Journal of Statistical Physics, Vol. 72, Nos. 5/6, 1993

On the Branching Structure of the Tree of States in Spin Glasses

Giorgio Parisi[1]

Received August 27, 1992; final April 29, 1993

We review some known results on the nature of the tree of states in spin glasses and we present new results on its topology. We pay particular attention to the so-called continuum limit in which the levels are labeled by a continuous variable x. We also study the dependence on the level x of the type of branching (bifurcation, trifurcation,...). We show that the statistics of the tree is universal in the continuum limit, i.e., it does not depend on the details of the algorithm used to generate the tree.

KEY WORDS: Spin glasses; equilibrium states; replica symmetry.

1. INTRODUCTION

Mean field theory of spin glasses predicts the existence of an infinite number of pure states. Each state carries a weight w_α, where α labels the state. For each pair α, β of states we can define a distance $d_{\alpha,\beta}$. It satisfies the ultrametricity inequality

$$d_{\alpha,\beta} < \max(d_{\alpha,\gamma}, d_{\gamma,\beta}) \quad \forall \gamma \tag{1.1}$$

This inequality implies that we can associate to each set of states and distances a branching tree, such that the leaves of the tree are the states. Each node of the tree is characterized by a level, and the distance of two states α and β is the height of the highest node we encounter going from α to β. The set of pure states will be therefore characterized by the w's and the d's or equivalently by the w's, by the topology of the tree, and by the levels of the nodes.

[1] Dipartimento di Fisica, INFN Sezione di Roma "Tor Vergata," II Università di Roma "Tor Vergata," Rome 00133, Italy.

822/72/5-6-1

In the mean field theory of spin glasses there is a probability law defined on trees, i.e., each tree (i.e., each set of w's and d's) may appear with a probability assigned by the theory. This probability is defined in a constructive way, i.e., one defines a process to construct it.

In the simplest case one introduces only a finite number of levels. One starts from the highest levels, by associating a weight to each node connected to the root; at the next step the descendants of the node are introduced and an extra weight (by following a specific rule) is assigned to the each of the descendants. The process is iterated once for each level we want to build. In most physical applications we have to consider the so-called *continuum limit* where the number of the levels goes to infinity. This limit is rather tricky and there are only a few descriptions of the resulting tree (including one remarkable paper of Ruelle[17]).

The aim of this paper is to present further investigations on the nature of such a tree which plays a crucial role in the mean field approach to spin glasses. In particular, we will discuss in detail the continuum limit and the topology of the tree. This detailed analysis may be interesting to mathematicians also because this infinite tree has remarkable properties from the probabilistic point of view. We will not be concerned with justifying the correctness of this approach for spin glasses: here we are only interested in describing the resulting theory.

After this introduction, in Section 2 we define equilibrium states (or phases). In Section 3 we present some simple results on the equilibrium states of spin glasses; we introduce ultrametricity and we construct the tree of states. In Section 4 we present some of the known results on the correlations among the distances of the states and their weights. In Section 5 we start to study the continuum limit, while in the following section we analyze the behavior at small x. Finally, in the last section we show how to perform numerical simulations of the tree and under which conditions the resulting probability distribution over the trees is universal in the continuum limit.

2. GENERAL CONSIDERATION ON THE EQUILIBRIUM STATES

In this section and in the next one we will present some simple results that have been obtained for the equilibrium states of the infinite-range spin glasses at low temperatures.[9,11,12] These results concern the structure of configurations which have a nonnegligible probability in the low-temperature region. We do not need to specify in detail the Hamiltonian of the system.

The system contains N spins σ_i, which may take only the two values

-1 or $+1$. Given two configurations σ and τ, we can define their distance as the percentage of the sites where the two configurations differ, i.e.,

$$d \equiv 1/(4N) \sum_{i=1,N} (\sigma_i - \tau_i)^2 \tag{2.1}$$

Such a distance belongs to the interval (0, 1), and it is zero only if the two configurations are equal. An interesting situation corresponds to the limit where the number N of sites goes to infinity. In this case two configurations may stay at a zero distance, if the number of differences remains finite when the total number N goes to infinity. The percentage of different σ's, not the absolute number, is relevant in this definition of the distance.

At a given temperature, the number of relevant configurations is extremely large when the number of variables N goes to infinity [it is proportional to $\exp(NS)$, S being the entropy density]. It is usual to assume that the set of configurations may be partitioned into subsets, called *states*.

States (or phases) are defined as follows.

(a) Each of the states contains an exponentially large number of configurations when N goes to infinity.

(b) More crucially, if we take two different generic configurations belonging to states A and B, respectively, their distance does not depend on the configurations, but only on the choice of A and B.

(c) It should also be true that the distance between two configurations of the same state should be strictly smaller than the distance between two configurations of different states. This last property may be written as

$$d_{A,A} < d_{A,B} \tag{2.2}$$

where $d_{A,B}$ is the distance between states A and B, i.e., the distance between two generic representatives of states A and B.

(d) The classification into states is the finest one which satisfies properties (a), (b), and (c).

The classification into states will depend on the temperature of the physical system. In many cases it can be rigorously proven that the classification into states is possible and unique. States are often called phases in the physical literature. It may be interesting to note that the definition of state may look very similar to the definition of species which is familiar to biologists.

The perplexed reader may find more useful an alternative definition of state. We consider a generic quantity A and we study its fluctuations

$$\langle A^2 \rangle - \langle A \rangle^2 = \langle (A - \langle A \rangle)^2 \rangle \tag{2.3}$$

(we indicate the expectation value of a quantity A by $\langle A \rangle$).

Intensive quantities may be defined as

$$1/N \sum_{i=1,N} \langle A_i(\sigma_i) \rangle \tag{2.4}$$

where the functions A_i depend only on the value of σ_i (they may depend on the site i).

The crucial question is, *do intensive quantities fluctuate*? Intuitively, we would like a negative answer, since intensive quantities are averages over the whole system. But, for example, this is not the case at a first-order transition point, where different phases coexist.

In a ferromagnet at low temperature the most likely configurations contain most of the spins up or down. In this case it is possible to classify configurations according to the direction into which the majority of spins points and to define an average restricted only to this kind of configuration. If we define these restricted averages, the percentage of spins directed in one direction does not fluctuate.

More precisely, if

$$\langle \sigma_i \rangle = \sum_{\{\sigma\}} \sigma_i \exp(-\beta H) \Big/ \sum_{\{\sigma\}} \exp(-\beta H) \tag{2.5}$$

and H is symmetric under the global transformation $\sigma \to -\sigma$, we have that

$$\langle \sigma_i \rangle = 0 \tag{2.6}$$

In this situation intensive quantities do fluctuate. Indeed, if we call Σ the intensive quantity corresponding to the spins (i.e., $\Sigma = 1/N \sum_{i=1,N} \sigma_i$), the expectation value of Σ is zero ($\langle \Sigma \rangle = 0$), while the expectation value of its square is nonzero ($\langle \Sigma \rangle^2 = 0$).

By classifying the configurations in two sets, we can define restricted averages $\langle \cdot \rangle_+$ and $\langle \cdot \rangle_-$ such that

$$\tfrac{1}{2}\langle A \rangle_+ + \tfrac{1}{2}\langle A \rangle_- = \langle A \rangle, \qquad \langle \sigma \rangle_+ = m, \qquad \langle \sigma \rangle_- = -m \tag{2.7}$$

In normal ferromagnetic systems it is possible to prove that intensive quantities do not fluctuate in $\langle \cdot \rangle_+$ and in $\langle \cdot \rangle_-$. This decomposition of a probability distribution in which intensive quantities fluctuate into the linear combination of restricted probability distributions in which the intensive quantities do not fluctuate can be done in many cases in statisti-

cal mechanics. These restricted probability distributions correspond to different states and the states will be identified by the expectation value of intensive quantities.

Generally speaking, we can thus write the decomposition in states as

$$\langle A \rangle = \sum_{\alpha} w_{\alpha} \langle A \rangle_{\alpha} \tag{2.8}$$

where α labels the species and w_{α} is the probability that a configuration belongs to the state α.

Although this decomposition into states can be done only in the infinite-volume limit, it may be useful to extend it to a finite (but large) system. In this case there will be some configurations (with very small probability, e.g., configurations with domains wall among positive and negative magnetization) which could not be classified into states.

Normally the classification into phases is quite poor. For usual materials, in the generic case there is only one phase and the classification is not very interesting. In slightly more interesting cases there may be two states: for example, if we consider the configurations of a large number of water molecules at 0°C we could classify them as water or ice: there are two states. In slightly more complex cases, if we choose carefully external parameters like the pressure or magnetic field we may have the coexistence of three or more phases (tricritical or multicritical points).

In all these cases the classification is rather simple and the number of states is quite small. It was really a surprise when it was discovered that in the infinite-range spin-glass model the number of states is very large: it goes to infinity with N (the number of spin variables) and a very interesting nested classification of states is possible. This behavior implies that the Gibbs rule is not valid for spin glasses. Indeed, the Gibbs rule states that in order to have the coexistence of n phases (n-critical point), we must tune n parameters. Here no parameter is tuned and the number of coexisting phases is infinite!

In the next section we examine the properties of the w's and d's in the mean field approach to spin glasses.

3. SOME RESULTS ON SPIN GLASSES

We now describe the results obtained in the study of spin glasses (which depend on the detailed model). We restrict our analysis to the infinite-range spin glass. As we have seen, it is convenient to introduce the distance d of two states:

$$d_{\alpha\gamma} \equiv = 1/(4N) \sum_{i=1,N} (\sigma_i - \tau_i)^2 \tag{3.1}$$

where τ and σ are two generic configurations belonging to states α and γ, respectively. The distance d is the usual Hamming distance between the two configurations. Other definitions of distance, are possible, e.g., $d^E_{\alpha\gamma} \equiv 1/N \sum_{i=1,N} (E_i^\sigma - E_i^\tau)^2$, where E_i^σ is the energy density of the σ configuration at site i, but no new information is carried by this distance (at least if we neglect the possibility of global sign reversal). In the mean field approach d^E is a function of d.

In spin glasses states do not have microscopically different properties and

$$d_{\alpha,\alpha} = d_{\gamma,\gamma} = D_S \tag{3.2}$$

The quantity D_S is related to the usual order parameter q_{EA} by

$$D_S = (1 - q_{EA})/2 \tag{3.3}$$

Indeed, introducing the state-dependent magnetization ($m_i^\alpha \equiv \langle \sigma_i \rangle_\alpha$), one finds that

$$d_{\alpha\gamma} = 1/(4N) \sum_{i=1,N} (m_i^\alpha - m_i^\gamma)^2 + D_S \tag{3.4}$$

and the order parameter q_{EA} is equal to $\sum_{i=1,N} (m_i^\alpha)^2$ (it does not depend on the choice of the state α). Sometimes a slightly different definition of distance is used (i.e., $\tilde{d}_{\alpha\gamma} = d_{\alpha\gamma} - D_S$) in such a way that $\tilde{d}_{\alpha\alpha} = 0$.

In the mean field approach to spin glasses the w's are random (not independent) variables.[9,13] In the following we will introduce the probability law which generates them.

First of all there is some kind of democracy between states in the sense that states may differ only in the number configurations which they contain and in the distances from the other states. There is no intrinsic difference, e.g., we have seen the variability inside states labeled by α (i.e., $d_{\alpha\alpha}$) does not depend on α: we have called it D_S.

Most interestingly, an explicit computation based on the replica-symmetry-breaking approach shows that the distance among states satisfies the ultrametricity inequality (1.1),[5,9,14] and consequently states may be classified in a taxonomic way.[16] Since natural taxonomy is familiar, it will be convenient to use in this section a biological terminology, and use the word *species* instead of *state*. If we want to recover the usual terminology, we should read *clusters*, *superclusters*, and *supersuperclusters* in place of *genera*, *families*, and *orders*.

For example, one may take a distance D_G, with $D_G > D_S$, and introduce genera $\alpha, \beta, \gamma, \ldots$, in such a way that all species belonging to the same genus are separated by distances less than or equal to D_G, while species

belonging to different genera have a distance greater than D_G. The value of D_G is arbitrary; however, if D_G is very close to D_S, the genera will be quite similar to species and they will contain only a few species, while if D_G is much greater than D_S, the genera classification will be very coarse. It is remarkable, however, that, independent of the value of D_G, the number of species in any genus is infinite (it goes to infinity with N, the number of variables).

In the same way one can introduce families by choosing a distance $D_F > D_G$ and by grouping together families at distances smaller than or equal to D_F. Families are similar to genera (they provide a coarser classification). In particular, each family contains a very large number of genera.

It should be evident to the reader that, if we are not satisfied with the classification into species, genera, and families, we can introduce orders by choosing a distance $D_O > D_F$ and by grouping together genera at distances smaller than or equal to D_O. Orders, too, contain a very large number of families.

We could go on for a while by introducing more levels of classification. Let us stop here and try to clarify some key points. The number of levels and the distances at which levels are defined (D_G, D_F, D_O, and so on) are obviously arbitrary: we could also introduce a classification which is coarser than genera, but finer than families.

Many more results are known if we take into account the number of configurations inside each given species. Let us call w_A the probability that a generic configuration belongs to the species A; the sum over all the species of w_A gives obviously one; w_A is the population in species A, divided by the total population.

If one counts the number of species with $w_A > p$, which we call $N_S(p)$, one finds that for small p it goes like

$$N_S(p) \approx p^{-x_S} \tag{3.5}$$

where x_S is a given number, which characterizes the tree. A power law distribution of frequencies is a quite common phenomenon in nature and it is gratifying that we have found that it appears in this more abstract setting.

In a similar way,

$$\begin{aligned} N_G(p) &\approx p^{-x_G} \\ N_F(p) &\approx p^{-x_F} \\ N_O(p) &\approx p^{-x_O} \end{aligned} \tag{3.6}$$

where $N_G(p)$ is the number of genera having a percentage of population greater than p, the population of a genus being defined as the sum of the populations of the species that belong to that genus. $N_F(p)$ and $N_O(p)$ are defined in an analogous way.

Mathematical consistency requires

$$x_S \leqslant x_G \leqslant x_F \leqslant x_O \tag{3.7}$$

The situation may be similar to the one described in Fig. 1, where we represent the classification into species, genera, and families as a tree.

In Fig. 1 the leaves of the tree are the species and the distance between different species is represented by the level we have to reach to find a path joining the two species. If the quantities D_G and D_F are the ones represented by the two thin horizontal lines, the classification into genera and families is represented by the wide horizontal lines below the tree. The 42 species are thus classified into 15 genera and 7 families. It is evident that the levels which we have chosen to define genera and families are arbitrary.

The tree of species in spin glasses is much more dense and branched than the one of Fig. 1; it is indeed an infinitely branched tree. Branches are present at any level and there is an infinite number of branches in any interval of distance. The very mathematical existence of such an object is far from being trivial and it has been proved only recently by Ruelle.[17]

The full characterization of such a tree is given by a function $x(D)$ such that

$$x_G = x(D_G) \tag{3.8}$$

This function tells us how the exponent x changes when one changes the definition of genera (the same function plays the equivalent role for families and orders).

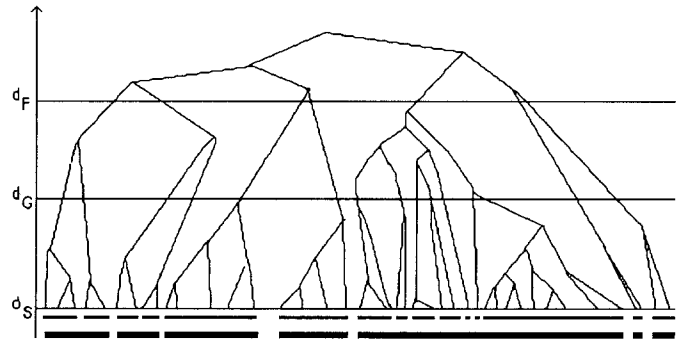

Fig. 1. An example of a tree, with the associated classification.

Different trees have different functions $x(D)$, as they depend on the precise nature of the Hamiltonian. In simple cases $x(D)$ can be computed: for example, this has been done in the infinite-range spin-glass model[9] and it has been found that $x(d)$ depends on the temperature.

This infinitely branched tree, in spite of its apparent complications, is the simplest mathematical structure for an infinitely branched classification and it is rather likely to be relevant [with different forms of the function $x(D)$] in many physical and possibly biological problems.

4. MORE INFORMATION ON THE TREE

In this section we provide some explicit formulas describing the distribution of the weights. Starting from them, the formulas of the previous section were derived. First we introduce some variables f_α for each species α, which we may call the free energy of the species α. The w's can be computed from the f's as follows:

$$w_\alpha = \exp(-f_\alpha) \Big/ \sum_\gamma \exp(-f_\gamma) \tag{4.1}$$

It is evident that if we shift all the f's by the same amount, the w's are left invariant.

In the mean field approach to spin glasses the f's are independent random variables and the probability of finding a state with a free energy in the interval $f, f + df$ is given by

$$dP = \exp[x_S(f - \tilde{f})]\, df \tag{4.2}$$

Equation (4.2) gives the probability distribution of the frequency of species and is essentially equivalent to Eq. (3.5).

We must now state the probability distribution of the d's and the correlations between the d's and the w's. Moreover, we would like to know if the species can be classified in some useful manner. It turns out that the following inequality is satisfied:

$$d_{\alpha,\gamma} \leqslant \max(d_{\alpha,\beta}, d_{\gamma,\beta}) \quad \forall \beta \tag{4.3}$$

Condition (4.3) tells us that the space of species is ultrametric; it amounts to saying that, for any choice of the number D, species may be classified into nonoverlapping clusters (of diameter D) such that α and γ belong to the same cluster if and only if $d_{\alpha\gamma} \leqslant D$. Clusters may be decomposed into subclusters of smaller diameter and so on. The whole distribution of states is characterized if we specify the clusters and their overlaps.

Ultrametricity is the crucial property that allows a taxonomic classification. A space is ultrametric if two *spheres* of the same diameter coincide if they have at least one point in common: if α belongs to the same family as β ($d_{\alpha,\beta} < D_F$) and β belongs to the same family as γ ($d_{\beta,\gamma} < D_F$), then α belongs to the same family as γ ($d_{\alpha,\gamma} < D_F$).

The description of the combined probability distribution of distances and frequencies can be obtained first in the simplest case where we suppose that the distances may assume a finite number of values, which we indicate by d_j, $j = 0, 1, \ldots, k$ (for simplicity we suppose that the d's are an increasing sequence). We will write down the formulas for $k = 3$; the generalization to arbitrary k is evident. At the end we will perform the limit $k \to \infty$, where the distance will take a continuous range of values.

If $k = 3$, the distance among an arbitrary pair of species is at maximum D_F; states at a smaller distance belong to the same family: if the distance is D_G (not D_F), they belong to the same genus; their distance may be equal to $D_S \equiv 0$ only if they coincide.

We now describe the frequency distribution of species and of genera inside the families and of the species inside the genera: the rules are so simple that they can be easily generalized to an infinite hierarchy.

The number of families diverges with N: we label them with an index α_3 and we associate to each of them a *free energy* f_{α_3}. The f's are random independent variables, whose probability distribution is

$$\exp[x_F(f_{\alpha_3} - \tilde{f})]\, df_{\alpha_3} \tag{4.4}$$

In each of the families there are infinitely many genera of diameter D_G; we label them with a pair of indices α_3, α_2 and we associate to each of them a "free energy" f_{α_3,α_2}: the f's are random independent variables, whose probability distribution is

$$\exp[x_G(f_{\alpha_3,\alpha_2} - f_{\alpha_3})]\, df_{\alpha_3,\alpha_2} \tag{4.5}$$

The parameter x_G satisfies $x_G < x_F$.

Finally the states are labeled with three indices $(\alpha_3, \alpha_2, \alpha_1)$ and the probability distribution of the associated free energies (which are once again random independent variables) is

$$\exp[x_S(f_{\alpha_3,\alpha_2,\alpha_1} - f_{\alpha_3,\alpha_2})]\, df_{\alpha_3,\alpha_2,\alpha_1} \tag{4.6}$$

where $x_S < x_F < x_G$. Integrating over the free energies of the families and of the genera, one finds that the probability distribution of the free energies of the states is given by Eq. (4.2).

It can be proved that these last equations imply Eq. (3.6).

This construction may be generalized to any value of k. In this way we obtain a distribution of the probability of the states and of their distances, the only parameters being the D's and the x's. We finally send k to infinity in such a way that the D's become dense in the interval D_S, D_M. In this way we obtain a monotonically decreasing function $x[D]$, which is defined in the interval $D_S - D_M$: the distribution of phases and of their distances depends only on this function $x[D]$.

It is not evident whether this construction survives in the limit k going to infinity, where an infinite number of levels is present. Fortunately, a detailed analysis by Ruelle shows that this construction really defines an infinitely branched tree. In the next section we will discuss in greater detail the properties of this infinitely branched tree.

Before starting this task, we remark that the notations we have used up to now are slightly different from the usual one to which we will shift in the next section.

We first introduce a factor β in the free energy [$w \propto \exp(-\beta f)$]. Equation (4.2) is not changed, but the new value of x_S is obtained by dividing the old value by a factor β.

Instead of the function $x(d)$, we will consider its inverse $d(x)$ [x being a variable defined in the interval (0, 1)], so that we will identify the clusters not by their distance, but by the corresponding value of x. We notice that changing the function $d(x)$ is just a redefinition of the distance, so that the specific form of the function $d(x)$ is essentially irrelevant for all the properties of the tree. Without loss of generality we can assume that the root of the tree is located at $x = 0$ and the states at $x = 1$. If the true value of x_S (which is usually called x_M) is smaller than 1 [i.e., $x(d) \leqslant x_S$] and the states are located at $x = x_S$, the function $d(x)$ will be flat (i.e., equal to D_S) for $x_M < x < 1$.

5. HOW TO DEAL WITH INFINITELY BRANCHED TREES

The limits in which the number of levels goes to infinity seem rather complex and very difficult to control analytically. However, the number of states with $w > p$ increases as p^{-x_M}. If we observe the system with resolution ε, in the sense that we neglect states whose weight is less than ε, the total number of observed states increases as a power of ε, while the total weight of these states is very close to 1 when ε goes to zero, i.e., it behaves as $1 - \text{const} \times \varepsilon^{(1-x_M)}$.

After the introduction of the resolution ε, the tree acquires only a finite number of branches. We can ask now more precise questions about the probability of having a branching of order k ($k = 1$ is no branching, $k = 2$ is a bifurcation,...). It clear that these probabilities are ε dependent and

have a singular limit when ε goes to zero. This problem of dealing with the limit ε going to zero, where an infinite number of states is present, has (quite amusingly) many technical points in common with the resummation of infrared singularities in quantum electrodynamics and in QCD.

A certain number of quantities may be computed by analytic methods, although in certain cases the computations may be very involved.

The formalism of the previous section is good enough for doing many computations, but it is not the most appropriate for many purposes. The main drawback is that the reference free energy of a cluster is not related in a deterministic way to the weight of a cluster, which is defined as the sum of the weights of all the states belonging to that cluster. Indeed, clusters which have the same reference free energy will have different total weight. Moreover, if we change the value of x_M by an infinitesimal amount, the weight of the clusters changes by a finite amount.

It turns out that there is an alternative procedure to construct the tree in which the weight at each level is fixed by construction at the intermediate stages and not computed at the end.

The first step consists in defining the free energy of the clusters as in the previous case; the difference with the previous approach is that the weights of the clusters are computed at the beginning: $w_c \propto \exp(-\beta f_c)$. (We recall that the weight of a cluster is the sum of the weight of all the states which belong to that cluster.)

We now suppose that the value of x corresponding to the clusters is c and that the level of the are subclusters s, with $c < s$ as usual. We want to compute the probability distribution of the weights of the subclusters. To this end one introduces M variables f_i, for $i = 1, M$. These f's are not uncorrelated as in the previous case and their conjoint probability distribution is proportional to

$$\exp\left(\beta s \sum_{i=1,M} f_i\right) \Big/ \left[\sum_{i=1,M} \exp(-\beta f_i)\right]^c \tag{5.1}$$

If we set $c = 0$, we recover an independent distribution. The weights of the subclusters are given by

$$w_i = w_c \exp(-\beta f_i) \Big/ \left[\sum_{i=1,M} \exp(-\beta f_i)\right] \tag{5.2}$$

The f_i have the physical meaning of the free energy of the subclusters at fixed value of the free energy of the cluster.

It can be seen that the denominator in Eq. (5.1) is essential to recovering an independent distribution of the f's at level s, as far as it cancels the effect of the correlation of the weights at level c.

It is a rather surprising fact that the two methods give exactly the same probability distribution, as can be verified by explicit and long computations. The first and second methods were introduced in refs. 8 and 7, respectively. The first method is simpler from an abstract point of view, while the second one has the advantage that the weight of the clusters we introduce is the sum of the weights of the states belonging to the cluster, not a reference value for extracting the weights of the states. The second method is much more appropriate if we introduce a cutoff in the weights of the states, because this induces a cutoff in the weights of the clusters. Indeed, in the second method it is possible to neglect clusters with a small weight, because all the states inside this cluster must have a small weight (less than that of the cluster). In the first method it is not legal to keep at an intermediate step only a fixed number of clusters (i.e., those with smallest free energy) when we go to the continuum limit, because in the generic case the states which have the largest weight do not belong to the clusters with lowest free energy.

Equation (5.2) can be derived from the following equations for the probability distribution of the w's. Indeed, the probability distributions of a finite number of w's have a simple form. An explicit computation using the replica method shows that the probability for finding k subclusters ($i=1, k$) inside the cluster with weight w_i is given by

$$s^{(k-1)}\Gamma(1-c)\,\Gamma(k-c/s)/[\Gamma(1-c)^k\,\Gamma(ks-c)\,\Gamma(1-c/s)] \times \prod_{i=1,k} wr_i^{-1-s}(1-wr_T)^{-1+ks-c} \tag{5.3}$$

where the wr's are the weights of the subclusters normalized to the cluster ($wr_i = w_i/w_c$) and wr_T is given by $\sum_{i=1,k} wr_i$. Of course the probability vanishes in the region where wr_T is greater than 1. Indeed, Eqs. (5.1) and (5.2) were proved starting from (5.3).

If we set $c=0$, we obtain the probability of finding k states ($i=1, k$) with weight w_i:

$$[s^{(k-1)}/\Gamma(ks)] \prod_{i=1,k} w_i^{-1-s}(1-w_T)^{-1+ks} \tag{5.4}$$

It may be interesting to note[4] that for reasons that are beyond our command the probability distribution (5.3) of the random mapping model[4] is given by (5.3) in the strange case $c=-1/2$, $s=0$.

Starting from (5.1)–(5.3), we can generate the tree hierarchically beginning from the root and going down to the leaves.

The first question is what happens from the probabilistic point of view when we go to the continuum limit in which infinitely many levels are

present, but we keep the resolution ε nonzero so that the total number of states is finite. The situation is rather clear from the mathematical point of view, as far as the total number of leaves is finite.

To this end it is interesting to consider the case in which the levels are separated by an infinitesimal spacing δ. In other words, we have to study the case in which the level of subclusters s is very close to the one of the clusters, i.e., $s = c + \delta$. We should find that (in the case of finite resolution) with a probability which is equal to $1 - A\delta$, there is only one subcluster; this is absolute necessary if the final tree contains a finite number of branches and consequently we want to define a transition probability which is proportional to the interval of x.

The key factor in expression (5.3) is $\Gamma(k - c/s)/\Gamma(1 - c/s)$, which is proportional to $s - c$ (i.e., δ) as soon k is greater than one. The branching probability in two or more states is given by the integral of (5.3) over the w's. If we assume that a cutoff is present and thus the integral over the w's is finite, the final result vanishes linearly with δ. This conclusion is confirmed by the distribution of the single wr, which according to the previous formula is given (neglecting high-order terms in δ) by

$$\delta wr^{-1-s}(1-wr)^{-1+\delta} = \delta(wr-1) + O(\delta) \tag{5.5}$$

The reader should note carefully that Eq. (5.5) is correct only in the distribution sense. Moreover, if the cutoff is removed, the branching probability is no longer proportional to δ because there is a divergence arising from the integral over the w's, which diverge in the limit ε going to zero.

The natural question to ask is whether all branching processes are bifurcations or whether elementary trifurcations (or higher-order branches) are also present. We can answer this question in the following way. If there are only bifurcations, the probability for having at least three different subclusters should be proportional to δ^2. One can readily see that this is not true: such a probability contains terms linear in δ.

The same conclusions could be reached by considering the previous probability distributions. If bifurcations make the leading contribution, the probability distribution should be symmetric in the exchange $wr \leftrightarrow 1 - wr$. This is not true. An explicit computation for the probability distribution for finding two substates with weights wr_1 and wr_2 shows that it is given by (at the leading order in δ)

$$[\delta/\Gamma(1-c)]\, wr_1^{-1-c} wr_2^{-1-c}(1-wr_T)^{-1+c} \tag{5.6}$$

and it is not concentrated at $wr_T = 1$, as would happen if only bifurcations were present.

We have thus shown that as soon as a branching happens bifurcations are not a privileged process. This can be intuitively understood by looking to the probability distribution in Eq. (5.1).

Let us assume that there is one state which has free energy much smaller than the others (let us take $i=1$ for simplicity); in this region the probability distribution simplifies to

$$\exp(\beta\delta f_1)\exp\left(\beta s \sum_{i=2,M} f_i\right) \tag{5.7}$$

We thus expect that f_1 is of order $-1/\beta\delta$, while the other f's are of order $-1/\beta s$. Therefore in most of the cases f_1 will be much smaller than the others and no branching will be seen (apart from branches with exponentially small weights). On the other hand, when f_1 is comparable to the others f's (which have a spacing among themselves which is of order $1/\beta s$) many states will give a nonnegligible contribution.

It may be interesting to note that in the region where c is close to zero, bifurcations become the dominant process, as can be seen by the approximate symmetry of the function in Eq. (5.5) and by the fact that the spacing among the levels is of order $1/c$. This result is confirmed by the appearance in Eq. (5.6) of a delta function of argument $(1-wr_T)$ in the limit c going to zero.

6. THE LOW-x EXPANSION

The phenomenon which makes the whole approach rather complex is the divergence in the total number of states so that the tree contains an infinite number of branches. This is very similar to what happens in quantum electrodynamics, where the total number of emitted photons diverges at low energy. In order to reach a better understanding, it is convenient to study the system in the approximation where only bifurcations are present, as happens at low values of x. Before removing this approximation (as we shall see in the next section), it is convenient to study carefully the results obtained neglecting trifurcations.

It is interesting to use this approach to perform some analytic computations in this situation. This may be considered as a preliminary step for doing more complex computations in the region where high-order processes like trifurcations cannot be neglected.

We assume that the probability for having a bifurcation of a state into two, the initial state having weight I, the two final states wI and $(1-w)I$, respectively, is

$$dB_2(w) = dx\,[b_2(w) + b_2(1-w)] \tag{6.1}$$

The function b_2 can be found directly by computing the probability for finding two states with weights w_1 and w_2 inside a cluster of weight 1 at x. From (5.3) we have

$$dP(w_1, w_2) = \{dx/[\Gamma(1-x)\,\Gamma(x)]\}\; w_1^{-1-x} w_2^{-1-x} (1-w_1-w_2)^{-1+x}$$
$$\approx dx\; w_1^{-1} w_2^{-1} \delta(1-w_1-w_2) \tag{6.2}$$

where we have neglected terms which vanish when x goes to 0.

We thus find that

$$b_2(w) = 1/w \tag{6.3}$$

We can check the correctness of the result as follows. For small x we should have that the probability of having a state with weight w is given by

$$p_1(w, x) = x[b_2(w) + b_2(1-w)] \tag{6.4}$$

Previous equations tell us that

$$p_1(w, x) = [1/\Gamma(x)]\; w^{-1-x}(1-w_T)^{-1+x} \tag{6.5}$$

in perfect agreement with Eq. (6.3).

Let us assume for the time being that Eq. (6.1) is correct for all x and let us use it to compute the corresponding function $p_1(w, x)$. It is easy to check that, if only bifurcations are present, the function $p_1(w, x)$ satisfies the following differential equation:

$$\partial p_1(w, x)/\partial x = -C p_1(w, x) + \int_w^1 dy/y\; p_1(y, x)\, b(w/y)$$
$$= -C p_1(w, x) + \int_0^1 dz/z\; p_1(w/z, x)\, b(z) \tag{6.6}$$

where the constant C represents the total probability for having bifurcations and $b(z) = 1/z(1-z)$ is a compact notation for $b_2(z) + b_2(1-z)$. As far as C is infinite, it is convenient to write down Eq. (6.6) as

$$\partial p_1(w, x)/\partial x = -C' p_1(w, x)$$
$$+ \int_0^1 dz/z\; [p_1(w/z, x)\, b(z) - p_1(w, x) z/(1-z)] \tag{6.7}$$

the integral over z being finite. The constant C' may be computed requiring that the average of w is 1 (the sum of the w's is indeed 1):

$$\int_0^1 dw\; w p_1(w, x) = 1 \tag{6.8}$$

It turns out that $C' = 0$.

Equations similar to (6.7) are very familiar in high-energy physics[14] and they are conveniently studied using the Mellin transform.

We define

$$\tilde{p}_1(s, x) \equiv \int_0^1 dw\, w^s p_1(w, x)$$

$$\tilde{b}(s) \equiv \int_0^1 dz\, z^s [b(z) - 1/(1-z)] \tag{6.9}$$

$$= -\gamma(s) + \gamma(1) = -\sum_{n=1, s-1} 1/n$$

where the function γ is the logarithmic derivative of the Euler Γ function.

It easy to see (using the convolution theorem for the Mellin transform) that $\tilde{p}_1(s, x)$ is simply given by

$$\tilde{p}_1(s, x) = \exp[x\tilde{b}(s)]/(s-1) \tag{6.10}$$

The function p_1 can be obtained by inverse Mellin transform. It is easy to check that $p_1(w, x)$ is proportional to $(1-w)^{-x}$ for w near to 1 (in agreement with the exact results), while it is less singular than $w^{-1-\varepsilon}$ for any positive ε for w near to zero (in disagreement with the exact results).

We have just seen that the approximation of neglecting trifurcations is enough to reproduce the leading singularity of the distribution functions close to $w=1$, while it is not enough to control the behavior near $w=0$, where trifurcations and high-order processes will play an important role.

In a similar way we can include the contribution of trifurcations and high-order processes. Indeed, at the leading order in x the probability of finding n states at $x+dx$ inside a cluster at level x is given by

$$dx\, x^{(n-1)} \Gamma(n-1) \prod_{i=1,n} w_i^{-1-x} \delta\left(\sum_{i=1,n} w_i - 1\right) \tag{6.11}$$

One finally finds that an n-furcation process has a probability proportional to x^{n-1}. It is therefore possible to develop a consistent expansion at low x neglecting n-furcation processes with large n. In particular it is possible to extend the computation which we have just done of the function $p_1(w, s)$ and to include the presence of trifurcations.

Another interesting computation we can perform is the following. For each realization of the tree and level x we can define the quantity

$$Y(x) \equiv \sum w_c^2 \tag{6.12}$$

where the sum is done over all the clusters at level x, w_c being the probability of each cluster. The probability distribution of $Y(x)$ [i.e., $P(Y)$] has been investigated in detail. Some of the moments have been explicitly computed[00] (e.g., $\langle Y \rangle = 1 - x$) and closed formulas exist for all the moments.[10] It has also shown by Derrida that the function $P(Y)$ is singular at $Y = 1/n$ for any integer n, so that its properties are rather complex.

We now compute the probability distribution $P(Y)$ for small x using the approach of the previous section. As far as only one bifurcation may have happened for small x, if w is the weight of a state, $1 - w$ should be the weight of the other states. In this situation Y is given by $w^2 + (1-w)^2$.

We thus find that

$$\langle Y^k \rangle = 1 - C'x + x \int_0^1 dw\, (1/w)\{w^2 + (1-w)^2]^k - 1\} \tag{6.13}$$

where the constant C' is fixed by the condition that $\langle Y \rangle = 1 - x$ and it turns out to be equal to 0. It is easy to verify that this result agrees up to $k = 6$ with that quoted in Mezard *et al.*[00]

7. ON UNIVERSALITY

At first sight it seems useless to do numerical simulations of the tree (i.e., to produce explicit realizations of it) in the presence of the large number of quantities that we can compute analytically. However, it is certainly interesting to discover if pictures of the tree like the one presented in Fig. 1 are realistic or not. Moreover, there are many quantities that at the present time we are not able to compute analytically.

Let me mention an example of an analytically hard computation which arises in the theory of directed polymers. We consider a function $f(x)$ which diverges as $x^{-\lambda}$ when x goes to zero. For each realization of the tree the function f evaluated on the tree is equal (by definition) to

$$\sum_{\alpha,\gamma} w_\alpha w_\gamma f(x_{\alpha\gamma}) \tag{7.1}$$

where the sum runs over the states and $x_{\alpha\gamma}$ is the level at which the two states α and γ separate [in other words $x_{\alpha\gamma} = x(d_{\alpha\gamma})$].

The evaluation of the expectation value of f averaged on all the trees is rather simple to compute and one gets

$$\langle f \rangle = \int_0^1 dx\, f(x) \tag{7.2}$$

In a similar way we get that

$$\langle f^2 \rangle = 1/3 \int_0^1 dx\, f^2(x) + 2/3 \left[\int_0^1 dx\, f(x) \right]^2 \tag{7.3}$$

At the present moment closed formulas for $\langle f^k \rangle$ are not available for generic k.

In the case in which $1 < \lambda < 2$, the expectation value of $\langle f \rangle$ is divergent; however, a close analysis shows that for most of the trees f will be finite[6] and the divergence of $\langle f \rangle$ arises as an effect of the average over the trees. The question of the computation of the most likely value of f, or equivalently of the average of its logarithm ($\langle \ln f \rangle$), is well posed and interesting, but it is not easy to answer analytically.

In this paper we will not present any numerical simulation of the tree, but we will discuss the general setting in which it can be done. We will find that the tree that we have obtained in the continuum limit is universal, i.e., its statistical properties do not depend on the way it is generated.

There is a relatively easy way to generate the tree (if the maximum value of x is not too large). One introduces a small step δx, and for each δx one generates M free energies according to the distribution (5.1). This should not be too difficult and it can be done, for example, using a Monte Carlo technique with an appropriate starting condition on the f's [e.g., $M-1$ free energies are uncorrelated and the last one is generated at a distance of order $1/(\delta x)$ from the smallest of the first $M-1$ free energies].

We introduce a cutoff and we remove all the branches with probability less than ε. If M is sufficiently large (this condition may be crucial for x close to 1, but not for small x; for x near to zero we can take $M=2$) and δx and ε are small, we can obtain very accurate results. The computational time is polynomial in all the parameters: the number of branches which are relevant at each stage of the computation is certainly smaller than ε^{-1}, and it would be much smaller than ε^{-1} in most of the cases.

If we follow this prescription it seems that there are no problems. It is, however, interesting to see what happens if we use the other method (in which the free energies are used only as reference point) to generate the tree. We introduce a step δx and for each step we introduce M branches. In this way we obtain a Cayley tree with coordination number $M+1$ and diameter $L = 1/\delta x$.

The states are on the surface of the Cayley tree and for each point of the surface (for each state) there is a unique path joining it to the origin (the root). On each node of the tree we extract a free energy f (the normalization of the free energy is slightly different from the previous one because we have absorbed a factor x in it), which is randomly distributed.

The f's are uncorrelated random variables; the probability of finding one in the interval $f - f + df$ is given by

$$dp(f) = \exp(f)\, df \tag{7.4}$$

for f smaller than a cutoff value.

The states (labeled by α) are at the surface of the Cayley tree and their weight is just proportional to

$$w_\alpha = \exp\left(- \sum_{i=1,L} f_i / x_i \right) \tag{7.5}$$

where $x_i = i\,\delta x$, and f_i is the free energy of the ith node which we have to cross from the origin of the tree to the state.

Summarizing, we have a Cayley tree (i.e., the substratum) which has an exponentially large number of branches. After that a weight of the branches is introduced according to Eq. (7.5) and we cut off all the branches with weight less than ε; we are left with a tree with a finite number of branches, which is the object of our study.

The reader should note that we have two trees:

(a) The substratum, which has infinitely many leaves and is fixed.

(b) The final probabilistic tree, which may be well approximated by a tree with a finite number of leaves.

In principle we have to stick to the form of $p(f)$ given by (7.4) and send M to infinity and δx to zero (L to infinity). In this way the number of points on the surface of the Cayley tree would be exponentially large, i.e., M^L, and we do not have the shortcut of considering only a small subset, because we cannot judge from the free energies f near the root which would be the weight of the states on the surface.

Although this method is not very efficient, it is rather interesting because, as we shall see now, the results in the limit where δx goes to zero do not depend on the form of $p(f)$ and on the value of M. In other words, the final probability distribution for the tree in the continuum limit does not depend on the details of the substratum and it is therefore a universal property of a wide class of processes which lead to a tree in the continuum.

The formulas we have just written for the weights describe a random directed polymer on a Cayley tree with a temperature (x) which depends on the distance from the origin.

As a preliminary computation let us consider the case where we remove the dependence of the x's on i and we set $x_i = m$. We recover the usual random directed polymer on a Cayley tree at fixed temperature, which has been carefully studied by Derrida and Spohn.[2] They find by

explicit computations that the statistics of the weights in the limit L going to infinity is independent of L and is given by Eqs. (4.1)–(4.2) for an appropriate value of x_S, which depends on m, M, and $p(f)$.

One point to which we should pay attention is that the states are no longer defined as the leaves of the tree, but all the leaves that are separated by a finite amount of branches (in the limit when L goes to infinity) belong to the same state. States are therefore obtained by putting together all leaves at a finite distance. In other words, for random directed polymers on a Cayley tree the thermodynamic limit of the distribution of the weights of the endpoint is independent of the shape of the tree, of the distribution of the randomness, and of the microscopic temperature (m). Everything combines in the parameter x_S.

The same results may be obtained using the replica method and showing that the model is described by a order parameter matrix Q with one-step replica symmetry breaking.[3]

In the same way we can extend these computations to the case where the temperature x_i is weakly dependent on i. One would find the same results as before with an x which depends on i/L. The only difference would be that the resulting value of x would no longer be the microscopic one (i.e., i/L), but a different one ($x = Ai/L$), where the constant A depends on M and $p(f)$ and can be computed using the formulas of Derrida and Spohn.[2]

It is rather amusing that we can use the replica method to solve a statistical model of random directed polymers which is the generalization of the tree of states obtained from the replica method itself.

We conclude by observing that the behavior of the probabilistic tree does not depend on the properties of the substratum and in this sense it is a universal properties of all models in which the continuum limit is obtained as a very large sequence of small steps.

ACKNOWLEDGMENTS

It is a pleasure to thank M. Mezard for many useful and interesting discussions and E. Marinari for a careful reading of the manuscript.

REFERENCES

1. B. Derrida, *J. Phys. Lett.* (Paris) **46**:401 (1985).
2. B. Derrida and H. Spohn, *J. Stat. Phys.* **51**:817 (1988).
3. B. Derrida and E. Gardner, *J. Phys. C* **19**:5787 (1988).
4. B. Derrida, M. Mezard, G. Parisi, and N. Sourlas, unpublished (1991).
5. S. Franz, G. Parisi, and M. A. Virasoro, Rome preprint (1992).
6. M. Mezard and G. Parisi, *J. Phys. I* (Paris) **1**:809 (1991).

7. M. Mezard, G. Parisi, and M. A. Virasoro, *J. Phys. A* **46**:L223 (1985).
8. M. Mezard, G. Parisi, and M. A. Virasoro, *Europhys. Lett.* **1**:77 (1986).
9. M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
10. M. Mezard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, *J. Phys.* (Paris) **45**:843 (1985).
11. G. Parisi, in *Field Theory and Statistical Mechanics*, J. B. Zuber and R. Stora, eds. (North-Holland, Amsterdam, 1984).
12. G. Parisi, *Physica Scripta* **35**:123 (1987).
13. G. Parisi, in *Chance and Matter*, J. Soulietie, R. Stora, and J. Vannimeneus, eds. (North-Holland, Amsterdam, 1988).
14. G. Parisi, *Field Theory, Disorder and Simulations* (World Scientific, Singapore, 1992).
15. G. Parisi, F. Ritort, and F. Slanina, Rome preprint (1992).
16. R. Rammal, G. Toulouse, and M. A. Virasoro, *Rev. Mod. Phys.* **58**:765 (1986).
17. D. Ruelle, *Commun. Math. Phys.* **48**:351 (1988).