Collective degrees of freedom and multiscale dynamics in spin glasses

N. Persky* and S. Solomon†
Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel
(Received 1 March, 1996; revised manuscript received 11 June 1996)

We relate the long-range, long-time correlations during the simulation of disordered complex systems to the relevant macroscopic effective collective degrees of freedom. We prove that in systems that have an ultrametric space of ground states, the tunneling between vacuums cannot be expressed in terms of spatially disjoint clusters or in terms of spatial multiscale hierarchies. We relate this to the ultraslow convergence difficulties of multiscale-cluster algorithms in such systems. On the contrary, in the case of finite connectivity (dilute) spin glasses, we are able to find multiscale-cluster algorithms that are much more efficient than the usual methods. We relate their efficiency explicitly to their action on specific collective degrees of freedom. These degrees of freedom are responsible for the slowing down of the usual algorithms. [S1063-651X(96)07510-1]

PACS number(s): 02.70.−c, 0.5.45.+b

I. COMPLEX SYSTEMS AND MACROS

One of the main characteristics of complex systems is their computational difficulty: the time necessary for their investigation and/or simulation grows very fast with their size [1]. The systematic classification of the difficulty and complexity of computational tasks is a classical problem in computer science [2].

In physical systems, the emergence of long-time scales is often related to multiscale spatial structures within the system. Long-range and long-time scale hierarchies (multiscale slowing down) are usually related to collective degrees of freedom (which we call herein *macros*) characterizing the effective dynamics at each scale.

The physical understanding of complex macroscopic phenomena is then often expressed through identifying the relevant macros and their effective dynamics (e.g., hadrons in the theory of quarks and gluons, Cooper pairs in superconductors, phonons in crystals, vortices in superfluids, flux tubes, and instantons, solitons, and monopoles in gauge theories). One can entertain the hope that many complex systems display some kind of universal multiscaling exponents generalizing the scaling critical exponents of the critical systems. One could hope for the existence of some kind of multiscale-universality classes generalizing the universality classes of renormalization-group theory. Such a situation would have a significant unifying effect on a very wide range of phenomena spreading over most of the contemporary scientific fields.

In the absence of a rigorous theoretical basis for such a hope, its investigation relies, for the moment, mainly on the use of computers. In particular one uses ''first-principles'' simulations, which implement directly and without the intermediary of *ad hoc* approximations the fundamental physics of the systems under study.

Usually, it is the dynamics of the macros during simulations that produces the multiscale slowing down and, reciprocally, the slow modes of the simulation dynamics project out the relevant macros [3]. Therefore, a better theoretical understanding of the multiscale structure of the system enables one to construct better algorithms by acting directly on the relevant macros. Reciprocally, understanding the success of a certain algorithm yields a deeper knowledge of the relevant degrees of freedom of the system (see, for example, the projection by a parallel transported multigrid of exact lattice Atyiah-Singer modes [4].

The present paper implements this point of view into the study of spin glasses. Section II introduces the basic notions of multiscale-cluster algorithms (MCA's). Section III describes the difficulties in applying MCA's to generic frustrated systems. Section IV contains rigorous results that forbid macros in ultrametric systems. Section V identifies the relevant macros and their role in constructing MCA's for dilute spin glasses. Section VI demonstrates numerically the efficiency of the resulting MCA. Section VII summarizes the conclusions. The Appendix contains the proofs of the results stated in Sec. IV. We interpret the negative results in Secs. III and IV and the positive results in Secs. V and VI as supporting, in both directions, the relation between macros and the efficiency of MCA's.

II. MULTISCALE-CLUSTER ALGORITHMS

An example of multiscale effective dynamics and its related multiscale slowing down is the critical slowing down at second-order phase transitions. There the relaxation time τ diverges with the systems size L as

$$\tau \sim L^z,$$

where z (~ 2) is the dynamical critical exponent. Consequently, the typical time needed to produce a large Boltzmann set of decorrelated configurations diverges and the standard local Monte Carlo methods become inefficient.

It was shown that when the detailed knowledge on the relevant macros is included in the simulation algorithms, the value of z can be reduced dramatically (down to 0) [5]. These algorithms, which we will call generically here multiscale-cluster algorithms, allow the very fast and precise computation of the equilibrium thermodynamic properties of

*Electronic address: nathanp@vms.huji.ac.il
†Electronic address: sorin@vms.huji.ac.il, http://shum.huji.ac.il/usorin

the systems. However, their main importance is to guide and validate by objective means (lowering of z) the intuitive guesses on the physically relevant macros and their macroscopic dynamics [3].

We are treating the various MCA's in a conceptually unified way: as expressions of the macros appearing at various scales. In fact, many of the explanations in Sec. V on the dynamical relevance of the spatial structures manipulated directly by the macros reduction algorithm can be given equally in the language of the dynamical algebraic multigrid [6] as well as in the framework of the cluster algorithm (CA) [7].

In statistical mechanics systems, the objective of MCA's is to generate as fast as possible a representative sample of configurations. This is realized by acting directly on the macroscopically relevant macros (in contrast to the usual local algorithms, which act on the microscopic elementary degrees of freedom).

The typical CA works according the following general principles.

(i) One selects a particular subset of allowed changes for the degrees of freedom associated with each site $i=1,\ldots,N$ of the system. [For example, in finite-temperature SU(2) gauge theory, the SU(2) matrix degree of freedom on the time-like links is allowed to change only its sign during a MCA step [8]. This is an algorithmic expression of the physical understanding that it is the center of the group that is the relevant degree of freedom.] This reduces the system to an Ising-like system $S=\{s_1,\ldots,s_N\}$, where the Ising variables s_k can take the values ± 1. A configuration is a specific assignment of one of these values for each s_k. The interaction energy

$$E(S)=\frac{1}{2}\sum_{i,j} J_{i,j}(1-s_i s_j)$$

is parametrized by the *link* parameters $J_{i,j}$ associated with each pair of sites i and j. (For notational convenience, we use in this and the following section a definition of the total energy that differs by an overall additional constant $E_0=\frac{1}{2}\Sigma_{i,j}J_{i,j}$ from the definition used in the rest of the paper.)

(ii) One constructs and updates a system of clusters that preserves the macroscopic dynamical properties of the initial spin system. The system of clusters and its dynamics is obtained by modifying the (link) parameters $J_{i,j}$ between the pairs of spins (i,j). More precisely the link (i,j) is either ''frozen'' $J_{i,j}=\infty$ or ''deleted'' $J_{i,j}=0$ based on the following classifications.

Consider the current values s_i^C and s_j^C of the two spins and their current energy $E(S_{i,j}^C)=\frac{1}{2}J_{i,j}(1-s_i^C s_j^C)$. (a) If the spins are in the low-energy state $E(S_{i,j}^C)=0$ the link (i,j) is called saturated or *satisfied*. Otherwise, the link is said *unsatisfied*. (b) If the difference between the satisfied and unsatisfied energies of the link $|J_{i,j}|$ is large, the link is called *strong*. Otherwise it is *weak*. Often the strong-weak label is given relative to the actual temperature of the system. For instance, at very high temperature $(T\gg|J_{i,j}^C|)$ all the links can be considered weak, while at low temperature $(T\ll|J_{i,j}^C)|)$ most of the links may be acting as strong. With this terminology, the cluster generating procedure is (a) freeze (with high probability) the strong satisfied links (links with low energy), (b) delete the strong unsatisfied links (links with high energy), (c) give for weak links an appropriate stochastic chance to both options (frozen or deletion) to arise, and (d) flip the relative signs of spins that belong (by the link deletions) to different clusters.

In the following we will call *loop* a closed chain of links

$$\{(i_k,i_{k+1})|k=1,\ldots,n;i_{n+1}\equiv i_1\}.$$

If the product

$$\prod_{k=1,n} J_{i_k,i_{k+1}}<0$$

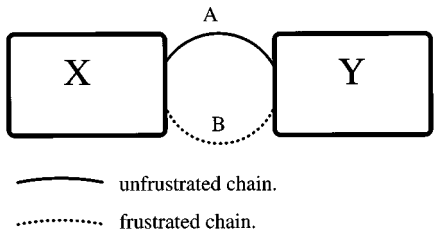
is negative the loop is said to be *frustrated*. If a loop is frustrated there exists no spin configuration for which all the links of the loop are satisfied.

III. MCA DIFFICULTIES IN FRUSTRATED SYSTEMS

The problem of the applicability of MCA to frustrated systems arises quite early because most of the cases in which the MCA *did not work* were situations in which the first step of Sec. II reduced the system to a frustrated one [9,3].

Some of the most important families of frustrated systems are the randomly frustrated systems such as spin glasses (SG's). A typical SG system presents a complex energy landscape consisting of many local minima, separated by huge barriers that scale with the size of the system. This is expressed by the emergence of an ultrametric structure of the ground-state space and an infinite hierarchy of exponentially divergent relaxation times [10].

To understand the difficulties that occur when applying MCA's to frustrated systems let us analyze in detail a simple scenario. Suppose that the first configuration (C1) in the figure



is an energy ground state (GS). Consider the case that the subsystems X and Y are linked by two chains of links A and B (as shown in the figure).

Suppose that in C1 the total energy of the links belonging to A is $E_A^1=0$, while the energy of chain B is $E_B^1=e$. Chain B is therefore unsatisfied, while chain A is satisfied. For the simplicity of the argument let as assume that the system is at very low temperature $T\ll e$, though the conclusion is valid at higher temperatures $(T<e)$ as well.

Assume now that a second GS configuration C2 is obtained by flipping in C1 the relative sign of the spins belonging to the regions X and Y such that $E_B^2=0$ and $E_A^2=E_B^1=e$. This means that in the ground-state C2 it is chain A that is satisfied, while B is unsatisfied. The total energy of C2 is equal to the total energy of C1. An efficient

algorithm should allow one to easily obtain C2 from C1 and vice versa.

However, with the usual cluster algorithm, this condition is not fulfilled. Indeed, if chain A is unsatisfied (as it happens in C2) so that the chain eventually would be cut by the CA, then automatically B is satisfied and all its links will be eventually frozen by the CA. Conversely, if B is cut, A is uncut (this happens in C1). Consequently, in any case, at least one of the chains A and B is frozen. Therefore X and Y are always included by the CA in the same cluster and there is no way to get from C1 to C2.

Therefore it appears that a frustrated disordered system is not compatible with an efficient updating algorithm. This argument becomes even stronger if X and Y are linked by several chains. We will present in Sec. IV some rigorous results extending these intuitions to a wider class of systems.

In a few special cases one can overcome these problems. If the problem is local and the structure of the links is completely known, then a ''two-bond deletion'' might help, such as in the fully frustrated system on a square lattice [11]. The two-bond deletion technique can be extended in other systems to three- and four-bond deletions [12] and even to an n-bond procedure. However, in general, the problem would still revert to exponentially combinatorial complexity if one has no *a priori* knowledge about which ''special'' subset of links one should combine.

The simulated annealing technique, which helped some systems from getting stuck in local minima, failed to provide a complete solution in the SG case. A related direction is the Swenden-Wang (SW) replica algorithm [7] and its modifications [13], which use simultaneously various replica of the system in order to identify large spatial regions that act coherently. This might work in a few simple cases, in low dimension (until now only two dimensions). However, one cannot expect such an algorithm to work for a general frustrated system because (i) in general (e.g., for spin glasses [14]) one can find for each GS exponentially many metastable states with energy close to the ground states. As a result, one needs in general an exponential number of replica in order to fully capture the structure of the system. (ii) The spatial regions that have to flip in order to turn various ground states into one another cannot in general be identified and manipulated as independent entities (see in Sec. IV the relevant theorems for ultrametric systems). These limitations indicate that one cannot get away from the combinatorial complexity of the general randomly frustrated problem. The CA logic of constructing clusters is based on *local* (see, e.g., [15]) features, implying local criteria such as deleting the unsatisfied links which define the block boundaries in the ferromagnet case. However, in a general case the feature of the cluster is only apparent in a *global* view, without any local signs (see Theorem 3).

The situation can be compared with having to find one's way in a labyrinth in phase space: each small local change in the position of the potential-energy labyrinth walls determines large unpredictable changes of the solution route depending on details scattered across the entire phase space. Consequently, we are discerning three main complexity cases.

(i) In very simple cases the pattern of complexity is reducible (maybe by an iterative multiscale procedure) and the MCA's capturing this reducible complexity are an efficient computational and conceptual tool.

(ii) In the general case one has to put an exponential computational effort to fully ''understand'' the structure of the system. This situation is similar to understanding the architecture of a labyrinth and is expressed by the theorems of Sec. IV.

(iii) In some cases the system contains certain macros that are ''irreducibly complex.'' Yet the interactions between these macros are tractable by MCA's or other algorithms. In these cases, MCA's can help reduce the ''less complex'' part of the dynamics leaving the ''irreducible cores'' for a separate treatment.

The last possibility has been exploited in the parallel transported multigrid [16,4] treatment of the fermions in gauge field background where the complexity related to the gauge freedom was eliminated at the multigrid level, while for macros related to frustration and topology (e.g., Atyiah-Singer zero modes) one has developed a method [3] for implicit identification, manipulation, and elimination of irreducibly complex macros. Similar intuitions are at the basis of the successful algorithms for diluted spin glasses described in Secs. V and VI. Recognizing the irreducibly complex parts of a complex system (rather than trying vainly to solve them by multiscale means) might be a very important aspect both conceptually and computationally.

IV. ULTRAMETRIC SYSTEMS DO NOT HAVE INDEPENDENT MACROS

As explained above, the SG systems present a certain hierarchy in their energy landscape that is responsible for the hierarchy of time scales characterizing their multiscale slowing down. This rugged energy landscape is also the origin of the ultrametric properties of their ground-state space [25].

One could hope to make some relation between the presence of an ultrametric hierarchy and the existence of an effective representation of the dynamics in terms of a spatial multiscale hierarchy of independent macros. This in turn would become the basis of an efficient MCA.

It turns out that the case is exactly the opposite: the ultrametric hierarchy characterizing SG's *ensures* the *nonexistence* of a representation of the effective macroscopic dynamics of the complex system in terms of their macroscopic disjoint subsets (i.e., a complex ultrametric system is not effectively reducible to a set of subsystems).

Let us express this in a more precise and rigorous way. In the Appendix we present some more details and the proofs of the theorems. We will consider Ising-like systems consisting of spins $s_k = \pm 1$:

$$\Omega = \{s_k | k = 1, \ldots, N\}, \tag{1}$$

where N is the arbitrary size of the system, which in the thermodynamic limit is taken to be infinity. A configuration is a specific assignment of the spins, e.g., $\Omega^D = \{s_1^D, \ldots, s_N^D\}$.

The system has an Ising-like bilinear Hamiltonian that defines an extensive energy. For example, in Sherrington-Kirkpatrick [17] -like model

$$E(\Omega^D)\equiv\langle\Omega^D,\Omega^D\rangle=-\frac{1}{2\sqrt{N}}\sum_{i,j}J_{i,j}s_i^Ds_j^D. \tag{2}$$

The metric in the configurations space is defined by the following distance. If two configurations Ω^0 and $\Omega^{\mathcal{A}}$ differ only by the sign of the spins belonging to a subset $\mathcal{A}$, then their distance is

$$d(\Omega^0,\Omega^{\mathcal{A}})=\frac{1}{4N}\sum_i(s_i^0-s_i^{\mathcal{A}})^2\equiv\rho(\mathcal{A}). \tag{3}$$

We shall call a system ultrametric (UM), if with the above metric the space of its GS's is an ultrametric space. Namely, for any three GS's Ω^1, Ω^2, and Ω^3 one has

$$d(\Omega^1,\Omega^2)\leqslant\max[d(\Omega^1,\Omega^3),d(\Omega^3,\Omega^2)]. \tag{4}$$

Note that for real systems this condition is fulfilled, probably up to measure zero of violations, and up to some small $\epsilon\sim d/N$ [18]. Those limitations do not affect our final conclusions, though one should be aware of their existence.

The first theorem expresses the fact that one cannot hope to travel between various GS's of an ultrametric system by just identifying and flipping independently various subsets (macros, i.e., collective objects). [Theorem 1 below is independent on the minimal energy property: it holds for any system with an UM subset of configurations singled out by some arbitrary property (e.g, GS's, high magnetization, and blue color).] More precisely, we have the following.

Theorem 1. Consider two subsets $\mathcal{A}$ and $\mathcal{B}$ of an UM system Ω that has a GS Ω^0. Assume that the states $\Omega^{\mathcal{A}}$ and $\Omega^{\mathcal{B}}$ obtained by flipping the spins of sets $\mathcal{A}$ and $\mathcal{B}$, respectively, in Ω^0 are GS's too. Then

$$\rho(\mathcal{A}\cap\mathcal{B})\geqslant\tfrac{1}{2}\min[\rho(\mathcal{A}),\rho(\mathcal{B})]. \tag{5}$$

Namely, the smaller of the sets $\mathcal{A}$ and $\mathcal{B}$ shares at least half of itself with the larger one. Moreover, if

$$\rho(\mathcal{A})>\rho(\mathcal{B})$$

then

$$\rho(\mathcal{A}\cap\mathcal{B})=\tfrac{1}{2}\rho(\mathcal{B}). \tag{6}$$

Theorem 1 means that at least half of the spins of one of the sets ($\mathcal{A}$ or $\mathcal{B}$) belong to the other set. This is hardly one's idea of two independent sets. In fact, generally, a point belongs to an infinite number of strongly overlapping clusters. This implies that locally one has no criterion for constructing the relevant macros. Those can be identified only from a global view.

In conclusion, in ultrametric systems it is ruled out that various regions of the system can be treated as independent collective degrees of freedom (macros). This picture can be extended to finite but small temperatures with the help of the ''pure state'' concept [1].

The failure of separability has conceptual implications in the sense that one cannot understand the complex system by analyzing its parts. In this sense an ultrametric system is conceptually irreducible to simpler entities. We will see in Sec. V that optimal global algorithms in fact reduce a system to its irreducible core.

One is tempted to conclude that the entire discussion of reductionism can be reformulated in terms of irreducible complex systems, i.e., in place of *assuming* ultrametricity and deducing the inexistence of independent dynamical subobjects, one can propose this *dynamical inseparability* as the fundamental property underlying irreducible complexity. Theorem 1 suggests therefore that one should engage in the systematic study of the systems that have families of GS's differing by strongly overlapping subsets. The topology induced in the system by these subsets might have interesting properties. In the rest of this section, we will give further characterization of the subsets $\mathcal{A}$ that relate (by their flipping) different ground states.

Theorem 2. If Ω^0 and $\Omega^{\mathcal{A}}$, differing by the sign of the spins in the region $\mathcal{A}$, are both GS's of a (not necessarily ultrametric) system Ω, then their actual spin arrangement Ω^0, restricted to the system $\mathcal{A}$ alone (we denote this particular configuration of $\mathcal{A}$ by $\mathcal{A}^0$), is a GS of the system $\mathcal{A}$ considered as isolated from the rest of Ω.

Note that this statement holds not only in low dimensions (where the surface energy is not extensive). To see in what respect this statement is nontrivial, note that in the presence of the system $\bar{\mathcal{A}}$ (the complement of $\mathcal{A}$ in Ω), the spins in system $\mathcal{A}$ are submitted to the influence of the external (to $\mathcal{A}$) action of the spins in $\bar{\mathcal{A}}$. For a general subset $\mathcal{A}$ of Ω, this will bring the spins of $\mathcal{A}$ into positions that are not necessarily optimal in terms of the internal $\mathcal{A}$ interactions alone. They would in general be in a position that strikes a compromise between minimizing the internal $\mathcal{A}$ energy and the interactions with the rest of the system ($\bar{\mathcal{A}}$). Theorem 2 finds conditions in which the action of $\bar{\mathcal{A}}$ can be ignored. This property has interesting unique consequences on the subsets that by flipping signs connect the GS's of ultrametric systems.

Theorem 3. Let Ω be a spin system of the type described by the energy [Eq. (2)]. Let the space of GS's of Ω be ultrametric [Eq. (4)].

Let Ω^0 and $\Omega^{\mathcal{A}}$ be two configurations of Ω that differ by the sign of the spins in a region $\mathcal{A}$. Assume that both Ω^0 and $\Omega^{\mathcal{A}}$ are GS's of Ω (which, according to Theorem 2, implies that $\mathcal{A}^0$ is the GS of $\mathcal{A}$). Then $\mathcal{A}^0$ is the unique GS of $\mathcal{A}$ (up to a global flipping of $\mathcal{A}$). This theorem throws some ironic light on the properties of GS's in ultrametric systems. *A posteriori* there is something qualitatively special in the sets that connect the GS's of UM systems: the uniqueness of their GS's. These sets are very special and do not share at all the proliferation of vacuums characteristic of typical SG subsystems. In fact, these subsets are not UM systems by themselves: Theorem III implies (among other things) that these subsets cannot constitute a multiscale hierarchy of UM subsystems included recurrently one in the other. Also the UM system is not self-similar in this sense.

V. EFFICIENT MCA'S IN DILUTE SPIN GLASSES

The discouraging arguments of Sec. III had a rather depressing effect on the expectations of the practitioners in the field on the performance of MCA in frustrated systems. We will see below that these arguments and even the theorems presented in Sec. IV still allow for a significant contribution of MCA's in frustrated systems as long as they possess macros.

As opposed to fully connected models such as the SK model [17], the geometry of the diluted models includes topological structures capable of engendering such macros. The CA's can then locate and act on large regions of the configuration that are weakly linked to the rest of it. In addition to CA's we construct a macros reduction algorithm (MRA) that acts explicitly on the same macros on which CA acts stochastically. In this way we make explicit the role of the macros in both algorithms. The MRA has a structure very similar to the dynamical algebraic multiGrid (DAMG) of [6]. Since its action is more direct, the MRA is more efficient than the CA in the models for which it was designed (see Sec. VI for numerical details). However, the CA is more versatile. Both are shown in Sec. VI to reach the real GS of the system in contrast to the local dynamics.

We consider again the Ising-like system

$$H=-\frac{1}{2}\sum_{i\neq j} J_{ij}s_i s_j , \tag{7}$$

with the probability distribution for $J_{i,j}$,

$$P(J_{ij})=(1-c/N)\,\delta(J_{ij})+(c/N)f(J_{ij}), \tag{8}$$

where $f(J_{ij})$ is the distribution of the surviving links after the dilution. This model is known as the highly diluted SG with finite average connectivity c $[=O(1)]$. The probability for a spin to have connectivity k in such a system follows the Poisson distribution $c^k\exp(-c)/k!$.

Many geometric properties of this system are well understood [19,20]. In particular, the system undergoes a percolation transition at $c=1$. The maximal cluster is of $O(\log N)$ for $c<1$, $O(N^{2/3})$ at $c=1$, and $O(N)$ for $c>1$, where its size is explicitly given by $P=1-P_0=1-\exp(-cP)$. The finite connectivity models at low temperature are directly connected to the graph partitioning problem [19,20] (dividing a graph into subgraphs with minimum connections between them).

A. Macro reduction algorithm

In the Introduction it was claimed that the very existence of an efficient MCA may help identify relevant macros in the system. To achieve this we construct a MRA that freezes explicitly spins into macros and by doing so reproduces (and improves) the performance of the efficient CA. The MRA consists of the following iterative steps.

(i) Access the points i of the system iteratively starting with the ones with lowest connectivity.

(ii) For an accessed point i find its strongest connection $J_{i,j}$, defined by $|J_{i,j}|>|J_{i,k}|$ for all $k\neq j$.

(iii) Freeze s_i and s_j into a macro such that $-s_iJ_{i,j}s_j$ is minimal. From now on the value of s_j alone labels the state of the macro and s_i is just a ''slave'' determined by it.

(iv) As seen in the detailed explanations below, when all points (macros) with one neighbor are exhausted, all trees shrink to points.

(v) When all points (macros) with two neighbors are exhausted, all linear chains in the system shrink to length 1.

(vi) After that, some regular Monte Carlo (MC) (either local or global) that acts on the new system can be used.

Alternatively, one can keep coarsening the lattice along the lines above while at the same time performing MC sweeps. This procedure is reminiscent of the DAMG presented in [6] and may involve V and W multigrid cycles [21].

As in [6], the reduced degrees of freedom are decided according to the geometry of the lattice not according to the current configuration. Therefore freezing degrees of freedom into such reduced objects does not interfere with the detailed balance.

The efficiency of the MRA relies on the following observations.

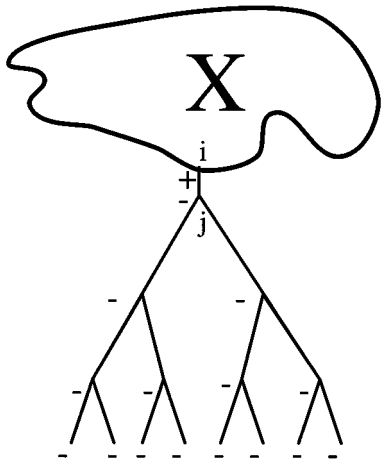
(i) The reduction stage takes negligible time (few MC steps) compared to the relaxation part.

(ii) The resulting reduced system is much smaller than the original one. For the lattices we have used $(c=2)$ in the Gaussian case, its volume was 4 times smaller than the initial volume.

(iii) The reduction in the volume has two independent effects: the cost of each sweep over the lattice is reduced by a factor proportional to the reduction of the volume and the number of sweeps necessary to converge to a ground state is reduced as a function of the volume depending on the slowing down characteristics. For spin glasses this might even exceed a polynomial dependence. Let us see now the detailed mechanisms by which MCA's such as the CA, algebraic multigrid, and MRA work where the local algorithm does not.

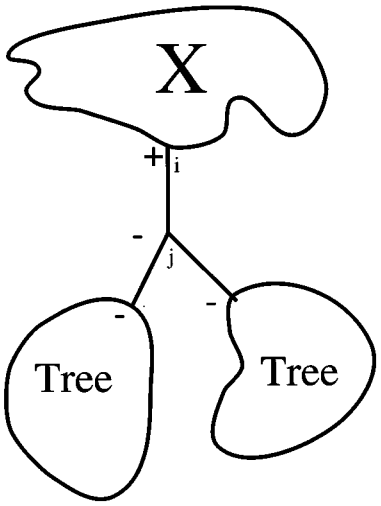
B. Trees

Consider a configuration.

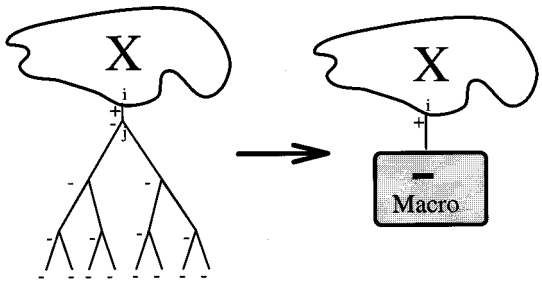


Assume that X is minimal if $s_i=+1$. Then the energy is minimized if all s's on the tree sites are $+1$. However , a simple MC algorithm might have problems in reaching this

minimal energy configuration. For instance, if one has



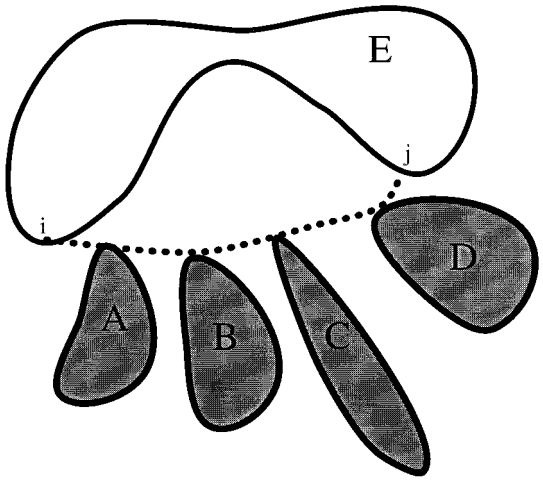
then even after ensuring $s_i=+1$, s_j will have two neighbors with $s=-1$ and will refuse flipping to $s_j=+1$. In contrast, the CA will first freeze all the links belonging to the tree, delete the (i,j) link, and only then perform a flip of the obtained macro in one painless step.



Similarly, MRA will identify the leaves and transform its sites into a macro. Eventually, the algorithm will create a macro standing in for the entire tree. Clearly one can now update the macro of the new system in one step and then return to the explicit microscopic representation.

C. Linear chains in Gaussian distributions

Consider a configuration of the type



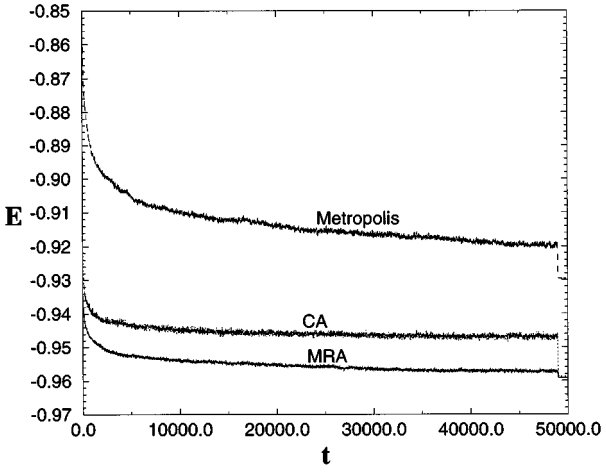


FIG. 1. $E(t)$, the energy per spin as a function of the running time, for Metropolis dynamics, the cluster algorithm (CA), and the macro reduction algorithm (RA) for a $c=2$, $N=5000$ lattice at $T\sim 0.3T_c$. In the last 1000 MCS we set $T=0$.

The subsets A,B,CD (''drops'') can be considered as macros and minimized separately. Assume that the block E minimization strongly prefers a particular position for i and j. For definitiveness take $s_i=+1$ and $s_j=+1$ and assume that the product of the links joining i to j is -1. The minimal energy configuration is therefore frustrated. In order to find the minimum one has to reach the configuration with only one unsatisfied link in the entire chain and make it such that the frustration resides on the link with the lowest $|J|$ in the chain.

As we mentioned in Sec. V B, the CA mainly freezes the strong satisfied links, flips the sign of the strong frustrated ones, and keeps trying the weak ones. This quickly approaches the configuration with the frustration on the weakest link.

The MRA also puts together the satisfied strong links into macros. Once the entire ij line is transformed into just one macro (with the strength of its lowest link) one can compare the price of its frustration to the price of frustrating E (and decide which of the two should remain unsatisfied).

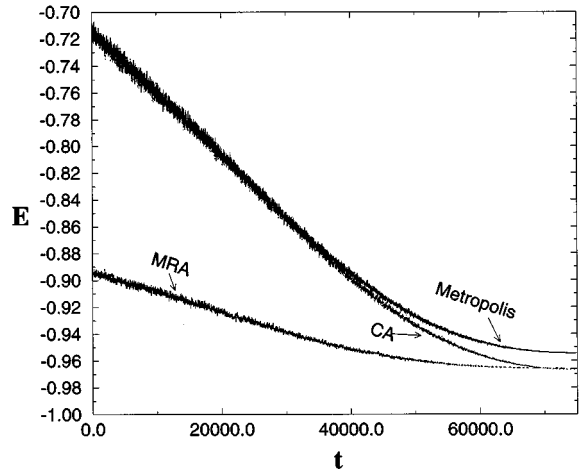
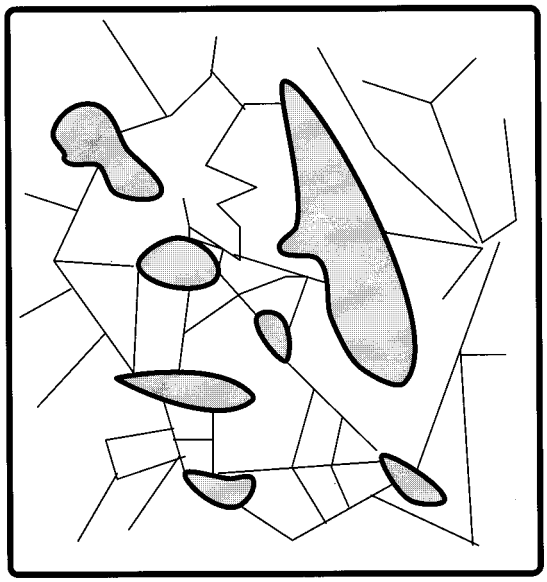


FIG. 2. Same as Fig. 1, but with an annealing schedule over the range $T\in 2.0-0$.

D. Strongly coupled islands

Consider a situation in which there are islands of sites related by very strong links submerged in a sea of links that are much weaker.



The filled zones in the figure are the islands of strong links and we did not draw their sites explicitly. (Beware. In many aspects, this figure might be quite misleading, especially in the infinite-dimensional case.)

Suppose the strong links are of lower density and possibly below the percolation. Suppose that the energy of the spins within the strong islands was somewhat minimized. The equilibration of the relative signs between the various islands is very inefficient by local updating because none of the spins of an island would accept flipping without the entire island doing so.

To check a simple example of such a case, we considered a model (the weak-strong model) in which the link distribution is given by

$$f(J)=\frac{a}{2}[\delta(J-\epsilon)+\delta(J+\epsilon)] + \frac{(1-a)}{2}[\delta(J-1)+\delta(J+1)]. \quad (9)$$

By defining each island as a macro, the MRA can of course realign an entire island in one step and with quite high acceptance. Then CA is efficient too insofar as it freezes the strong links and it allows the others to reach an equilibrium (especially when annealing is applied). However, in a general case, the MRA is more difficult to apply, while the CA is still quite efficient in implicitly identifying macros in a stochastic manner. As we report in Sec. VI, we found indeed for the weak-strong model a huge gap between the satisfied energy per spin achieved by a MCA on one side and by a local Metropolis dynamics on the other.

VI. COMPARING THE LOCAL MC TO THE MRA AND CA

The results in this section show that the multiscale-cluster algorithms not only converge faster than the usual methods,

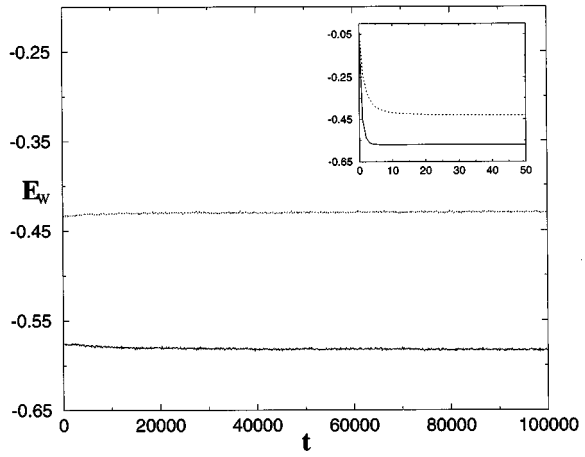


FIG. 3. Weak link energy $E_W(t)$ as a function of the running time, for the weak-strong model, with $c=2$, $a=0.7$, $N=5000$, and $T=0.5J_W$. The strong and the weak links were scaled to 100 and 1, respectively. The solid and the dotted lines indicate cluster and Metropolis dynamics, respectively. Inset: the first 50 steps.

but they reach ground states that are unaccessible to the local methods in any practical time. As seen below, this is especially true when one uses an annealed schedule, in which the temperature is gradually lowered. The temperature variation enables the CA to act on clusters at various scales, corresponding to the different temperatures, and to address and freeze first the stronger links into small but very strongly coupled islands. At lower energies larger, loosely coupled islands are formed and acted upon. In addition, the implementation of the MRA is shown to be more efficient than the CA both on the rate of convergence and in the effective size of the system (as explained in Sec. V, each sweep in the MRA is actually smaller than the CA sweep, due to the reduced size of the system in the MRA case).

A. Dilute SG model in the Gaussian case

Simulations on the model defined by Eqs. (7) and (8) with a Gaussian distribution for the links $f(J)\propto\exp(-J^2)$ were carried out comparatively using local dynamics (Metropolis), the cluster algorithm [22] and the macros reduction algorithm.

The simulations were carried out for various connectivity values c and at temperatures below the glassy transition T_c. The size of the system was between 1000 and 5000 sites. The results were averaged over at least ten different samples. A typical result is presented in Fig. 1. In Fig. 1 one sees the evolution of the energy of the system ($n=5000$) monitored during its computer simulation [up to 50 000 Monte Carlo steps (MCS) per spin]. One can clearly see a gap between the energies reached with the global dynamics (the CA and MRA) and the local dynamics. The MRA gets exactly the same energy level as the CA, but it converges much faster.

In order to emphasize those features, we performed measurements using simulated annealing for the above dynamics. In Fig. 2 one can see a similar picture of the gap between the energies reached with the global dynamics (the CA and MRA) and the local dynamics. This suggests that the CA and MRA converge faster than the usual methods and reach the

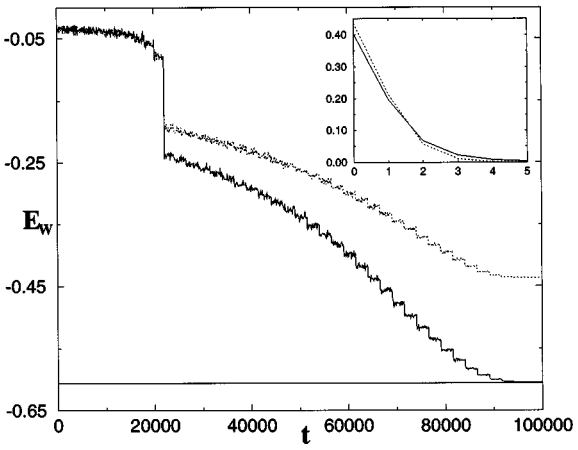


FIG. 4. $E_W(t)$ for the weak-strong model, with $c=2$ and $a=0.7$. The annealing schedule range is $T\in 123-0$, with $\Delta T=10$ for $T>3$ and $\Delta T=0.1$ for $T<3$. The solid and the dotted lines indicate cluster and Metropolis dynamics, respectively. The horizontal line denotes the analytical GS. Inset: P_l for the weak-strong case (solid line) and the $J=\pm 1$ case (dotted line).

true ground states which are unaccessible to the local methods in any practical time.

As seen in Figs. 1 and 2, the MRA converges faster than the CA. This reduction in the slowing down is related to the fact that the reduced system has a much smaller volume than the initial one. In fact, the gain in CPU time is much larger than seen from the graphs because the graphs show the energy evolution measured in MCS per spin. More precisely, due to the reduction of the volume, a sweep over the reduced lattice takes about 1/4 of the CPU time necessary to sweep the initial lattice (we measured that for $c=2$ the MRA relaxation steps need to act only on about 25% of the initial system).

B. Weak-strong model

The simulations of the weak-strong model [Eq. (9)] were compared with the analytical results in [23]. In [23] the self-consistent description of the low temperatures of the weak-strong model was based on the probability distribution of the local field defined by $h_i\equiv T\tanh^{-1}\langle S_i\rangle_T$. Physically, this field is the first excitation, namely, in the limit $T\to 0$, and $|h_i|$ is the minimum energy cost for flipping the ith spin from its GS by the ''best'' reorganization of the system. This local field is in truth an oxymoron, since it depends on global properties of the cluster. The exchange field $\Sigma_j J_{ij}m_j$, on the other hand, is truly a local property depending on the local connectivity (note that $|h_i|\leqslant|\Sigma_j J_{ij}m_j|$). It was found that within the replica symmetry assumption the GS energy of the weak links is given by

$$E_W=-\frac{1}{2}caP_0^2\epsilon+\frac{1}{2}c(1-a)\sum_{k=0}^{\infty}(1-4\sigma_k^2)\epsilon-\overline{h}, \quad (10)$$

where $\sigma_k=P_0/2+\Sigma_{l=1}^{k}P_l$ and $\overline{h}=\Sigma_{h=-\infty}^{\infty}|h|P(h)$. Note that the energy of the strong links $E_S=-\frac{1}{2}c(1-a)$ is eliminated from Eq. (10).

The explicit value of E_W depends on the local field $P(h)$, which in general is difficult to calculate. Nevertheless, after some work, $P(h)$ can be determined from the equation

$$P(h)=e^{-cQ}\int_{-\infty}^{\infty}\frac{dy}{2\pi}\exp\left[-iyh+\frac{cQa}{2}(e^{iy\epsilon}+e^{-iy\epsilon})+c(1-a)\sum_{l=1}^{\infty}P_l(e^{iyl\epsilon}+e^{-iyl\epsilon})\right]. \quad (11)$$

The resulting $P(h)$ can then be compared with the usual random $J=\pm 1$ result $P_l=\exp(-cQ)I_{|l|}(cQ)$ [23], where $I_l(x)$ is the modified Bessel function. The graphs of the two P_l's are presented in the inset of Fig. 3. Note that after scaling ϵ to 1, the P_l for the two cases are very close. However, the values of the *exchange field* are very different: in the $J=\pm 1$ case the exchange field of a spin is usually not far from its local field (around the number of its neighbors), while in the weak-strong model the local field is $O(\epsilon)$ vs the exchange field, which is $O(1)$. The fact that P_l is much smaller, raises the hope that a global dynamics will be superior to the local one. This is confirmed by the experiments below.

C. Performance of the local dynamics vs the CA in the weak-strong model

In the first set of runs we choose the connectivity c (the average number of neighbors) and the fraction of the strong links $(1-a)$ such that the density of the strong links by themselves is below the percolation threshold $c_S=c(1-a)<1$. It is clear that in this situation all the strong links are unfrustrated and the frustration is located only on the weak links. Therefore only the energy of the weak links E_W is to be considered. In Fig. 3 one can see a large steady difference (35%) in the energy between the local dynamics and that of the cluster dynamics.

Weak-strong mode with simulated annealing

In the second set of runs on the weak-strong model, we performed measurements using simulated annealing for both global and local dynamics (Fig. 4). This was performed by cooling over a wide range of temperatures. At larger temperatures, there were the strong links that reached their minimal energy and only then, at lower temperatures, the weak links adapted to the strong links environment.

The clear energy difference between the local and cluster cases is due to the fact that the local dynamics is totally stuck since the probability of flipping a cluster consisting of strong links is practically zero for the local algorithm. On the other hand, the CA ''knows'' how to deal with the strong link structures by treating them as only ''one degree of freedom'' for each cluster. In other words, the CA is extremely efficient in solving the problem of how to arrange the weak links in the environment of the strong links.

The mean-field solution for such models [23] is known to be unstable [24]. However, in Fig. 4 one can see that the analytical GS [23], is in very good agreement with the averaged GS energy obtained by the CA.

VII. OUTLOOK

Identifying the nature and the dynamics of the macros may help understand the complex multiscale nature of a system. The techniques for identifying the relevant macros lead to a deeper understanding of the way the macroscopic description of our world arises in the study of complex systems composed of simple microscopic elements.

The macros can be multiscale reducible, but in many cases there might exist complex irreducible cores. While such irreducible macros might have fortuitous characteristics such as lack generality and present nongeneric properties, they might be very important if the same set of cores appears recurrently in biological, neurological, or cognitive systems in nature.

In such situations, rather than trying to understand the macros structure, dynamics, and properties on general (multiscale, analytic) grounds as collections of their parts, one may have to recognize the unity and uniqueness of these macros and resign oneself in just making an as intimate as possible acquaintance with their features. One may still try to treat them by the implicit elimination method [3] where the complex objects are presenting, isolating, and eliminating themselves by the very fact that they are projected out by the dynamics as the slow-to-converge modes.

ACKNOWLEDGMENTS

The research reported in this paper has been supported in part by grants from the Germany-Israel Foundation and from the Israeli Academy of Sciences and Humanities. This paper was completed during a visit at ICTP and SISSA. Discussions with D. Kandel, I. Kanter, G. Mack, and M. Virasoro are gratefully acknowledged.

APPENDIX:

We present in this Appendix the proofs of the theorems. We will consider Ising-like systems consisting of spin $s_k = \pm 1$:

$$\Omega = \{s_k | k = 1, \ldots, N\}, \tag{A1}$$

where N is the arbitrary size of the system, which in the thermodynamic limit is taken to infinity. A configuration is a specific assignment of the spins, e.g., $\Omega^D = \{s_1^D, \ldots, s_N^D\}$.

The system has an Ising-like bilinear Hamiltonian that defines an extensive energy. For example, in a SK [17] -like model (with $J_{i,j} \in \mathcal{R}$)

$$E(\Omega^D) \equiv \langle \Omega^D, \Omega^D \rangle = -\frac{1}{2\sqrt{N}} \sum_{i,j} J_{i,j} s_I^D s_J^D . \tag{A2}$$

Note that in spite of this shorthand ''scalar product'' notation, the ''norm'' it defines is not positive; we will use only its bilinear properties.

The metric in the configurations space is defined by the following distance: If two configurations $\Omega^0$and $\Omega^{\mathcal{A}}$ differ (only) by the sign of the spins belonging to a subset $\mathcal{A}$, then their distance [note that in a finite system $\rho(\mathcal{A}) =$ (the number of elements in $\mathcal{A})/N$] is

$$d(\Omega^0, \Omega^{\mathcal{A}}) = \frac{1}{4N} \sum_i (s_i^0 - s_i^{\mathcal{A}})^2 \equiv \rho(\mathcal{A}). \tag{A3}$$

A configuration $-\Omega^D = \{-s_1^D, \ldots, -s_N^D\}$ differing from Ω^D only by a global change of sign is considered to be identical to Ω^D. This brings the support of the function $d(\Omega^1, \Omega^2)$ into the interval $[0, 0.5]$.

We shall call a system ultrametric if with the above metric the space of its GS's is an ultrametric space. [Actually, the theorem below is independent of the minimal energy property: it holds for any system with an UM subset of configurations singled out by some arbitrary property (e.g, GS's, high magnetization, and blue color).] Namely, for any three GS's Ω^1, Ω^2, Ω^3 one has

$$d(\Omega^1, \Omega^2) \leqslant \max[d(\Omega^1, \Omega^3), d(\Omega^3, \Omega^2)]. \tag{A4}$$

For the proof of Theorem 1 we have the following.

Proof. (i) By definition $d(\Omega^{\mathcal{A}}, \Omega^0) = \rho(\mathcal{A})$, $d(\Omega^{\mathcal{B}}, \Omega^0) = \rho(\mathcal{B})$, and $d(\Omega^{\mathcal{B}}, \Omega^{\mathcal{A}}) = \rho(\mathcal{A}) + \rho(\mathcal{B}) - 2\rho(\mathcal{A} \cap \mathcal{B})$. (ii) On the other hand, the UM system (15) implies, in our case, either (a) all the distances are equal $d(\Omega^0, \Omega^{\mathcal{A}}) = d(\Omega^0, \Omega^{\mathcal{B}}) = d(\Omega^{\mathcal{A}}, \Omega^{\mathcal{B}})$ or (b) two of the distances are equal and the third is shorter. (iii) Considering (i) and (ii) for Ω^0, $\Omega^{\mathcal{A}}$, and $\Omega^{\mathcal{B}}$ and (without loss of generality) assuming that $\rho(\mathcal{A}) \geqslant \rho(\mathcal{B})$, one has one of the following possibilities: (iia) is valid, in which case $\rho(\mathcal{A}) = \rho(\mathcal{B}) = 2\rho(\mathcal{A}) - 2\rho(\mathcal{A} \cap \mathcal{B}) \Rightarrow \rho(\mathcal{A} \cap \mathcal{B}) = 1/2\rho(\mathcal{A}) \rho(\mathcal{A}) = \rho(\mathcal{B}) = 2\rho(\mathcal{A}) - 2\rho(\mathcal{A} \cap \mathcal{B}) \Rightarrow \rho(\mathcal{A} \cap \mathcal{B}) = 1/2\rho(\mathcal{A}) = 1/2\rho(\mathcal{B})$; (iib) is valid and $\rho(\mathcal{A}) = \rho(\mathcal{B}) > 2\rho(\mathcal{A}) - 2\rho(\mathcal{A} \cap \mathcal{B})$, in which case $\rho(\mathcal{A} \cap \mathcal{B}) > 1/2\rho(\mathcal{A}) = 1/2\rho(\mathcal{B})$; or (iib) is valid and $\rho(\mathcal{A}) = \rho(\mathcal{A}) + \rho(\mathcal{B}) - 2\rho(\mathcal{A} \cap \mathcal{B})$, in which case $\rho(\mathcal{A} \cap \mathcal{B}) = 1/2\rho(\mathcal{B})$.

In any of these three possibilities the desired relation [Eq. (A5)] is fulfilled:

$$\rho(\mathcal{A} \cap \mathcal{B}) \geqslant \tfrac{1}{2} \min[\rho(\mathcal{A}), \rho(\mathcal{B})].$$

□

Here we list further notations. In the following manipulations we will use the definitions and conventions below. Assume that Ω is a system with Eq. (A1) as its energy and $\mathcal{M}$ is a subset $\mathcal{M} \subset \Omega$. Ω^0 is the initial configuration of Ω: $\Omega^0 = \{s_i^0\}$. $\mathcal{M}^D$ is the restriction of some arbitrary configuration Ω^D to the subsystem $\mathcal{M}$, namely, $\mathcal{M}^D = \{s_i^D | s_i \in \mathcal{M}\}$. The energy associated with $\mathcal{M}^D$ excluding its interaction with the rest of Ω is

$$E(\mathcal{M}^D) \equiv \langle \mathcal{M}^D, \mathcal{M}^D \rangle = -\frac{1}{2\sqrt{N}} \sum_{i,j(s_i, s_j \in \mathcal{M})} J_{i,j} s_i^D s_j^D . \tag{A5}$$

Let $\overline{\mathcal{M}}$ be the complementary of $\mathcal{M}$ in Ω ($\overline{\mathcal{M}} \cap \mathcal{M} = \varnothing$ and $\mathcal{M} \cup \mathcal{M} = \Omega$). Its energy in the configuration Ω^D is

$$E(\overline{\mathcal{M}^D}) = \langle \overline{\mathcal{M}^D}, \overline{\mathcal{M}^D} \rangle = -\frac{1}{2\sqrt{N}} \sum_{i,j(s_i, s_j \in \overline{\mathcal{M}})} J_{i,j} s_i^D s_j^D . \tag{A6}$$

The interaction energy between such disjoint parts of a configuration will be noted by

$$\langle \overline{\mathcal{M}^D}, \mathcal{M}^D \rangle + \langle \mathcal{M}^D, \overline{\mathcal{M}^D} \rangle \tag{A7}$$

$$= -\frac{1}{2\sqrt{N}} \sum_{s_i \in \mathcal{M}, s_j \in \overline{\mathcal{M}}} J_{i,j} s_i^D s_j^D \tag{A8}$$

$$-\frac{1}{2\sqrt{N}} \sum_{s_i \in \overline{\mathcal{M}} s_j \in \mathcal{M}} J_{i,j} s_i^D s_j^D = 2\langle \overline{\mathcal{M}^D}, \mathcal{M}^D \rangle. \tag{A9}$$

We will use the formal notation

$$\Omega^D = \overline{\mathcal{M}^D} + \mathcal{M}^D, \tag{A10}$$

which ensures that

$$\langle \overline{\mathcal{M}^D} + \mathcal{M}^D, \overline{\mathcal{M}^D} + \mathcal{M}^D \rangle \tag{A11}$$

$$= \langle \mathcal{M}^D, \mathcal{M}^D \rangle + \langle \overline{\mathcal{M}^D}, \overline{\mathcal{M}^D} \rangle + 2\langle \mathcal{M}^D, \overline{\mathcal{M}^D} \rangle. \tag{A12}$$

We will use similar sum notations for any configuration that is specified in terms of specific spin configurations of disjoint subsets covering Ω. The scalar product $\langle , \rangle$ (A2) is distributive with respect to this sum. Let $\Omega^{\mathcal{M}}$ be a configuration obtained from Ω^0 by flipping $\mathcal{M}$, namely, $s_i^{\mathcal{M}} = -s_i^0$ if $s_i \in \mathcal{M}$ and $s_i^{\mathcal{M}} = s_i^0$ otherwise. We will use, in this case, the notation

$$\Omega^{\mathcal{M}} = \overline{\mathcal{M}^0} - \mathcal{M}^0. \tag{A13}$$

This notation has the advantage that, according to (A2),

$$E(\Omega^{\mathcal{M}}) \equiv \langle \overline{\mathcal{M}^0} - \mathcal{M}^0, \overline{\mathcal{M}^0} - \mathcal{M}^0 \rangle \tag{A14}$$

$$= \langle \mathcal{M}^0, \mathcal{M}^0 \rangle + \langle \overline{\mathcal{M}^0}, \overline{\mathcal{M}^0} \rangle - 2\langle \mathcal{M}^0, \overline{\mathcal{M}^0} \rangle. \tag{A15}$$

Definition. Let $\mathcal{A}$ be a subset of the system Ω or Ω itself ($\mathcal{A} \subseteq \Omega$). We say a configuration $\mathcal{A}^D$ is a C ground state of $\mathcal{A}$, [in short $S_G(\mathcal{A},C)$ from now on] if

$$E(\mathcal{A}^D) - E(\mathcal{A}^g) \equiv \langle \mathcal{A}^D, \mathcal{A}^D \rangle - \langle \mathcal{A}^g, \mathcal{A}^g \rangle < C, \tag{A16}$$

where $\mathcal{A}^g$ is the state with the absolute minimum energy in $\mathcal{A}$ and C is a positive constant. In other words, $E(\mathcal{A}^D)$ is up to C a minimum of the energy of the subset $\mathcal{A}$, $E(\mathcal{A}^K) \forall K$.

Theorem 2. If Ω^0 and $\Omega^{\mathcal{A}}$ differing by the sign of the spins in the region $\mathcal{A}$ are both $GS(\Omega,C)$ in a (not necessarily ultrametric) system Ω, then $\mathcal{A}^0$ is $GS(\mathcal{A},C)$.

Proof. Note that the states Ω^0 and Ω^A can be written

$$\Omega^0 = \overline{\mathcal{A}^0} + \mathcal{A}^0 \tag{A17}$$

and, respectively,

$$\Omega^A = \overline{\mathcal{A}^0} - \mathcal{A}^0. \tag{A18}$$

Accordingly, their energies are

$$E[\Omega^0] \equiv \langle \Omega^0, \Omega^0 \rangle = \langle \mathcal{A}^0, \mathcal{A}^0 \rangle + \langle \overline{\mathcal{A}^0}, \overline{\mathcal{A}^0} \rangle + 2\langle \mathcal{A}^0, \overline{\mathcal{A}^0} \rangle, \tag{A19}$$

$$E[\Omega^A] \equiv \langle \Omega^A, \Omega^A \rangle = \langle \mathcal{A}^0, \mathcal{A}^0 \rangle + \langle \overline{\mathcal{A}^0}, \overline{\mathcal{A}^0} \rangle - 2\langle \mathcal{A}^0, \overline{\mathcal{A}^0} \rangle. \tag{A20}$$

Without loss of generality one can assume $E[\Omega^0] \geqslant E[\Omega^A]$ (otherwise one can interchange the names of Ω^0 and Ω^A). This implies

$$4\langle \mathcal{A}^0, \overline{\mathcal{A}^0} \rangle \geqslant 0. \tag{A21}$$

Now assume that $\mathcal{A}^0$ is not a $S_G(\mathcal{A},C)$. Then there exists a subset $\mathcal{X} \subset \mathcal{A}$ that can be flipped to bring the system $\mathcal{A}$ to the absolute minimum $[S_G(\mathcal{A},0)]$. Using the notation (A13) this state can be written

$$\mathcal{A}^{\mathcal{X}} = \overline{\mathcal{X}^0_A} - \mathcal{X}^0, \tag{A22}$$

where $\overline{\mathcal{X}_A}$ is the complement of $\mathcal{X}$ in $\mathcal{A}$. The assumption that $\mathcal{A}^0$ is not $S_G(\mathcal{A},C)$ then reads

$$E[\mathcal{A}^0] - E[\mathcal{A}^{\mathcal{X}}] \equiv \langle \overline{\mathcal{X}^0}_A + \mathcal{X}^0, \overline{\mathcal{X}^0}_A + \mathcal{X}^0 \rangle - \langle \overline{\mathcal{X}^0}_A - \mathcal{X}^0, \overline{\mathcal{X}^0}_A - \mathcal{X}^0 \rangle \tag{A23}$$

$$\equiv 4\langle \mathcal{X}^0, \overline{\mathcal{X}^0}_A \rangle \rangle C. \tag{A24}$$

Note that if $\mathcal{A}^{\mathcal{X}} = \overline{\mathcal{X}^0}_A - \mathcal{X}^0$ is a ground state of $\mathcal{A}$, so is $-\mathcal{A}^{\mathcal{X}} = -\overline{\mathcal{X}^0}_A + \mathcal{X}^0$. However, by including the subset $\mathcal{A}$ in the whole system Ω, the sign of its interaction with the rest of the system $\overline{\mathcal{A}^0}$ has opposite sign in the two cases.

This leads to the following total energies in terms of the configurations of the subsets $\mathcal{X}$, $\overline{\mathcal{X}_A}$ and $\overline{\mathcal{A}}$:

$$E[\Omega^0] \equiv \langle \Omega^0, \Omega^0 \rangle \equiv \langle \overline{\mathcal{A}^0} + \overline{\mathcal{X}^0}_A + \mathcal{X}^0, \overline{\mathcal{A}^0} + \overline{\mathcal{X}^0}_A + \mathcal{X}^0 \rangle, \tag{A25}$$

$$E[\Omega^{\mathcal{X}}] \equiv \langle \Omega^{\mathcal{X}}, \Omega^{\mathcal{X}} \rangle \equiv \langle \overline{\mathcal{A}^0} + \overline{\mathcal{X}^0}_A - \mathcal{X}^0, \overline{\mathcal{A}^0} + \overline{\mathcal{X}^0}_A - \mathcal{X}^0 \rangle, \tag{A26}$$

$$E[\Omega^{\overline{\mathcal{X}_A}}] \equiv \langle \Omega^{\overline{\mathcal{X}_A}}, \Omega^{\overline{\mathcal{X}_A}} \rangle \equiv \langle \overline{\mathcal{A}^0} - \overline{\mathcal{X}^0}_A + \mathcal{X}^0, \overline{\mathcal{A}^0} - \overline{\mathcal{X}^0}_A + \mathcal{X}^0 \rangle. \tag{A27}$$

Consequently, taking the combination (A25) and (A28) plus ((A25)–(A27) one gets

$$(E[\Omega^0] - E[\Omega^{\mathcal{X}}]) + (E[\Omega^0] - E[\Omega^{\overline{\mathcal{X}_A}}]) = 4(\langle \mathcal{X}^0, \overline{\mathcal{A}^0} \rangle + \langle \mathcal{X}^0, \overline{\mathcal{X}^0}_A \rangle) + 4(\langle \overline{\mathcal{X}^0}_A, \overline{\mathcal{A}^0} \rangle + \langle \overline{\mathcal{X}^0}_A, \mathcal{X}^0 \rangle)$$

$$= 4\langle \mathcal{A}^0, \overline{\mathcal{A}^0} \rangle + 8\langle \mathcal{X}^0, \overline{\mathcal{X}^0}_A \rangle. \tag{A28}$$

According to (A21) and (A24), the right-hand term is larger than $2C$, therefore, the same should hold for the left-hand term. This means that at least one of the brackets in the left-hand term is larger than C:

$$(E[\Omega^0] - E[\Omega^2]) > C \quad \text{or} \quad (E[\Omega^0] - E[\Omega^2]) > C.$$

However, according to (A16) this contradicts the initial statement that Ω^0 is a $S_G(\Omega,C)$. The contradiction can only be

attributed to our assumption that $\mathcal{A}^0$ is not a $S_G(\mathcal{A},C)$. This proves the theorem by reduction to the absurd. ■

Note that one can take $C=0$ and then the whole theorem can be referred to as absolute ground states.

Theorem 3. Let Ω be a spin system of (A2) type. Let the space of $GS(\Omega,3C)$ be ultrametric (A4). Let Ω^0 and $\Omega^{\mathcal{A}}$ be two configurations that are differing by the sign of the spins in a region $\mathcal{A}$. Assume that both Ω^0 and $\Omega^{\mathcal{A}}$ are $GS(\Omega,C)$'s which, according to Theorem 2, implies that $\mathcal{A}^0$ is a $GS(\mathcal{A},C)$. Then, $\mathcal{A}^0$ is the unique $GS(\mathcal{A},C)$ of $\mathcal{A}$ (up to a global flipping of $\mathcal{A}$).

Proof. If $\mathcal{A}$ has another $S_G(\mathcal{A},C)$, that state can be obtained by flipping a certain subset $\mathcal{X}\subset\mathcal{A}$. The expression of the new $S_G(\mathcal{A},C)$ in terms of the configurations on the subsets $\mathcal{X}$ and $\mathcal{X}_{\mathcal{A}}$ (its complement in $\mathcal{A}$) will be

$$\mathcal{A}^{\mathcal{X}}=\overline{\mathcal{X}_A}^{\,0}-\mathcal{X}^0. \tag{A29}$$

The assumption that both $\overline{\mathcal{X}_A}^{\,0}-\mathcal{X}^0$ and $\mathcal{A}^0\equiv\overline{\mathcal{X}_A}^{\,0}+\mathcal{X}^0$ are $S_G(A,C)$ implies

$$-C<\langle\overline{\mathcal{X}_A^0}-\mathcal{X}^0,\overline{\mathcal{X}_A^0}-\mathcal{X}^0\rangle-\langle\mathcal{A}^0,\mathcal{A}^0\rangle<C, \tag{A30}$$

i.e.,

$$-C<-4\langle\mathcal{X}^0,\overline{\mathcal{X}_A^0}\rangle<C. \tag{A31}$$

The total energies of the full Ω configurations that can be formed with the various configurations on the subsets $\mathcal{X}$, $\overline{\mathcal{X}_A}$, and $\overline{\mathcal{A}^0}$ are

$$E[\Omega^0]\equiv\langle\Omega^0,\Omega^0\rangle\equiv\langle\overline{\mathcal{A}^0}+\overline{\mathcal{X}_A^0}+\mathcal{X}^0,\overline{\mathcal{A}^0}+\overline{\mathcal{X}_A^0}+\mathcal{X}^0\rangle, \tag{A32}$$

$$E[\Omega^{\mathcal{A}}]\equiv\langle\Omega^{\mathcal{A}},\Omega^{\mathcal{A}}\rangle\equiv\langle\overline{\mathcal{A}^0}-\overline{\mathcal{X}_A^0}-\mathcal{X}^0,\overline{\mathcal{A}^0}-\overline{\mathcal{X}_A^0}-\mathcal{X}^0\rangle, \tag{A33}$$

$$E[\Omega^{\mathcal{X}}]\equiv\langle\Omega^{\mathcal{X}},\Omega^{\mathcal{X}}\rangle\equiv\langle\overline{\mathcal{A}^0}+\overline{\mathcal{X}_A^0}-\mathcal{X}^0,\overline{\mathcal{A}^0}+\overline{\mathcal{X}_A^0}-\mathcal{X}^0\rangle, \tag{A34}$$

$$E[\Omega^{\overline{\mathcal{X}_A}}]\equiv\langle\Omega^{\overline{\mathcal{X}_A}},\Omega^{\overline{\mathcal{X}_A}}\rangle\equiv\langle\overline{\mathcal{A}^0}-\overline{\mathcal{X}_A^0}+\mathcal{X}^0,\overline{\mathcal{A}^0}-\overline{\mathcal{X}_A^0}+\mathcal{X}^0\rangle. \tag{A35}$$

Taking into account that both (A32) and (A33) represent $S_G(\Omega,C)$ energies [and assuming without loss of generality that (A33) is lower] one gets for their difference [(A32) and (A33)] the inequalities

$$0<4\langle\overline{\mathcal{A}^0},\overline{\mathcal{X}_A^0}+\mathcal{X}^0\rangle<C. \tag{A36}$$

The configurations in (A34) and (A35) are not guaranteed to be $S_G(\Omega,C)$, so for the combinations (A32)–(A34) and (A32)–(A35) respectively, one gets inequalities in only one direction

$$E[\Omega^0]-E[\Omega^{\mathcal{X}}]\equiv4\langle\overline{\mathcal{A}^0}+\overline{\mathcal{X}_A^0},\mathcal{X}^0\rangle<C, \tag{A37}$$

$$E[\Omega^0]-E[\Omega^{\overline{\mathcal{X}_A}}]\equiv4\langle\overline{\mathcal{A}^0}+\mathcal{X}_A^0,\overline{\mathcal{X}^0}\rangle<C, \tag{A38}$$

respectively. By considering the combinations (A37) plus (A31) and (A38) plus (A31) one gets

$$4\langle\overline{\mathcal{A}^0},\mathcal{X}^0\rangle<2C \tag{A39}$$

and, respectively,

$$4\langle\overline{\mathcal{A}^0},\overline{\mathcal{X}_A^0}\rangle<2C. \tag{A40}$$

Substituting (A39) and (A40), respectively, in the left relation of (A36) one gets

$$4\langle\overline{\mathcal{A}^0},\mathcal{X}^0\rangle>-2C \tag{A41}$$

and

$$4\langle\overline{\mathcal{A}^0},\overline{\mathcal{X}_A^0}\rangle>-2C. \tag{A42}$$

This in turn implies [using (37)]

$$4\langle\overline{\mathcal{A}^0}+\overline{\mathcal{X}_A^0},\mathcal{X}^0\rangle>-3C \tag{A43}$$

and, respectively,

$$4\langle\overline{\mathcal{A}^0}+\mathcal{X}_A^0,\overline{\mathcal{X}^0}\rangle>-3C. \tag{A44}$$

But (A43) and (A44) are, according to the identities in (A37) and (A38), the difference between the energies of the configurations Ω^0 and $\Omega^{\mathcal{X}}$ (Ω^0 and $\Omega^{\overline{\mathcal{X}_A}}$). Therefore, one gets the bounds

$$E[\Omega^0]-E[\Omega^{\mathcal{X}}]<-3C, \tag{A45}$$

$$E[\Omega^0]-E[\Omega^{\overline{\mathcal{X}_A}}]<-3C, \tag{A46}$$

i.e., the states obtained by flipping $\mathcal{X}$ and $\overline{\mathcal{X}_A}$ are both $S_G(\Omega,3C)$. However, given that $S_G(\Omega,3C)$ space is UM and according to Theorem 1, this would oblige the sets $\mathcal{X}$ and $\overline{\mathcal{X}_A}$ to share at least half of the spins of the smaller of them. Since the two sets are by definition disjoint, we reached a contradiction. This reduces to absurd our assumption that there exists more than one $S_G(\mathcal{A},C)$. ■

[1] G. Parisi, Nucl. Phys. B **26**, 181 (1992).

[2] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).

[3] S. Solomon, in *Annual Reviews of Computational Physics II*, edited by D. Stauffer (World Scientific, Singapore, 1995), pp. 243–294; M. Baeker, Ph.D. thesis, DESY, 1995 (unpublished).

[4] M. Harmatz, P. Lauwers, S. Solomon, and T. Wittlich, Nucl. Phys. B (Proc. Suppl.) **30**, 192 (1993).

[5] H.-G. Evertz, M. Hasenbush, M. Marcu, K. Pinn, and S. Solomon, Phys. Lett. B **254**, 185 (1991); M. Agishtein, R. Ben-Av, A. Migdal, and S. Solomon, Mod. Phys. Lett. A **6**, 1115 (1991); N. Persky, R. Ben-Av, and S. Solomon, Phys. Rev. B **51**, 6100 (1995).

[6] D. Stenhill, S. Solomon, and K. Wolowelsky, Comput. Phys. Commun. **83**, 23 (1994).

[7] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[8] R. Ben-Av, H. G. Evertz, M. Marcu, and S. Solomon, Phys. Rev. D **44**, 2953 (1991).
[9] D. Kandel and E. Domany, Phys. Rev. B **43**, 8539 (1991); A. D. Sokal, in *Quantum Fields on the Computer*, edited by M. Creutz (World Scientific, Singapore, 1992), p. 211.
[10] H. Sompolinsky and A. Zippelius, Phys. Rev. B **25**, 6860 (1982).
[11] D. Kandel, R. Ben-Av, and E. Domany, Phys. Rev. Lett. **65**, 941 (1990).
[12] S. Wiseman and E. Domany (unpublished).
[13] N. Kawashima and M. Suzuki, J. Phys. A **XX**, 1055 (1992).
[14] A. J. Bray and M. A. Moore, J. Phys. C **13**, L469 (1980); J. Phys. A **14**, L377 (1980).
[15] N. Kawashima and J. E. Gubernatis, Phys. Rev. B **51**, 1547 (1995).
[16] A. Brandt, R. Ben-Av, and S. Solomon, Nucl. Phys. B **329**, 193 (1990); R. Ben-Av, M. Harmatz, P. G. Lauwers, and S. Solomon, Nucl. Phys. B **405**, 623 (1993).
[17] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **32**, 1792 (1975).
[18] P. Baldi and E. B. Baum, Phys. Rev. Lett. **56**, 1598 (1986).
[19] I. Kanter and H. Sompolinsky, Phys. Rev. Lett. **58**, 164 (1987), and references therein.
[20] M. Mezard and G. Parisi, Europhys. Lett. **3**, 1067 (1987).
[21] A. Brandt, R. Ben-Av, M. Harmatz, E. Katznelson, P. G. Lauwers, S. Solomon, and K. Wolowesky, Phys. Lett. B **253**, 185 (1991).
[22] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[23] N. Persky, I. Kanter, and S. Solomon, Phys. Rev. E **53**, 1212 (1996).
[24] J. R. L de Almeida, C. De Dominicis, and P. Mottishaw, J. Phys. A **21**, L693 (1988).
[25] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).