0.084±0.094	0.53±0.34	0.833(8)
	0.5	4	0.31±0.21	0.062±0.056	0.49±0.32	0.56(1)
		5	0.26±0.18	0.052±0.043	0.49±0.33	0.57(1)
		6	0.25±0.16	0.046±0.046	0.47±0.33	0.52(1)
		8	0.22±0.15	0.035±0.034	0.47±0.31	0.55(1)
4	0.2	3	0.74±0.19	0.107±0.105	0.62±0.34	0.840(6)
		4	0.73±0.19	0.083±0.092	0.53±0.34	0.830(6)
		5	0.73±0.19	0.082±0.098	0.51±0.34	0.77(1)
	0.8	3	0.15±0.15	0.036±0.031	0.47±0.31	0.298(9)
		4	0.14±0.12	0.025±0.029	0.54±0.31	0.37(1)
		5	0.14±0.11	0.020±0.025	0.57±0.29	0.38(2)

Table 4.2: The normalized sizes of the domains $\mathcal{G}_1$ and $\mathcal{G}_2$, and the correlation between them. 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 presented. In each entry of the table we give the average over the disorder $\{J\}$ and the standard deviation $[x]_J \pm \sqrt{[x]_J{}^2 - [x^2]_J}$.

D	L	$\lvert\mathcal{G}_3\rvert/N$	$\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2)$	$P(\mathcal{G}_3 \neq \emptyset)$	$\lvert\mathcal{G}'_3\rvert/N$	$P(\mathcal{G}'_3 \neq \emptyset)$
3	4	0.048±0.060	0.55±0.32	0.914(4)	0.087±0.176	0.834(6)
	5	0.046±0.060	0.52±0.32	0.914(4)	0.085±0.184	0.882(5)
	6	0.043±0.060	0.48±0.30	0.924(3)	0.081±0.183	0.896(4)
	8	0.036±0.052	0.43±0.29	0.905(5)	0.076±0.175	0.905(5)
4	3	0.045±0.062	0.56±0.32	0.928(3)	0.094±0.216	0.838(6)
	4	0.037±0.061	0.48±0.30	0.908(4)	0.061±0.149	0.920(3)
	5	0.034±0.062	0.43±0.31	0.84(1)	0.072±0.187	0.865(8)

Table 4.3: The size of the spin domain $\mathcal{G}_3$ and $\mathcal{G}'_3$, and the correlation of $\mathcal{G}_3$ with $\mathcal{G}_1 \cup \mathcal{G}_2$. All results are taken for realizations where the domain concerned does not vanish. We give also the probability of this to happen. All data was taken for $T = 0.2$. We present the average over $\{J\}$ ± the standard deviation.

D	L	$\lvert\mathcal{G}_3 \cap \mathcal{G}'_3\rvert/\lvert\mathcal{G}_3\rvert$	$\lvert\mathcal{G}_3 \cap \mathcal{G}'_3\rvert/\lvert\mathcal{G}'_3\rvert$	$P(\mathcal{G}_3 \neq \emptyset$ and $\mathcal{G}'_3 \neq \emptyset)$
3	4	0.23±0.38	0.19±0.38	0.772(8)
	5	0.15±0.32	0.13±0.31	0.818(7)
	6	0.19±0.35	0.15±0.34	0.832(6)
	8	0.16±0.32	0.13±0.30	0.827(8)
4	3	0.25±0.39	0.21±0.38	0.782(8)
	4	0.16±0.33	0.12±0.30	0.844(6)
	5	0.19±0.33	0.14±0.30	0.73(1)

Table 4.4: The part of $\mathcal{G}_3$ and $\mathcal{G}'_3$ which is common to both of them. Results taken for realizations where $\mathcal{G}_3 \neq \emptyset$ and $\mathcal{G}'_3 \neq \emptyset$. All data was taken for $T = 0.2$. We present the average over $\{J\}$ ± the standard deviation.

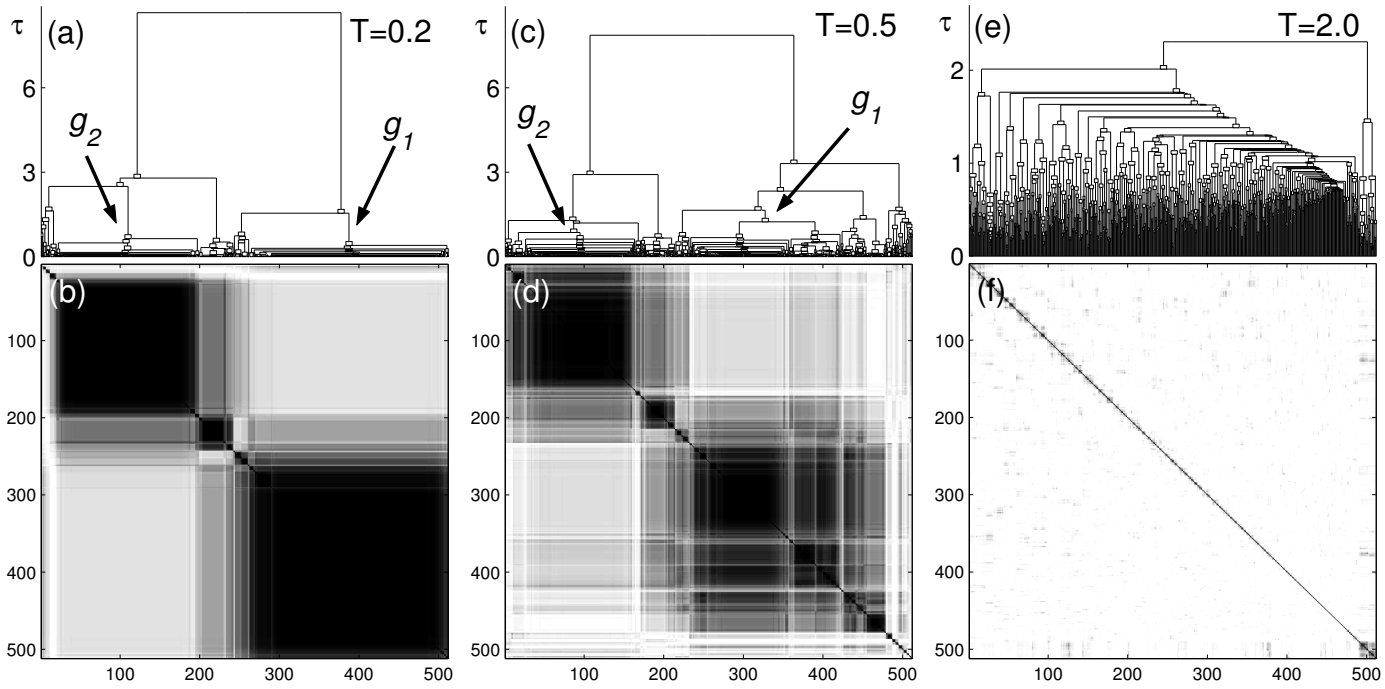

Figure 4.7: **(a)** The spin dendrogram $\mathcal{D}$ for the data of Fig. 3.1(a) produced by Ward's algorithm. **(b)** The spin distance matrix d of this realization realization. The spins are ordered according to their clusters in $\mathcal{D}$. Darker shades correspond to smaller distances and higher 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).

D	T	L	$\mathcal{S}(g_1, \mathcal{G}_1)$	$\mathcal{S}(g_2, \mathcal{G}_2)$
3	0.2	4	0.996±0.063	0.896±0.273
		5	1.000±0.000	0.971±0.128
		6	1.000±0.000	0.986±0.083
		8	1.000±0.000	0.985±0.089
	0.5	4	0.944±0.227	0.916±0.166
		5	0.964±0.184	0.952±0.140
		6	0.969±0.172	0.965±0.129
		8	0.980±0.135	0.962±0.145
4	0.2	3	0.998±0.038	0.902±0.269
		4	1.000±0.000	0.983±0.090
		5	1.000±0.000	0.987±0.079
	0.8	3	0.862±0.333	0.870±0.187
		4	0.930±0.252	0.937±0.144
		5	0.965±0.184	0.983±0.087

Table 4.5: Similarity of $\mathcal{G}_a$ to its most similar cluster $g_a \in \mathcal{D}$. The numbers are the averages over all realizations ± the standard deviation. For $a = 2$ realization with $\mathcal{G}_2 = \emptyset$ were not considered.

5. State overlap

So far we offered a full description of the system in its low T phase relating state space behavior to the microscopic structure in spin space. This description suffices to understand the thermodynamic characteristics of the system. Most works so far, however, did not measure directly the microscopic features of the system (with the exception of [11]), but examined their indirect implications on other parameters, such as the widely addressed overlap distribution $P(q)$. To make contact with the literature we turn now to demonstrate how our picture provides a microscopic interpretation to the observed $P(q)$, and show how our methods allow a decomposition of this function into its physically relevant constituent parts.

Most works [7, 12, 21] presented results for the average over the disorder, $P(q) = [P_J(q)]_J$. We focus here on $P_J(q)$, the overlap distribution for a specific realization $\{J\}$ of the bonds. We show that this is a compound quantity, i.e. the sum of two main parts. One part, $P_J^i(q)$, is the overlap distribution *within* a pure state (and between a pure state and its spin reversed counterpart), which converges to $\delta(|q| - q_{\rm EA})/2$. The second part, $P_J^o(q)$, is the overlap distribution between states that belong to *two different* pure states. $P_J^o(q)$ is also the sum of several contributions, corresponding to different pairs of pure states. The transition between every such pair of pure states can be associated with flipping a specific set of spin domains and, accordingly, with different levels in the states' hierarchy.

In Fig. 5.1(a) we present the total overlap distribution for a specific realizations. A clear structure of several peaks appears in each case. We will relate each peak to flipping one or more spin domains. In this regard our interpretation resembles the RSB picture [2] which also relates the peaks of $P(q)$ [7] to the free energy barriers between the pure states. We will present a direct relation between these peaks and the spin domains $\mathcal{G}_a$, which are the cores of contiguous spin clusters with low (microscopic) surface tension [11].

We use the hierarchical cluster structure of the state space to investigate the specific contributions to $P_J(q)$ by each particular spin domain $|\mathcal{G}_a|$. We do this by calculating the contribution to $P_J(q)$ from state pairs of different clusters at the level a of the hierarchy. We will focus here mainly on two components of $P_J(q)$, related to the first and the second levels in the state hierarchy.

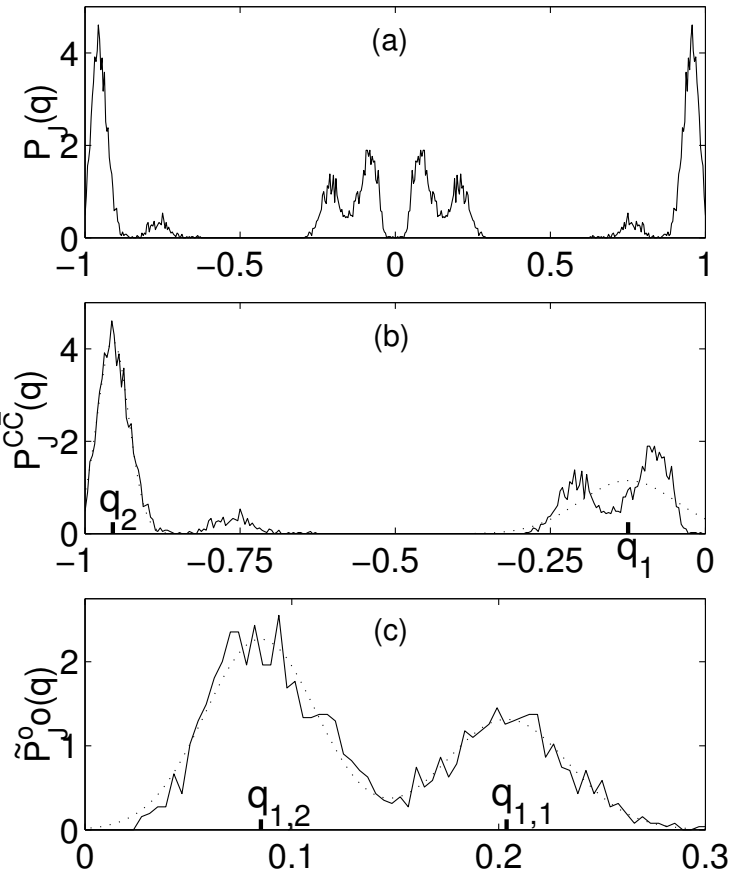

Figure 5.1: **(a)** The distribution $P_J(q)$ for the same realization whose data presented in Fig. 3.1. **(b)** The partial distribution $P_J^{\mathcal{C}\bar{\mathcal{C}}}(q)$ (see text) for the same realization. The estimates q_1 and q_2 for the peak locations are marked. **(c)** The partial distribution $\tilde{P}_J^o(q)$ (see text) for the same realization. The estimates $\tilde{q}_1$ and $\tilde{q}_2$ for the peak locations are marked. In both (b) and (c) the dotted line represents the best fit obtained using (5.10).

5.1. Pairs of states from $\mathcal{C}$ and $\bar{\mathcal{C}}$

Consider first the contribution to $P_J(q)$ from pairs of states $\mu \in \mathcal{C}$ and $\nu \in \bar{\mathcal{C}}$. Denote the distribution of $q_{\mu\nu}$ over this subgroup of pairs as $P_J^{\mathcal{C}\bar{\mathcal{C}}}(q)$. These state pairs contribute half of $P_J(q)$, including a part of the trivial $P_J^i(q)$ around $-q_{\rm EA}$. Since $\mathcal{C}$ and $\bar{\mathcal{C}}$ are related by spin-flip symmetry, we can obtain $P_J(q)$ from $P_J^{\mathcal{C}\bar{\mathcal{C}}}(q)$ by

$$P_J(q) = P_J^{\mathcal{C}\bar{\mathcal{C}}}(q) + P_J^{\mathcal{C}\bar{\mathcal{C}}}(-q) \tag{5.1}$$

For a realization with $|\mathcal{G}_1| > N/2$ we have

$$P_J^{\mathcal{C}\bar{\mathcal{C}}}(q) \simeq \begin{cases} 0 & q \geq 0 \\ P_J(q) & q < 0 \end{cases} \tag{5.2}$$

Consider the set $\mathcal{G}_{\mu\nu}$, defined in (4.3), of spins that flip when passing from a state $\mu \in \mathcal{C}$ to a state $\nu \in \bar{\mathcal{C}}$. Let $\mathcal{G}_{\mu\nu}^{\rm cont}$ be the largest contiguous part of $\mathcal{G}_{\mu\nu}$. We calculated $p_{\rm cont}$ - the average over the state pairs and the disorder of $|\mathcal{G}_{\mu\nu}^{\rm cont}|/|\mathcal{G}_{\mu\nu}|$;

$$p_{\rm cont} = \left[\frac{1}{|\mathcal{C}||\bar{\mathcal{C}}|} \sum_{\mu\in\mathcal{C}} \sum_{\nu\in\bar{\mathcal{C}}} |\mathcal{G}_{\mu\nu}^{\rm cont}|/|\mathcal{G}_{\mu\nu}| \right]_J . \tag{5.3}$$

We found that for most realizations and all system sizes $\mathcal{G}_{\mu\nu}$ is contiguous. For $D = 3$ we have $1 - [p_{\rm cont}]_J = 1.7 \times 10^{-5}$ for $L = 8$, and even smaller values for $L < 8$; for $D = 4$ we have $1 - [p_{\rm cont}]_J = 0.004$ for $L = 3$, 1.5×10^{-6} for $L = 4$ and 3.4×10^{-7} for $L = 5$. Thus the sets $\mathcal{G}_{\mu\nu}$ are, in almost all $\{J\}$, the macroscopic excitations discussed in [11]. Here we will investigate their statistical significance and estimate their size distribution.

For the sake of demonstrating our point, we refer to the states obtained for the realization whose $P_J(q)$ was presented in Fig. 5.1(a). For this realization we clearly identified *three* spin domains.

The overlap $q_{\mu\nu}$ between states μ and ν is related to the size of $\mathcal{G}_{\mu\nu}$ by

$$q_{\mu\nu} = 1 - 2|\mathcal{G}_{\mu\nu}|/N . \tag{5.4}$$

Note that for most state pairs $\mu \in \mathcal{C}$ and $\nu \in \bar{\mathcal{C}}$ our definition of $\mathcal{G}_1$ yields $\mathcal{G}_{\mu\nu} \supset \mathcal{G}_1$ so $q_{\mu\nu} < 1 - 2|\mathcal{G}_1|/N$.

Assuming $\mathcal{G}_2$ does not vanish, we classify these state pairs into two types:

1. Pairs in which $\mathcal{G}_1$ flips when moving from μ to ν (i.e. changes configuration from $\Uparrow_1$ to $\Downarrow_1$ or vice versa), but $\mathcal{G}_2$ remains fixed. The definition of $\mathcal{G}_2$ yields that in most such cases $\mathcal{G}_2 \cap \mathcal{G}_{\mu\nu} = \emptyset$. Thus $|G_{\mu\nu}| < N - |\mathcal{G}_2|$, and we must have $2|\mathcal{G}_2|/N - 1 \leq q_{\mu\nu} \leq 1 - 2|\mathcal{G}_1|/N$.

2. Pairs in which both $\mathcal{G}_1$ and $\mathcal{G}_2$ flip when moving from state to state. For those pairs in most cases $\mathcal{G}_{\mu\nu} \supset \mathcal{G}_1 \cup \mathcal{G}_2$ so $q_{\mu\nu} \leq 1 - 2(|\mathcal{G}_1| + |\mathcal{G}_2|)/N$.

Disregarding the splitting induced by $\mathcal{G}_3$ and $\mathcal{G}_3'$ we identify two main peaks that dominate $P_J^{\mathcal{C}\bar{\mathcal{C}}}(q)$, as seen in Fig. 5.1(b). One, centered at q_1, is attributed to state pairs of the first type, and hence

$$2|\mathcal{G}_2|/N - 1 \leq q_1 \leq 1 - 2|\mathcal{G}_1|/N . \tag{5.5}$$

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

$$q_2 \leq 1 - 2(|\mathcal{G}_1| + |\mathcal{G}_2|)/N , \tag{5.6}$$

5.2. Are the two peaks due to finite size?

We present now arguments and results that show that these two peaks survive for large system sizes. We do not expect the peak at q_2 to vanish since it includes the negative half

of $P_J^i(q)$, and in particular the trivial symmetry transition from a state to its inverse. This part of $P^i(q)$ converges to a delta function at $-q_{\rm EA}$.

There are three distinct mechanisms by means of which the picture could change qualitatively and approach a single peak as L increases: (a) if the fraction of state pairs in which $\mathcal{G}_1$ and $\mathcal{G}_2$ flip separately tends to zero; (b) if the normalized size $|\mathcal{G}_2|/N$ goes to zero and (c) if the width of each peak is much larger than $|\mathcal{G}_2|/N$ and the two peaks are indistinguishable. In the first case the size of the peak at q_1 decreases until it disappears as $L \to \infty$; in case (b) and (c) the two peaks merge. We discuss now each of these eventualities separately.

(a) Buildup of correlations:
The weight of the excitations contributing to the peak at q_1 will vanish if $\mathcal{G}_1$ and $\mathcal{G}_2$ always flip together, i.e. if they are fully correlated. As we have seen in Chapter 4 the average correlation $\bar{c}_{12}$ between the spins of $\mathcal{G}_1$ and $\mathcal{G}_2$ decreases (as L increases) and seems to approach 0.5. This means that $\mathcal{G}_1$ can flip independently for about 1/4 of the state pairs and, hence, about 1/4 of the excitations $\mathcal{G}_{\mu\nu}$ will be of the first first kind.

(b) The size of $\mathcal{G}_2$:
$|\mathcal{G}_2|/N$ does decrease for the system sizes measured, as seen in Tab. 4.2. Nevertheless, Figs. 4.4 and 4.4 strongly indicate that the size distributions for both $|\mathcal{G}_1|/N$ and $|\mathcal{G}_2|/N$ converge to distributions with non-zero mean and variance. Since $\mathcal{G}_2$ is macroscopic, any pair of states μ and ν of the first kind belong to different pure states (i.e. not related by spin-flip symmetry). Hence these state pairs will contribute to $P_J^o(q)$.

(c) Indistinguishable peaks:
Since in general $2|\mathcal{G}_2|/N - 1 < 1 - 2(|\mathcal{G}_1| + |\mathcal{G}_2|)/N$ this does seem possible. In order to be certain that this is not the case, we measured for each realization the following quantities:

$$r_1 = (1 - q_1)N/2|\mathcal{G}_1| \tag{5.7}$$
$$r_2 = (1 - q_2)N/2|\mathcal{G}_1 \cup \mathcal{G}_2| \tag{5.8}$$
$$r_3 = (1 - q_1)N/2(N - |\mathcal{G}_2|)\ . \tag{5.9}$$

The values q_1 and q_2 were found, for each $\{J\}$, by fitting $P_J^{\mathcal{C}\bar{\mathcal{C}}}(q)$ to a sum of two Gaussians,

$$P_J^{\mathcal{C}\bar{\mathcal{C}}}(q) = b_1 \exp[(q - q_1)^2/{a_1}^2] + b_2 \exp[(q - q_2)^2/{a_2}^2]\ , \tag{5.10}$$

with a_i, b_i and q_i as fit parameters, as plotted in 5.1(b). We discarded the realizations for which we could not identify clearly two separate Gaussians, and kept only those cases for which the following criteria were satisfied:

$$\begin{aligned} (q_1 - q_2) &\geq \max\{a_1, a_2\}/4\ ; \\ -1 \leq q_1, q_2 &\leq 1\ ; \\ a_1, a_2 &\geq 2/N\ ; \\ b_1, b_2 &\geq 1/M\ . \end{aligned} \tag{5.11}$$

These conditions are satisfied by a fraction p of the realizations. We present in Tab. 5.1 the values of r_1, r_2, r_3 and r_2/r_1, averaged over the realizations that satisfy (5.11). Note

that for all but the smallest size $r_1/r_2 \approx 1$. Since we have already shown that $\mathcal{G}_2 \neq \emptyset$, this means that $q_1 \neq q_2$. We also see that, as expected on the basis of the inequalities (5.5) and (5.6), in general $r_1 > 1$, $r_2 > 1$ and $r_3 < 1$.

5.3. Pairs of states from $\mathcal{C}_1$ and $\mathcal{C}_2$

A more rigorous argument to support the claim that the peaks do not merge if $\mathcal{G}_2$ does not vanish and $\bar{c}_{12} < 1$ goes as follows. According to our definition of the spin domains, all states where $\mathcal{G}_1$ and $\mathcal{G}_2$ are in configurations $\Uparrow_1, \Uparrow_2$ belong to $\mathcal{C}_1$, and all states where they are in configurations $\Uparrow_1, \Downarrow_2$ belong to $\mathcal{C}_2$. For each pair of spins $i \in \mathcal{G}_1$ and $j \in \mathcal{G}_2$ we estimate

$$c_{ij} \simeq \frac{|\mathcal{C}_1|}{|\mathcal{C}|}[\Uparrow_1]_i[\Uparrow_2]_j + \frac{|\mathcal{C}_2|}{|\mathcal{C}|}[\Uparrow_1]_i[\Downarrow_2]_j = [\Uparrow_1]_i[\Uparrow_2]_j \frac{|\mathcal{C}_1| - |\mathcal{C}_2|}{|\mathcal{C}_1| + |\mathcal{C}_2|} \ . \tag{5.12}$$

Using this relation we can estimate from the measured $\bar{c}_{12}$ the size of $\mathcal{C}_2$:

$$|\mathcal{C}_2| \simeq \frac{M}{4}\left(1 - \sqrt{\bar{c}_{12}}\right) \ . \tag{5.13}$$

Thus, the probability to sample a pair of states $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$,

$$2|\mathcal{C}_2||\mathcal{C}_1|/0.5M(M-1) \simeq \frac{1-\bar{c}_{12}}{8} \ , \tag{5.14}$$

does not vanish when $\bar{c}_{12} < 1$. For each such pair we expect $\mathcal{G}_2 \subset \mathcal{G}_{\mu\nu}$, and thus $q_{\mu\nu} < 1 - 2|\mathcal{G}_2|/N$. Our results indicate that as L increases $\mathcal{G}_2$ remains macroscopic and $\bar{c}_{12}$ approaches 0.5. Therefore distribution of overlaps between pairs of states $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$ will be non-trivial and will have a finite weight in the total $P_J(q)$.

The states $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$, i.e. pairs taken from state clusters that appear at the second level of the states' hierarchy, contribute a non-vanishing part of $P_J(q)$. Denote their contribution to $P_J(q)$ by $\tilde{P}^o_J(q)$ (see fig 5.1(c)). Since the states μ and ν belong to (in most cases, when $\mathcal{G}_2$ is macroscopic) different pure states, the sub-distribution $\tilde{P}^o_J(q)$ constitutes a major part of $P^o_J(q)$[1] and serves as a lower bound to it. In Fig. 5.2 we present $\tilde{P}^o_J(q)$ and $P_J(q)$ for four realizations $\{J\}$. The figure shows clearly that the separation is natural, and not just an artifact of our analysis.

In some realizations, such as the ones that yield Figs. 5.2(a) and 5.2(d), $\tilde{P}^o_J(q)$ exhibits two peaks; this splitting is due, as mentioned above, to spin domains $\mathcal{G}_3$ and G'_3. We analyzed $\tilde{P}^o_J(q)$ in the same way as we did for $P^{\mathcal{C}\bar{\mathcal{C}}}_J(q)$, using the same fit (5.10), with $\tilde{q}_{1,1}$ and $\tilde{q}_{1,2}$ denoting the centers of the two Gaussians. We look for the fraction $\tilde{p}$ of the realizations, for which $\tilde{P}^o_J(q)$ fulfills the criteria (5.11). We define $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$ in the same manner as r_1, r_2 and r_3 were defined, but now with respect to $\mathcal{G}_2$ and $\mathcal{G}_3$. We chose to use $\mathcal{G}_3$ which is associated with $\mathcal{C}_1$ and therefor has better statistics than G'_3 which is associated

[1] The peak of $P^{\mathcal{C}\bar{\mathcal{C}}}_J(q)$ at q_1 (see Fig. 5.1(b)) is due to related pairs, taken from $\mathcal{C}_1$ and $\bar{\mathcal{C}}_2$.

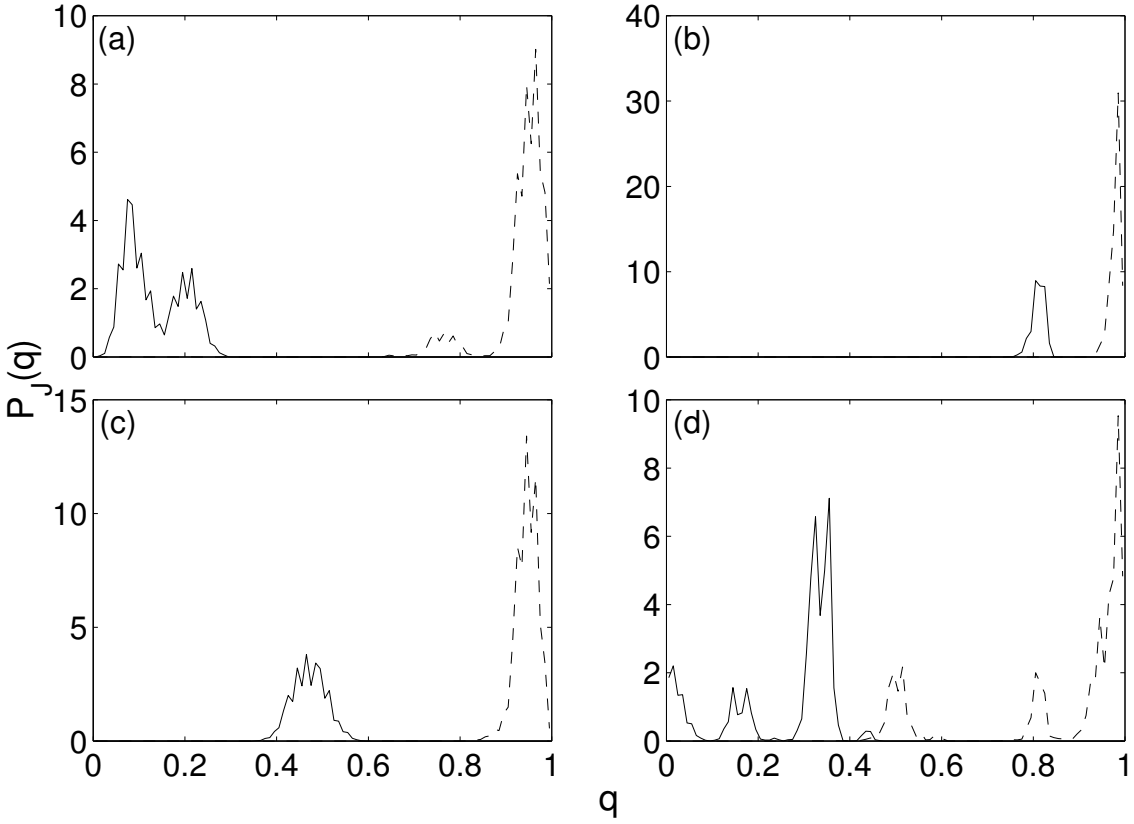

Figure 5.2: The distribution $P_J(q)$ for four realizations of $\{J\}$ at $T = 0.2$. The distribution in (a) is the same as in Fig. 5.1(a). The solid line describes $\tilde{P}^o_J(q)$. The dashed line plots the rest of the of the distribution $P_J(q) - \tilde{P}^o_J(q)$, and is dominated by the distribution $P^i_J(q)$ of overlaps inside the pure states.

with $\mathcal{C}_2$. The results, reported in Tab. 5.1, show a similar behavior to that of $P_{\mathcal{C}\bar{\mathcal{C}}}$. However, since the values of $\tilde{p}$ are low, the results are not conclusive.

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. Our picture, however, is *not* consistent with RSB: it yields a non-ultrametric state structure, as we show in Chapter 6.

5.4. Measuring $\tilde{P}^o(q)$

We now measure the contribution of $\tilde{P}^o(q) = [\tilde{P}^o_J(q)]_J$ to the average distribution $P(q) = [P_J(q)]_J$. In order to assure that $\tilde{P}^o_J(q)$ indeed represents the overlap between different pure states, we measured it only for those realizations in which $\mathcal{G}_2$ was relatively large, $|\mathcal{G}_2| > 0.05N$. For the rest of the realizations we set $\tilde{P}^o_J(q)$ to zero; hence our $\tilde{P}^o(q)$ is a lower bound of the true one. In Figs. 5.3 we show the distributions $P(q)$ and $\tilde{P}^o(q)$. For systems with Gaussian couplings $P^i(q)$ has a very small contribution at $|q| < 0.7$ and $\tilde{P}^o(q)$ 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 [28].

D	L	r_1	r_2	r_3	r_2/r_1	p
3	4	1.55±0.36	1.80±0.71	0.73±0.37	1.15±0.30	0.006
	5	1.43±0.65	1.31±0.12	0.78±0.17	0.99±0.18	0.038
	6	1.37±0.32	1.36±0.34	0.86±0.19	1.00±0.15	0.128
	8	1.37±0.24	1.33±0.20	0.86±0.13	1.00±0.10	0.248
4	3	1.51±0.28	1.34±0.18	0.71±0.09	0.91±0.21	0.010
	4	1.27±0.26	1.26±0.21	0.90±0.15	1.00±0.14	0.136
	5	1.26±0.33	1.24±0.26	0.93±0.13	1.00±0.07	0.260
		$\tilde{r}_1$	$\tilde{r}_2$	$\tilde{r}_3$	$\tilde{r}_2/\tilde{r}_1$	$\tilde{p}$
3	4	1.60±0.38	1.66±0.36	0.27±0.11	0.99±0.25	0.006
	5	1.61±0.49	1.60±0.50	0.30±0.12	0.87±0.24	0.026
	6	1.84±0.41	1.52±0.48	0.26±0.14	0.94±0.33	0.056
	8	1.62±0.27	1.66±0.39	0.24±0.13	0.98±0.25	0.090
4	3	1.30±0.00	1.36±0.00	0.39±0.00	0.67±0.00	0.002
	4	1.66±0.81	1.50±0.46	0.31±0.13	0.83±0.26	0.030
	5	1.71±0.52	1.70±0.70	0.21±0.16	1.06±0.44	0.080

Table 5.1: The relations between peak centers and the correlated domains $\mathcal{G}_a$, at $T = 0.2$. We present the average over $\{J\}$ and the standard deviation. The results confirm our prediction: $r_1, r_2 > 1$; $r_3 < 1$; $r_2/r_1 \approx 1$. In the last column we present the part p of the realizations considered, i.e. the realizations for which the criteria (5.11) have been met.

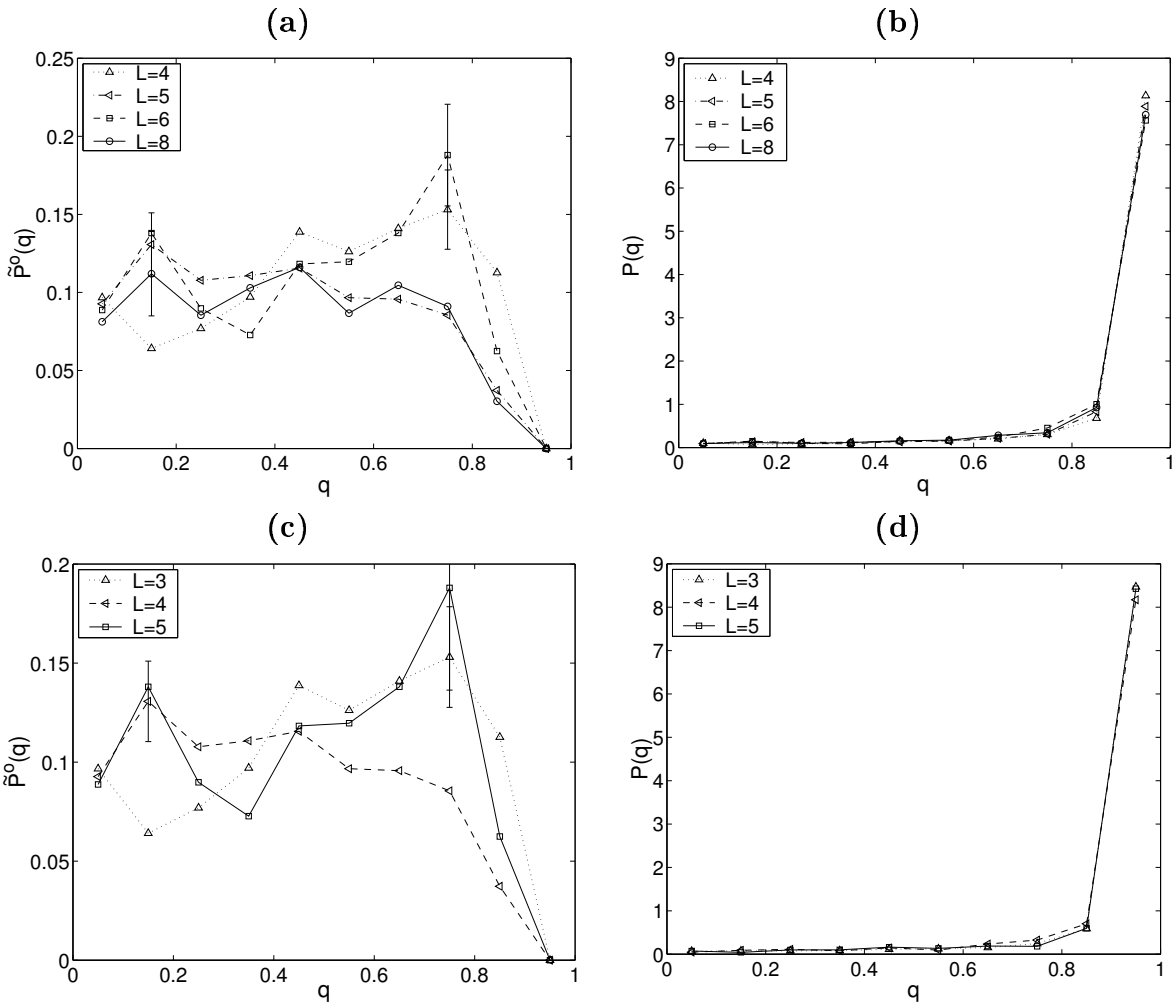

Figure 5.3: **(a)** The partial distribution $\tilde{P}^o(q)$ for $D = 3$ $L = 4, 5, 6, 8$. It is normalized so that $2\int_0^1 \tilde{P}^o(q)dq$ is its weight in the total $P(q)$. For each L the largest error bar is shown. **(b)** The distribution $P(q)$ for the same systems as in (a). The size of the error bars is of the order or less than the size of the symbols. **(c)** $\tilde{P}^o(q)$ for $D = 4$ $L = 3, 4, 5$. Same as in (a). **(d)** $P(q)$ for the same systems as in (c).

6. Ultrametricity

Ultrametricity is one of the main characteristics of the mean field RSB picture. Efforts to establish [9] or dismiss [29] the existence of ultrametricity in short range spin glasses did not yield any conclusive results. The main problem is that we can equilibrate only small systems, where ultrametricity is hindered by finite size effects. The problem with small systems is that the triplets of states which may present ultrametricity, have a low statistical weight, which increases with the size of the system. The special tool provided by our method of analysis allows us to overcome this problem. We allocate the triplets that should exhibit ultrametricity and base our analysis only the bases of those triplets. 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

$$D_{\alpha\beta} \leq \max\{D_{\alpha\gamma}, D_{\beta\gamma}\} \; . \tag{6.1}$$

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 pure state at $T < T_c$, since for two states μ and ν inside the pure state $P(D_{\mu\nu}) \to \delta(D_{\mu\nu} - q_{\rm EA})$.

The non-trivial result of RSB is that the pure states *themselves* are ultrametric. In order to investigate this claim, we have to focus on triplets of states, each chosen from a different pure state. For large systems with many pure states this does not require special care, since almost all triplets of states will belong to three different pure states. For small systems, however, a large fraction of the possible triplets will have at least two states from the same pure state. 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 Chapter 3 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 from these three clusters belong to three different pure states, since we have to pass a macroscopic barrier 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 Section 4.2). 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 pure states (determined by $\mathcal{G}_1$ and $\mathcal{G}_2$) and hence the the RSB picture clearly does not hold.

At first sight it may seem more reasonable to use the first two levels of the state hierarchy, where the domains $\mathcal{G}_1$ and $\mathcal{G}_2$ are both macroscopic. This is not possible for the present purpose, for the following reason. In order to check the state space for ultrametricity, we first have to divide it into two parts, related by global spin reversal, and use only triplets with all three states from the same part Otherwise ultrametricity breaks down in a trivial fashion, as we now show. Let μ and ν be two states sampled according to their Boltzmann weights. The state $\bar{\mu}$, given by $\mathbf{S}^{\bar{\mu}} = -\mathbf{S}^{\mu}$, has the same weight as μ and therefore there is reasonable probability that it or a state close to it (with a similar spin configuration) will be sampled. For the triplet of states $\mu\bar{\mu}, \nu$ we have $D_{\mu\nu} = 1 - D_{\bar{\mu}\nu}$ and $D_{\mu\bar{\mu}} = 1$, so if $D_{\mu\nu} > 0$ we have $D_{\mu\bar{\mu}} > \max\{D_{\mu\nu}, 1 - D_{\bar{\mu}\nu}\}$, which violates ultrametricity. Our analysis helps us to avoid such triplets of states, since it provides us with a way to identify the natural symmetric partition of state space into $\mathcal{C}$ and $\bar{\mathcal{C}}$, as discussed in Chapter 3. We examine only triplets with all three states in $\mathcal{C}$, with $\mathcal{G}_1$ in configuration $\Uparrow_1$.

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

$$K_{\mu\nu\rho} = \frac{D_{\mu\nu} - D_{\mu\rho}}{D_{\nu\rho}} \ . \tag{6.2}$$

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 $\{J\}$. We found that in all systems we find with high probability that $K_{\mu\nu\rho} = 1$ exactly (see Table 6.1). This is, however, clearly a finite size effect; as L increases the probability $P(K = 1)$ decreases dramatically. We would like to disregard this part of the distribution. This cannot be done if it represents a peak that flattens as L increases. This, however, is clearly not the case: we present in Table 6.1 the probability $P(0.9 \leq K < 1)$, and show that its increase 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

$$P(K|K < 1) = \begin{cases} P(K) \,/\, P(K < 1) & K < 1 \\ 0 & K = 1 \end{cases} \ . \tag{6.3}$$

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. 6.1. In Table 6.1 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

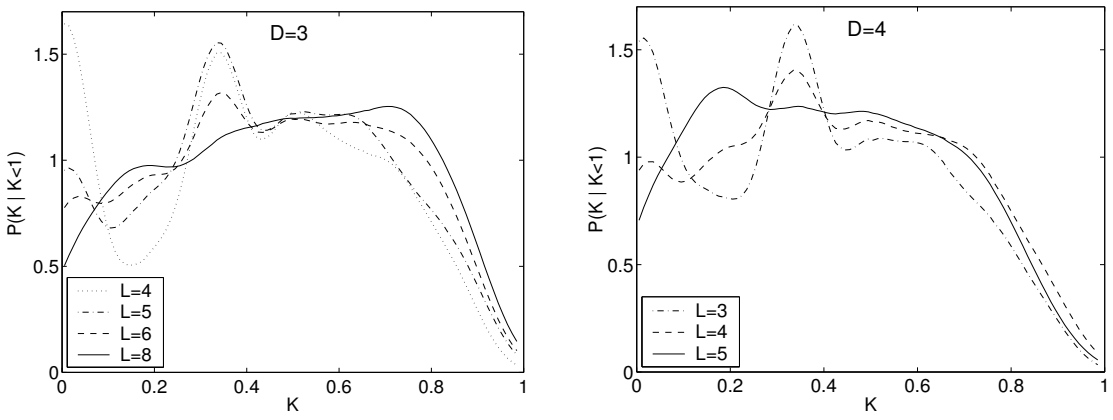

Figure 6.1: The distribution $P(K|K<1)$ of $K_{\mu\nu\rho}$, for $\mu \in \mathcal{C}_2$, $\nu \in \mathcal{C}_{1a}$ and $\rho \in \mathcal{C}_{1b}$. All systems are sampled at $T = 0.2$.

with non-vanishing mean and variance, indicating breakdown of ultrametricity for the three pure states studied.

We still 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 $L \to \infty$. We deduced in Chapter 5 that since the average correlation $\bar{c}_{12}$ between $\mathcal{G}_1$ and $\mathcal{G}_2$ does not approach one, $|\mathcal{C}_2|/N$ does not vanish. 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 than both $|\mathcal{C}_{1a}|/N$ and $|\mathcal{C}_{1b}|/N$ do not vanish. Therefore the probability to sample such a state triplet remains finite, and the system does not exhibit ultrametricity.

We must remark that for the small system sizes measured we could not extract conclusive numerical evidence to the effect that $\mathcal{G}_3$ remains macroscopic. 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 $\mathcal{G}_1 \cup \mathcal{G}_2$ decreases as L increases, but it is not conclusive. Nevertheless, we gave here the prediction made according to our picture, and have demonstrated the power of our analysis in probing specific parts of state space to obtain a clearer picture of a specific attribute and eliminate finite size effects.

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 6.1: 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 $T = 0.2$.

7. Ground states of binary Ising spin glasses

There has been an upsurge of interest in the Edwards-Anderson model (1.1) with binary couplings, $J_{ij} = \pm 1$, at 3 dimensions. Several papers investigated the overlap distribution $P(q)$ at finite and zero temperature. Most works so far concluded that $P(q)$ is non-trivial at finite T but is trivial at $T = 0$, as in the case of a Gaussian spin glass. However, for a binary spin glass $q_{\rm EA} \neq 1$ [21], that is, there is a single pair of pure states with finite entropy per spin [30]. We apply our analysis to the ground states of this model. Our conclusion is that $P(q)$ *is non-trivial at* $T = 0$.

Berg et al. [22] addressed the issue directly by generating ground states $\mathbf{S}^\mu = (S_1^\mu, S_2^\mu, ...S_N^\mu)$, for 512 realizations $\{J\}$ of systems with sizes $L = 4, 6, 8$ (and for 7 realizations of $L = 12$). For each $\{J\}$ they computed the overlap distribution function $P_J(q)$, where the overlap $q^{\mu\nu}$ is calculated between all pairs of ground states μ, ν. They studied the function obtained by averaging over all realizations, $P(q) = [P_J(q)]_J$. In particular, they evaluated $P(0)$; the second moment of the distribution $\sigma^2(q)$, and the quantity $x_{1/2}$, where x_a is defined by

$$x_a = 2 \int_0^a P(q) dq \tag{7.1}$$

If $P(q) \to \delta(|q| - q_{\rm EA})/2$ for large L, all these quantities should extrapolate to zero (provided one uses $a < q_{\rm EA}$). Berg et al. found that all three quantities decrease as L grows; they could, however, fit the data to L^{-y}, with $y = 0.72 \pm .12$, as well as to $A + BL^{-3}$, indicating consistency with extrapolation to both vanishing and non-vanishing limiting values.

Hartmann [21] also studied the size-dependence of $x_{1/2}$ and found that it behaves as L^{-y} with $y = 1.25 \pm .05$, indicating a trivial $P(q)$; the same conclusion was reached by Hatano and Gubernatis [31] who studied $P(0)$ at finite temperatures. Krzakala and Martin [32] presented arguments that also support a trivial $P(q)$. Finally, very recently Palassini and Young [20] evaluated $P(q)$ for a sequence of temperatures and sizes $L = 4, 6, 8, 10$. They evaluated $x_{1/2}$ as function of L and T and demonstrated that the data are consistent with a scaling form. According to their scaling, for fixed $T > 0$ and sizes $L \gg L_c(T)$, $x_{1/2}$ goes to a constant, $x^\infty(T) \propto T$; hence they find that $P(q)$ is non-trivial at $T > 0$ and trivial at $T = 0$.

We argue here that all the studies mentioned measured a compound quantity, $x_{1/2}$, which is the sum of two parts; one which is relevant to the question asked, and another which is irrelevant. Furthermore, for some of the sizes studied, the irrelevant part is as large as the relevant one.

In order to demonstrate this point we generate an unbiased sample of ground states of the model (1.1) with binary couplings in 2, 3 and 4 dimensions. The methods we use to generate those samples are described in Section 7.1.

In Section 7.2 we relate to the observable measured in previous works for 3 dimensional systems. We analyze the unbiased samples in the same way we analyzed the data for Gaussian spin glasses in Chapters 3 and 4. We have found that the structure of both spin and state space seems qualitatively identical to that of a Gaussian spin glass in *finite* temperature. As in the latter, this structure persists as $L \to \infty$.

In Section 7.3 we use our method of analysis to isolate the relevant part of $P(q)$ for 3 dimensions, and present the results obtained when this is done. The results verify the conclusion of Section 7.2.

7.1. Creating an unbiased sample of ground states

We applied the *genetic cluster exact approximation* (GCEA) algorithm [33] (see Appendix B) to generate *biased* [34] samples of ground states of the model (1.1) with the values $J_{ij} = \pm 1$ assigned to each bond with equal probabilities[1]. Out of those samples, which are *biased* [34], we generate unbiased ones, of $M = 2 \times 500$ states each[2] - 500 states and their spin inverted images.

The ground states of a $\pm J$ spin glass can be divided into *valleys* [30]. A valley V consists of all the ground states that can be traversed by flipping one spin at a time without changing the energy. At $T = 0$, under Metropolis dynamics, each valley constitutes an ergodic component of the phase space. For a different choice of the dynamics few valleys my be joined to one ergodic component.

We cannot identify a valley with a pure state, but we are sure that all the states within a valley are in the same pure state for the following reason. In sub-section 7.1.3 we find that the density profile inside a valley is smooth and homogeneous, that is, the number of states inside a ball of radius r in state space depends only on r, the valley size $|V|$ and the system size N. Since this density profile is smooth as a function of r we conclude there is no cluster structure inside a valley, and thus no sub-partition into different pure states.

We observe correlated spin domains in the system, as the ones in a system with Gaussian couplings. If a domain wall could be advanced through these domains, connecting two pure states by a series of zero energy spin flips, the density profile inside the valley would be non-homogeneous, in contradiction to our findings.

The samples obtained by the GCEA are biased. Fortunately, the bias over-represents the smaller valleys, though larger valleys still have a higher probability to be sampled [33]. The probability to miss small valleys by this method is *lower* then by an unbiased method, such as Monte Carlo. Thus, we are assured to sample all the statistically significant valleys.

[1] In fact we ensured to have equal numbers of positive and negative bonds.

[2] except for realizations with less than 1000 ground states.

It is simple to overcome this bias for *small systems*, of less than 2×10^5 ground states per valley. For such systems the probability that the GCEA will miss a valley is entirely negligible [33]. For these systems we can find all the ground states. The GCEA generated sample contains one or more ground state from each valley (of non-negligible statistical weight). Given a ground state g_0, we can obtain all the states that reside in its valley V using the following algorithm:

1. Initialize two ordered lists of states $\mathcal{L}$ and $\mathcal{L}'$. (we used alphabetic order for the states in each list).
2. Place g_0 in $\mathcal{L}'$.
3. While $\mathcal{L}'$ is not empty, for each state $g \in \mathcal{L}'$ do
    1. for each spin i, if flipping i does not increase the energy of g then
        - generate a new state g' from g by flipping i.
        - Look for g' in $\mathcal{L}$ and $\mathcal{L}'$. If it is not found, add g' to $\mathcal{L}'$.
    2. Move g from $\mathcal{L}'$ to $\mathcal{L}$.
4. Return $\mathcal{L}$.

This algorithm requires $|V|$ iterations, one for each ground state in V. In each iterations it tries N spin flips, and if a new ground state is found it performs a search in $\mathcal{L}$, whose complexity is $\mathcal{O}[\log |\mathcal{L}|]$, and a search in $\mathcal{L}'$. The total complexity of this algorithm is $\mathcal{O}[N|V| \log |V|]$.

For systems with less than 2×10^5 ground states in each valley we obtained all the ground states, using the algorithm described above. After obtaining all the states with this method of exact enumeration, we randomly selected M of them to obtain an unbiased sample, with each valley represented according to its weight.

For systems with valleys larger then than 2×10^5 ground states our computer resources did not suffice to obtain all ground states. For those systems we have generated an unbiased sample by simulated tempering [35] (see Appendix C). For $L = 8,\ D = 3$ the computer time required to generate a large sample exceeds the limit of our resources, and we managed to obtain a sample only for 60 realizations. Thus, we used another method for this system size, based on a relation found by Hartmann [36] between the valley size $|V|$ and an estimate of its "diameter" $l_{\max}$. The numbers of realizations sampled by each method are give in table 7.1

We implemented a variant of the method suggested by Hartmann [36]. We start from the states produced by the GCEA and generate an unbiased sample of the states *within* a valley. The first step is to sort the samples into valleys, using ballistic search [36]. The second is to estimate the size $|V|$ of each valley, using an empirical scaling relation. In the last step we generate a sample of M ground states, in which each valley has the proper weight (proportional to $|V|$).

D	L	exact enumeration	ST	$l_{\max}$
2	7	988		
	8	995		
	10	872	13	
	14		832	
3	4	975		
	5	797		
	6	899	42	
	8		60	878
4	3	988		
	4	452	27	

Table 7.1: The number of realizations we have tested for each system size L and dimension D. For each L and D the table shows for how many realizations the ground states were sampled by each method: by exact enumeration; by using simulated tempering; and by using the empirical relation between the valley size and $l_{\max}$.

7.1.1. Sorting ground states into valleys

We used the a method developed by Hartmann, called ballistic search [36]. In order to determine if two ground states μ and ν are in the same valley we try to find a *path* from one to the other by flipping spins without increasing the energy. Let Δ be the group of spins which have different values in μ and ν. We start from μ. Iteratively a free spin is selected randomly from Δ, flipped, and removed from Δ. If we have reached ν then a path exist and μ and ν are in the same valley. If we fail, i.e. we reach a point where no spin in Δ is free, it is still possible that a path exists. Hartmann found the the probability p_f to find a path if one exists depends on the size of Δ and the system size. For $N = L^3$ he found that for $|\Delta| \leq 40$ $p_f > 0.8$ and for $|\Delta| \leq 80$ $p_f > 0.2$.

We have sorted the states using the following procedure:

1. Choose a state μ_0 from the sample. Initialize a valley $V_1 = \{\mu_0\}$.
2. Initialize a list of valleys $\mathcal{L} = \{V_1\}$.
3. For each state μ
    1. For each valley $V_k \in \mathcal{L}$ repeat 3 times
        - Choose a state $\nu \in V_k$.
        - Try to find a path from μ to ν as described above.
        - If successful
            - If a path was found to a previous valley $V_{k'}$, merge V_k with $V_{k'}$.
            - Else, add μ to V_k.

2. If no path was found to any valley, initialize a new valley $V = \{\mu\}$, and add it to $\mathcal{L}$.

4. For each two valleys $V_k, V_{k'} \in \mathcal{L}$ repeat 10 times

    1. Choose randomly two states $\mu \in V_k$ and $\nu \in V_{k'}$.
    2. Try to find a path from μ to ν.
    3. If successful, merge V_k and $V_{k'}$.

5. return $\mathcal{L}$.

This procedure ensures that the probability of an error, i.e. the probability to associate two states from one valley to different valleys, is less than $(1 - p_f)^{13}$. For most pairs of states and most valleys $p_f > 0.5$. Thus, the probability to 'split' a valley is negligible.

7.1.2. Generating an unbiased sample within a valley

All the states of a valley have the same energy, and hence the same thermodynamic weight. In the sense of Boltzmann weights, any group of M states from the valley will have the same probability to be sampled. Our definition of an unbiased sample is one over which the average value of spin i, $\langle S_i \rangle_M$, is equal (or close) to the thermodynamic average $\langle S_i \rangle_V$ over all the states in the valley V. We obtain such a sample by starting several random walks inside the valley from distant (see below) starting points and stop when the values of $\langle S_i \rangle$ for the different walks are close.

We start with the group of ground states from a valley V. It may happen that the set represents only a 'corner' of the valley, i.e. some spins that are fixed over the sample are not fixed over the whole valley. In order to avoid this problem we expand the set. For each state in the set we generate a 'distant' state in the following way. We start from the state, list all the free spins, and flip one of them randomly. We continue to do so, updating the list of free spins as we go, without flipping a spin twice. When all the currently free spins were already flipped once we stop and add the generated final state to the set. This way we produce an expanded set of states $C_{\rm exp}$.

To generate an ensemble of points in V we perform random walks in state-space (zero-temperature Monte Carlo). Rather than using all the states of $C_{\rm exp}$ as starting points for these random walks, we use only a subset Λ from $C_{\rm exp}$. On the one hand, to save computing time this subset should be as small as possible. On the other hand, we want to generate as many independent distant states as possible. To satisfy these conflicting demands, we try to identify Λ, the smallest subset of $C_{\rm exp}$ which has the same number of *unlocked* spins as the whole expanded set. Say there are $\zeta(C_{\rm max})$ spins whose sign is fixed in all the states of $C_{\rm exp}$; then $N - \zeta(C_{\rm exp})$ spins are *unlocked*. The set Λ is constructed iteratively, initialized by a randomly chosen first state. If at any point in our iterative process $\zeta(\Lambda)$ is the number of spins that are fixed over Λ, we identify from $C_{\rm exp}$ that particular state μ which, when added to Λ, gives maximal reduction of ζ: $\mu = {\rm a}rgmin_\nu[\zeta(\Lambda \cup \{\nu\})]$. We add this μ to

Λ. The procedure stops when ζ cannot be decreased any further, i.e. when the number of fixed spins in Λ equals $\zeta(C_{\rm exp})$.

We generate now our unbiased sample Γ, of states within the valley. First place each state μ_k of $\Lambda = \{\mu_1, \ldots, \mu_m\}$ in its ensemble Γ_k. Next, use each of these m states as the starting configuration for a random walk in the valley. This is done as a zero temperature Monte Carlo simulation with Metropolis single-spin-flip dynamics. Each walk k is sampled after every t steps, with $t = 2 + 0.1N(N - \zeta(\Lambda))^2$, and the state of the system at step t is added to an ensemble Γ_k. We now perform measurements for each spin i which is *non-fixed* in the set of states Λ. For each such spin i we calculate $\langle S_i \rangle_k$, the average value of the spin in the ensemble Γ_k, and the variance $v_i = var_k(\langle S_i \rangle_k)$ over the m ensembles. We stop the random walks (and addition of new states to the ensembles Γ_k) when $\bar{v}$, the average variance (averaged over all the non-fixed spins), is less than 0.01. We assume that when this condition is met, our zero-temperature simulation has ran long enough to give us an unbiased sample of the valley. We join all the ensembles Γ_k into one list of states $\Gamma_V = \bigcup_k \Gamma_k$. The states in Γ_V constitute our unbiased sample of the valley V.

As opposed to most Monte Carlo procedures, we do not discard the states sampled at the beginning of each walk, since each starting configuration has the same weight as any other configuration sampled.

7.1.3. Estimation of valley size

We tried two methods for estimating the valley size. The first method uses a new scaling relation which relates the density profile inside a valley to its size. This method gives a new insight on the structure of the sub-space in state space which constitutes a valley or a pure state, but is not useful because of the large computer resources it requires. The second method is a variant of the method used by Hartmann [36]. It uses an empirical relation between the valley size and its "diameter".

The first method for estimation of the valley size $|V|$ is based on a scaling relation which relates $|V|$ to the result of a local measurement, n_r. by local measurement we mean a procedure that starts at some point μ in state-space and wanders away, stopping at some finite distance from it.

Define $n_r(\mu)$ as the number of states in V that differ from a given state μ by less than r spin flips. We measure $n_r(\mu)$ for $K = \max\{100, |\Gamma_V|\}$ states in V and calculate the average $n_r = \sum_{\mu=1}^{K} n_r(\mu)/K$. We found the following empirical relation:

$$\frac{\ln n_r}{\ln |V|} = 1 - a \exp\left(-(b + c \ln N)\frac{r \ln r}{\ln |V|}\right) \tag{7.2}$$

with $a = 0.825$, $b = 0.492$ and $c = 0.124$. Fig. 7.1 shows the extent to which (7.2) holds.

When we encounter a valley V whose size we wish to measure, we first measure n_r and then fit it to the form of (7.2), using $|V|$ as a fit parameter. Estimating a specific valley size using this relation does not yield an exact result. We give here the standard error SE,

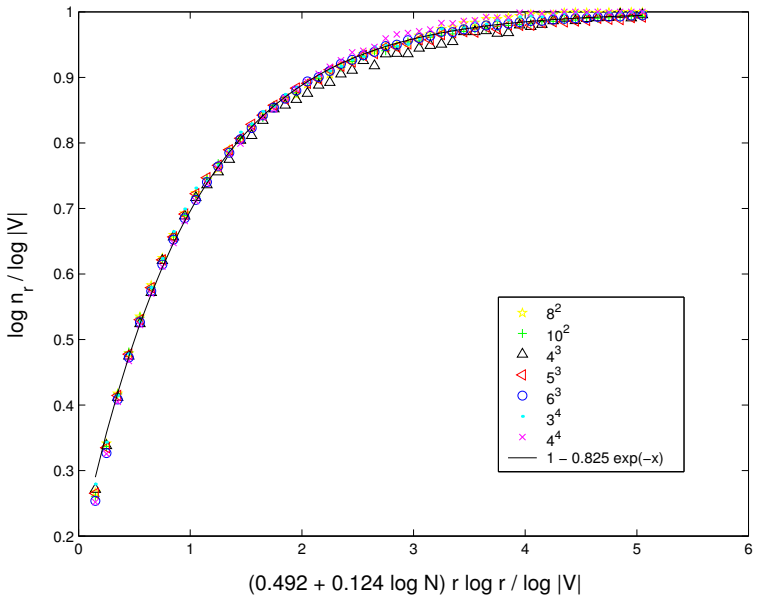

Figure 7.1: Scaling plot of n_r as a function of r. The plot is scaled by the valley size $|V|$ and the system size N. Each point is the average in a bin of size 0.1 on the x axis. For each system size and dimension data triplets of $\{r, n(r), |V|, \}$ (with $r = 3 \ldots 20$ for dimensions $D = 2, 3$ and $r = 3 \ldots 12$ for $D = 4$) were taken from all the valleys that were calculated using exact enumeration, which provided the exact value of $|V|$. The error bars are smaller then the size of the symbols. The solid line corresponds to $y = 1 - a \exp(x)$ with $a = 0.825$. The values for the other parameters are $b = 0.492$ and $c = 0.124$.

estimated by

$$\mathrm{SE}(L, D)^2 = \mathrm{avg}_V \left(\frac{|V| - s_V}{|V|} \right)^2 \tag{7.3}$$

where s_V is the estimated size of valley V. The average in (7.3) is over all the valleys V, with $|V| > 20$ in system of dimension D and linear size L for which we have exact enumeration of the states (see Table 7.1). We estimate the standard error using different values of r to obtain s_V. We have $\mathrm{SE}(6, 3) = 0.29$ for $r = 8$, 0.40 for $r = 6$ and 0.69 for $r = 4$ (estimated over 3934 valleys). For lower values of r we have a larger error. For other dimensions the error was larger: for $r = 4$ we have $\mathrm{SE}(10, 2) = 0.75$ (estimated over 6164 valleys) and $\mathrm{SE}(4, 4) = 0.98$ (estimated over 1700 valleys).

For large valleys in the regime $r \ll |V|$, n_r is exponential in r, and calculating n_r for $r > 4$ is computationally expensive, and inefficient (it is faster to sample by simulated tempering). For such valleys this method is not useful. It does, however, give us the density profile within a valley. If there had been a sub structure of clusters inside a valley (e.g. if a valley contained more than one pure state) we would have observed jumps in the density profile each time r had reached an inter-cluster distance. We do not observe such behavior. Thus we conclude that that there is no substructure inside a valley, and therefor all the

states inside a valley belong to one pure state.

The second method is a variant of the one used by Hartmann in [36]. It is based on an estimate $l_{\max}$ of the "diameter" of the valley. We start from a randomly chosen ground state μ in valley V, and flip iteratively randomly chosen free spins, without flipping a spin twice. We stop when all the spins that are currently free were already turned. The number spins we flipped is denoted by $l_{\max}(\mu)$. We repeat this procedure $K = 1000$ times for each valley, and average over the results to obtain $l_{\max} = (1/K) \sum_{\mu=1}^{K} l_{\max}(\mu)$. We estimate the valley size by

$$|V| = \exp(b \, {l_{\max}}^{\alpha}) \; , \tag{7.4}$$

where $b = 0.72$ and $\alpha = 0.96$ for $D = 3, 4$. For $D = 2$ we use $b = 0.75$ and $\alpha = 0.92$.

The standard errors, as defined by (7.3), when s_V is estimated using this method, are $\mathrm{SE}(6,3) = 0.51$, $\mathrm{SE}(10,2) = 0.45$ and $\mathrm{SE}(4,4) = 0.40$.

The approximation used is uncontrolled, and thus might lead into wrong results. We will demonstrate how that might happen in the next sub-section.

7.1.4. Uncontrolled approximations may yield false trivial overlap

Our method of approximation for the valley size, as well as other methods [30, 36] usually aimed to estimate the valley entropy $\log |V|$. In general such a method will yield an estimation of $\log |V|$ within a certain standard deviation $\sigma(L, D)$, which depends on the system dimension D and increases with the system size L. The estimated size s_V of a valley V is a random variable given by

$$s_V = \lambda_V |V| \; , \tag{7.5}$$

where λ_V is a random variable with variance Λ^2 which increases exponentially with $\sigma(L, D)$.

The overlap distribution is the sum of the overlap distribution within pure states $P^i(q)$ which converges to $\delta(|q| - q_{\mathrm{EA}}/2)$, and the overlap distribution between pure states $P^o(q)$ which is non-trivial, i.e. has a finite support for $|q| < q_{\mathrm{EA}}$. In case there are only two pure states $P^o(q)$ vanishes, and $P(q)$ is trivial. This happens if all spin domains $\mathcal{G}_a$ are fully correlated. We will show that if $\sigma(L, D)$ increases too fast with L the estimated correlation between each two domains will approach one, even if the true correlation is saturated at a smaller value.

We will demonstrate our point on the correlation $\bar{c}_{12}$ (defined in (4.6)), between domains $\mathcal{G}_1$ and $\mathcal{G}_2$, but the same considerations can be applied to any two spin domains. For clarity and simplicity, we will consider a simple scenario, in which $\bar{c}_{12} = 0$. In this case the number of states in $\mathcal{C}_1$, in which $\mathcal{G}_1$ and $\mathcal{G}_2$ are in configurations $\Uparrow_1 \Uparrow_2$ is equal to the number of states in $\mathcal{C}_2$, in which the two domains are in $\Uparrow_1 \Downarrow_2$.

We assume each of the clusters $\mathcal{C}_\alpha$ consists of K valleys, of equal size v. The estimate of $|\mathcal{C}_\alpha|$ based on the approximation is

$$\sum_{V \in \mathcal{C}_\alpha} s_V = v \sum_{V \in \mathcal{C}_\alpha} \lambda_V = \lambda_\alpha |C_\alpha| \; , \tag{7.6}$$

where $\lambda_\alpha = K^{-1}\sum_{V\in\mathcal{C}_\alpha}\lambda_V$ is a random variable with expectation value one (if our approximation is tuned correctly) and variance Λ^2/K.

The relation between $\bar{c}_{12}$ and the cluster sizes $|\mathcal{C}_\alpha|$, is given in (5.12). Using this relation we find that, under our approximation, the estimated value of $\bar{c}_{12}$ is

$$\tilde{c}_{12} = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 . \tag{7.7}$$

In order to calculate the expectation value of $\tilde{c}_{12}$ we define two new random variables

$$\lambda_+ = \lambda_1 + \lambda_2 \tag{7.8}$$

$$\lambda_- = \lambda_1 - \lambda_2 . \tag{7.9}$$

They both have a variance of $2\Lambda^2/K$. The expectation values are $\overline{\lambda_+} = 2$ and of $\overline{\lambda_-} = 0$. All moments of distribution of λ_- and λ_+ are the same, except the first one. Thus $\overline{{\lambda_+}^2} - \overline{\lambda_+}^2 = \overline{{\lambda_-}^2} - \overline{\lambda_-}^2$, and the expectation values of ${\lambda_+}^2$ and ${\lambda_-}^2$ maintain the relation

$$\overline{{\lambda_+}^2} = 4 + \overline{{\lambda_-}^2} . \tag{7.10}$$

In order to calculate the expectation value $\overline{{\lambda_-}^2}$ we need to have all the moments of distribution of λ_-. If all other moments but the second vanish, i.e. λ_- is normally distributed, then $\bar{\lambda_-}^2$ will be $2\Lambda^2/K$. This value serves as a lower bound for $\bar{\lambda_-}^2$. The expectations value of $1/{\lambda_+}^2$ is bounded from below by $1/\overline{{\lambda_+}^2}$ (since ${\lambda_+}^2 \geq 0$ equality is achieved only if all moments of distribution, except the first, vanish). The expectation value of $\tilde{c}_{12}$ is bounded from below by

$$\tilde{c}_{12} \simeq \overline{\left(\frac{{\lambda_-}^2}{{\lambda_+}^2}\right)} \geq \frac{\overline{{\lambda_-}^2}}{\overline{{\lambda_+}^2}} \geq 1 - \frac{2}{2 + \Lambda^2/K} . \tag{7.11}$$

If $\Lambda^2/2K \to \infty$ as $L \to \infty$, then $\tilde{c}_{12}$ will extrapolate to one, while $\bar{c}_{12}$ remains zero.

Let us examine the implication of this fact for the commonly studied $x_{1/2} = \int_{-1/2}^{1/2} P(q)dq$. As we will show in Section 7.3, the main contribution to $P(q)$ in the interval $0 \leq q < 1/2$ comes from the overlap of states if $\mathcal{G}_1$ is in the same configuration and $\mathcal{G}_2$ is flipped, or vice versa. The weight of the contribution of these state pairs to $P(q)$ is $(1 - \bar{c}_{12})/2$. Thus, when $\Lambda^2/K > 1$ is large enough, our estimate of $x_{1/2}$ will be $\tilde{x}_{1/2} \sim K/\Lambda^2$.

Let us assume that the standard deviation $\sigma(L, D)$ grows linearly with L, $\sigma(L, D) \sim aL$, so that $\Lambda^2 \sim \exp(2aL)$. We assume also that the number of valleys in the system is exponential in L, so $K \sim \exp(bL)$. If $b < 2a$ the approximation will yield $\tilde{x}_{1/2} \sim \exp((b - 2a)L)$ even if $x_{1/2}$ remains finite.

In [37] Hartmann studied the number of valleys as a function of the system size for dimensions $D = 2, 3, 4$. It not clear that this number does increase exponentially with L, and for $D = 4$ it seems more like a power law dependence. For this case we need $\sigma(L, D)$ to increase not faster than $\log(L)$. For $D = 2, 3$ exponential dependence does not look improbable, with $b \leq 2$. Thus, in order to have a reliable approximation we

must have $a < b/2$. Since our approximation is uncontrolled we have no means to assure $\sigma(L, D)$ satisfies this requirement. Thus, results that point to a trivial overlap using such an approximation method are unreliable.

Our results, presented in Section 7.2, show that $\bar{c}_{12}$ for $D = 2, 3$ extrapolates to a value smaller than one. This value should not be regarded as the correct value of $\bar{c}_{12}$ at the thermodynamic limit, but only as an upper bound to it. Our main conclusion, that $\bar{c}_{12} < 1$ at $L = \infty$ and therefor the domains $\mathcal{G}_1$ and $\mathcal{G}_2$ do not merge, still holds.

7.2. State hierarchy and spin domains

7.2.1. State hierarchy

The first step of our analysis is to cluster the states. The resulting dendrogram for a specific realization of 6^3 spins is presented Fig. 7.2(a). The state distance matrix, ordered according to the dendrogram, is presented in Fig. 7.2(b). The relative height of each node in the dendrogram and the distinct patterns in the distance matrix show that there is a clear hierarchical cluster structure of the ground states.

We present in Table 7.2 the average distance (3.2) inside each of the clusters $\mathcal{C}$, $\mathcal{C}_1$ and $\mathcal{C}_2$, and the width (3.3) of the distance distribution inside these clusters. The results look similar to those obtained for Gaussian spin glasses at $T = 0.2$. The average distance $D(\mathcal{C})$ within $\mathcal{C}$ is about 9 times larger then the average distance $1 - D(\mathcal{C})$ between $\mathcal{C}$ and $\bar{\mathcal{C}}$. We also see that the average distance $D(\mathcal{C}_1, \mathcal{C}_2)$ between $\mathcal{C}_1$ and $\mathcal{C}_2$ is more than twice (or three times for $D = 3, 4$) the average distance inside each of them.

We also see that the average distances inside clusters decrease as the dimension of the system increases. For $D = 3, 4$ these distances decrease as L increases, but $D = 2$ shows the opposite behavior.

The line in Table 7.2 marked '8 ST' presents the results for the 60 realizations of size 8^3 we sampled by simulated tempering. These results are unbiased, but since the number of realizations is small the results are highly noisy. We compare them with the results for the 878 realizations sampled by valley size approximation. Taking into account the large error expected because of the small number of realizations measured by simulated tempering, the results are in reasonable agreement.

7.2.2. Spin domains

We extracted the correlated domains using the procedure described in Chapter 4. The size distributions for the largest two domains $\mathcal{G}_1$ and $\mathcal{G}_2$ are plotted in Fig. 7.3. We see that for dimensions $D = 2, 3, 4$ the distribution seem to converge (over most of the domain) already for the small system sizes analyzed. The averages and the widths of these distributions, presented in Table 7.3, also point to a convergence of the distribution.

For dimension $D = 2$ the domain $\mathcal{G}_1$ consist of about 50% of the spins, and is noticeably smaller than for dimensions $D = 3, 4$, where it consist of about 75% of the spins. The second

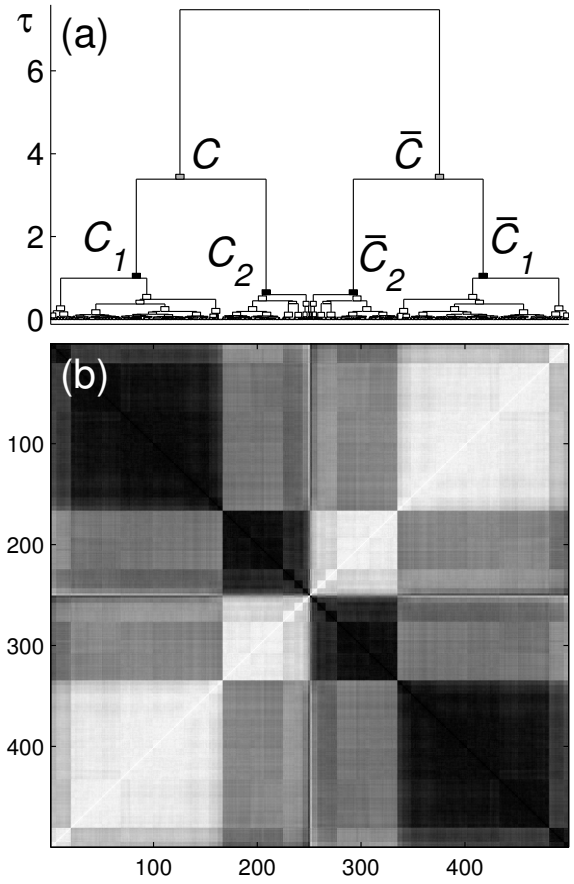

Figure 7.2: **(a)** The dendrogram obtained by clustering all the ground states, of a particular realization $\{J\}$, of size $N = 6^3$. **(b)** When the states are ordered according to the dendrogram, a clear block structure is seen in $D_{\mu\nu}$, the distance matrix of the ground states. Darker shades correspond to shorter distances.

domain is also smaller for $D = 2$ ($|\mathcal{G}_2| \sim 0.07N$) than for $D = 3, 4$ ($|\mathcal{G}_2| \sim 0.1N$). Thus, in two dimensional systems we expect higher entropy, since the number of free spins or small spin domains, which are left out of $\mathcal{G}_1$ and $\mathcal{G}_2$, is larger than in higher dimensions. Note that still we do not see any *qualitative* difference between $D = 2$ and $D = 3, 4$.

We have seen that the domains do not vanish. Now we have to assure they do not merge, i.e. that their correlation does not approach one. In table 7.3 we also present the results for the average correlation $\bar{c}_{12}$ (defined by (4.6)) between the domains $\mathcal{G}_1$ and $\mathcal{G}_2$.

For $D = 2$, $\bar{c}_{12}$ decreases as the system size L increases. For $D = 3$ it increases, and in order to determine if it approaches one we carried out fits of the form

$$\bar{c}_{12}(L) = \bar{c}_{12}(\infty) - AL^{-\phi} \tag{7.12}$$

with A and ϕ as fit parameters. The minimum of χ^2 is 1.7×10^{-4} for $\bar{c}_{12}(\infty) = 0.54(17)$ (with $A = 1.82(1.86)$ and $\phi = 1.20(87)$). For $\bar{c}_{12}(\infty) = 1$ we have $\chi^2 = 3.4 \times 10^{-4}$ ((with

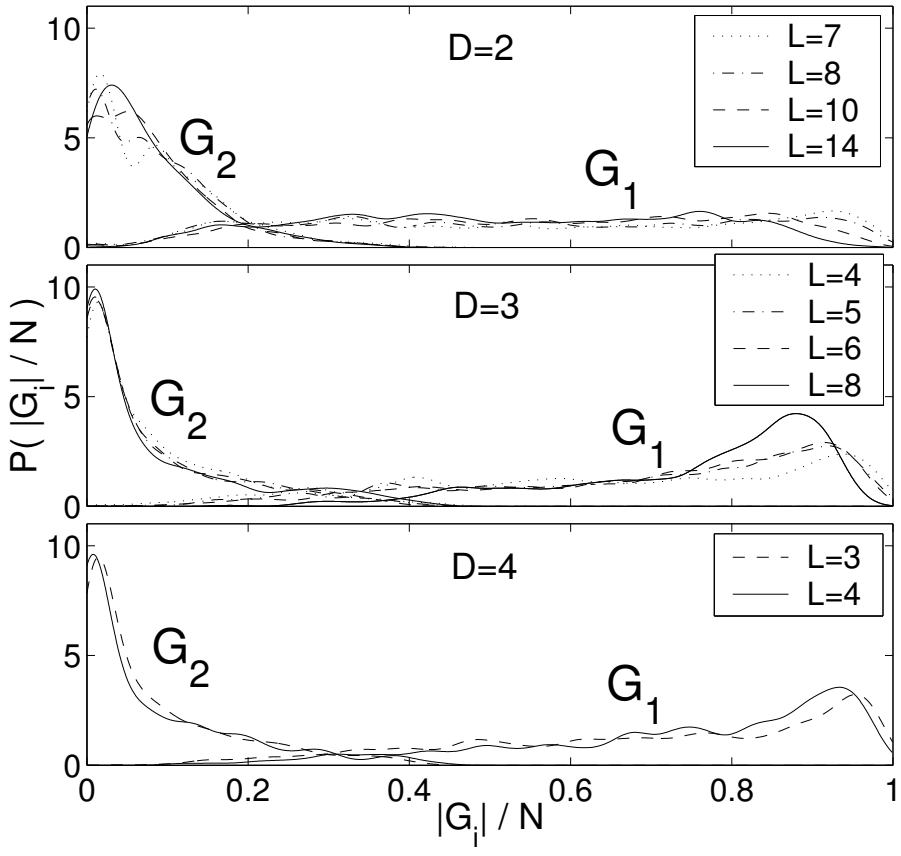

Figure 7.3: Normalized size distributions of the spin domains $\mathcal{G}_1$ and $\mathcal{G}_2$ for systems with different dimensions D and sizes L.

$A = 1.40(9)$ and $\phi = 0.40(4)$). Our result $\bar{c}_{12}(\infty) < 1$ is not conclusive and in itself does not give a convincing evidence for a non-trivial overlap. A much more convincing evidence will be given in Section 7.3.

For $D = 4$ we have only two data points, and it is useless to extrapolate.

When $\mathcal{G}_2$ vanishes we cannot determine if it is because its size is zero, or because $\bar{c}_{12} \approx 1$ and it merges with $\mathcal{G}_1$. Therefor we have calculated the results for $|\mathcal{G}_2|/N$ and $\bar{c}_{12}$ only for those realizations where $\mathcal{G}_2$ does not vanish. We report in the last column of Table 7.3 the probability $P(\mathcal{G}_2 \neq \emptyset)$ for $\mathcal{G}_2$ not to vanish.

The results obtained by simulating tempering are in agreement with the ones obtained by valley size approximation. In particular, the results for $|\mathcal{G}_2|/N$ and $\bar{c}_{12}$, by which we distinguish a trivial overlap distribution from a non-trivial one, are in reasonable agreement (particularly when taking into account the large statistical error of the simulated tempering results due to the small number of realizations sampled).

We have also measured the size of $\mathcal{G}_3$, the largest correlated spin domain which is unlocked over $\mathcal{C}_1$. The results are given in Table 7.4. We again give the results only for

realizations on which $\mathcal{G}_3$ does not vanish, and give in the right column of the table the probability $P(\mathcal{G}_3 \neq \emptyset)$. The 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 $\mathcal{G}_1 \cup \mathcal{G}_2$) increases but has small values. We cannot determine whether for large L it approaches one or a smaller value.

The size of $\mathcal{G}_3$ decrease with the system size. We carried out fits of the form

$$|\mathcal{G}_3(L)|/N = G_3(\infty) + AL^{-\phi} \tag{7.13}$$

with A and ϕ as fit parameters. The results are given in Table 7.5. For $D = 2$ the minimum of χ^2 is 1.0×10^{-3} for $G_3(\infty) = 0.43(8)$, and for $D = 3$ the minimum of χ^2 is 2.8×10^{-3} for $G_3(\infty) = 0.40(4)$. For both dimensions imposing $G_3(\infty) = 1$ yields in a much higher χ^2. For $D = 4$ we have only two data points and it is useless to extrapolate.

7.2.3. Spin space structure

After identifying the spin domains from the structure of the state space, we would like to see if this structure is apparent in spin space. As in Chapter 4 we cluster the spins, using the distance matrix (4.10) as an input to Ward's clustering algorithm. The resulting dendrogram for a specific realization is presented in Fig. 7.4(a). In Fig. 7.4(b) we present the spin distance matrix reordered according to the dendrogram.

Both Figs. 7.4(a) and 7.4(b) suggest there is a non-trivial structure in spin space, with large highly correlated spin clusters, seen as wide nodes with low τ values in 7.4(a) and indicated by the large dark squares in 7.4(b).

We turn now to identify those clusters g_1, g_2 in our spin dendrogram, which can be associated most naturally with the domains $\mathcal{G}_1$ and $\mathcal{G}_2$. For $a = 1, 2$ we find the cluster g_a in the spin dendrogram which is most similar to $\mathcal{G}_a$ using the similarity measure $\mathcal{S}$ defined in (4.11). The results in Table 7.6 show that for all dimensions when L is not too small the similarity is large. This suggest that the spins domains we identified have a meaningful physical role in spin space for binary spin glasses just as for spin glasses with Gaussian couplings.

7.3. Correct extrapolation of the overlap distribution

The overlap distribution $P_J(q)$ for a particular realization has, at $T = 0$, the typical form presented in Fig. 7.5. It has a large peak centered at some $q_0 \approx q_{\rm EA}$, and one or more smaller peaks. The largest peak is due to the overlap of pairs of states that belong to the *same* pure state. Denote the overlap distribution of such pairs by $P_J^i(q)$. The other peaks, at lower q, are due to the overlap between states that belong to two *different* pure states. The corresponding overlap distribution is $P_J^o(q)$ and we have

$$P_J(q) = P_J^i(q) + P_J^o(q) \qquad \text{and} \qquad P(q) = P^i(q) + P^o(q) \tag{7.14}$$

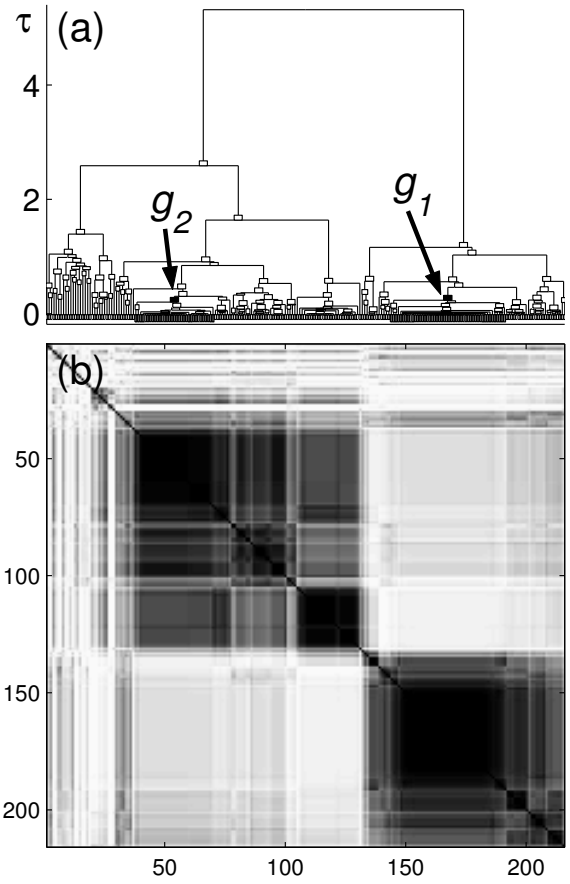

Figure 7.4: **(a)** The dendrogram obtained by clustering the spins of the system of Fig. 7.2. For this realization $g_a = \mathcal{G}_a$ for both $a = 1, 2$. **(b)** When the spins are ordered according to the dendrogram, a structure of correlated spin domains emerges; darker shades correspond to shorter distances and higher correlations.

where the second equation is the average of the first over all realizations. The observable x_a, defined in (7.1), can be written as the sum

$$x_a = x_a^i + x_a^o = 2\int_0^a P^i(q) + 2\int_0^a P^o(q) \tag{7.15}$$

Irrespectively of whether $P(q)$ is trivial or non-trivial, one expects that the width of $P^i(q)$ decreases with increasing size, since $P^i(q) \to \delta(|q| - q_{\rm EA})/2$ as $L \to \infty$. Therefore as L increases, the integral x_a^i decreases towards 0;

$$x_a^i \approx L^{-y_i} \tag{7.16}$$

On the other hand, the behavior of $P^o(q)$ (and x_a^o) *does* distinguish a trivial $P(q)$ from a non-trivial one; in the first case $P^o(q) \to 0$, while in the non-trivial case $P^o(q)$ and x_a^o

D	L	$D(\mathcal{C})$	$w(\mathcal{C})$	$D(\mathcal{C}_1)$	$w(\mathcal{C}_1)$
2	7	0.158±0.095	0.094±0.057	0.102±0.068	0.053±0.035
	8	0.162±0.094	0.093±0.055	0.108±0.067	0.053±0.034
	10	0.158±0.086	0.086±0.051	0.108±0.062	0.048±0.031
	14	0.168±0.080	0.086±0.051	0.118±0.055	0.047±0.029
3	4	0.118±0.081	0.078±0.057	0.070±0.051	0.039±0.031
	5	0.101±0.072	0.064±0.054	0.062±0.040	0.031±0.026
	6	0.094±0.067	0.060±0.054	0.059±0.037	0.028±0.026
	8	0.078±0.050	0.049±0.050	0.050±0.021	0.018±0.017
	8 ST	0.087±0.058	0.053±0.046	0.057±0.025	0.024±0.021
4	3	0.098±0.074	0.068±0.056	0.057±0.043	0.033±0.029
	4	0.081±0.063	0.056±0.054	0.048±0.033	0.023±0.025
		$D(\mathcal{C}_2)$	$w(\mathcal{C}_2)$	$D(\mathcal{C}_1,\mathcal{C}_2)$	$w(\mathcal{C}_1,\mathcal{C}_2)$
2	7	0.084±0.065	0.047±0.036	0.247±0.150	0.063±0.040
	8	0.093±0.067	0.049±0.037	0.250±0.146	0.064±0.039
	10	0.098±0.061	0.047±0.033	0.244±0.137	0.058±0.035
	14	0.113±0.058	0.048±0.032	0.259±0.133	0.057±0.033
3	4	0.057±0.053	0.036±0.035	0.204±0.147	0.043±0.031
	5	0.054±0.045	0.030±0.031	0.181±0.142	0.034±0.026
	6	0.056±0.044	0.030±0.033	0.178±0.143	0.029±0.023
	8	0.048±0.028	0.021±0.026	0.162±0.135	0.019±0.015
	8 ST	0.059±0.039	0.028±0.033	0.173±0.130	0.027±0.023
4	3	0.046±0.045	0.029±0.033	0.177±0.142	0.035±0.028
	4	0.044±0.037	0.025±0.032	0.164±0.142	0.023±0.019

Table 7.2: The average distances within and between the state clusters. The numbers are the averages over all realizations ± the standard deviation, i.e. $[x]_J \pm ([x^2]_J - [x]_J{}^2)^{1/2}$.

do not vanish as $L \to \infty$. We believe that previous analysis was hindered by the lack of ability to decompose $P(q)$ and $x_{1/2}$ into its two constituent parts; our method enables us to perform this task.

We calculate the distribution $\tilde{P}^o_J(q)$, defined in Chapter 5, to which only pairs of states $\mu \in \mathcal{C}_1$ and $\nu \in \mathcal{C}_2$ contribute. This function is a lower bound to $P^o_J(q)$, since we might have for some realizations a third macroscopic cluster, in which case $\mathcal{C}_1$ contains states from more than one pure state. When this happens, some pairs of states, both taken from $\mathcal{C}_1$, contribute to $P^o_J(q)$, and we do not include them in $\tilde{P}^o_J(q)$. As in Chapter 4, in order to assure that $\mathcal{C}_1$ and $\mathcal{C}_2$ indeed do not belong to one pure state, we consider only those realizations for which $|\mathcal{G}_2| > 0.05N$. Otherwise, we set $\tilde{P}^o_J(q) = 0$.

It is important to stress the fact that the main result of the present study, that there are states whose overlap contribution should be separated from the self-overlap peak and does not vanish in the thermodynamic limit, does *not* depend qualitatively on the way

D	L	$\lvert\mathcal{G}_1\rvert/N$	$\lvert\mathcal{G}_2\rvert/N$	$\bar{c}_{12}$	$P(\mathcal{G}_2 \neq \emptyset)$
2	7	0.55±0.28	0.095±0.078	0.82±0.19	0.871(3)
	8	0.54±0.27	0.093±0.072	0.80±0.21	0.846(4)
	10	0.55±0.24	0.097±0.069	0.77±0.23	0.796(6)
	14	0.51±0.23	0.088±0.070	0.74±0.24	0.876(4)
3	4	0.64±0.25	0.093±0.096	0.20±0.23	0.873(4)
	5	0.69±0.23	0.090±0.096	0.27±0.27	0.828(5)
	6	0.70±0.22	0.093±0.095	0.34±0.29	0.894(5)
	8	0.75±0.16	0.098±0.107	0.39±0.31	0.800(5)
	8 ST	0.73±0.18	0.090±0.087	0.40±0.32	0.78(2)
4	3	0.67±0.23	0.088±0.094	0.22±0.24	0.914(3)
	4	0.74±0.20	0.104±0.103	0.36±0.29	0.750(9)

Table 7.3: Normalized domain sizes of $\mathcal{G}_1$ and $\mathcal{G}_2$, and the correlation between them. All measurements were taken only when the domain concerned is not empty. The numbers are the averages over all realizations ± the standard deviation, i.e. $[x]_J \pm ([x^2]_J - [x]_J{}^2)^{1/2}$.

D	L	$\lvert\mathcal{G}_3\rvert/N$	$\bar{c}(\mathcal{G}_3, \mathcal{G}_1 \cup \mathcal{G}_2)$	$P(\mathcal{G}_3 \neq \emptyset)$
2	7	0.082±0.183	0.20±0.22	0.733(6)
	8	0.061±0.112	0.21±0.21	0.746(6)
	10	0.057±0.043	0.23±0.21	0.568(8)
	14	0.047±0.038	0.24±0.21	0.673(8)
3	4	0.079±0.195	0.22±0.23	0.783(5)
	5	0.042±0.089	0.24±0.23	0.637(8)
	6	0.045±0.069	0.29±0.25	0.563(8)
	8	0.038±0.067	0.34±0.27	0.520(8)
	8 ST	0.043±0.060	0.34±0.26	0.67(3)
4	3	0.070±0.182	0.22±0.24	0.828(5)
	4	0.045±0.062	0.34±0.25	0.46(1)

Table 7.4: Normalized domain size for $\mathcal{G}_3$, and its correlation with $\mathcal{G}_1 \cup \mathcal{G}_2$. All measurements were taken only when the domain measured is not empty. The numbers are the averages over all realizations ± the standard deviation, i.e. $[x]_J \pm ([x^2]_J - [x]_J{}^2)^{1/2}$.

the state clusters are determined. In fact, any method, which projects out a particular contribution to $P_J(q)$ and has a non vanishing weight in the $L \to \infty$ limit, will lead to the same conclusion. The only requirements are that the method is applied for all system sizes in the same way and the contribution is measured in absolute weights with respect to the total $P_J(q)$.

In the present case $\tilde{P}^o_J(q)$ also has a clear physical meaning. It is the distribution of

D	$G_3(\infty)$	A	ϕ	χ^2
2	0.43(8)	24.(44.)	3.0(1.4)	1.0×10^{-3}
	impose 0	2.5(6)	0.82(15)	3.7×10^{-3}
3	0.40(4)	$4.(43.) \times 10^5$	10.(8.)	2.8×10^{-3}
	impose 0	3.9(3.1)	1.2(5)	$26. \times 10^{-3}$

Table 7.5: Fit of the size of the domain $\mathcal{G}_3$ for dimensions $D = 2, 3$ to the form $x(L) = G_3(\infty) + AL^{-\phi}$, with $G_3(\infty)$ as a fit parameter and when we impose $G_3(\infty) = 0$.

D	L	$\mathcal{S}(g_1, \mathcal{G}_1)$	$\mathcal{S}(g_2, \mathcal{G}_2)$
2	7	0.98±0.09	0.85±0.22
	8	0.98±0.10	0.89±0.19
	10	0.98±0.09	0.97±0.09
	14	0.98±0.07	0.98±0.06
3	4	0.99±0.03	0.83±0.23
	5	0.99±0.02	0.88±0.19
	6	0.99±0.03	0.94±0.14
	8	0.99±0.02	0.97±0.06
	8 ST	0.99±0.02	0.96±0.07
4	3	0.99±0.02	0.84±0.23
	4	0.99±0.03	0.94±0.12

Table 7.6: The similarity $\mathcal{S}(g_a, \mathcal{G}_a)$ of the spin domain $\mathcal{G}_a$ and the cluster g_a most similar to it in the spins dendrogram, for $a = 1, 2$. All measurements were taken only when the domain concerned is not empty. The numbers are the averages over all realizations ± the standard deviation, i.e. $[x]_J \pm ([x^2]_J - [x]_J{}^2)^{1/2}$.

overlaps between pairs of states on the two sides of the second largest free energy barrier in the system.

We have shown in Sec. 7.2 that, since $\mathcal{G}_2$ remains macroscopic and $\bar{c}_{12}$ does not extrapolate to one, the overlap distribution at the $L \to \infty$ limit is not trivial. In order to demonstrate explicitly that this indeed is the case, we studied $\tilde{P}^o(q)$ and $P(q)$. The function $\tilde{P}^o(q) = [\tilde{P}^o_J(q)]_J$, presented in Fig. 7.6, is a conservative estimate, and a lower bound, for $P^o(q)$. For comparison, we also present the full $P(q)$ in Fig. 7.7.

Having obtained $\tilde{P}^o(q)$ we denife $\tilde{x}^o_a$, which serves as a lower bound on x^o_a, by

$$\tilde{x}^o_a = 2 \int_0^a \tilde{P}^o(q) dq \; . \tag{7.17}$$

We found that the rate of convergence of $\tilde{P}^o(q)$ to its limiting large-L form is non-uniform. For the sizes studied, convergence (with increasing L) is much slower in the

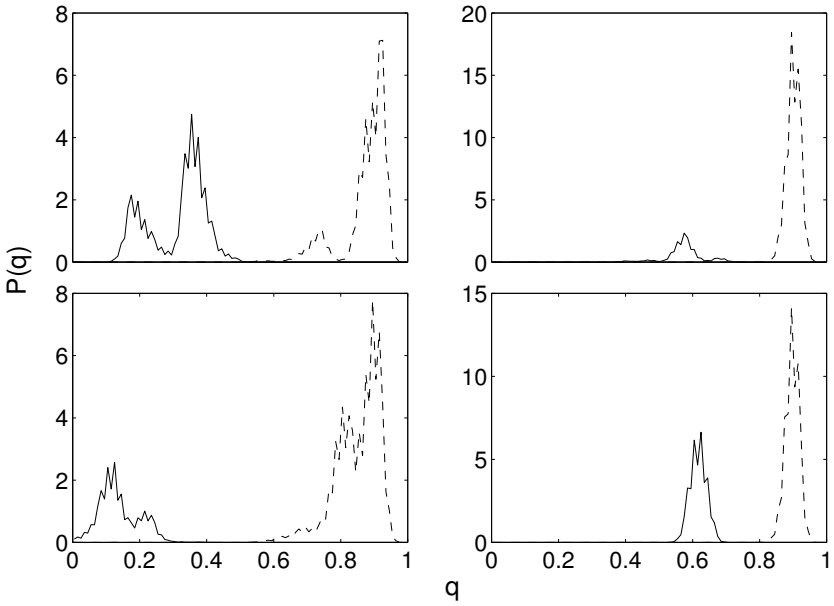

Figure 7.5: The state overlap distribution $P(q)$ for four different realizations $\{J\}$ for system size $L = 8$. Each distribution is divided to its components. The partial distribution $\tilde{P}^o_J(q)$ (see text), is represented by a solid line. The rest of the distribution, $P_J(q) - \tilde{P}^o_J(q)$, which includes $P^i_J(q)$, is represented by a dashed line.

interval $0 \leq q \leq 0.4$, where $\tilde{P}^o(q)$ has low values and large relative errors. On the other hand, in the interval $0.4 \leq q \leq 0.7$ it seems to have converged. Therefore we chose this range for our analysis, and calculated the integrals

$$x^* = 2 \int_{0.4}^{0.7} P(q) dq \, , \tag{7.18}$$

$$\tilde{x}^{*o} = 2 \int_{0.4}^{0.7} \tilde{P}^o(q) dq \, . \tag{7.19}$$

The values obtained for $L = 4, 5, 6, 8$ are presented in Table 7.7. Perhaps the most direct evidence for our claim is the manner in which the values of $\tilde{x}^{*o}$ level off as the size increases, at 0.047. On the other hand, those of x^* decrease with size. We performed a fit of the latter to the form

$$x^* = A + BL^{-y} \tag{7.20}$$

The results of several attempts to fit the data to this form are summarized in Table 7.8. The best fit (with $\chi^2 = 1.0 \times 10^{-5}$) was obtained for $y = 2.06(49)$ and $A = 0.042(15)$, which is close to 0.047. Imposing this value, i.e. setting $A = 0.047$ and fitting B and y, we had a somewhat larger $\chi^2 = 1.1 \times 10^{-5}$; imposing $A = 0$ yields a worse fit, with $\chi^2 = 4.2 \times 10^{-5}$. We believe that these results clearly show that $P(q)$ is non-trivial.

To make contact with previous analysis we also calculated $x_{1/2}$ and performed similar fits, the results of which are also presented in Table II. As discussed above, in this range

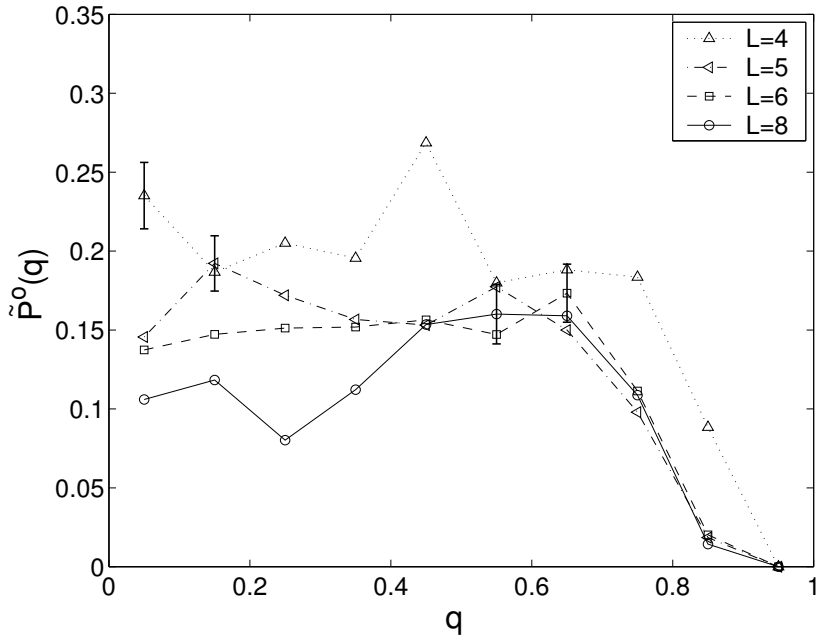

Figure 7.6: The partial distribution $\tilde{P}^o(q)$ for $L = 4, 5, 6, 8$. It is normalized so that $2\int_0^1 \tilde{P}^o(q)dq$ is its weight in the total $P(q)$. For each L the largest error bar is shown.

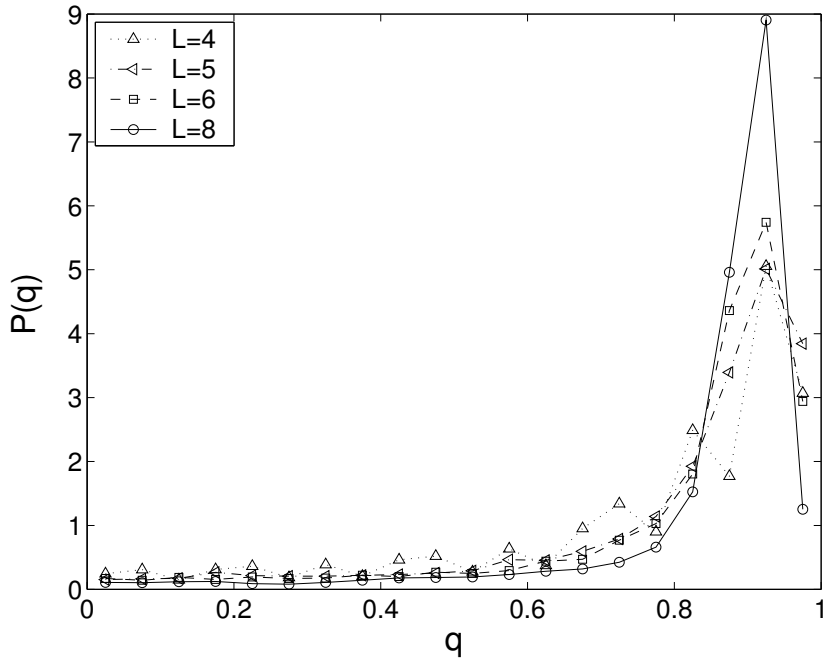

Figure 7.7: The distribution $P(q)$ for $L = 4, 5, 6, 8$. The size of the error bars is of the order or less than the size of the symbols.

L	x^*	$\tilde{x}^{*o}$	$x_{1/2}$	$\tilde{x}^o_{1/2}$
4	0.161(5)	0.064(10)	0.157(7)	0.109(14)
5	0.115(5)	0.048(11)	0.105(6)	0.082(15)
6	0.096(5)	0.048(10)	0.095(5)	0.074(14)
8	0.070(4)	0.047(12)	0.062(4)	0.057(14)

Table 7.7: Values of the observable (defined in (7.1), (7.17), (7.18) and (7.19)) for different system sizes.

of q the function $\tilde{P}^o(q)$ has larger statistical fluctuations, and is decreasing with size (to a limiting value that is expected to be small, albeit non-zero). Indeed the best fit for $x_{1/2}$ is attained for $y = 2.07(1.51)$ and $A = 0.036(47)$, with $\chi^2 = 1.0 \times 10^{-4}$. Note that this χ^2 is 10 times the value obtained when fitting x^*. Since our estimated value of A, as well as the estimates of others [21, 22] is much smaller then the values of $x_{1/2}$ used to perform the fit, it is hard to distinguish, by means of this extrapolation, between $A = 0$ and a small positive A. Indeed, when we impose $A = 0$ or both $A = 0, B = 1$ we get fits of comparable quality, with an exponent which is consistent with Hartmann's estimate.

Finally, we attempted to fit the data for $\tilde{x}^o_{1/2}$. The values of $\tilde{x}^o_{1/2}$ for the system sizes used are smaller and noisier the the results for x^*. Nevertheless, using the same fit for $\tilde{x}^o_{1/2}$ yields minimum of $\chi^2 = 1.5 \times 10^{-5}$ for $y = 1.94(1.02)$ and $A = 0.040(20)$, quite consistent with the results obtained for x^*.

fit	A	B	y	χ^2
x^*, best	0.042(15)	2.06(1.16)	2.06(49)	1.0×10^{-5}
x^*, imposed A	0.047	2.45(36)	2.21(10)	1.1×10^{-5}
x^*, imposed A	0	0.88(13)	1.24(9)	4.2×10^{-5}
$x_{1/2}$, best	0.036(47)	2.10(3.68)	2.07(1.51)	1.0×10^{-4}
$x_{1/2}$, imposed A	0	0.97(26)	1.33(17)	1.3×10^{-4}
$x_{1/2}$, imposed A, B	0	1	1.35(2)	1.3×10^{-4}
$\tilde{x}^o_{1/2}$, best	0.040(20)	1.00(1.16)	1.94(1.02)	1.5×10^{-5}

Table 7.8: Best fit parameters for $x(L) = A + BL^{-y}$.

8. Summary

We presented a new picture of the spin glass-phase in finite dimensional systems. This picture is consistent with recent numerical findings of a non-trivial overlap distribution [7,9,12] and macroscopic spin domains with microscopic surface tension [11]. It is consistent with the droplet picture [6], and inconsistent with the mean-field RSB picture [2].

In the spin glass phase the system consists of macroscopic spin domains of variable sizes. Each union of these domains is a core of contiguous spin clusters with microscopic surface tension. These domains correspond to free energy barriers of variable sizes. This variability give rise to hierarchical structure in state space. Each level in the hierarchy is associated with a domain: the first (highest) level is associated with the largest domain, the next level is associated with the second largest domain and so on. In each level the state clusters are partitioned with respect to free energy barriers associated with the corresponding domain.

For each level of the hierarchy, except the first one, there are different domains correlated over each state cluster, e.g. the domain $\mathcal{G}_3$ is correlated over $\mathcal{C}_1$ while a different domain $\mathcal{G}'_3$ is correlated over $\mathcal{C}_2$. $\mathcal{G}_3 \neq \mathcal{G}'_3$, although in general they 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 $L \to \infty$.

This hierarchical picture is *not* consistent with RSB. In the RSB scenario if we take clusters from a certain level of the hierarchy, e.g. $\mathcal{C}_1$ and $\mathcal{C}_2$, than the partial distance matrix $\tilde{D}_{ij}$ for $i \in \mathcal{C}_1$ and $j \in \mathcal{C}_2$ is uniform. We have concluded in Chapter 3 that width of the distribution $P(\tilde{D}_{ij})$of values in $\tilde{D}_{ij}$ does not vanish.

Furthermore, we presented evidence that the existence of a macroscopic $\mathcal{G}_3$, separating $\mathcal{C}_1$ into two pure states (or sets of pure states) leads to lack of ultrametricity.

In Chapters 5,6 and 7 we demonstrated how, by using our methods we can separate the state space into its components. This way we can examine only a chosen part of this space, thus obtaining more reliable numerical results and reducing finite size effects.

In particular, the power of our analysis is demonstrated in resolving the overlap distribution $P(q)$ of binary spin glasses at $T = 0$. Other papers extrapolated the total $P(q)$ and concluded that it becomes trivial in the thermodynamic limit. We found that a relevant part of $P(q)$ converges to a non-trivial distribution.

Clustering analysis can be applied to other systems with a non-trivial phase space structure, which have several pure states unrelated by any apparent symmetry, such as random field models or other models with random anisotropy [38]. It can help not only in the investigation of the macroscopic properties of the system, but also in understanding the micro-structure that give rise to these properties.

A. Why must SPC be a short-range algorithm?

SPC is a hierarchical clustering algorithm which works by mapping the data into a q-state Potts model (typically $q = 20$) with the Hamiltonian

$$\mathcal{H} = -\sum_{i<j} J_{ij}\delta_{\sigma_i\sigma_j} \ . \tag{A.1}$$

Each data point i is mapped onto a spin σ_i and the distances d_{ij} are translated into the interactions

$$J_{ij} = \exp(-{d_{ij}}^2/2a^2) \ , \tag{A.2}$$

where a is a characteristic distance which we will discuss later. The model is then simulated for different values of the temperature T. If for a certain T the correlation $c_{ij} = \langle\delta_{\sigma_i\sigma_j}\rangle$ is larger than 0.5 then the points i and j are in the same cluster.

At $T = 0$ the spins are fully correlated and they all form one cluster. At high T the correlation vanishes, and each spin is in its own cluster. There is an intermediate range in which groups of spins which are closer to each other, i.e. have stronger interactions, will be correlated among themselves, but they will not be correlated with each other. This is the *super-paramagnetic phase*, in which each correlated spin cluster is like a super-spin which is. in its turn, uncorrelated with the other super-spins. In this case we will have a non-trivial cluster structure and thus obtain a non-trivial partition of the data.

Since the simulation time of SPC increases linearly with the number of bonds, it is seemed reasonable to consider only the distances for the K nearest neighbors of each spin and omit the other bonds. The characteristic distance a is taken as the average of these distances. Since the interactions decrease exponentially with a, it indeed seems reasonable to discard bonds J_{ij} for which $d_{ij} \gg a$.

This method of nearest neighbors works well on a noisy data (such as biological [39] or image processing [40] data) where the clusters are connected via a noisy background of points scattered around them [41]. In the special case where there is no such background, and where the distances between points inside a cluster are much smaller than the distance between points of different clusters, as in the data for the states or spins of a spin-glass, all the nearest neighbors of a spin will be in the same cluster. The determination of the bonds will divide the data into the lowest level clusters, but the upper hierarchical structure will not be recovered by the SPC, since there will be no bonds between the clusters to give information on their organization.

For this reason we have to use SPC without discarding any bond. The distance a will have to be determined so that all stages in the hierarchy can be found by the SPC. This in itself constitutes a problem. The interactions J_{ij} are sensitive to the distance d_{ij} only if $d_{ij} \sim a$. If $d_{ij} \gg a$ then $J_{ij} \approx 0$ and if $d_{ij} \ll a$ then $J_{ij} \approx 1$. By setting the value of a we will determine which stages in the hierarchy will be found by the SPC.

Even if the dependence of the interactions on the distances were not exponential, we would have encountered another problem. In order for the algorithm to detect the cluster hierarchy we need the distances between clusters to be distinguishable by the SPC. On the other hand, if the bonds between the clusters are not very week, the super-paramagnetic phase will not appear. We will prove this claim for a simple mean-field model and than study its implications on realistic finite dimensional data.

A.1. Condition for the existence of a super-paramagnetic phase

We consider a q-state Potts model of MN spins, divided into M clusters of N spins each [41]. The state of the system is described by $\{\sigma_i^a\}$ - the values of spins $i = 1 \dots N$ in clusters $a = 1 \dots M$. The interaction between two spins of the same cluster is J_1/N, and the interaction between two spins of different clusters is J_2/MN. Both interactions were normalized so that their contribution to the energy per spin will remain finite in the $N, M \to \infty$ limit. The Hamiltonian of this system is

$$\mathcal{H} = -\frac{1}{N} J_1 \sum_{a=1}^{M} \sum_{i<j} \delta_{\sigma_i^a \sigma_j^a} - \frac{1}{MN} J_2 \sum_{a<b} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_{\sigma_i^a \sigma_j^b} \; . \tag{A.3}$$

We now define for each cluster a and spin state α the variable x_α^a, as the fraction of spins in cluster a that are in the state α,

$$x_\alpha^a = \frac{1}{N} \sum_i \delta_{\sigma_i^a \alpha} \; . \tag{A.4}$$

$\mathcal{H}$ can be written as

$$\mathcal{H} = -\frac{N}{2} J_1 \left(\sum_a \sum_\alpha {x_\alpha^a}^2 - \frac{M}{N} \right) - \frac{N}{M} J_2 \sum_{a<b} \sum_\alpha x_\alpha^a x_\alpha^b \; . \tag{A.5}$$

At the paramagnetic phase all the variables x_α^a are equal. In the low T phase the spins are correlated, and most spins in cluster a will have the same value, which we refer to as the *super-spin* of cluster a, and denote by $\hat{\sigma}_a$. In this phase we expect $x_{\hat{\sigma}_a}^a$ to be much higher than the other x_α^a.

Following the solution presented in [42], the free energy is minimized when x_α^a is of the form

$$\begin{aligned} x_{\hat{\sigma}_a}^a &= \frac{1}{q}[1 + (q-1)s] \; ; \\ x_i^a &= \frac{1}{q}(1-s), \quad i \neq \hat{\sigma}_a \; , \end{aligned} \tag{A.6}$$

where $0 \leq s \leq 1$ is the order parameter. High s yields strong correlation between the spins within a cluster. Substituting (A.6) into (A.5) we obtain the energy per spin

$$\begin{aligned}\tfrac{1}{MN}\mathcal{H} = &-\frac{1}{2}J_1\left(\frac{q-1}{q}s^2+\frac{1}{q}-\frac{1}{N}\right)-\frac{M-1}{2qM}J_2[(1-2q)s^2+(2q-2)s+1]\\ &-\frac{1}{M^2}J_2 s^2\sum_{a<b}\delta_{\hat{\sigma}_a\hat{\sigma}_b}\ .\end{aligned} \tag{A.7}$$

The first and second terms of (A.7) correspond to the interaction between the spins within and between clusters. The third term describes the interaction *between the super-spins*. It has the same form of the simple Potts Hamiltonian (A.1) with equal bonds. Thus, the state which minimizes the free-energy of the super-spins is identical to that of (A.6). In analogy to the variables x_α^a we now define $\hat{x}_\alpha$ as the fraction of super-spins in state α. The solution which minimizes the free energy is of the form

$$\begin{aligned}\hat{x}_1 &= \frac{1}{q}[1+(q-1)\hat{s}];\\ \hat{x}_i &= \frac{1}{q}(1-\hat{s}), \quad i\neq 1\ ,\end{aligned} \tag{A.8}$$

where $\hat{s}$ is the order parameter for the super-spins, in analogy to s. Thus

$$\sum_{a<b}\delta_{\hat{\sigma}_a\hat{\sigma}_b}=\sum_\alpha\frac{1}{2}M\hat{x}_\alpha(M\hat{x}_\alpha-1)=\frac{1}{2}M^2\left(\frac{q-1}{q}\hat{s}^2+\frac{1}{q}-\frac{1}{M}\right)\ . \tag{A.9}$$

We have now

$$\begin{aligned}\tfrac{1}{MN}\mathcal{H} = &-J_1\frac{q-1}{2q}s^2-\left(1-\frac{2}{M}+\frac{1}{2qM}\right)J_2\,s^2-\frac{M-1}{M}J_2\frac{q-1}{q}s\\ &-J_2\frac{q-1}{2q}s^2\hat{s}^2+\mathrm{C}\,,\end{aligned} \tag{A.10}$$

where C is a constant independent of s and $\hat{s}$. Removing the constant and taking the limit $M,N\to\infty$ we obtain

$$\frac{1}{MN}\mathcal{H}=-J_1\frac{q-1}{2q}s^2-J_2\,s^2-J_2\frac{q-1}{q}s-J_2\frac{q-1}{2q}s^2\hat{s}^2\ , \tag{A.11}$$

Note that when $s>0$ the contribution of the $\hat{s}^2$-term to the energy per spin (A.11) is finite in the thermodynamic limit.

The entropy per spin of the system is

$$\frac{1}{MN}\mathcal{S}=\mathcal{S}(s)+\frac{1}{N}S(\hat{s})\ , \tag{A.12}$$

where

$$\mathcal{S}(s) = -\frac{1+(q-1)s}{q}\ln\left(\frac{1+(q-1)s}{q}\right) - \frac{(q-1)(1-s)}{q}\ln\left(\frac{(q-1)(1-s)}{q}\right) . \tag{A.13}$$

Note that the contribution of the super-spins to the entropy per spin vanish as $N \to \infty$. The free energy per spin f is given as a function of s and $\hat{s}$ by

$$MN\beta f(s,\hat{s}) = \beta\mathcal{H} - \mathcal{S} . \tag{A.14}$$

For $s = 0$ there is no contribution of the super-spins to the energy, and the entropy will determine $\hat{s} = 0$. For $s > 0$ the contribution of the super-spin energy to f will be much larger then that of the entropy, and the minimizing the energy term will require $\hat{s} > 0$. Thus, as soon as the spins within the clusters are ordered the super-spins will be ordered. We will have no super-paramagnetic phase in which $s > 0$ and $\hat{s} = 0$.

We want to find the temperature T_1 of the phase transition $(s = 0) \to (s = s_c)$ and the temperature T_2 of the transition $(\hat{s} = 0) \to (\hat{s} = \hat{s}_c)$. It is clear that $T_1 \geq T_2$ since the energy term of $\hat{s}$ vanishes above T_1, and the entropy term of $\hat{s}$ is maximized when $\hat{s} = 0$.

The critical parameters T_1 and s_c are solved jointly from $\partial f/\partial s = 0$ and $f(0,0) = f(s_c,0)$. We will assume

$$\beta J_2 \ll \min\{\beta J_1, 1/q\} . \tag{A.15}$$

Then, up to a constant, we have

$$\beta f(s,0) = -\beta J_1 \frac{q-1}{2q} s^2 - \mathcal{S}(s) + \mathcal{O}[\beta J_2] . \tag{A.16}$$

Keeping the only the first two terms we have the mean field free energy of the Potts model [42], for which the transition parameters are

$$\beta_1 J_1 = K_q , \tag{A.17}$$

$$s_c = (q-2)/(q-1) , \tag{A.18}$$

where $\beta_1 = 1/T_1$ and $K_q = 2(q-1)\ln(q-1)/(q-2)$. Note that $\beta_1 = K_q/J_1$. Thus, for the above approximation (A.15) to hold it is enough that

$$J_2 \ll J_1/q . \tag{A.19}$$

We can find T_2 and $\hat{s}_c$ in the same manner. Since $T_1 \geq T_2$ we assume $s \geq s_c$. The free energy per spin is given by

$$\beta f(s,\hat{s}) = -\beta J_2 \frac{q-1}{2q} s^2 \hat{s}^2 - \frac{1}{N}\mathcal{S}(\hat{s}) + \bar{\mathrm{C}} , \tag{A.20}$$

where the term $\bar{C}$ is independent of $\hat{s}$. This free energy is again similar to the free energy of the simple Potts model. The transition parameters are

$$\beta_2 N J_2 s^2 = K_q \ , \tag{A.21}$$

$$\hat{s}_c = (q-2)/(q-1) \ . \tag{A.22}$$

The above parameters were deduced assuming $s \neq 0$. If we find $\beta_2 < \beta_1$ the transition of $\hat{s}$ will occur as soon as $s = s_c$, i.e. at $\beta = \beta_1$. In this case we have $\beta_2^{\rm real} = \beta_1$. This means that there is a single phase transition in the system, from a ferromagnetic to a paramagnetic phase. *There will be no super-paramagnetic phase.*

The condition for the existence of a super-paramagnetic phase is $\beta_2 > \beta_1$, which yields

$$J_1/J_2 > s^2 N \geq {s_c}^2 N \ . \tag{A.23}$$

We have already limited this calculation to the case where (A.19) holds. We see that even under this assumption J_2 is not weak enough to ensure $\beta_2 > \beta_1$. Higher values of J_2 will only cause β_2 to decrease as $1/J_2$, while β_1 will not change significantly.

A.2. Implication on a finite dimensional data

Now we can consider the limitations on the data we want to analyze with the SPC. We assume to have one level hierarchy. The characteristic distance between points inside the clusters is d_1, and between points outside the cluster is d_2. The corresponding interactions are defined by

$$J_i(d) = \exp(-{d_i}^2/2a^2), \quad i = 1, 2 \ . \tag{A.24}$$

If we want to have a super-paramagnetic phase, condition (A.23) must be fulfilled. In terms of the distances d_i the condition is

$$\exp\left(\frac{{d_2}^2 - {d_1}^2}{2a^2}\right) > {s_c}^2 N \ . \tag{A.25}$$

Assuming $s_c \approx 1$ and ${d_2}^2 \gg {d_1}^2$ we have the condition $d_2/a > \sqrt{\ln(N)}$.

A typical data is taken out of a d-dimensional space. Let $d_1 = 1$ be a characteristic small distance. If d_2 is a characteristic macroscopic distance we can estimate $d_2 \sim N^{1/d}$. The condition we have on a is

$$a < \frac{N^{1/d}}{\sqrt{\ln N}} \ . \tag{A.26}$$

If we want the SPC to consider all the distances in the data, a cannot be smaller then the small distance d_1, and it probably have to be larger if we want the SPC to consider also distances on the scale of d_2. So we must set $a \geq d_1 = 1$. This imposes that

$$N^{2/d} > \ln N \ . \tag{A.27}$$

d	minimal N
9	41,831
10	332,106
11	2,764,920
12	24,128,092
13	220,290,009
14	2,099,467,159

Table A.1: The minimal number of points N that fulfills condition (A.27) for a given dimension d.

In Table A.1 we give the minimal N values that fulfill the condition (A.27) for some values of the dimension d. We can see that N increases exponentially with d. For the data sets of states and spins of spin glasses we have $d > 100$, and the number of points N is far from satisfying (A.27).

B. Genetic cluster exact approximation

The cluster exact algorithm (CEA) [43] is an efficient way to generate ground states of a binary spin glass. The genetic CEA [33] is an enhancement of the CEA by a hybrid genetic optimization algorithm [44].

B.1. Cluster exact approximation

The CEA consists of the following stages. It finds in the system contiguous ferromagnetic domains, i.e. domains which have no frustration. It 'freezes' all the spins outside these domains. Since there is no frustration inside the domains the constrained system can be mapped to a ferromagnet with local fields. A ground state of this ferromagnetic system is found [45]. The resulting configuration is mapped back to the original spin glass system.

The 'ferromagnetic' domains are obtained in a greedy manner. The first spin in the domain is chosen at random. At each iteration a spin is chosen among the neighbors of the domain. If the inclusion of the spin in the domain leads to frustration inside the domain it is discarded and marked. If not, it is added to the domain. The algorithm will not try to join a marked spin again. When all the neighbors of all the spins in the domain are marked, a new, unmarked, spin is chosen as the seed of a new domain and the process starts again. The process stops when all the spins are either included in ferromagnetic domains or marked.

A gauge transformation is performed so that all the bonds connecting two spins in the same ferromagnetic domain will be positive, i.e. each domain is mapped to a ferromagnet. The marked spins are frozen. Each marked spin is now inducing a constant magnetic field on its neighbors. The current system is then a ferromagnet with local magnetic fields.

A ground state of a ferromagnet with local random fields can be found with a small computational effort [45]. The system is mapped into a graph. The vertices are the spins. Between each two neighboring spins i and j there is an edge whose capacity is equal to the interaction strength J_{ij}. Two edges are added to the graph, denoted b_+ and b_-. Let B_i be the local field on spin i. If $B_i > 0$ an edge with capacity B_i is added between b_+ and i, and if $B_i < 0$ an edge with capacity $-B_i$ is added between b_- and i.

In order to find the ground state of the ferromagnet we have to partition the graph into two components: one which includes b_+ and in which all the spins will be set to $+1$, and the other which includes b_- and in which all the spins will be set to -1. In general, in each such partition there have to be a group of unsatisfied bonds J_{ij}. In other words: we have to find the set of bonds with the minimal sum of capacities that when removed

from the graph there will be no path between b_- and b_+. This assignment is performed by the *min-cut* [46] algorithm. Then the spins in the two components are set accordingly. The new state of the ferromagnet is a ground state.

Note that the new configuration of the spins depends only on the choice of the ferromagnetic domains and on the values of the frozen spins. It does not depend on the previous values of the spins inside the domains. Thus, it is impossible to use this process in a way that maintains equilibrium. Indeed, this algorithm presents a bias [34], which we discuss below.

We inverse the gauge transformation to map the ferromagnet back to the spin glass system. Note that since the ferromagnet and spin glass are related by gauge transformation, and we have not increased the energy of the ferromagnet, the energy of the system have not increased.

B.2. The genetic algorithm

A genetic algorithm is an optimization algorithm based on the concept of Darwinian evolution. We start with a population of replicas of the same realization, each with a random configuration of the spins. In a hybrid algorithm [44] we generate two children from two members of the populations by combining their configuration, applying mutations, and than performing a local optimization procedure, such as CEA or steepest decent. We then compare each child with a parent and keep in the population only the fittest one, i.e. the replica with the smallest energy.

We start with a population of M_i replicas of the same realization. These replicas are ordered on a ring, so that each replica has two neighbors. In each iteration we choose at random two neighboring replicas. Then we create two children by a *triadic crossover*: first we create the children as copies of the parents; a mask is created from a distant (on the other side of the ring) state by reversing a fraction 0.1 of its spins; we select the spins which have the same values in the mask and in the first parent, and swap the values of those spins between the children. We apply a mutation to each child by choosing at random a fraction p_m of its spins and flipping them.

Next we perform a local optimization of each child by applying CEA on it n_{opt} times. We match a child-parent pairs, so that the sum of distances (3.1) in the pairs will be minimal. If the energy of the child is not higher than the energy of the matching parent, the child replace the parent in the population.

The above procedure is repeated $n_o \times M_i$ times. Then the population is cut by half: from each pair of neighbors we discard the one with the higher energy. We continue the process, setting $M_i \leftarrow M_i/2$ until $M_i = 4$. Then the replica with the lowest energy is taken as the result of the calculation.

B.3. Bias of the results

As we wrote above the probability to sample a certain ground state by the GCEA algorithm is biased [34]. Still, this bias does *not* affect our results. As we describe in Section 7.1, we use the results of the GCEA only to obtain representatives of each valley and then generate an unbiased sample by generating a sample in each valley and estimating the valley size.

The only way the bias of the GCEA could have undermined our results is if it had increased the probability to miss a certain valley. *This does not happen.* On the contrary, smaller valleys have higher probability to be sampled by GCEA then by an unbiased method such as Monte-Carlo [33]. The probability to sample a certain valley increases with its size, so we are in on danger of missing large valleys either.

C. Simulated tempering

For systems with a complex low energy landscape, such as spin glasses, we cannot use canonical Monte-Carlo methods to generate an unbiased sample at $T < T_c$. In order to sample all of the low energy valleys, i.e. all the regions in phase space with energy $E \lesssim k_{\rm B}T$, we have to go through the macroscopic free energy barriers separating those regions. The time it takes a single spin flip Monte-Carlo process to pass such a barrier grows exponentially with $\beta = 1/T$. Several methods [14, 15, 47, 48] were developed in order to avoid this problem. Simulated tempering [35] is one of these.

Simulated tempering is a Monte-Carlo method in which the temperature is a dynamic variable. The configuration space for this method is a Cartesian product of the state space and a discrete temperature space, with the inverse temperatures $\beta_1, \ldots, \beta_M$, i.e. each point in this space is of the form $\{\mathbf{S}^\mu, m\}$, where $\mathbf{S}^\mu$ is a spin configuration and m the temperature index. If $\beta_1 < \beta_c$ then when $m = 1$ the system is not confined by any (macroscopic) barriers in state space, and is free to move between valleys. The simulation does not go through the barrier, but over it. In order to obtain a sample at β we choose the set $\{\beta_m\}$ so that $\beta_M = \beta$ and take out of the generated sample $\{\mathbf{S}^\mu, m\}$ the sub-sample for $m = M$. The time it will take the simulation at $m = M$ to pass over a free energy barrier is just the time it takes it to go to $m = 1$ and back.

The weight of each configuration $\{\mathbf{S}^\mu, m\}$ is proportional to its Boltzmann weight at the corresponding temperature β_m:

$$P(\mathbf{S}^\mu, m) = \mathcal{Z}^{-1} \exp(-\beta_m \mathcal{H}[\mathbf{S}^\mu] + g_m) \ , \tag{C.1}$$

where g_m is an adjustable constant. The partition function for the phase space is $\mathcal{Z} = \sum \mathcal{Z}_m \exp(g_m)$, where $\mathcal{Z}_m$ is the canonical partition function at β_m. The probability to be at a certain m is given by

$$P_m \propto \mathcal{Z}_m \exp(g_m) = \exp(-\beta_m f_m + g_m) \ , \tag{C.2}$$

where f_m is the free energy corresponding to β_m. If we choose $g_m = \beta_m f_m$, the probabilities of all β_m are equal.

This method is implemented by combining a canonical (isothermal) Monte Carlo with a Metropolis dynamics in the $\{m\}$ space. After every canonical Monte-Carlo sweep over the system (at the current β_m) one Metropolis step which allows m to increase or decrease by one. We set $m' = m + \delta m$, with $\delta m = \pm 1$ with equal probabilities. Let $\mathbf{S}$ be the current

configuration of the system. The probability for m to change to m' is given by:

$$P[\,(\mathbf{S},m)\to(\mathbf{S},m')\,]=\begin{cases} P(\mathbf{S},m')/P(\mathbf{S},m) & P(\mathbf{S},m')<P(\mathbf{S},m) \\ 1 & P(\mathbf{S},m')\geq P(\mathbf{S},m) \\ 0 & m'<1 \ \text{ or } \ m'>M \end{cases} \tag{C.3}$$

C.1. Sampling ground states of a binary spin glass

We use simulated tempering to generate unbiased samples of the ground states of binary spin glasses. For a given realizations we generate samples of states for all β_m. For each m the weights of all ground states are equal, since their Boltzmann weights are equal. Thus, each ground state is sampled with equal probability, and the sub-sample of ground states is unbiased.

We identify the ground state using our previous knowledge of the ground state energy, obtained using GCEA [33] (see Appendix B). In order to assure that the samples are statistically independent, we label the ground states according to the time they were obtained: $\mathbf{S}^1$, $\mathbf{S}^2$, $\mathbf{S}^3$, ... Between the time $\mathbf{S}^t$ was sampled and the time $\mathbf{S}^{t+1}$ was sampled there might have been other states sampled, but they were not ground states. We then find a time τ_i which is longer than the decorrelation time of spin i, $\langle S_i^t S_i^{t+\tau_i}\rangle_t < e^{-1}$. We then choose the sampling period as $\tau=\max\{\tau_i\}$, and the final sample for the analysis is $\mathbf{S}^{\tau}$, $\mathbf{S}^{2\tau}$, $\mathbf{S}^{3\tau}$, ... For some systems with $N=8^3$ we have chosen the sampling period as $\tau/2$ instead of τ, in order to obtain a larger ground state sample. By doubling the sample size we did not double the information obtained by it, but it did supply us with more information, especially on the distribution within valleys.

The inverse temperatures used by the simulated tempering were $\beta_m=0.24+0.10m$, with $1\leq m\leq M=14$. Those values were used for all system sizes and dimensions. For dimensions $D=3$ ($T_c\approx 1.14J/k_{\rm B}$ [49]) and $D=4$ ($T_c\approx 2.0J/k_{\rm B}$ [50]) we indeed have $\beta_1<T_c^{-1}$ and $\beta_M\gtrsim 0.2T_c^{-1}$. For two dimensions $T_c=0$ [27, 50]), but still at β_M a large portion of the states sampled were ground states.

Each Monte-Carlo iteration included 20 Metropolis sweeps and one m-change Metropolis step. A state was sampled every k iterations, with $30\leq k\leq 300$. If it was a ground state it was recorded.

Since we have no previous knowledge on the free energy of each realization we have to obtain the variables g_m in an iterative way, trying to equalize the probabilities P_m. We start with a test set of the variables $\{g_m\}$ (usually taken from a previously analyzed realization). We run simulated tempering using this set, and evaluate the probabilities $P_m\{g_m\}$. We then create the next test set by

$$g'_m=g_m-\log(P_m/4+1)+\mathrm{C}\ , \tag{C.4}$$

where C is a constant tuned to have $\sum g'_m=0$. We repeat this process until $\mathrm{var}\{g_m\}\leq 0.05m^{-1}$.

Bibliography

[1] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35:1792–1796, 1975.

[2] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. A. Virasoro. Replica symmetry breaking and the nature of the spin glass phase. *J. Physique*, 45:843–854, 1984.

[3] M. A. Moore, H. Bokil, and B. Drossel. Evidence for the droplet picture of spin glasses. *Physical Review Letters*, 81:4252–4255, 1998.

[4] M. Palassini and A. P. Young. Triviality of the ground state structure in ising spin glasses. *Physical Review Letters*, 83:5126–5129, 1999.

[5] S. Franz and G. Parisi. Non trivial overlap distribution at zero temperature. cond-mat/0006188.

[6] D. S. Fisher and D. A. Huse. Equilibrium behavior of the spin-glass ordered phase. *Physical Review B*, 38:386–411, 1988.

[7] E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo. Phase structure of the three-dimensional adwards-enderson spin glass. *Physical Review B*, 58:852–863, 1998.

[8] J. Houdayer and O. C. Martin. A geometrical picture for finite dimensional spin glasses. *Europhysics Letters*, 49:794–800, 2000.

[9] S. Franz and F. Ricci-Tersenghi. Ultrametricity in three-dimensional edwards-anderson spin glasses. *Physical Review E*, 61:1121–1124, 2000.

[10] D. A. Huse and D. S. Fisher. Pure states in spin glasses. *J. Phys A*, 20:L997, 1987.

[11] F. Krzakala and O. C. Martin. Spin and link overlaps in 3-dimensional spin glasses. *Physical Review Letters*, 85:3013, 2000.

[12] H. G. Katzgraber, M. Palassini, and A. P. Young. Spin glasses at low temperatures. cond-mat/0007113.

[13] C. M. Newman and D. L. Stein. Realistic spin glasses below eight dimensions: A highly disordered view. *Physical Review E*, 63:16101, 2001.

[14] K. Hukushima and K. Nemoto. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65:1604, 1996.

[15] E. Marinari. Optimized monte carlo methods. In J. Kertész and I. Kondor, editors, *Advances in computer simulations*, page 50. Springer-Verlag, Berlin, 1998. (cond-mat/9612010).

[16] C. M. Newman and D. L. Stein. Spin-glass model with dimension-dependent ground state multiplicity. *Physical Review Letters*, 72:2286, 1994.

[17] S. Franz, G. Parisi, and M. A. Virasoro. The replica model on and off equilibrium. *Journal de Physique I*, 2:1869–1880, 1992.

[18] G. Hed, A. K. Hartmann, D. Stauffer, and E. Domany. Spin domains generate hierarchical ground state structure in $J = \pm 1$ spin glasses. cond-mat/0007356.

[19] F. Barahona, R. Maynard, R. Rammal, and J. P. Uhty. Morphology of ground-states of two-dimentional frustration model. *J. Phys A*, 15:673, 1982.

[20] M. Palassini and A. P. Young. The $\pm J$ spin glass: Effects of ground state degeneracy. cond-mat/0012161.

[21] A. K. Hartmann. How to evaluate ground-state landscapes of spin glasses thermodynamical correctly. *Eur. Phys. J. B*, 13:539, 2000.

[22] B. A. Berg, U. E. Hansmann, and T. Celik. Ground-state properties of the three-dimensional ising spin glass. *Physical Review B*, 50:16444, 1994.

[23] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data.* Prentice–Hall, Englewood Cliffs, 1988.

[24] M. Blatt, S. Wiseman, and E. Domany. Data clustering using a model granular magnet. *Neural Computation*, 9:1805–1842, 1997.

[25] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci.*, 96:6745–6750, 1999. in fact we used a variant of this method, enssuring that the dendrogram is symmetric.

[26] R. D. Reed and R. J. Marks. *Neural Smithing*, page 299. The MIT Press, 1999.

[27] R. N. Bhatt and A. P. Young. Numerical studies of ising spin glasses in two, three, and four dimensions. *Physical Review B*, 37:5606, 1988.

[28] G. Hed, A. K. Hartmann, and E. Domany. Correct extrapolation of overlap distribution in spin glasses. cond-mat/0012451.

[29] A. K. Hartmann. Are ground states of 3d $\pm J$ spin glasses ultrametric? *Europhysics Letters*, 44:249, 1998.

[30] S. Kirkpatrick. Frustration and ground-state degeneracy in spin glasses. *Physical Review B*, 16:4630–4641, 1977.

[31] N. Hatano and J. E. Gubernatis. The $\pm J$ spin glass: Effects of ground state degeneracy. cond-mat/0008115; see, however, E. Marinari, G. Parisi, F. Ricci-Tersenghi and F. Zuliani, cond-mat/0011039.

[32] F. Krzakala and O. C. Martin. Discrete energy landscapes and replica symmetry breaking at zero temperature. cond-mat/0010010.

[33] A. K. Hartmann. Analysis of the statistical behavior of genetic cluster-exact approximation. *Physica A*, 275:1, 1999.

[34] A. W. Sandvik. Comment on: "evidence for nontrivial ground-state structure of 3d $\pm J$ spin glasses". *Europhysics Letters*, 45:745, 1999.

[35] E. Marinari and G. Parisi. Simulated tempering: a new montecarlo scheme. *Europhysics Letters B*, 19:451, 1992.

[36] A. K. Hartmann. A new method for analyzing ground-state landscapes: ballistic search. *J. Phys. A*, 33:657, 2000.

[37] A. K. Hartmann. Ground-state clusters of two, three and four-dimensional $\pm J$ ising spin glasses. *Physical Review E*, 63:16106, 2001.

[38] R. Fisch. Critical behavior of randomly pinned spin-density waves. *Physical Review B*, 51:11507–11514, 1995.

[39] G. Getz, E. Levin, E. Domany, and M. Q. Zhang. Super-paramagnetic clustering of yeast gene expression profiles. *Physica A*, 279:57, 2000.

[40] E. Domany, M. Blatt, Y. Gdalyahu, and D. Weinshall. Super-paramagnetic clustering of data: application to computer vision. *Comp. Phys. Comm.*, 121-122:5, 1999.

[41] S. Wiseman, M. Blatt, and E. Domany. Super–paramagnetic clustering of data. *Physical Review E*, 57:3767–3783, 1998.

[42] F. Y. Wu. The potts model. *Reviews of modern physics*, 54:235–268, 1982.

[43] A. K. Hartmann. Cluster-exact approximation of spin glass ground states. *Physica A*, 224:480, 1996.

[44] K. F. Pál. The ground state energy of the edwards-anderson ising spin glass with a hybrid genetic algorithm. *Physica A*, 223:283–292, 1996.

[45] A. K. Hartmann and K. D. Usadel. Exact determination of all ground states of random field systems in polynomial time. *Physica A*, 214:141–152, 1995.

[46] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1994.

[47] R. H. Swendsen and J. Wang. Replica monte carlo simulation of spin-glasses. *Physical Review Letters*, 57:2607, 1986.

[48] B. A. Berg and T. Neuhaus. Multicanonical algorithms for first order phase transitions. *Physics Letters B*, 267:9, 1991.

[49] H. G. Ballesteros, A. Cruz, L. A. Fernández, V. Martín-Mayor, J. Pech, J. J. Ruiz-Lorenzo, A. Taracón, P. Télez, C. L. Ullod, and C. Ungil. Critical behavior of the three-dimensional ising spin glass. cond-mat/0006211.

[50] R. R. P. Singh and S. Chakravarty. Critical behavior of an ising spin-glass. *Physical Review Letters*, 57:245, 1986.