Ultrametricity for physicists

R. Rammal

Centre de Recherches sur les Très Basses Températures, Centre National de la Recherche Scientifique, BP 166 X, 38042 Grenoble-Cédex, France

G. Toulouse

Ecole Supérieure de Physique et Chimie Industrielles, 10 rue Vauquelin, 75231 Paris, France

M. A. Virasoro

Dipartimento di Fisica, Universita di Roma I, I-00185 Roma, Italy

Ultrametricity is a simple topological concept, but its appearance in the language of physicists is recent. This review provides all the elementary background (from mathematics, taxonomy, and statistical physics) and surveys the main fields of development (spin glasses, optimization theory). Static and dynamic aspects are covered. From present knowledge, one can already draw some tentative conclusions on the common causes for the occurrence of ultrametric structures in nature. Some perspectives on unresolved problems in physics and biology are also presented.

CONTENTS

I. INTRODUCTION

The first mention, to our knowledge, of the word ultrametricity in a physics journal is quite recent (Mézard *et al.*, 1984a,1984b). The occasion was an investigation of the mean-field theory of spin glasses, and the discovery of an ultrametric structure for the distribution of pure states in configuration space came as a surprise. None of the authors had previous knowledge of this mathematical concept, despite its basic simplicity and the fact that it had been maturing in mathematics since the turn of the century.

Several times in the past (in the case of relativity, quantum mechanics, and topological defects, to name a few) a similar surprise has occurred and intrigued physicists and laymen alike: the observation that mathematical concepts and tools, useful for the development of physics, had already been invented and elaborated by mathematicians, for purely theoretical reasons and without any preconception of their possible applications.

The advent of ultrametricity in physics is perhaps a relatively minor event, compared to other breakthroughs. No hard mathematical formalism is involved. But perhaps for this reason, because the concepts are easy to visualize and pervasive in many sciences (statistical physics, optimization theory, biology) (Toulouse, 1984), it is an amusing story to tell. And some readers, like Monsieur Jourdain, will acquire a new appreciation of already familiar lore.

Because ultrametricity is a newcomer in physics, this review gives unusual attention to history and perspectives. The mathematical background is introduced through its historical development (Sec. II). Perception of the practical relevance of ultrametric structures emerged first in taxonomy (Sec. III). Emphasis is given to the theory of spin glasses, which is an archetype of the physics of frustrated disordered systems, and which revealed the importance of nontrivial ultrametricity. Accordingly, Sec. IV is our "pièce de résistance." Combinatorial optimization problems, of which the traveling salesman problem is an

archetype, and other complex problems in engineering and biology have energy landscapes that are amenable to similar analysis (Sec. V). Basic concepts of classification theory and measurement of deviations from ultrametricity are presented in Sec. VI. Random walks on regular ultrametric spaces and simulated annealing dynamics are discussed in Sec. VII. Section VIII offers perspectives on some open biological problems. Finally, in Sec. IX, the sources for ultrametricity in the natural sciences are summarized.

II. ULTRAMETRICITY IN MATHEMATICS

A. Definitions

A metric space is a space endowed with a distance. A distance, in general, obeys the triangular inequality

$$d(A,C) \leq d(A,B)+d(B,C) , \tag{2.1}$$

where A,B,C are any three points of the space.

The *ultrametric inequality* is a stronger inequality:

$$d(A,C) \leq \mathrm{Max}\{d(A,B),d(B,C)\} . \tag{2.2}$$

A distance that satisfies the ultrametric inequality is called an *ultrametric distance.* A space endowed with an ultrametric distance is called an *ultrametric space.*

B. History

The following historical sketch owes a great deal to the *Abrégé d'histoire des mathématiques,* written under the direction of Jean Dieudonné, one of the founders of Bourbaki (Dieudonné, 1978). Useful insights from Benzécri (1984) and invaluable guidance from Amice (1975) are gratefully acknowledged.

A natural starting point for the present review is the article published in 1897 by Kurt Hensel (1861—1941), introducing the notion of p-adic numbers. This was part of algebraic number theory. Kronecker (a professor at Berlin, under whom Hensel studied) had published in 1882 a famous memoir on the foundations of this recently revigorated branch of mathematics. However, Hensel's idea was so novel and unexpected that it remains in the history of mathematics as a famous example of work developed in almost complete isolation. Only fifteen years later did the situation begin to change, with the introduction of simple topological notions in the field of p-adic numbers. This evolution is representative of the invasion of many branches of mathematics by the language of geometry. Very much the same trend, perhaps less surprisingly, has taken place in physics.

Let us consider the formal sums

$$\pm \sum_{i=0}^{\infty} a_i p^i \quad \text{with } 0 \leq a_i \leq p-1 ,$$

where the number p and the coefficients a_i are natural integers. Such a sum represents an integer x if

$$x \equiv a_0 + a_1 p + \cdots + a_i p^i \pmod{p^{i+1}} \quad \text{for all } i .$$

Actually it can be verified that any integer, positive or negative, can be represented by a sum $\sum_{i=0}^{\infty} a_i p^i$, with $0 \leq a_i \leq p-1$. But an arbitrary infinite series does not always represent an integer.

Hensel realized that, with the inclusion of negative exponents, he could represent not only integers, but also rational numbers. Thus a rational number can be represented as a formal sum

$$\sum_{i=r}^{\infty} a_i p^i \quad \text{with } 0 \leq a_i \leq p-1, \quad a_r \neq 0 ,$$

where r is a relative integer (positive, negative, or zero); such a representation is unique. The set of all formal sums is denoted Q_p and is called the field of p-adic numbers; it contains Q, the usual field of rational numbers, but is distinct from it.

At the beginning of the century, the notations of topology were spreading and, in 1906, M. Fréchet introduced the general notion of metric space. A metric space E is a space for which a distance function $d(x,y)$ is defined for any pair of elements (x,y) belonging to E. The distance function takes non-negative values and satisfies three properties (obeyed by the usual Euclidean distance),

$$\begin{aligned} &\text{(i)} \quad d(x,y)=0 \Longleftrightarrow x=y , \\ &\text{(ii)} \quad d(y,x)=d(x,y) , \\ &\text{(iii)} \quad d(x,z) \leq d(x,y)+d(y,z) ; \end{aligned} \tag{2.3}$$

this last property is the triangular inequality (2.1).

An example of such a distance is the p-adic distance on the field Q of rational numbers. Given a fixed prime number p, for any relative integer x, $x \neq 0$, let $|x|_p = p^{-r}$ where p^r is the highest power of p dividing x.

Let us recall that an absolute value on a field K is a function φ, with non-negative values, such that

$$\begin{aligned} &\text{(i)} \quad \varphi(x)=0 \Longleftrightarrow x=0 , \\ &\text{(ii)} \quad \varphi(xy)=\varphi(x)\varphi(y) , \\ &\text{(iii)} \quad \varphi(x+y) \leq \varphi(x)+\varphi(y) . \end{aligned} \tag{2.4}$$

One sees by conditions (i) and (ii) that if some absolute value on Q is such that $\varphi(x)=|x|_p$, whenever x is a nonzero integer, it is unique. Then one shows that this φ also satisfies condition (iii): it is called the p-adic absolute value $|x|_p$.

In the case of the classical absolute value $|x| = \mathrm{Max}(x,-x)$, one has $|x+x| > |x|$ if $x \neq 0$, which constitutes the principle of Archimedes. With Hensel's definition of the p-adic absolute value, one has

$$|x+y|_p \leq \mathrm{Max}\{|x|_p, |y|_p\} ,$$

which is a more stringent inequality than (iii) above. More strikingly,

$$|x+x|_p \leq |x|_p ,$$

which violates the principle of Archimedes. For this reason, the p-adic absolute value is said to be non-Archimedean (or ultrametric).

Through the p-adic absolute value $|x|_p$, the p-adic distance $d_p(x,y)$ is defined by

$$d_p(x,y)=|x-y|_p .$$

From the classical absolute value on Q, one gets the field R of real numbers by completion, which amounts to creating new "numbers" as limits of Cauchy sequences that do not have a rational limit. The same construction, with a concept of Cauchy sequence relative to the p-adic absolute value $|x|_p$, gives a new complete field, analogous to R, which is precisely the field Q_p of p-adic numbers naturally provided by an extension of the p-adic absolute value, namely,

$$\text{if } x\in Q_p,\ x=\sum_{i=r}^{\infty} a_i p^i,\ a_r\neq 0,\ \text{then } |x|_p=p^{-r} .$$

As a consequence of Hensel's definition, two rational numbers are p-adically close if their difference is divisible by a high power of p. In order to give a feeling for the peculiarities of the p-adic distance, let us consider an example, $p=5$. Two numbers are 5-adically close if their difference is divisible by a high power of 5. Thus 6 is closer to 1 than it is to 7. Equidistant to 1, 11, 16, 21, or 26, 6 is closer to 31 and even closer to 131, closer still to 631. If, for example, one considers the triangle formed by the three numbers 1,6,7, one observes that it is an isosceles triangle with 7 equidistant to 1 and 6.

The natural geometrical ordering of p-adic numbers is not along the real line but on a hierarchical generating tree, as discussed in later sections. Actually, the coexistence of different distances for the same set of objects is not unfamiliar. For instance, between two human beings, besides the usual geometrical distance, which measures spatial separation, one can define a genetic distance measuring kinship relations.

The pioneer work of Hensel led to considerable developments in number theory. Coinage of the word ultrametric came much later and is due to Marc Krasner (1912—1985). In a note presented at the French Academy of Sciences on October 23, 1944, entitled "Nombres semi-réels et espaces ultramétriques," he elucidated the topological generality of ultrametric spaces, beyond the algebraic context in which the notion had appeared (Krasner, 1944).

Thus, as he noticed, in an ultrametric space, every point inside a ball [that is, every point B such that $d(A,B)\leq r$] is itself at the center of the ball, and the diameter of a ball is equal to its radius. Each ball is both open and closed, or "clopen" (Schikhof, 1984). Two balls are either disjoint or contained one within the other. In equivalent terms, the ultrametric inequality (2.2) implies that, in an ultrametric space, all triangles are either equilateral or isosceles with a small base (third side shorter than the two equal ones).

The examples discussed in the following sections concern systems with a large number of degrees of freedom, where a natural definition of distance among configurations already exists. One is then dealing with an infinite-dimensional metric space where a conveniently chosen sample of points may be organized ultrametrically. As a direct consequence of Eq. (2.2), such a set must be disconnected and sparse. In an ultrametric set, there are no points intermediate between any pair of points. In Sec. III.D we shall show that any such ultrametric set can be characterized quite generally as the final result of a "neutral evolution" process reminiscent of that found in biological evolution.

III. ULTRAMETRICITY IN TAXONOMY

A. Hierarchical trees

Taxonomy is the first field of science where the notion of ultrametricity appeared, outside mathematics. That occurred, twenty years after Krasner's note, during the sixties (Benzécri, 1965,1984; Hartigan, 1967; Jardine *et al.*, 1967; Johnson, 1967; Jardine and Sibson, 1971).

Intuitively, the idea is the following (mathematical definitions are postponed until Sec. VI). Suppose one has a collection of objects (living species, for instance) that one has managed to classify (species within genera, genera within families, say, or more generally, taxa within taxa within taxa). The classification can be represented as a dendrogram, or hierarchy, generally pictured as an inverted tree, where the objects (the leaves of the tree) are displayed horizontally on the bottom line. Going from the bottom up, several leaves (species) merge into a branch (genus); several such branches merge into a higher branch, etc. [Fig. 1(a)].

Higher taxa comprise a larger diversity of species than lower taxa. When this inclusion relation can be quantified, i.e., when a positive real number can be associated with each class, the hierarchy becomes an indexed hierarchy [Fig. 1(b)]. In the light of evolution theory, the natural classification of living species is just a family tree. Therefore the branching points can, in principle, be ordered and dated, and the vertical axis becomes equivalent to a time axis. Any natural definition of the distance between two extant species should be proportional to the age of their closest common ancestor, that is, the time elapsed since they branched off (hence the time during which

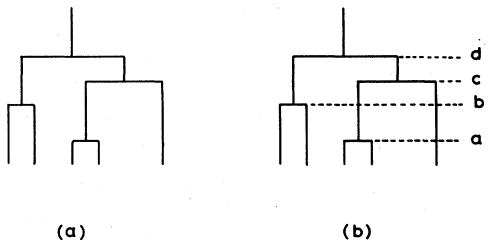


FIG. 1. Hierarchical trees: (a) hierarchy; (b) indexed hierarchy, $a<c<d$, $b<d$.

divergent evolution has separated the two lineages).

Armed with such a picture, one sees clearly the one-to-one relation between an indexed hierarchy and an ultrametric set. Any three elements of the set, i.e., any three objects on the bottom line of such a hierarchical tree, form a triangle that is either isosceles with a small base or equilateral. Conversely, any ultrametric set can be displayed as an inverted tree, where the vertical axis reveals the ordering of the objects into nested clusters of increasing radius.

B. Classifications and data analysis

A classification is an ordering of objects into groups on the basis of their relationships. There is an evident human compulsion toward categorization of any multitude. Today, this classificatory activity is also called data analysis (Gordon, 1981; Benzécri, 1984).

The logical starting point for a data analysis is a collection of N objects and an $N \times N$ distance matrix (symmetric matrix with vanishing diagonal elements). At this stage, a discussion of several caveats is useful.

The selection of a good similarity or dissimilarity index is a major decision. A good choice makes the difference between a natural order and an artificial system.

A dissimilarity index need not be a proper distance: it may violate the triangular inequality (2.1). Distances measured by air fares are a simple case in point: to fly directly from A to C may be more expensive than to go from A to B, then from B to C.

All proper classifications do not lead to trees. For instance, the classification of chemical elements leads to the Mendeleyev table, which is not a tree, though some grouping into clusters is cogent. See also the comparisons of multidimensional, tree-fitting, or clustering representations in proximity analyses of psychological data (Shepard, 1980).

Tree ordering should not automatically be identified with true ultrametricity. Consider, for instance, the teletypewriter code, which is conveniently displayed as a decision tree (Fig. 2). Starting from the trunk, a left turn means 0, a right turn means 1. Thus the "address" of the letter D, for instance, is 01101. One can define an ultrametric distance between two addresses as the 2-adic distance defined in Sec. II, on the 2-adic numbers obtained by reading from right to left the digits of each address. But this is an artificial distance, the natural one being the Hamming distance. The Hamming distance between two binary numbers is defined as the number of places in which the digits are different, and it measures the risk of confusion due to random errors. With respect to this natural distance, the encoded set of 26 letters is not at all ultrametric (an ultrametric subset of 4 letters, much sparser therefore, may be extracted from this five-dimensional hypercube).

When the distance matrix satisfies ultrametricity, it follows that a dendrogram can be unambiguously built. Still pending is the question of the meaning of branching

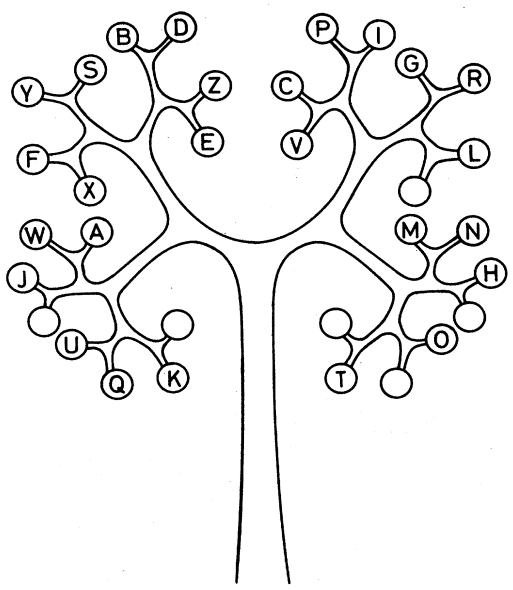


FIG. 2. Decision tree for the teletypewriter code: left=0, right=1. Example: D↔01101.

points. For biological classification, this question has been settled by evolution theory, in which natural classification has become phylogenetic: the branching points are full-fledged ancestors, to be reconstituted by palaeontology. However, for those who hold evolution to be a speculation, the branching points are just formal constructs. Certainly, an analyst confronted with data that are ultrametric, or nearly so, will be strongly tempted to find a "genetic" cause for this structure. We shall see how far this metaphor can be used in the case of spin glasses (Sec. IV).

It is a well-known biological fact that convergent evolution is a systematist's enemy, because similar morphological characteristics can be found in unrelated species (Lewin, 1985). For instance, dolphins are mammals, not fish, and the giant and lesser pandas are not the close relatives that their given names could make us believe (O'Brien *et al.*, 1985). Improved measurements of distance are supplied by molecular biology (Zuckerkandl and Pauling, 1965). Detailed comparison of the sequences (primary structures) of biomolecules, proteins or nucleic acids, allows for molecular, as distinct from morphological, phylogenetic reconstruction, which is less sensitive to convergence effects (Kimura, 1983; Ninio, 1983; Volkenshtein, 1984) (Fig. 3). An additional superiority of molecular phylogeny comes from the existence of molecular clocks (i.e., constancy in time of the rate of change of a given molecule), for which there is considerable empirical evidence. Indeed, at the molecular level, the evolution of a gene is to a large extent a random walk in sequence space, with a well-defined clock, for instance 5×10^{-9} substitutions per nucleotide site per year in the globin pseudogenes of mammals (Kimura, 1985). A branching diffusion process in an infinite-dimensional space is a

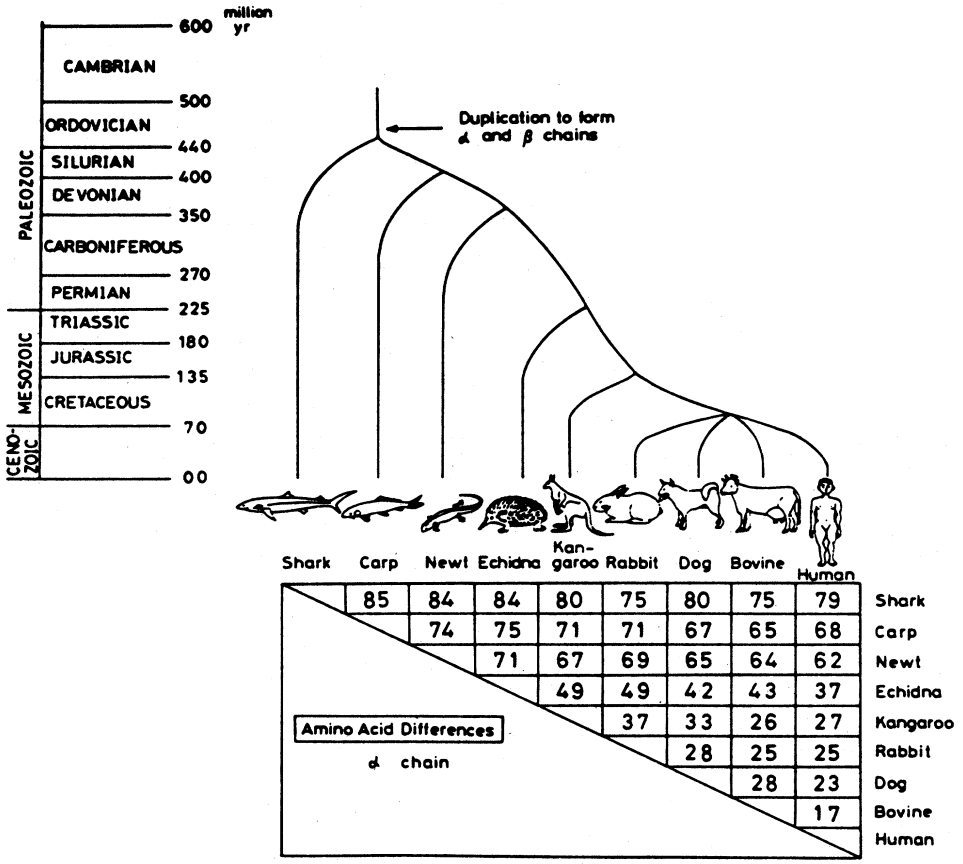


FIG. 3. A phylogenetic tree of several vertebrates together with their times of divergence (see the geological time scale at the left) and the numbers of amino acid differences between the hemoglobin α chains of each pair of these animals (see the table below the tree). Note that each comparison of the α hemoglobins involves a total of about 140 amino acid sites, so that, for example, the human and the carp have accumulated amino acid differences at about half (68/140) of the sites, since they diverged roughly 400 million years ago. If one uses a statistical method which corrects for superimposed amino acid substitutions, one gets roughly the evolutionary rate 10^{-9} per amino acid site per year, from most of the comparisons, suggesting the uniformity (constancy) of the rate of amino acid substitutions in evolution (from Kimura, 1985).

simple class of an ultrametricity-generating Markov process. However, the data for a given biomolecule are never ideally ultrametric, and this raises the question of reconstituting one evolution tree from sometimes ambiguous or conflicting data (Penny *et al.*, 1982; Wilson, 1985).

Two very simple procedures for tree construction have been devised by taxonomists, circa 1950 (Sørensen, 1948; Florek *et al.*, 1951), and later related to ultrametrics. They are ascending, bottom-to-top, aggregation procedures. Given a collection of objects with a distance matrix, one begins by aggregating the two closest objects into one cluster; then one defines the (renormalized) distances between this cluster and the remaining objects, and more generally between two clusters.

In one procedure, called single-linkage clustering (Florek *et al.*, 1951; Sokal and Sneath, 1963), the distance between two clusters is the minimum distance between any member of one cluster and any member of the other. This procedure, systematically pursued upward, leads to a unique tree and a unique renormalized distance matrix. Renormalized distances are smaller than or equal to the original distances. The renormalized distance matrix is ultrametric and it is called the subdominant ultrametric.

In the other procedure, called complete-linkage clustering (Sørensen, 1948), the distance between two clusters is the maximum instead of the minimum in the above definition. This procedure need not lead to a unique tree. Renormalized distances are larger than or equal to the original distances. The renormalized distance matrices are ultrametric.

Thus these two procedures provide lower and upper ultrametric bounds for a given set of data, and they suggest natural definitions for degrees of ultrametricity (Sec. VI) in problems that are not strictly ultrametric. It should be mentioned that many other techniques are now commonly used in data analysis, in order to extract information from all sorts of data, but they fall outside the scope of this review (Shephard, 1980; Gordon, 1981; Benzécri, 1984).

Coming back to Fig. 3, one can obtain a more concrete feeling for the assertion that ultrametricity means absence

of intermediates, as stated at the end of Sec. II. Indeed, naively one might have thought that dog was somewhat intermediate between rabbit and man, but no, the three form a closely equilateral triangle. Or that bovine was between kangaroo and man, but instead these form an isosceles triangle with kangaroo equidistant from man and bovine.

C. Distances and overlaps

A distance is a dissimilarity index. An overlap is a similarity (or proximity) index.

In many problems, the relationship between two objects is more naturally expressed as an overlap. With spin degrees of freedom, the overlap between two spin configurations is conveniently defined as the scalar product in configuration space:

$$q^{\alpha\beta}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_i^{\alpha}\cdot\mathbf{S}_i^{\beta},\quad -1\le q\le 1\ , \tag{3.1}$$

where N is the number of spins; classical spins are unit magnitude vectors with m components. Obviously, $q^{\alpha\alpha}=1$, $q^{\alpha\beta}=q^{\beta\alpha}$.

As far as the definition of ultrametricity is concerned, overlaps can be used just as well as distances. The ultrametric inequality (2.2) becomes

$$q^{\alpha\beta}\ge \mathrm{Min}\{q^{\alpha\beta},q^{\beta\gamma}\}\ . \tag{3.2}$$

But what is the metric inequality for overlaps?

For Ising spins ($m=1$), the overlap defined by Eq. (3.1) and the Hamming distance, defined above, are simply related by

$$q=1-2d\ , \tag{3.3}$$

from which a metric inequality for overlaps follows:

$$q^{\alpha\gamma}\ge q^{\alpha\beta}+q^{\beta\gamma}-1\ . \tag{3.4}$$

However, for vector spins ($m\ge 2$), the generalized Hamming distance defined by Eqs. (3.1) and (3.3) is not a proper distance, in general: neither Eq. (2.1) nor Eq. (3.4) need be satisfied with such definitions. A safer definition of the distance (Walstedt, 1983; Henley, 1985) is obtained from

$$q=1-2d^2 \tag{3.5}$$

because d is now a Euclidean distance in N-dimensional configuration space,

$$d^2=\frac{1}{N}\sum_{i=1}^{N}(S_i^{\alpha}-S_i^{\beta})^2\ , \tag{3.6}$$

and thus satisfies the metric inequality (2.1); but the corresponding metric inequality for overlaps is not pretty.

Equations (3.3) and (3.5), between overlap and distance, are not the only ones conceivable. An exponential relation such as

$$q=e^{-d} \tag{3.7}$$

is also legitimate (provided q is non-negative) and has indeed been found relevant, in the proximity analysis of many psychological data (Shepard, 1980; see also Doolittle, 1985). In this case, the metric inequality for overlaps is

$$q^{\alpha\gamma}\ge q^{\alpha\beta}\cdot q^{\beta\gamma}\ . \tag{3.8}$$

D. Ultrametric sets in infinite-dimensional systems

For systems with a large number N of degrees of freedom, ultrametricity, exact in the limit $N\to\infty$, is a natural type of organization. The understanding of this fact came after the analysis of the spin glass model (Mézard *et al.*, 1984a,1984b; Mézard and Virasoro, 1985) that will be discussed in the next section. However, because we now understand that some of these results are model independent and will emerge in other fields, we attempt in this section to describe the structure common to any ultrametric set in an infinite-dimensional space.

We therefore assume an embedding N-dimensional space (the configuration space), which is itself contained in $\mathbb{R}^N$: the space of N real component vectors. The natural metric is assumed to be defined on it. Ultrametricity is then a property of a conveniently chosen sparse subset.

A trivial example consists of M vectors (with $M/\sqrt{N}\to 0$ when $N\to\infty$) whose components are chosen by random, equally distributed, independent processes with finite average $\overline{v}$ and variance $\overline{v^2}$. Any two vectors $v_i^{(a)}$ and $v_i^{(b)}$ have overlap:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}v_i^{(a)}v_i^{(b)}=(\overline{v})^2,\ \text{if } a\ne b$$
$$=\overline{v^2},\ \text{if } a=b\ . \tag{3.9}$$

Therefore all triangles are equilateral.

To generate an ultrametric set corresponding to a nontrivial indexed hierarchy, we imagine a stochastic branching process acting simultaneously and independently on each component of the vector. The number of steps in this process is chosen equal to the number of levels of the tree. At the kth step the tree has been generated up to the $(k-1)$th level. At the end of each branch one then begins a new ramification, for instance in R branches, by drawing R times each component of a stochastic vector y_i^k with a probability distribution conditioned by the values of the same component of the previously drawn vectors on that branch (the ancestors) $y_i^{k-1},y_i^{k-2},\ldots,y_i^1$. The stochastic branching process is defined by the probability distributions,

$$P_k(y^k\,|\,y^{k-1},\ldots,y^1)\ . \tag{3.10}$$

There are L such functions, where L is the number of levels of the tree (the number of different values taken by the index of the hierarchy). After the last, Lth, step a

large set of vectors with different ancestry have been generated,

$$y_i^{L(a)} a = 1, \ldots, M \ , \tag{3.11}$$

with M finite.

In the limit $N \to \infty$ this set is ultrametric. In fact, the overlap between two vectors depends only on their ancestry and is larger the closer they are in parenthood. A simple probabilistic calculation shows that the overlap between vectors having their closest common ancestor at the Hth level is the following monotonically increasing function of H:

$$q_H = \int dy^1 dy^2 \cdots dy^{H-1} \prod_{k=1}^{H-1} P_k(y^k | y^{k-1}, \ldots, y^1) \left[\int dy^H \cdots dy^L \prod_{k=H}^{L} P_k(y^k | y^{k-1}, \ldots, y^1) y^L \right]^2 . \tag{3.12}$$

In this way we prove that any neutral evolution process acting on a sufficiently large vector space of features gives rise to an ultrametric set. Remarkably the converse also holds: given any ultrametric set, defined on an infinite-dimensional vector space and isotropic, i.e., such that the statistical properties of the vector components do not distinguish any particular orientation in the space, one can construct a stochastic branching process such that the ultrametric set is an unbiased realization of it.

As explained in previous sections, an ultrametric set always defines an indexed hierarchical tree. The elements of the set lie at the leaves of the tree, while the branching intersections are empty. In the theory of evolution, on the other hand, the intersections correspond to past species. In order to introduce an "evolutionary process" for an arbitrary ultrametric set, we introduce ancestors as vectors lying at the branching points and defined as the average of their lineage. To complete the proof we have to construct the stochastic process that generates the descendants from an ancestor. The assumed isotropy of the set allows us to sum powers of the components of the vector to calculate moments of the probability distribution. Assuming the necessary regularity and convergence, one can then reconstruct the probability distribution.

Notice, however, that the correspondence is not one to one. In particular, the branching stochastic process we have constructed is essentially embedded in $\mathbb{R}^N$. This may be unnatural. For instance, the neutral biological evolution described in the previous section acts on the space of biomolecule configurations, which is not $\mathbb{R}^N$: a past species is definitely not the average of its descendants.

We further observe that classifying a set by defining on it an indexed hierarchy does not amount to a complete description of the set. The information about the higher moments of the vector components as well as higher-order correlations between vectors at fixed distances from each other is not contained in a simple index. Rather, it is implicitly contained in the defining equations of the branching stochastic process, which, on the other hand, do not know about the ramification characteristics of the tree. How ramified the foliation is and how homogeneous the ramification depend on a survival rate of the mutants that seems to depend crucially on other features. In particular, we shall see in Sec. IV that the hierarchical tree of the spin glass system is far from regular. Therefore we conclude that a full description of an ultrametric set requires knowledge both of the branching process and of the parameters of the tree.

Another remarkable property of ultrametric sets in high-dimensional spaces is that every element of the set defines an ultrametric distance among different components of the vectors (Mézard and Virasoro, 1985). Loosely speaking, the distance between components i and j measures how far one has to go in the set to find another element with the values of the components v_i, v_j permuted. More precisely, one imagines a discretization of the y^k variables corresponding to each component. Then the two sequences

$$y_i^1 y_i^2 \cdots y_i^L \quad \text{at component } i \ ,$$

$$y_j^1 y_j^2 \cdots y_j^L \quad \text{at component } j \ ,$$

will coincide from a particular level up. The level of convergence defines the distance. The higher it is, the farther apart they are. In Sec. VII we shall discuss the peculiarities of a random walk in an ultrametric set. If one allows the elementary steps of the random walk to be distributed up to a certain $d_{\max}$, the lack of intermediate points characteristic of an ultrametric set (Sec. II) forces the system to remain in a ball in configuration space of diameter $d_{\max}$. In this case one can define a partition of the components of the vectors such that the average value of any integer power of the vector components inside each cell is constrained. Only intracell components can exchange information. This is reminiscent of the mechanism discussed in Palmer *et al.* (1984).

IV. SPIN GLASSES

Spin glasses are important because they provide the simplest model to analyze, endowed with a type of phase transition, and hence a type of organization at low temperature, that seems radically different from all previously known.

The crucial ingredients in these models are disorder and frustration. By disorder we mean that in the defining properties of the system (i.e., in the Hamiltonian) there are an infinite number of parameters, with nontrivial complexity. One in general assumes that they are the representative samples of a statistical ensemble, so as to be able to study many different systems by performing appropriate ensemble averages. Frustration generally means that there is competition among conflicting in-

teractions. As a consequence the system does not find one accommodation that satisfactorily complies with all constraints. Paradoxically it finds instead a multitude of almost satisfactory accommodations, and hence these systems turn out to be particularly rich in the number of equilibrium states at low temperature. The description of the rich topography of the energy surface (free energy at $T\neq 0$), in particular, the relative positions of the valleys in phase space (ultrametricity) and their relative depths, is the subject of this section.

The reader is directed to the review on spin glasses to appear in this journal (Binder and Young, 1986) for a more thorough presentation and references.

A. Mean-field theory

1. The history

Ultrametricity entered physics surreptitiously in 1979 (Parisi, 1980), at a time when phase transitions were habitually explained in terms of the spontaneous breaking of symmetries. In the context of the mean-field theory of spin glasses, Parisi was confronted with a Hamiltonian symmetric under the permutation of n variables. If permutational symmetry means that the n elements are to a certain extent equal, breaking the symmetry means that some of them are more equal than the others. In this way one is destined to end up with a hierarchy. However, as we shall shortly explain, the n elements are fictitious. In the end, n had to be continued analytically to zero, and it took some years to realize that ultrametricity had something to do with the physics of the problem.

The model to be solved was the Sherrington-Kirkpatrick mean-field version of a spin glass (Sherrington and Kirkpatrick, 1975). The Hamiltonian is

$$\mathcal{H}=-\sum_{i>j} J_{ij}S_iS_j \ , \tag{4.1}$$

where S_i is an Ising spin ($S_i=\pm 1$) on each of N sites ($1\leq i\leq N$), all of them coupled through a random symmetric matrix with elements J_{ij} that are statistically independent, Gaussian distributed, and have averages

$$\overline{J_{ij}}=0, \quad \overline{J_{ij}^2}=1/N \ .$$

Every set of J_{ij} defines a sample of the system, and one is interested in the thermodynamics of a typical unbiased sample, at sufficiently low temperatures. To solve this problem one has to calculate the average of the free energy,

$$\overline{\log Z}=\overline{\log \operatorname{Tr}_{\{S_i\}} e^{-\beta\mathcal{H}}} \ , \tag{4.2}$$

but this is difficult.

2. The replica trick

Edwards and Anderson (1975) proposed the use of the following formula:

$$\overline{\log Z}=\lim_{n\to 0}\frac{\overline{Z^n}-1}{n} \ . \tag{4.3}$$

For integer n, Z^n is simply the partition function of n identical systems with the same couplings J_{ij} (replicas). The average over the disorder can now be taken (it is a Gaussian integral) to obtain a site-permutation-invariant expression. It is then possible to decouple the N sites. The trace over the Ising spins becomes trivial, and one ends up with an integral over $n(n-1)/2$ variables:

$$\int \left[\prod_{a,b} dQ_{ab}\right] e^{NG(Q_{ab})} \tag{4.4}$$

with

$$Q_{ab}=\frac{1}{N}\sum_{i=1}^{N} S_i^{(a)}S_i^{(b)}, \quad 1\leq a,b\leq n, \quad a\neq b \ ,$$

where the use of the saddle-point technique is justified because $N\to\infty$.

Notice that $Q_{ab}\in[-1,1]$. The system, as defined in Eq. (4.1), is symmetric under the mapping $S_i\to -S_i$ (time reversal). The breaking of such symmetry causes the usual type of phase transition, whose effects could be confusing in the investigation of the glassy transition. We shall therefore always imagine an infinitesimal external field acting on the system. In that case $Q_{ab}\in[0,1]$.

Parisi's ansatz restricts the search for the minimum to a Q_{ab} invariant under a subgroup of the permutation group of n elements. If P_n is the group of permutations of n elements, consider the chain

$$P_n\supset (P_{m_1})^{n/m_1}\otimes P_{n/m_1} \ , \quad P_{m_1}\supset (P_{m_2})^{m_1/m_2}\otimes P_{m_1/m_2} \ ,$$

and so on. For instance, an 8×8 Q_{ab} matrix invariant under $(P_2)^4\otimes(P_{4/2})^2\otimes P_{8/4}$ is shown in Fig. 4.

The matrix Q_{ab} so generated is parametrized in terms of a single function $x(q)$ that measures the fraction of replica pairs such that their overlap is less than q. The analytic continuation of $G(Q_{ab})$ to $n=0$ can be done, and the saddle-point equations determine the function $x(q)$.

$$\begin{vmatrix}
0 & q_0 & q_1 & q_1 & q_2 & q_2 & q_2 & q_2 \\
q_0 & 0 & q_1 & q_1 & q_2 & q_2 & q_2 & q_2 \\
q_1 & q_1 & 0 & q_0 & q_2 & q_2 & q_2 & q_2 \\
q_1 & q_1 & q_0 & 0 & q_2 & q_2 & q_2 & q_2 \\
q_2 & q_2 & q_2 & q_2 & 0 & q_0 & q_1 & q_1 \\
q_2 & q_2 & q_2 & q_2 & q_0 & 0 & q_1 & q_1 \\
q_2 & q_2 & q_2 & q_2 & q_1 & q_1 & 0 & q_0 \\
q_2 & q_2 & q_2 & q_2 & q_1 & q_1 & q_0 & 0
\end{vmatrix}$$

FIG. 4. Typical example of an 8×8 Parisi matrix Q_{ab} for the spin glass problem.

The Sherrington-Kirkpatrick (SK) solution corresponded to the choice

$$x(q)=\theta(q-q_{\mathrm{EA}}) , \tag{4.5}$$

which is equivalent to the matrix $Q_{ab}=q_{\mathrm{EA}}$ for $a\neq b$ (Edwards and Anderson, 1975; Sherrington and Kirkpatrick, 1975). Symmetry among replicas remained unbroken. The saddle-point equations plus the stability criterion prove that replica symmetry is broken. In this case, as usual, one has to sum over all the different saddle-point matrices Q_{ab} that are obtained by an arbitrary permutation of the replica indices.

By 1983 the amount of circumstantial evidence in favor of Parisi's solution was growing. Several tests were successfully passed: (i) the entropy was positive, for all temperatures, in contradistinction to the previous attempts; (ii) numerical simulations agreed with the predictions from the solution; (iii) the saddle point was proved to be marginally stable under infinitesimal transformations.

However, a physical interpretation of the solution was lacking and the mathematical consistency of the analytic continuation was dubious.

3. The many valleys of a spin glass

The phenomenological characterization of the spin glass phase comprises the presence of hysteresis cycles, remanent magnetizations dependent on the past history of the sample, and long relaxation times. All of these effects suggest the existence of many stable or at least metastable equilibrium states. Numerical experiments (Mackenzie and Young, 1982) are more precise: they show that ergodicity is broken in the SK spin glass.

Breaking of replica symmetry and breaking of ergodicity are logically equivalent. It suffices to imagine two real replicas (Blandin *et al.*, 1980) to realize that a nontrivial $x(q)$ means that with finite probability the two systems will end up in two different states. In 1983, Parisi derived in this way a physical interpretation of $x(q)$.

Imagine two real replicas, with large but finite N, in thermal equilibrium. At regular time intervals Δt, the overlap of the spin configurations is measured and from this data a histogram is constructed with the frequency of occurrence of the overlap:

$$P_J(q)=\lim_{N\to\infty}\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\delta\left[\frac{1}{N}\sum_{i=1}^{N}S_i^{(1)}(t_0+k\cdot\Delta t)S_i^{(2)}(t_0+k\cdot\Delta t)-q\right] . \tag{4.6}$$

The unusual order of the limits guarantees that the consequences of the ergodic theorem hold. Therefore

$$P_J(q)=\operatorname*{Tr}_{S_i^{(1)},S_i^{(2)}}\left[\frac{e^{-\beta\mathcal{H}(S^{(1)})-\beta\mathcal{H}(S^{(2)})}}{Z^2}\delta\left[\frac{1}{N}\sum_{i=1}^{N}S_i^{(1)}\cdot S_i^{(2)}-q\right]\right] . \tag{4.7}$$

A modified version of the replica trick eliminates the Z^2 denominator:

$$P_J(q)=\lim_{n\to 0}\operatorname*{Tr}_{S_i^{(a)}}\left[\prod_a e^{-\beta\mathcal{H}(S^{(a)})}\delta\left[\frac{1}{N}\sum_{i=1}^{N}S_i^{(1)}S_i^{(2)}-q\right]\right],\quad a=1,2,\ldots,n . \tag{4.8}$$

The J subindex is there to remind us that the measurement is made on one sample, though only the average over J can be analytically calculated. To leading order in $1/N$, the contributions from all the saddle points dominate the integral. It follows that

$$\overline{P_J(q)}=\lim_{n\to 0}\sum_{\{Q\}}\delta(Q_{12}-q) . \tag{4.9}$$

The different saddle points are obtained by simple permutations in the replica indices. Therefore it is equivalent to consider one matrix Q_{ab} and average over all rows and columns (De Dominicis and Young, 1983),

$$\overline{P_J(q)}=\lim_{n\to 0}\frac{1}{n(n-1)}\sum_{\substack{a,b\\ a\neq b}}\delta(Q_{ab}-q) . \tag{4.10}$$

Remembering the definition of $x(q)$, one finally obtains

$$\overline{P_J(q)}=\frac{dx}{dq} . \tag{4.11}$$

4. The topography of the landscape. Ultrametricity

Parisi's simple interpretation of the order parameter turned out to be the key to a greater understanding of the spin glass phase in the mean-field approximation. The next, rather obvious, step was to calculate the distribution probability among the configurations. The calculation is a more cumbersome straightforward generalization of the previous one. Instead of Eq. (4.7), one finds

$$P_J(q_1,q_2,q_3)=\operatorname*{Tr}_{S_i^{(1)},S_i^{(2)},S_i^{(3)}}\left[\frac{e^{-\beta[\mathcal{H}(S^{(1)})+\mathcal{H}(S^{(2)})+\mathcal{H}(S^{(3)})]}}{Z^3}\delta\left[\frac{1}{N}\sum S_i^{(1)}S_i^{(2)}-q_1\right]\right.$$
$$\left.\times\delta\left[\frac{1}{N}\sum S_i^{(2)}S_i^{(3)}-q_2\right]\cdot\delta\left[\frac{1}{N}\sum S_i^{(1)}S_i^{(3)}-q_3\right]\right] \tag{4.12}$$

and instead of Eq. (4.10)

$$\overline{P_J(q_1,q_2,q_3)}=\lim_{n\to 0}\frac{1}{n(n-1)(n-2)}\sum_{\substack{a,b,c\\ a\neq b,a\neq c\\ b\neq c}}\delta(Q_{ab}-q_1)\delta(Q_{bc}-q_2)\delta(Q_{ac}-q_3)\ . \tag{4.13}$$

A simple inspection of the matrix Q_{ab} in Fig. 4 demonstrates ultrametricity. If $q_1\geq q_2\geq q_3$, then a,b are in an internal block with respect to c, and by construction $q_2=q_3$. Ultrametricity is therefore implicit in Parisi's ansatz. The evidence in its favor is just the evidence in favor of Parisi's solution.

B. Analysis of an ultrametric set

The alleged ultrametricity of the set of relevant configurations of an SK model opens the way for the specific analysis proposed in Sec. III. We should like to find the corresponding stochastic branching process and characterize the foliation of the tree. The Parisi function $P(q)$, as derived from the saddle-point equations, has its support in the interval $[0,q_{\rm Max}]$ with a δ-function contribution at $q_{\rm Max}$ and a smooth continuous distribution in the open interval. It follows that the tree has new branches at every value of q. It is more convenient to approximate $P(q)$ by a finite number of weights at discrete values of q. Then the ramification at each discrete level will always be in an infinite number of branches. There are weights associated with each of them. They reflect the Gibbs-Boltzmann weight of the canonical ensemble. The relative thickness of a particular branch is related to the sum of the weights of the configurations. The replica method, as defined, allows a full description only in these terms, though in other problems a different weighting may be more relevant. The microcanonical ensemble measure would give information on the entropy of each branch, and the analysis would not be very different from the one here described. In optimization problems and neural networks, the basis of attraction appears as a natural weight. This case will be discussed in Sec. V.

1. The branching diffusion process

Pursuing the line of analysis of Sec. IV.A we can consider an arbitrary number of replicas and try to calculate the expectation value,

$$\overline{\left\langle\frac{1}{N}\sum_{i=1}^{N}S_i^{(1)}S_i^{(2)}\cdots S_i^{(p)}\right\rangle_J}\ , \tag{4.14}$$

where the large angle brackets stand for the Gibbs-Boltzmann thermal trace over S_i performed on one particular realization of the J_{ij}, while the overbar indicates the quenched average over J_{ij}. In general the thermal average fluctuates with J_{ij} even in the thermodynamic limit (Mézard *et al.*, 1984a,1984b; Young *et al.*, 1984). However it can be proved that if the average is restricted to the replica configurations such that all overlaps

$$q_{ab}^J=\left\langle\frac{1}{N}\sum_{i=1}^{N}S_i^{(a)}S_i^{(b)}\right\rangle \tag{4.15}$$

are fixed, then the resulting expectation value is the same for every sample J_{ij} and therefore equal to its average ("self-averaging"; Mézard and Virasoro, 1985).

We discuss a particular choice of the mutual overlaps, which is the simplest allowing for a full disentangling of the structure involved. We choose s among the p replicas with mutual overlaps q, while one of them is surrounded by p-s replicas at a larger overlap q'. As indicated in Sec. III.D we define the ancestor at age q as the state defined by the unweighted average at each site of the local magnetization of an infinite number of configurations, all lying at overlap q of each other. Then

$$m_i^{(A_q)}=\lim_{R\to\infty}\frac{1}{R}\sum_{k=1}^{R}S_i^{(k)}\ , \tag{4.16}$$

where $(1/N)\sum_{i=1}^{N}S_i^{(a)}S_i^{(b)}=q$, for all pairs of descendants $(k_1=a,k_2=b)$, so that

$$\frac{1}{N}\sum_{i=1}^{N}(m_i^{(A_q)})^k=\frac{1}{R^k}\sum_{\substack{\{a_l\}\\ a_l\neq a_{l'}}}\frac{1}{N}\sum_{i=1}^{N}S_i^{(a_1)}S_i^{(a_2)}\cdots S_i^{(a_k)}\ . \tag{4.17}$$

Therefore the calculation of Eq. (4.14) would allow us to know the correlations between an ancestor and its descendants:

$$\frac{1}{N}\sum_{i=1}^{N}(m_i^{(A_q)})^s(m_i^{(A_{q'})})^{p-s}\ . \tag{4.18}$$

The calculation of Eq. (4.14) by the replica method is tedious and complicated. In this review we quote the result (see Mézard and Virasoro, 1985, for the details). It can be written in terms of a fundamental kernel $\mathcal{P}_{q,q'}(y,y')$:

$$\frac{1}{N}\sum_{i=1}^{N}(m_i^{(A_q)})^s(m_i^{(A_{q'})})^{p-s} = \int_{-\infty}^{\infty} dy\, \mathscr{P}_{0,q}(0,y)[m_q(y)]^s \times \int_{-\infty}^{\infty} \mathscr{P}_{q,q'}(y,y')[m_{q'}(y')]^{p-s}dy' \quad (4.19)$$

with

$$m_q(y) = \int_{-\infty}^{\infty} \mathscr{P}_{q,1}(y,y')th(\beta y')dy' . \quad (4.20)$$

The kernel is defined by a diffusionlike equation in a time q flowing from 1 to 0,

$$-\frac{\partial \mathscr{P}}{\partial q} = \frac{1}{2}\frac{\partial^2 \mathscr{P}}{\partial y^2} + \frac{\partial \ln z(q,y)}{\partial y}\frac{\partial \mathscr{P}}{\partial y} , \quad (4.21)$$

with the initial condition

$$\lim_{q\to q'} \mathscr{P}_{q,q'}(y,y') = \delta(y-y'), \quad q \le q' \quad (4.22)$$

and boundary condition

$$\lim_{y\to\pm\infty} \mathscr{P}_{q,q'}(y,y') = \lim_{y'\to\pm\infty} \mathscr{P}_{q,q'}(y,y') = 0 , \quad (4.23)$$

while the auxiliary function $z(q,y)$ satisfies the nonlinear equation

$$-\frac{\partial z}{\partial q} = \frac{1}{2}\frac{\partial^2 z}{\partial y^2} - \frac{P(q)}{x(q)} z \ln z , \quad (4.24)$$

with initial condition

$$z(1,y) = th\beta y . \quad (4.25)$$

$P(q)$ is determined from the self-consistency condition:

$$q = \int_{-\infty}^{\infty} dy\, \mathscr{P}_{0,q}(0,y)[m_q(y)]^2 . \quad (4.26)$$

Therefore the solution of the SK model is reduced to the problem of solving Eqs. (4.21) and (4.24), parametrically as a function of $P(q)$, and then adjusting the latter so that Eq. (4.26) is satisfied (for solutions near the critical temperature, see Parisi, 1980; for solutions within an approximation, see Vannimenus *et al.*, 1981; and for solutions near $T=0$, see Sommers and Dupont, 1984).

Once the solution has been made explicit, the meaning of Eq. (4.19) is straightforward. By reconstruction we derive

$$\frac{1}{N}\sum_{i=1}^{N}\delta(m_i^{(A_q)}-m)\delta(m_i^{(A_{q'})}-m') = \int_{-\infty}^{\infty} dy\, dy' \mathscr{P}_{0,q}(0,y)\delta[m-m_q(y)]\cdot\mathscr{P}_{q,q'}(y,y')\delta[m'-m_{q'}(y')]$$
$$= \left|\frac{dy}{dm_q(y)}\right| \cdot \left|\frac{dy'}{dm_{q'}(y')}\right| \cdot \mathscr{P}_{0,q}(0,y)\mathscr{P}_{q,q'}(y,y')\Big|_{\substack{m_q(y)=m \\ m_q(y')=m'}} . \quad (4.27)$$

It follows that (i) the probability that the ancestor A_q has magnetization in the range $m, m+\Delta m$ at a site is

$$\left|\frac{dy}{dm_q(y)}\right| \cdot \mathscr{P}_{0,q}(0,y)\Big|_{m_q(y)=m} \cdot \Delta m ; \quad (4.28)$$

(ii) the probability that at the same site the descendant $A_{q'}$ has magnetization in the range $m', m'+\Delta m$ is

$$\left|\frac{dy'}{dm_{q'}(y')}\right| \cdot \mathscr{P}_{q,q'}(y,y')\Big|_{\substack{m_{q'}(y')=m' \\ m_q(y)=m}} \cdot \Delta m . \quad (4.29)$$

Therefore the knowledge of the function $\mathscr{P}_{q,q'}(y,y')$ amounts to the full determination of the stochastic branching process, which turns out to be Markovian in this case.

Although low-energy configurations are particular realizations of the stochastic branching process, it is not true in general that every realization is a low-energy configuration. For instance, expectation values that include the J_{ij} have to be considered, and they restrict the possible configurations.

A clearer picture of the set arises if we discretize the range of magnetization in a certain number of intervals. Then the sites can be partitioned into cells (at a scale q) where the magnetization of the ancestor A_q is constant. Each cell can then be partitioned into subcells (at a scale $q'>q$) according to the magnetization of the ancestor $A_{q'}$. In this way every configuration defines an ultrametric structure among the sites as anticipated in Sec. III.D.

2. The depth of the valleys and the foliation of the tree

Every configuration is weighed with the Gibbs-Boltzmann factor. Therefore the relevant ones are those with energies obeying

$$\mathscr{H}(S^{(a)}) - \mathscr{H}(S^{(b)}) = \text{finite} .$$

The leading contribution to the energy goes linearly with N, is the same for all the relevant configurations, and can be proved to be self-averaging (Khanin and Sinai, 1979). The nonleading contributions are not, and they determine the Gibbs-Boltzmann weight. Most of the results discussed in this section concern quantities that fluctuate from sample to sample.

The explicit calculation of $\overline{P_J(q_1,q_2,q_3)}$ gives the result

$$P(q_1,q_2,q_3) = \tfrac{1}{2}P(q_1)x(q_1)\delta(q_1-q_2)\delta(q_1-q_3) + \tfrac{1}{2}P(q_1)P(q_2)\theta(q_1-q_2)\delta(q_2-q_3) + \tfrac{1}{2}P(q_2)P(q_3)\theta(q_2-q_3)\delta(q_3-q_1) + \tfrac{1}{2}P(q_3)P(q_1)\theta(q_3-q_1)\delta(q_1-q_2) . \quad (4.30)$$

For any $q_c < q_{\rm Max}$ (the δ-function region requires some care) the integrated contributions in the volume $0 \le q_i \le q_c$ of all four terms are equal. It follows that, among all the triangles, $\frac{1}{4}$ are equilateral and $\frac{3}{4}$ isosceles. The relatively large number of equilateral triangles is a noticeable feature of this kind of ultrametricity.

Integrating Eq. (4.30) over q_3, we obtain

$$P(q_1,q_2)=\tfrac{1}{2}P(q_1)\delta(q_1-q_2)+\tfrac{1}{2}P(q_1)P(q_2)\ . \quad (4.31)$$

This equation shows that the distribution of valleys away from a particular one is conditioned by the fact that this one lies at a fixed distance from a third. There are two sources of correlations. On one side, the fluctuations with J_{ij} imply

$$\overline{P_J(q_1)P_J(q_2)}-\overline{P_J(q_1)}\cdot\overline{P_J(q_2)} = \tfrac{1}{3}[P(q_1)\delta(q_1-q_2)-P(q_1)P(q_2)]\ , \quad (4.32)$$

while on the other side, in a fixed sample,

$$\overline{P_J(q_1,q_2)}-\overline{P_J(q_1)P_J(q_2)} = \tfrac{1}{6}[P(q_1)\delta(q_1-q_2)-P(q_1)P(q_2)]\ . \quad (4.33)$$

This shows that the environment of a valley varies considerably from one to another. The sign of the effect reflects ultrametricity. Notice, however, that Eqs. (4.32) and (4.33) are zero above the critical temperature [where $P(q)$ reduces to a δ function]. This is a hint that the varying environment is a direct consequence of the Gibbs-Boltzmann weights. Both quantities (4.32) and (4.33) are easy to obtain in numerical simulations and can serve as a valuable gauge for comparing different models. We shall discuss them further in Sec. V.

There is a more direct way of seeing the variation of the environment. For instance, one can define

$$W_I=\sum_{\{S_i\}}\frac{e^{-\beta\mathcal{H}(S_i)}}{Z}\theta\left[q-\frac{1}{N}\sum_{i=1}^{N}S_iS_i^{(I)}\right]\ , \quad (4.34)$$

where $S_i^{(I)}$ is a selected configuration. A partition in clusters is thereby introduced. An exact expression (Mézard *et al.*, 1984a,1984b) has been given for the average number of clusters with weights W_I falling in the interval W, $W+dW$:

$$f(W,q)dW=\frac{W^{-1-x(q)}(1-W)^{-1+x(q)}}{\Gamma(1-x(q))\Gamma(x(q))}dW\ . \quad (4.35)$$

Corresponding formulas can be derived for the average number of simultaneous occurrences of n clusters. The explanation of these results required the analysis of the distribution probability for the energies of configurations and clusters. A simple scenario emerged (Mézard *et al.*, 1986; see also the related paper on the random energy model, Derrida and Toulouse, 1985).

(i) The energies of the relevant configurations inside a cluster at scale $q_{\rm Max}$ are independent, equally distributed random variables. The average number in the interval E, $E+dE$ is

$$d\mathcal{N}_I(E)=C\,e^{\beta E}dE\ . \quad (4.36)$$

The normalization constant C provides a convenient definition of the Ith cluster energy,

$$C=e^{-\beta E_I}\ . \quad (4.37)$$

In Parisi (1983) it is proved that the clusters at $q_{\rm Max}$ can be identified with the pure states. Then Eq. (4.36) is simply the usual definition of entropy,

$$d\mathcal{N}_I(E)=e^{S(E)}dE=e^{S(E_I)+\beta(E-E_I)}dE\ , \quad (4.38)$$

and therefore the cluster energy is in fact the free energy of the corresponding state.

(ii) The cluster energies inside a higher-order cluster at scale q are independent, equally distributed random variables. The average number in the interval E, $E+dE$ is

$$d\mathcal{N}_K(E)=e^{\beta x(q)(E-E_K)}dE \quad (4.39)$$

whose normalization implicitly determines E_K, the higher-order cluster energy.

The picture of the hierarchical tree that emerges from this analysis shows a certain degree of scale invariance. Notice, for instance, that all the dependence on q in Eqs. (4.34)—(4.39) is controlled by $x(q)$. The ramification at each branching point is similar, with a few dominating branches accounting for a large piece of the total weight, while an infinite number shares a small part of it. The tree is definitely very far from being regular. This explains the lack of self-averaging. One can hardly use statistical averages over a few dominating contributions. A different situation is expected when one uses the attraction basins as weights, because they seem to vary less strongly than the Gibbs-Boltzmann weight (see Sec. V).

C. Numerical tests

Although the derivation of ultrametricity explicitly uses the Gibbs-Boltzmann weights, the conclusion is valid with greater generality. In the thermodynamic limit ($N\to\infty$) the probability of occurrence of certain triads of relevant configurations is zero. Even in the paramagnetic phase, ultrametricity of a trivial kind (all triangles equilateral) holds. For infinite temperature it reduces to the property of orthogonality of randomly chosen vectors (see Sec. III.D).

To test ultrametricity numerically, it is important to understand finite volume effects. In the case of randomly chosen vectors we know that the corrections are of order $1/\sqrt{N}$: if $v_i^{(1)},v_i^{(2)}$ are independent, equally distributed random variables with average $\overline{v}$, then

$$\frac{1}{N}\sum_{i=1}^{N}v_i^{(1)}v_i^{(2)}=\overline{v}^{\,2}+O(1/\sqrt{N})\ . \quad (4.40)$$

Using the result (Secs. III.D and IV.A) that an ultrametric set on an infinite-dimensional space can be seen

TABLE I. Moments of the difference between two sides of the triangles at a fixed value of one of them [see Eq. (4.41)]. Ising spin glass with infinite range.

	$M_{12}(q_3)$		$M_{32}(q_3)$		$M_{13}(q_3)$	
q_3	$N=32$	$N=64$	$N=32$	$N=64$	$N=32$	$N=64$
0.5	1.79×10^{-5}	3.51×10^{-6}	7.29×10^{-5}	1.55×10^{-5}	3.28×10^{-5}	6.11×10^{-6}
0.625	1.37×10^{-5}	2.46×10^{-6}	5.01×10^{-5}	1.78×10^{-5}	1.72×10^{-5}	9.48×10^{-6}
0.75	3.69×10^{-5}	3.76×10^{-6}	1.80×10^{-4}	8.86×10^{-5}	1.01×10^{-4}	7.41×10^{-5}
0.875	2.27×10^{-5}	1.68×10^{-6}	9.46×10^{-4}	2.39×10^{-4}	8.64×10^{-4}	1.77×10^{-4}

as generated by a stochastic branching process, one can estimate the finite volume correction. If the number of generations remains bounded when $N\to\infty$, then statistical techniques show that the correcting terms will again be of order $1/\sqrt{N}$ with coefficients which now depend on the overlaps.

The spin glass case is more difficult because the number of generations is not bounded when $N\to\infty$. Analytic methods are not available, and one has to rely on Monte Carlo methods to estimate the corrections (Bhatt and Young, 1986).

Another obstacle in testing ultrametricity at finite N is the triangle inequality. Given two sides of a triangle with lengths $a<b$, the third side c must fall inside the interval $(b-a,b+a)$ [see Eq. (2.3)]. In addition, smoothness of the distribution near the end points makes the final picture qualitatively similar to that expected from an approximate version of ultrametricity. One must therefore rely on the available information on the correction terms and in particular its N dependence to distinguish between the two hypotheses.

In Parga *et al.* (1984), ultrametricity was tested on the local minima of the energy, with Gibbs-Boltzmann weights corresponding to a temperature $T=0.2$ (critical temperature $T_c=1$) and $N=32$ and 64. Two different tests were considered. Given $q_1\le q_2\le q_3$, the first one consisted in the calculation at fixed q_3 of the following moments:

$$M_{12}=\langle(q_1-q_2)^2\rangle\ ,$$
$$M_{13}=\langle(q_1-q_3)^2\rangle\ , \tag{4.41}$$
$$M_{23}=\langle(q_2-q_3)^2\rangle\ .$$

The results are summarized in Table I. The moment M_{12} seems to converge more rapidly to zero than M_{13} or M_{23}.

In the second test, ultrametricity was pitted against the "null" hypothesis that the distribution of q_1 between zero and q_2 is flat inside the interval allowed by the triangle inequality. The ratio

$$U=\frac{\int_0^{q_2}dq_1P(q_1,q_2,q_3)(q_1-q_2)^2}{\int_0^{q_2}dq_1P(q_1,q_2,q_3)}\cdot\frac{\int_0^{q_2}dq_1}{\int_0^{q_2}dq_1(q_1-q_2)^2} \tag{4.42}$$

was measured at $q_3=2[N/3]$ and $2[N/6]\le q_2\le 2[N/3]$ (see Table II). An overall convergence to zero is evident, though the rate seems to vary dramatically with q_3.

The distribution probability $P(q_1,q_2,q_3)$ could be obtained from the full information on frequencies of occurrence of every triplet of configurations. It is convenient to consider a particular two-dimensional section [for instance, by integration over q_3—see Eqs. (4.31), (5.1), and (5.2)] or projection (Sourlas, 1984). The latter reference eliminates one degree of freedom of the triplets by normalizing the sum of the distances,

$$d_i^R=\frac{3d_i}{d_1+d_2+d_3},\quad i=1,2,3\ . \tag{4.43}$$

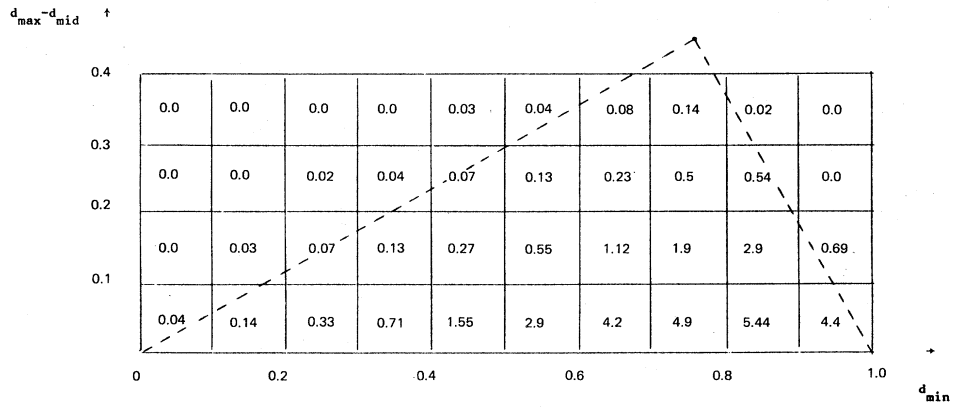
Then every triplet is represented by a point inside an equilateral triangle, the distances from the sides being the normalized d_i^R. This domain is fully symmetric among the three configurations, a clear advantage over other choices. A possible problem could arise if the data to be folded together had different correction terms in $1/N$. The triangle inequality can be easily visualized. Sourlas used this test on a three-dimensional Ising spin glass. One of the histograms is shown in Table III. The boundaries are the limits of the triangular inequalities. There is a clear sign of ultrametricity. However we should point out that the number of spins is large ($16\times16\times16$), and in the SK model the same degree of ultrametricity seems to appear in smaller systems.

Bhatt and Young (1986) have investigated the SK model through a Monte Carlo simulation. They consider $N=32$, 128, and 512, at a temperature $T=0.6$, and study the probability distribution of the difference between the two smaller overlaps when the larger one is fixed at $q=0.5$. The results are shown in Fig. 5. We consider this simulation convincing evidence in favor of ultrametricity. The width of the distribution decreases as $N^{-0.3}$, a reasonable rate, given the fact that the tree be-

TABLE II. Comparison between the hypothesis of ultrametricity and a null hypothesis (uniform distribution in q_1-q_2, at fixed q_3, with $q_3>q_2>q_1$) [see Eq. (4.42)]. Ising spin glass with infinite range.

	U	
q_2	$N=32$	$N=64$
0.3125	1.09	1.06
0.375	0.87	0.71
0.4375	0.47	0.24
0.500	0.33	0.23
0.5625	0.34	0.093
0.625	0.23	0.068

TABLE III. Density of triangles generated in a Monte Carlo simulation of a $(16)^3$, $J=\pm 1$. Ising spin glass, with nearest-neighbor couplings, for $\beta=0.8$ and a small external magnetic field $h/J=0.2$. The horizontal axis represents the smallest side of the triangle $d_{\min}$, while the vertical axis represents the difference between the two other sides $d_{\max}-d_{\mathrm{mid}}$ [distances are scaled according to Eq. (4.43)]. Exact ultrametricity would correspond to zero density, except along the horizontal axis. The dashed lines represent the triangle inequality constraints. Corresponding figures, obtained for a $(12)^3$ system, with the same parameters, show much less ultrametricity (unpublished data from N. Sourlas).



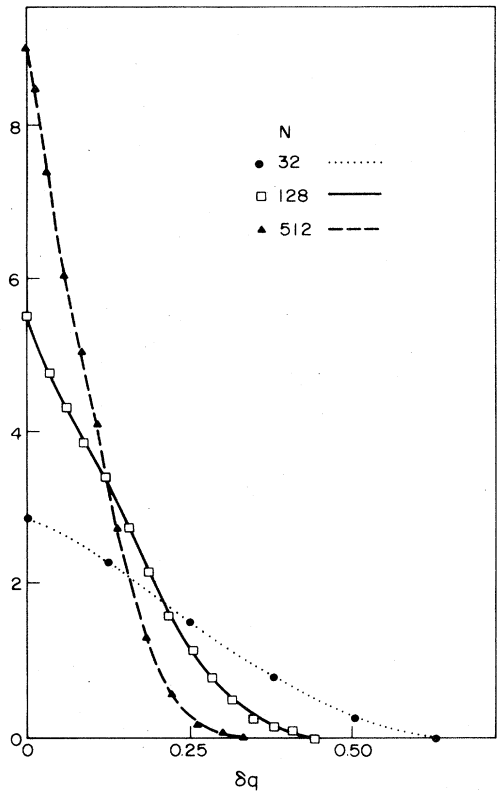


FIG. 5. Probability distribution of the difference between the two smaller overlaps ($\delta q = q_{\mathrm{mid}} - q_{\min}$) for a fixed larger overlap value $q=0.5$. The temperature is 0.6. In the $N\to\infty$ limit, the distribution is expected to become a δ function at the origin (from Bhatt and Young, 1986).

comes continuous in the $N\to\infty$ limit.

The growing consensus around the Parisi solution inevitably leads to a lack of motivation for lengthy numerical calculations. A word of caution is perhaps in order. Most of the properties we have discussed are implicit in Parisi's ansatz. Ultrametricity is there and so is the exponential distribution of energies (they are a consequence of the property that each row in the Q_{ab} matrix is a permutation of the others). It seems very hard to prove rigorously that the ansatz is correct. Therefore we must rely on numerical simulations, perhaps with larger N, to test this interesting solution effectively.

V. COMPLEX PROBLEMS

The combinatorial optimization problems of computer science, whose archetype is the traveling salesman problem (TSP), and the frustrated disordered systems of statistical physics, whose archetype is the spin glass, have many features in common. The analogy, clearly recognized not long ago (Kirkpatrick, 1981), is best revealed by the introduction of a temperature, as a new control parameter of the optimization search.

This analogy can be made operational by examining the freezing phenomena due to frustration (conflicting constraints), and through the efficient use of the method of simulated annealing, the notion of energy landscape in configuration space, and the analysis of the distribution of nearly degenerate local minima.

A. Presentation

The problem of the traveling salesman is simple: Find the shortest route for visiting N points. A table of point-

to-point distances defines one instance of the TSP. Combinatorial optimization in general refers to problems of minimizing a cost function with a large number of variables. This has turned into a major field of applied mathematics and computer science, with the development of algorithmic complexity theory (Garey and Johnson, 1979; Papadimitriou and Steiglitz, 1982). Besides the TSP, the problems of matching and of graph partitioning or coloring are also worth mentioning, among hundreds.

These well-defined problems, however, are only a subset of the many optimization problems, often loosely defined, that may be encountered in engineering and biology and that are gathered together under the heading of complex problems. This gets us into computer-aided design and artificial intelligence, which are fields of practical importance. We have now to explain how ultrametricity may be relevant.

First, it has been proved that the method of simulated annealing (Kirkpatrick *et al.*, 1983; Černý, 1985), as developed for spin glasses, is useful both for combinatorial optimization problems, such as the TSP, and for complex problems, such as partitioning, placement, and wiring of electronic circuits.

Thereafter, the similarity with spin glasses instigated numerical and theoretical analyses of energy landscapes, including searches for ultrametricity, in traveling salesman problems (Kirkpatrick and Toulouse, 1985), graph partitioning (Kirkpatrick, 1984; Fu and Anderson, 1985), matching (Bachas, 1985; Mézard and Parisi, 1985), placement (Solla *et al.*, 1986), and graph coloring (Bouchaud and Le Doussal, 1985).

The main practical issue is, can one get help, in solving a given optimization problem, from a knowledge of the typical energy landscape in a family of similar problems? Or, can configuration-space analysis guide us in a choice of algorithm? There are several related questions: How common is the occurrence of ultrametric distributions of local minima? What is a proper weighting for the local minima? If a problem is ultrametric, what is the physical significance of this ultrametricity, and what next? If it is nearly ultrametric, how is the deviation to be characterized? If it is far from ultrametric, what else can be said? Is there a link between simulated annealing and ultrametricity? Is the traditional classification of problems, according to algorithmic complexity, physically sound? Only partial answers are available at present, and even less will be said here. A few basic concepts are presented merely as background for a field with many ramifications.

B. Configuration-space analysis

Given an optimization problem, one starts with the set of all possible configurations (all possible routes for a TSP). With a notion of proximity (similarity, overlap) between two configurations, this set acquires a topology and becomes what will be called the configuration space. Then the cost (energy) function of the problem covers this configuration space with an energy landscape. If this landscape presents one global minimum, and no other local minimum, the problem is simple, and an iterative improvement algorithm (gradient energy descent) will be an effective search algorithm. At the other extreme is the problem in which many nearly degenerate local minima coexist. Note that some seemingly hard problems may be eventually simplifiable by the choice of a better topology or by an enlargement of configuration space, which would destabilize the previous local minima without creating new ones.

Once the problem is well defined, the first step in the analysis of configuration space consists in finding the overlap probability distribution $P(q)$, or pair statistics for local minima. A choice of weights must be made at this stage, on which the whole subsequent analysis may be very dependent. Boltzmann weighting is the natural choice in statistical mechanics (Sec. IV). Attractor basin size appears to be a more appropriate weight in the context of content-addressable memories (Sec. VIII). Equal weight below some energy cutoff has been chosen in several cases for numerical convenience. The shape of $P(q)$, as well as its size dependence, is often very instructive by itself.

Detection of ultrametricity, however, requires triangle statistics. The probability distribution $P(q_1,q_2,q_3)$ has been studied numerically for the TSP (Kirkpatrick and Toulouse, 1985) by considering for a fixed value of the smallest side the statistics of the comparison of the two other sides. Since this analysis is similar to that discussed in Sec. IV.C, it is not repeated here. Rather, we introduce two other tests, not yet mentioned, which involve a systematic elimination of one variable in the triangle statistics. The residual information is contained in a unique function of two variables, which allows for easier visualization.

(i) Integration over one variable in $P(q_1,q_2,q_3)$,

$$P(q_1,q_2)=\int P(q_1,q_2,q_3)dq_3 ,$$

gives the probability that two sides of a triangle have lengths (overlaps) q_1 and q_2. The subtracted probability function

$$C(q_1,q_2)=P(q_1,q_2)-P(q_1)P(q_2)$$

vanishes for a homogeneous distribution of points, which provides a convenient null limit. But it vanishes also for a regular tree, so it is not testing pure ultrametricity. Instead, it is sensitive to a combination of ultrametricity and heterogeneity.

In general, for ultrametric sets, $C(q_1,q_2)$ will exhibit some peaking along the diagonal ($q_1=q_2$), but it will also contain nondiagonal contributions coming from the short side of isosceles triangles. See Eq. (4.31) for an example.

Comparable data for $C(q_1,q_2)$ have been obtained numerically for random-distance traveling salesman problems (Fig. 6), Ising spin glasses (Fig. 7), and Heisenberg spins on fully frustrated cubic lattices (Kirkpatrick and Toulouse, 1985; Lallemand *et al.*, 1985).

(ii) Solla *et al.* (1986) have looked for a more direct test

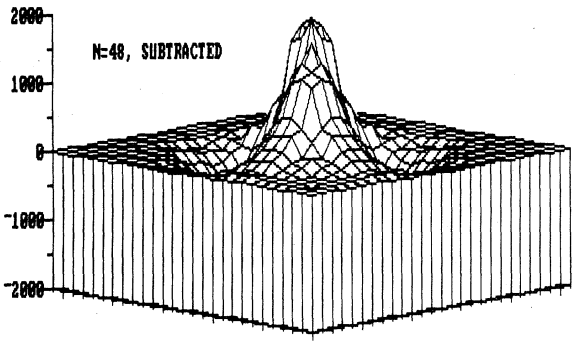


FIG. 6. Subtracted statistics of pairs of bonds for the traveling salesman problem ($N=48$). This figure shows the data looking along the $q_{ij}=q_{ik}$ axis from the large q end (Kirkpatrick and Toulouse, 1985). The vertical scale is arbitrary.

of ultrametricity and defined another subtracted distribution:

$$\tilde{C}(q_1,q_2)=\tilde{P}(q_1,q_2)-\tilde{P}(q_1)\tilde{P}(q_2)\ ,$$

where $\tilde{P}(q_1,q_2)$ is the probability that the two longer sides of a triangle have lengths (overlaps) q_1 and q_2, and $\tilde{P}(q_1)=\int \tilde{P}(q_1,q_2)dq_2$ (Fig. 8). The advantage of $\tilde{C}(q_1,q_2)$ is that it is more directly tuned to ultrametricity (for an ultrametric set, all positive values of the function are concentrated on the diagonal). The drawback is that it gives a positive test on a homogeneous set. So the functions $C(q_1,q_2)$ and $\tilde{C}(q_1,q_2)$ do give complementary information, and there is a price to pay in each case for the loss of information from the full $P(q_1,q_2,q_3)$.

Several other ultrametricity tests for complex problems have been, or can be, conceived of, but the most instructive at this stage appears to be an empirical comparison of different energy landscapes, using the same test. Section VI contains a detailed case study, which bears on the relation between ultrametricity and sparseness, with systematic use of one test.

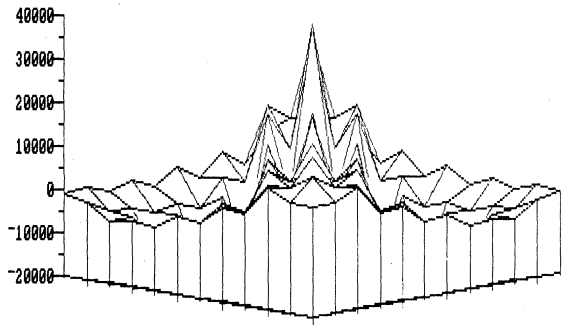


FIG. 7. Subtracted statistics of pairs of bonds evaluated for a Sherrington-Kirkpatrick spin glass model with $N=24$ spins (Kirkpatrick and Toulouse, 1985). The vertical scale is arbitrary. The local minima are similarly weighted in Figs. 6 and 7, which are thus comparable.

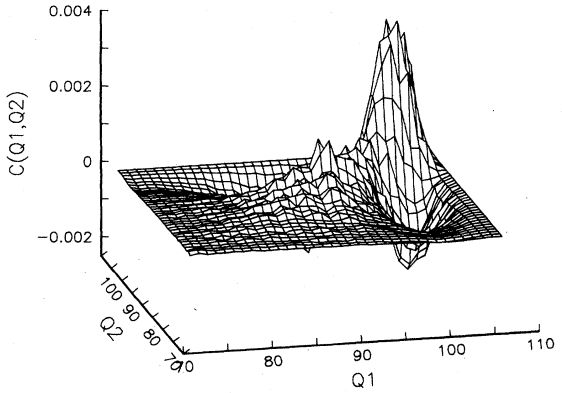


FIG. 8. Subtracted statistics of pairs of bonds for a one-dimensional placement problem (Solla *et al.*, 1986). See text for the difference between these statistics and those of Figs. 6 and 7.

C. The significance of ultrametricity

It was mentioned above (Sec. III) that there exists a trivial kind of ultrametricity, with all triangles equilateral. Nontrivial ultrametricity requires an appreciable weight for isosceles triangles, i.e., an overlap distribution function $P(q)$ that is not reduced to one δ function. This shows that evidence for ultrametricity does not end the study. Subsequent determination of the relative weights for isosceles and equilateral triangles may eventually lead to distinctions among subclasses of ultrametricity.

In any energy landscape, it is possible to define an ultrametric distance between two local minima by the minimal energy barrier separating them. Correlation between this ultrametric distance and the natural distance (viz., number of moves needed to go from one point to another) may provide a physical explanation for the occurrence of ultrametricity. On such a basis, Solla *et al.* (1986) have suggested that the simulated annealing algorithm is most efficient for ultrametric problems.

D. Other landscapes

Symmetry-breaking transitions (ferromagnets, etc.) give energy landscapes with a regular distribution of equivalent valleys. Spin glasses have drawn attention to the important case of hierarchical distributions. This certainly does not exhaust the list of interesting possibilities, and the ultrametric tree should not hide the forest. Another case that comes to mind is an energy landscape, mostly flat, but with a number of excluded islands (as in a model of immunology, where all antigens would evoke a response except those corresponding to self). Controlled modifications of an energy landscape, with feedback from configuration space to parameter space, are discussed in Sec. VIII in the context of a model for memory and learning.

VI. DEVIATIONS FROM ULTRAMETRICITY

The goal of this section is threefold. First, we shall introduce some precise definitions relative to various concepts discussed in this review. Second, the notion of the subdominant ultrametric $d^<$ will be introduced and a simple method for its construction given. A measure of the deviation from exact ultrametricity is implied from the definition of $d^<$. The end of this section is devoted to illustrative examples and to general remarks.

A. Basic definitions

Let Ω be a finite set. A *hierarchy* $H=\{h\}$ on Ω is a set of parts of Ω, such that (i) $\Omega\in H$, (ii) for every $\omega\in\Omega$, $\{\omega\}$ belongs to H, and (iii) for each pair $h\in H$, $h'\in H$: $h\cap h'\neq\varphi\Rightarrow h\subset h'$ or $h'\subset h$. Here φ refers to the standard notation of the empty set.

An *indexed hierarchy* on Ω is a pair (H,f), where H denotes a given hierarchy on Ω and f is a positive function, defined on H and satisfying the following conditions.

(i) $f(h)=0$ if and only if h is reduced to a single element of Ω (singleton) and (ii) if $h\subset h'$, then $f(h)<f(h')$. Note that for a given h, $f(h)$ corresponds to the "level" of aggregation (Sec. III), where the elements of h have been aggregated for the first time.

In order to build up a hierarchy on a given set Ω, it is generally useful to introduce a dissimilarity index or aggregation index δ, between the elements of Ω. In general, δ associates with each pair (h_1,h_2) of parts in Ω a positive real number $\delta(h_1,h_2)\geq 0$ such that $\delta(h_1,h_2)=\delta(h_2,h_1)$. For instance, if Ω is a metric space with a distance d, a large number of aggregation indices can be defined using d. For a recent survey, see the review paper of Murtagh (1983). We mention just two examples, previously discussed in Sec. III.

(a) Single-linkage index δ_1, defined as the Hausdorff distance between two parts h_1 and h_2 in Ω:

$$\delta_1(h_1,h_2)=\min\{d(x,y)\mid x\in h_1, y\in h_2\}\ . \tag{6.1}$$

(b) Complete-linkage index δ_2, defined similarly by

$$\delta_2(h_1,h_2)=\mathrm{Max}\{d(x,y)\mid x\in h_1, y\in h_2\}\ . \tag{6.2}$$

These indices are actually examples in a large class of aggregation indices widely used in systematic classification. In general, given Ω and an aggregation index δ, the construction of the corresponding hierarchy can be carried out through the search of a sequence of partitions of Ω, starting with the finest partition and ending with the class Ω containing all elements of the set Ω. Such an upward hierarchical classification has been described in Sec. III using the single-linkage index δ_1.

B. Hierarchies and ultrametrics

Note first that any given partition of the set Ω: $\Omega=\cup_i \Omega_i$ induces a large number of trivial ultrametrics: $d(x,x)=0$, $d(x,y)=1$ if $x\in\Omega_i$, $y\in\Omega_j$ $(i\neq j)$, and $d(x,y)=a$ if $i=j$, $0<a<1$. The general connection between indexed hierarchies and ultrametrics, which is clearly visible on the classification trees, was rigorously proved by Benzécri (1965,1984). This result states that there is a one-to-one correspondence between indexed hierarchies and ultrametrics both defined on the same set. Indeed, associated with each indexed hierarchy (H,f) on Ω is the following ultrametric:

$$\sigma(x,y)=\min_{h\in H}\{f(h)\mid x\in h, y\in h\}\ . \tag{6.3}$$

This means that the distance $\sigma(x,y)$ between two elements x and y in Ω is given by the smallest element in H, which contains both x and y (rule of the closest common ancestor). Inversely, each ultrametric σ is associated with one and only one indexed hierarchy.

The previous equivalence between the set $\mathscr{H}$ of indexed hierarchies and the set $\mathscr{U}$ of ultrametrics on Ω leads to the natural question of how we find the best hierarchy on Ω. The answer is to optimize $\Delta(d,\delta)$ over the set $\mathscr{U}$ of ultrametrics where $\delta\in\mathscr{U}$ and d is a given metric on Ω. Here Δ refers to a measure of adequacy between d and δ. Such an optimization problem is actually not very well defined, because of two intrinsic difficulties. The first originates in the choice of the adequacy measure $\Delta(d,\delta)$ and the second is that of the unicity of the optimized hierarchy. Among various proximity measures between two metrics d and δ, the following are just a few examples:

$$\Delta_0(d,\delta)=\max_{x,y\in\Omega}|d(x,y)-\delta(x,y)|\ , \tag{6.4}$$

$$\Delta_1(d,\delta)=\sum_{x,y\in\Omega}|d(x,y)-\delta(x,y)|\ , \tag{6.5}$$

and more generally (Minkowski distances)

$$\Delta_\alpha(d,\delta)=\left[\sum_{x,y\in\Omega}|d(x,y)-\delta(x,y)|^\alpha\right]^{1/\alpha},\quad \alpha>0\ . \tag{6.6}$$

More complicated proximity measures are given, for instance, by (Hartigan, 1967)

$$\Delta(d,\delta)=\sum_{x,y\in\Omega}p(x)p(y)[d(x,y)-\delta(x,y)]^2\ , \tag{6.7}$$

where $p(x)$ is a weight function on the elements of Ω.

Given a proximity measure Δ, there is in general the possibility of two or more optimal hierarchies. Remarkably, a simple though partial solution to this degeneracy problem is provided by the notion of the subdominant ultrametric.

C. Subdominant ultrametric

Instead of considering the whole set $\mathscr{U}$ of ultrametrics on Ω, one limits the search for δ to $\mathscr{U}^<=\{\delta\in\mathscr{U}\mid\delta\leq d\}$, which is the set of ultrametrics on Ω which are lower than d [δ is lower than d if $\delta(x,y)\leq d(x,y)$ for all pairs (x,y)]. The subdominant ultrametric $d^<$ is defined as the

upper limit of $\mathscr{U}^<$. This is the maximal element in $\mathscr{U}^<$, and by definition

$$d^<(x,y)=\mathrm{Max}\{\delta(x,y)\mid \delta\in\mathscr{U},\delta\le d\}\ . \tag{6.8}$$

This somewhat abstract definition of $d^<$ becomes clearer in terms of the associated optimization problem. If $\Delta(d,\delta)$ denotes a proximity measure, such as Δ_0, Δ_1, or Δ_α introduced before, then $d^<$ is the only solution that realizes the minimum of $\Delta(d,\delta)$ (Jardine *et al.*, 1967; Jardine and Sibson, 1968; Gower and Ross, 1969). More precisely,

$$\Delta(d,d^<)=\min\{\Delta(d,\delta)\mid \delta\in\mathscr{U},\delta\le d\}\ . \tag{6.9}$$

In addition to its simplicity and remarkable properties, the subdominant ultrametric $d^<$ is actually very easy to obtain. Usually (Benzécri, 1984), the hierarchy (H,f) is constructed first and then $d^<$ is deduced from H. For the sake of clarity, we shall describe another equivalent method, giving directly the output $d^<$ from the input d. This is the minimal-spanning-tree construction method (Gordon, 1981; Murtagh, 1983). Because of the transparent interpretation of this method we shall outline the basic steps in this construction. The minimal-spanning-tree (MST) method can be readily described in graph theoretic terms (Papadimitriou and Steiglitz, 1982). Associated with the metric space (Ω,d) is a simple, nondirected graph, with the elements of Ω as vertices. The edge (x,y) has a length equal to $d(x,y)$. The main step in obtaining $d^<$ is the construction of a minimal spanning tree on the connected graph so obtained. Recall that the MST is a tree A, having the same vertices as Ω, but of minimal total length. Note that A is not uniquely defined, and more than one MST can be constructed on Ω. Despite this nonunicity of A, $d^<$ as obtained is unique. When there is an MST on Ω, the distance $d^<(x,y)$ between two elements x and y in Ω is given by

$$d^<(x,y)=\mathrm{Max}\{d(w_i,w_{i+1}),1\le i\le n-1\}\ , \tag{6.10}$$

where $C_{xy}=\{(w_1,w_2),(w_2,w_3),\ldots,(w_{n-1},w_n)\}$ denotes the unique chain in A, between x and y ($w_1=x,w_n=y$).

This procedure defines in very precise terms the subdominant ultrametric $d^<$. A simple method for constructing an MST is provided by Kruskal's algorithm (Papadimitriou and Steiglitz, 1982). An illustrative example is shown in Fig. 9 for a set Ω of five elements.

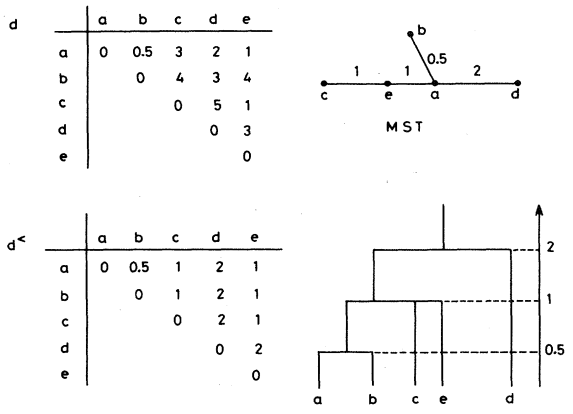


FIG. 9. An illustrative example of the construction of the subdominant ultrametric $d^<$ on the metric space (Ω,d) of five elements, using the method of the minimal-spanning tree (MST).

D. Examples

As was shown before, given a finite metric space (Ω,d), we can obtain from the subdominant ultrametric $d^<$ a simple solution of the optimization problem posed in Sec. IV.B. Furthermore, $d^<$ can readily provide clear information on the degree of ultrametricity of (Ω,d). In particular, $d^<$ provides an answer to the following question: how far is d from being an ultrametric? In other words, what is the "minimal distortion" of d in order to become an ultrametric? A possible answer is given by the comparison of d with $d^<$. A measure of the proximity between d and $d^<$ will be given by Δ_0, Δ_1, or Δ_α, as defined in Sec. IV.B. Thus a quantitative expression of the relative distortion will be given, for instance, by

$$\mathscr{D}=\sum_{x,y\in\Omega}(d(x,y)-d^<(x,y))\Big/\sum_{x,y\in\Omega}d(x,y)\ . \tag{6.11}$$

Here d denotes the input metric on Ω, and $d^<$ is the associated subdominant ultrametric. In general, $0\le\mathscr{D}\le 1$ vanishes if d is already an ultrametric (i.e., $d^<=d$) and provides a quantitative measure of ultrametricity. Small values of $\mathscr{D}$ would indicate that d is not very far from being an ultrametric and inversely.

Example 1. A simple example where the MST construction is a trivial task is provided by $\Omega=\{x_1,x_2,\ldots,x_n\}$, where Ω denotes n points on the real line, and $d(x_i,x_j)=|x_i-x_j|$ is the usual Euclidean metric. In the case where $x_i=i$, for instance ($1\le i\le n$), the subdominant ultrametric $d^<$ reduces to the trivial ultrametric $d^<(x_i,x_j)=1$: all triangles are equilateral. In this example, $\mathscr{D}=1-3/(n+1)$, and for large n, $\mathscr{D}\sim 1$ in agreement with the simple intuition: Euclidean spaces are far from being ultrametric spaces.

Example 2. A less trivial example is given by the case where Ω is a set of M binary words, taken randomly from among the 2^B possible words, of B bits each. The M elements of Ω can also be viewed as M particular Ising spin configurations of a system of B spins. This is a subset of the B-dimensional hypercube, and the distance between two words

$$u=(u_1,\ldots,u_B),\quad v=(v_1,\ldots,v_B)\ ,$$
$$u_i\in\{0,1\},\quad v_i\in\{0,1\}$$

can be defined by the Hamming metric

$$d(u,v)=\sum_{i=1}^{B}|u_i-v_i|\ .$$

It is clear that for $M=2^B$, $d^<$ reduces to the trivial ultrametric and $\mathscr{D}_B(x=1)=1-2/B\sim 1$ at large B. Here

$x = M/2^B$ refers to the filling factor of the hypercube. In the other limit $x \ll 1$, at $M \gtrsim 3$, $\mathscr{D}_B(x) \sim B^{-1/2}$, and $\mathscr{D}_B(x)$ is actually an increasing function of both x and B. Numerical calculations (Rammal *et al.*, 1985) show that $\mathscr{D}_B(x)$ assumes the staircase behavior shown in Fig. 10. Instead of a smooth function, $\mathscr{D}_B(x)$ exhibits a series of stairs and jumps between adjacent plateaus at very precise values x_k of x, $2 \le k \le B/2$. Moreover, the values taken by $\mathscr{D}_B(x)$ are quantized and can be labeled by an integer k. Indeed the stairs are well represented by

$$\mathscr{D}_B(x_k) = 1 - 2k/B \quad \text{(level of the stair } k\text{)}, \tag{6.12}$$

$$x_k = 1 \Big/ \sum_{i=0}^{k-1} \binom{B}{i}, \quad k \ge 2$$

$$\text{(abscissa of the center of the stair } k\text{)}. \tag{6.13}$$

This puzzling behavior of the relative distortion $\mathscr{D}_B(x)$ can be understood with geometrical considerations. In fact, each of the M words in Ω can be considered as the center of a k sphere, of volume

$$\omega_k = \sum_{i=0}^{k-1} \binom{B}{i},$$

containing all points on the hypercube within a distance up to k ($1 \le k \le B/2$). For large B and M, but finite x, the probability of coverage of the hypercube by M such k spheres is given by $1-(1-\omega_k/2^B)^M \sim 1-\exp(-x\omega_k)$. The centers of the stairs correspond to $x = x_k \sim 1/\omega_k$. The jumps of $\mathscr{D}_B(x)$ are due to the competition between two hierarchies associated with k and $k-1$, each of which corresponds to a regular coverage of the hypercube.

Note that for fixed but large B, $\mathscr{D}_B(x)$ approaches zero as x goes to zero. This means that ultrametricity is actually a natural property of large spaces. In contrast, $\mathscr{D}_B(x)$ reduces to a step function for B strictly infinite.

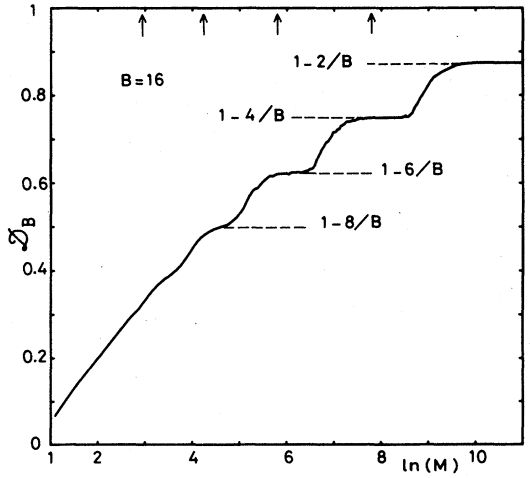


FIG. 10. Steplike behavior of the degree of ultrametricity $\mathscr{D}_B$ for M words taken randomly from among the set of 2^B B-bit words. Arrows indicate the centers of stairs (Rammal *et al.*, 1985).

E. Remarks

Example 2 above shows how an ultrametric set can be generated by a random process, in the sparse limit, where the collection of selected objects is a very small fraction of the total number of possible objects.

In the case of the teletype code (Fig. 2), the 26 letters of the alphabet densely fill the space of $2^5 = 32$ binary words of 5 letters. The deviation from ultrametricity, using the Hamming distance, is consequently large (upper plateau of Fig. 10).

It is a useful generalization to consider "words" made of B "letters," extracted from an alphabet of size R. For binary words, as above, $R = 2$. For nucleic acids, the repertoire contains four nucleotides, $R = 4$. For proteins, twenty amino acids, $R = 20$. For spoken words, there are around forty phonemes, $R = 40$. Increase of R, which is the ramification number of the decision tree, increases the number of possible words, which is equal to R^B. In biomolecules as in languages, the selected words are very sparse sets.

The analysis in Sec. VI.D, example 2, generalized to $R \ge 2$, leads to

$$x_k = 1 \Big/ \sum_{i=0}^{k} \binom{B}{i} (R-1)^i$$

and (6.14)

$$\mathscr{D}_B(x_k) = 1 - \frac{k}{B}\frac{R}{R-1}.$$

This means that for increasing R, $\mathscr{D}_B(x)$ increases for the same value of k.

When compared with triangle statistics, the measurement of the degree of ultrametricity with $\mathscr{D}$ suffers actually from two disadvantages: the chaining effect and sensitivity to fluctuations. Indeed, two points x and y that are very far from each other for d [i.e., large $d(x,y)$] will become very close for $d^<$ if there is a chain of points between x and y such that every successive link is very small in comparison with $d(x,y)$. This is the so-called chaining effect of the subdominant ultrametric. This effect is at the origin of the strong sensitivity of $\mathscr{D}$ to fluctuations, as can be seen on the jumps shown in Fig. 10. Triangle statistics, in contrast, do not exhibit such a sensitivity because of their statistical nature. For these reasons, the measure of $\mathscr{D}$ is probably best adapted to sparse spaces that are nearly ultrametric.

Finally, what can be the use of a degree of ultrametricity? Clearly, it is open to criticism. One mere number, like a degree of ultrametricity (or a quantity of information or a fractal dimension), may be a very poor characterization of a complex problem. Moreover, there is some arbitrariness in the definition of $\mathscr{D}$. There are two answers to this criticism.

First, a degree of ultrametricity can be helpful as a function of the parameters in a given problem. Its increase or decrease may then be meaningful. The variation of the entropy with temperature, pressure, magnetic field, etc., provides a valuable example. Second, it is plausible

that in a number of problems, as is the case for the mean-field theory of spin glasses, Sec. IV, ultrametricity will be found to hold in some extreme simplifying limit (uncorrelated random exchange interactions in the SK model, neutral selection in evolution theory). At the next level of analysis, what will appear significant may be the deviations from ultrametricity (due to correlated interactions, in spin glass theories of learning, or due to convergent evolution in evolution theory). Then the deviation from ultrametricity may indeed become a measure of the quantity of relevant information.

VII. DYNAMICAL ASPECTS

In this section, two types of time evolution in an ultrametric space are considered. These two extremes do not exhaust all possible dynamical processes, but they offer instructive limiting cases. One motivation for the study of random walks in ultrametric spaces (A) comes from the slow relaxation effects in disordered systems (glasses, spin glasses, proteins, etc.). In these materials, there is growing evidence for a hierarchical structure in configuration space, and the tree is certainly a heterogeneous one. Since the topic of dynamics of disordered systems is much too vast to be covered here, attention is restricted to random walks on homogeneous ultrametric spaces. These toy models exhibit a variety of anomalous diffusive behaviors and possibilities of transitions between them. Thus they add to our knowledge of diffusion on periodic or fractal lattices. For contrast, the time evolution during a temperature descent is reviewed (B). The motivation in this case comes from simulated annealing, which is an optimization search strategy.

A. Random walks in ultrametric spaces

A walker, restricted to steps of a given length, does not diffuse in an ultrametric space, because he remains always within the same distance from his starting point (Mézard *et al.*, 1984b). This localization effect is a direct consequence of the ultrametric inequality; it is just another expression of the absence of intermediates.

Greater richness of behaviors, however, obtains if one allows for variable-range hopping.

1. The pure model

The simplest formulation of the problem assumes a regular hierarchical tree with p levels and a fixed branching ratio K (Fig. 11). The total number of points is $N=K^p$. The distance R between two points is taken to be equal to the height m, $m=1,\ldots,p$, of their closest common ancestor level.

Temperature-assisted hopping is assumed, with potential barriers $\Delta(m)$ which are increasing functions of the level m. The probability $W(m)$ of jumping over a barrier of level m is thus taken as

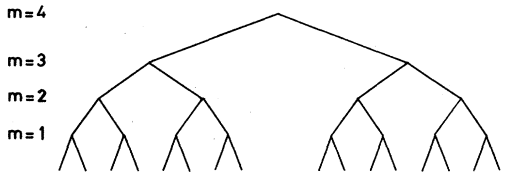


FIG. 11. Regular hierarchical tree with $p=4$ levels and a fixed branching ratio $K=2$. The distance R between two points at the bottom line is given by the height m, $m=1,\ldots,p$ of their closest common ancestor level.

$$W(m)\sim e^{-\Delta(m)/T} . \tag{7.1}$$

The time needed to move a distance m apart is unaffected by smaller steps:

$$t\sim e^{\Delta(m)/T} , \tag{7.2}$$

and the walker's displacement $\langle R(t)\rangle$ is given implicitly in terms of the barrier height function $\Delta(m)$. Two examples will serve as illustrations (Ogielski and Stein, 1985; Paladin *et al.*, 1985).

(a) If $\Delta(m)$ is linear in m, $\Delta(m)=\Delta m$, then Eq. (7.2) implies

$$\langle R(t)\rangle\sim\frac{T}{\Delta}\ln t , \tag{7.3}$$

which is an anomalously slow diffusion law. Normal diffusion, in Euclidean spaces of any dimensionality, has the following asymptotic behavior:

$$\langle R^2(t)\rangle\sim Dt . \tag{7.4}$$

Slow anomalous diffusion is commonly found on self-similar lattices (Alexander and Orbach, 1982; Rammal and Toulouse, 1983),

$$\langle R^2(t)\rangle\sim t^{\tilde{d}/\bar{d}} , \tag{7.5}$$

where $\tilde{d}$ and $\bar{d}$ are, respectively, the fractal and spectral dimensions; $\tilde{d}\le\bar{d}\le d$, where d is the embedding Euclidean space dimension. For a Euclidean lattice, $\tilde{d}=\bar{d}=d$, hence Eq. (7.4). Our ultrametric space has an infinite fractal dimension, hence Eq. (7.3).

Another quantity of interest is the probability of return to the origin $P_0(t)$, which can be easily estimated from the inverse of the number of points accessible within time t. From Eq. (7.3),

$$P_0(t)\sim K^{-R(t)}\sim t^{-(T/\Delta)\ln K} , \tag{7.6}$$

to be compared with the general expression for self-similar lattices,

$$P_0(t)\sim t^{-\tilde{d}/2} . \tag{7.7}$$

Identification gives

$$\tilde{d}=\frac{2T}{\Delta}\ln K , \tag{7.8}$$

showing that the spectral dimension is temperature dependent. As a consequence, a transition from compact to noncompact diffusion occurs at $T=\Delta\ln K$. Indeed (Ram-

mal and Toulouse, 1983), for $\tilde{d}<2$, the mean number of distinct visited points $S(t)$ varies as

$$S(t)\sim t^{\tilde{d}/2}, \tag{7.9}$$

whereas, for $\tilde{d}>2$, $S(t)$ is linear in t. Such a transition, as a function of temperature, has been observed numerically by Blumen *et al.* (1985).

(b) Another interesting case corresponds to the choice $\Delta(m)=\Delta \ln m$. These barriers increase less rapidly as a function of distance, and the diffusion is expected to be faster than above. Indeed, Eq. (7.2) leads to

$$\langle R(t)\rangle \sim t^{T/\Delta} \tag{7.10}$$

and

$$P_0(t)\sim e^{-(\ln K)t^{T/\Delta}}, \tag{7.11}$$

which is a Kohlrausch (Frauenfelder, 1984), or stretched exponential, law. Both fractal and spectral dimensions are infinite. Convergence for the sum of the probabilities (7.1) limits the temperature variations to $0<T<\Delta$. A new transition, from slow to fast anomalous behavior, appears at $T=\Delta/2$. It corresponds to entry into a Lévy flight regime.

2. A mixed model

The first-studied model for ultrametric diffusion (Huberman and Kerszberg, 1985) is actually a diffusion model of mixed nature, because it contains a finite-dimensional (one-dimensional, in fact) character. The distribution of potential barriers is represented in Fig. 12, assuming $\Delta(m)=\Delta m$ and $K=2$.

Because of the one-dimensional character, the relation between R and m is now

$$R=K^m. \tag{7.12}$$

The effective jump probability over a barrier of level m is

$$W^{\text{eff}}(m)\sim e^{-\Delta(m)/T}\cdot\frac{1}{R}, \tag{7.13}$$

which includes an entropy contribution coming from the constraint of step-by-step, one-dimensional motion. Hence

$$t\sim e^{\Delta(m)/T}\cdot R, \tag{7.14}$$

entailing, for $\Delta(m)=\Delta m$,

$$\langle R(t)\rangle \sim t^{T\ln K/(T\ln K+\Delta)} \tag{7.15}$$

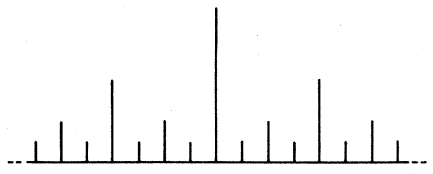


FIG. 12. Hierarchical array of barriers for a mixed model of diffusion. Particles diffuse from box to neighboring box, which gives a one-dimensional character to the random walk.

and

$$P_0(t)\sim t^{-T\ln K/(T\ln K+\Delta)}. \tag{7.16}$$

Equality of the exponents in Eqs. (7.15) and (7.16) is a consequence of $\tilde{d}=d=1$.

These expressions are only valid for $T<\Delta/\ln K$. Indeed, it is clear that fast anomalous diffusion cannot possibly occur in this model (Teitel and Domany, 1985). For $T>\Delta/\ln K$, normal diffusion sets in. Similar results are obtained with randomly placed barriers (Alexander *et al.*, 1981), and thus the regular hierarchical structure is not really a crucial feature.

3. Other hierarchical models

Very fast anomalous diffusion, $\langle R^2(t)\rangle \sim t^3$, is observed in turbulent flows. A hierarchical model (more complex than those given above) in which the ancestor levels of the hierarchy also represent physical states, has been solved and shown to encompass the possibility of describing fast diffusion (Grossmann *et al.*, 1985).

B. Simulated annealing

A random walk is a blind way of exploring a space. This fact is aggravated for diffusion in an ultrametric space, because trapping in potential wells slows down the search.

The method of simulated annealing (Kirkpatrick *et al.*, 1983) obviates this effect, which plagues conventional isothermal Monte Carlo methods, by making a temperature descent, whereby major decisions are made first, at higher temperatures, and minor decisions later, at lower temperatures.

This temperature control thus inverts, in a way, the evolution process of a random walk, which wastes a lot of time in relatively minor decisions before addressing more important ones.

VIII. SOME OTHER OPEN PROBLEMS

Recently, the concept of ultrametricity has made its appearance in other problems of biological interest, which are briefly described here.

A. Neural networks

Longstanding interest in neural networks originated with the modeling of brain functions and with the design of computing devices. A neural network, with the standard neuronal dynamics, can be associated with an energy landscape, provided the connections (synapses) are symmetric. It was stressed by Hopfield (1982) that an energy landscape provides the simplest model for a content-addressable memory. The valleys are the stored memory patterns, and retrieval occurs by gradient dynamics. Suitable learning rules allow for adaptive energy landscapes.

Much subsequent work has been concerned with the maximal storage capacity, assuming random uncorrelated input patterns. Within the mean-field framework, it is possible to obtain a detailed theory (Amit *et al.*, 1985) that interpolates between the limits of the standard ferromagnet and the pure spin glass. This is a step forward in statistical mechanics, because it shows the possibility of describing analytically a very complex behavior of metastable states, and it fills a gap between the usual symmetry-breaking transitions in the manner of Landau and the replica-symmetry-breaking transitions in the manner of Parisi (Sec. IV).

However, with these assumptions, the model does not provide the memory with a hierarchical structure, though this would clearly be desirable for categorization purposes. Two suggestions, which are in fact complementary, have been advanced recently for obtaining an ultrametric content-addressable memory. One approach starts from a spin-glass-like energy landscape and introduces strong learning constraints, so that the tree of configurations is pruned by selective learning, in a way that preserves its ultrametricity (Kirkpatrick and Toulouse, 1985; Toulouse *et al.*, 1986). The other approach introduces hierarchical categorization at the encoding level, with a layered perceptual architecture (Parga and Virasoro, 1985; Virasoro, 1986).

B. Protein freezing and folding

Proteins are strings of amino acids. A typical example is myoglobin, which has 153 monomer units (Frauenfelder, 1984). The secondary and tertiary structures of biological proteins show remarkable variations with temperature. Meanwhile the covalent linear sequence, or primary structure, is rigid and acts like a quenched disorder, which suggests an analogy with spin glasses. Below physiological temperatures, there is mounting evidence for a freezing transition into a hierarchy of conformational states and substates, with several tiers (Ansari *et al.*, 1985; Stein, 1985). Note the remarkable appearance of ultrametricity at two levels in protein theory, and for two distinct reasons: (i) in protein configuration space, because of a presumably hierarchical energy landscape (Sec. IV), (ii) in protein sequence space, because of a stochastic branching process, due to a largely neutral evolution (Sec. III).

At a higher level of protein organization, the folding transition, which is a capital issue in molecular biology (Wetlaufer, 1984), is an open challenge, at the horizon of optimization theory (Secs. V and VII).

C. Evolutive biotechnology

The close analogy between simulated annealing and biological evolution (at the morphological level), as optimization processes (Kirkpatrick *et al.*, 1983), has led to the suggestion that "simulated evolution" may be an efficient practical means of producing proteins with desired properties. In this analogy, the mutation rate plays the role of temperature. Accelerated evolution is obtained by control of the mutation rate as well as by rapid selective scanning, instead of the slow mechanism of survival of the fittest (Anderson, 1985; Eigen, 1985a,1985b).

IX. CONCLUSIONS

At the end of this survey, covering instances from physics and biology, some conclusions can be drawn as to the reasons for the occurrence of ultrametricity in the natural sciences.

Ultrametricity emerges as a consequence of randomness and of the laws of large numbers. Two distinct typical situations have been identified. The simplest is a purely stochastic branching process in a large space (e.g., the neutral evolution of pseudogenes). The second concerns the valley structure, i.e., the energy landscape, of disordered frustrated systems (spin glasses and many optimization problems).

In the case of evolution, both aspects, random walk and optimization, are almost always intermingled. The relative proportions vary according to the characteristics considered, from the molecular to the morphological level.

In the case of spin glasses, ultrametricity holds both in the high-temperature (paramagnetic) phase and in the low-temperature (spin glass) phase. In the paramagnetic phase, the coupling energy is smaller than the thermal motion, and the ultrametricity is a trivial one, with all triangles equilateral on average, in the thermodynamic limit. Below the spin glass transition, ergodicity breaking sets in and reveals a hierarchical tree of valleys; ultrametricity is preserved, albeit with a changed, richer content, since it contains information on the energy landscape due to the couplings. In the absence of a magnetic field, the spin glass transition breaks ultrametricity, inasmuch as time-reversal invariance is then spontaneously broken. If the paramagnetic ultrametricity went unnoticed earlier, it is because it stood in the shadow of the more conspicuous effects associated with the usual symmetry-breaking phase transitions.

As nontrivial ultrametricity is detectable by third-order statistics but invisible at second order, so may subtler structures be concealed by third-order statistics and revealed at higher orders.

ACKNOWLEDGMENTS

We express our gratitude to our colleagues, M. Mézard, G. Parisi, and N. Sourlas, with whom all our knowledge on ultrametricity has been shaped and shared. We are also thankful to B. Douçot, S. Kirkpatrick, and N. Parga for very fruitful collaborations. Kind permissions for figure reproductions, from R. Bhatt, M. Kimura, and S. Solla, are gratefully acknowledged.

REFERENCES

Alexander, S., J. Bernasconi, W. R. Schneider, and R. Orbach, 1981, Rev. Mod. Phys. **53**, 175.

Alexander, S., and R. Orbach, 1982, J. Phys. (Paris) Lett. **43**, L625.

Amice, Y., 1975, *Les nombres p-adiques* (P.U.F., Paris).

Amit, D., H. Gutfreund, and H. Sompolinsky, 1985, Phys. Rev. Lett. **55**, 1530.

Anderson, P. W., 1985, in *Emerging Syntheses in Science,* edited by D. Pines (Santa Fe Institute, Santa Fe, N.M.).

Ansari, A., J. Berendzen, S. F. Bowne, H. Frauenfelder, I.E.T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young, 1985, Proc. Natl. Acad. Sci. U.S.A. **82**, 5000.

Bachas, C., 1985, Phys. Rev. Lett. **54**, 53.

Benzécri, J. P., 1965, Lecture notes, reproduced in Benzécri, 1984.

Benzécri, J. P., 1984, *L'analyse des données* 1, La taxinomie (Dunod, Paris).

Bhatt, R. N., and A. P. Young, 1986, J. Magn. Magn. Mater. **54-57**, 191.

Binder, K., and A. P. Young, 1986, Rev. Mod. Phys., in press.

Blandin, A., M. Gabay, and A. T. Garel, 1980, J. Phys. C **13**, 403.

Blumen, A., J. Klafter, and G. Zumofen, 1985, J. Phys. A **19**, L77.

Bouchaud, J. P., and P. Le Doussal, 1985, Europhysics Lett. **1**, 91.

Černý, V., 1985, J. Optim. Theory Appl. **45**, 41.

De Dominicis, C., and A. P. Young, 1983, J. Phys. A **16**, 2063.

Derrida, B., and G. Toulouse, 1985, J. Phys. Lett. (Paris) **46**, L223.

Dieudonné, J., 1978, Ed., *Abrégé d'histoire des mathématiques* (Hermann, Paris).

Doolittle, R. F., 1985, Sci. Am. **253**, 74.

Edwards, S. F., and P. W. Anderson, 1975, J. Phys. F **5**, 965.

Eigen, M., 1985, in *Emerging Syntheses in Science,* edited by D. Pines (Santa Fe Institute, Santa Fe, N.M.), p. 25.

Florek, K., J. Lubaszewicz, J. Perkal, H. Stainhaus, and S. Zubrzycki, 1951, Colloq. Math. **2**, 282.

Frauenfelder, H., 1984, Helv. Phys. Acta **57**, 165.

Fréchet, M., 1906, in Dieudonné (1978).

Fu, Y., and P. W. Anderson, 1985, J. Phys. A **19**, 1605.

Garey, M. R., and D. S. Johnson, 1979, *Computers and Intractability* (Freeman, San Francisco).

Gordon, A. D., 1981, *Classification* (Chapman and Hall, London).

Gower, J. C., and G. L. S. Ross, 1969, Appl. Stat. **18**, 54.

Grossmann, S., F. Wegner, and K. H. Hoffmann, 1985, J. Phys. Lett. (Paris) **46**, L575.

Hartigan, J. A., 1967, J. Am. Stat. Assoc. **62**, 1140.

Henley, C., 1985, private communication.

Hensel, K., 1897, in Dieudonné (1978).

Hopfield, J. J., 1982, Proc. Natl. Acad. Sci. U.S.A. **79**, 2254.

Huberman, B., and M. Kerszberg, 1985, J. Phys. A **18**, L331.

Jardine, C. J., N. Jardine, and R. Sibson, 1967, Math. Biosci. **1**, 173.

Jardine, N., and R. Sibson, 1968, Math. Biosci. **2**, 465.

Jardine, N., and R. Sibson, 1971, *Mathematical Taxonomy* (Wiley, New York).

Johnson, S. C., 1967, Psychometrika **32**, 241.

Khanin, K. M., and Y. G. Sinai, 1979, J. Stat. Phys. **20**, 573.

Kimura, M., 1983, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, England).

Kimura, M., 1985, New Scientist, July 11, 41.

Kirkpatrick, S., 1981, in *Disordered Systems and Localization,* Lecture Notes in Physics Vol. 149, edited by J. Ehlers, K. Hepp, R. Kippenhan, H. A. Weidenmüller, and J. Zittartz (Springer, Berlin), p. 280.

Kirkpatrick, S., 1984, J. Stat. Phys. **34**, 975.

Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi, 1983, Science **220**, 671.

Kirkpatrick, S., and G. Toulouse, 1985, J. Phys. (Paris) **46**, 1277.

Krasner, M., 1944, C. R. Acad. Sci. **219**, tome II, 433.

Kronecker, L., 1882, in Dieudonné (1978).

Lallemand, P., H. T. Diep, A. Ghazali, and G. Toulouse, 1985, J. Phys. (Paris) Lett. **46**, L1087.

Lewin, R., 1985, Science **229**, 743.

Mackenzie, N. D., and A. P. Young, 1982, Phys. Rev. Lett. **49**, 301.

Mézard, M., and G. Parisi, 1985, J. Phys. (Paris) Lett. **46**, L771.

Mézard, M., G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, 1984a, Phys. Rev. Lett. **52**, 1156.

Mézard, M., G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, 1984b, J. Phys. (Paris) **45**, 843.

Mézard, M., G. Parisi, and M. A. Virasoro, 1985, J. Phys. (Paris) Lett. **46**, L217.

Mézard, M., G. Parisi, and M. A. Virasoro, 1986, Europhys. Lett. **1**, 77.

Mézard, M., and M. A. Virasoro, 1985, J. Phys. (Paris) **46**, 1293.

Murtagh, F., 1983, Comput. J. **26**, 354.

Ninio, J., 1983, *Molecular Approaches to Evolution* (Princeton University, Princeton, N.J.).

O'Brien, S. J., W. G. Nash, D E. Wildt, M. E. Bush, and R. E. Benveniste, 1985, Nature (London) **317**, 140.

Ogielski, A. T., and D. L. Stein, 1985, Phys. Rev. Lett. **55**, 1634.

Paladin, G., M. Mézard, and C. De Dominicis, 1985, J. Phys. (Paris) Lett. **46**, L985.

Palmer, R. G., D. L. Stein, E. Abrahams, and P. W. Anderson, 1984, Phys. Rev. Lett. **53**, 958.

Papadimitriou, C. H., and K. Steiglitz, 1982, *Combinatorial Optimization* (Prentice-Hall, Englewood Cliffs).

Parga, N., G. Parisi, and M. A. Virasoro, 1984, J. Phys. (Paris) Lett. **45**, L1063.

Parga, N., and M. A. Virasoro, 1986, J. Phys. (Paris), in press.

Parisi, G., 1980, J. Phys. A **13**, 1887.

Parisi, G., 1983, Phys. Rev. Lett. **50**, 1946.

Penny, D., L. R. Foulds, and M. D. Hendy, 1982, Nature (London) **297**, 197.

Rammal, R., J. C. Angles d'Auriac, and B. Douçot, 1985, J. Phys. (Paris) Lett. **46**, L945.

Rammal, R., and G. Toulouse, 1983, J. Phys. (Paris) Lett. **44**, L13.

Schikhof, W. H., 1984, *Ultrametric Calculus* (Cambridge University Press, Cambridge, England).

Shepard, R. N., 1980, Science **210**, 390.

Sherrington, D., and S. Kirkpatrick, 1975, Phys. Rev. Lett. **32**, 1792.

Sokal, R. R, and P. H. A. Sneath, 1963, *Principles of Numerical Taxonomy* (Freeman, San Francisco).

Solla, S. A., G. B. Sorkin, and S. R. White, 1986, in *Disordered Systems and Biological Organization,* NATO Advanced Study Institute Ser. F, Vol. 20, edited by E. Bienenstock, F. Fogelman, and G. Weisbuch (Springer, Berlin).

Sommers, H. J., and W. Dupont, 1984, J. Phys. C **32**, 5785.
Sørensen, T., 1948, Biol. Skr. **5**, 1.
Sourlas, N., 1984, J. Phys. (Paris) Lett. **45**, L969.
Stein, D., 1985, Proc. Natl. Acad. Sci. U.S.A. **82**, 3670.
Teitel, S., and E. Domany, 1985, Phys. Rev. Lett. **55**, 2176.
Toulouse, G., 1984, Helv. Phys. Acta **57**, 459.
Toulouse, G., S. Dehaene, and J. P. Changeux, 1986, Proc. Natl. Acad. Sci. U.S.A. **83**, 1695.
Vannimenus, J., G. Toulouse, and G. Parisi, 1981, J. Phys. **42**, 565.
Virasoro, M. A., 1986, in *Disordered Systems and Biological Organization,* NATO Advanced Study Institute Ser. F, Vol. 20, edited by E. Bienenstock, F. Fogelman, and G. Weisbuch (Springer, Berlin).
Volkenshtein, M. V., 1984, Usp. Fiz. Nauk **143**, 429 [Sov. Phys.—Usp. **27**, 515 (1984)].
Walstedt, R. E., 1983, in *Heidelberg Colloquium on Spin Glasses,* Lecture Notes in Physics, edited by J. L. van Hemmen and I. Morgenstern (Springer, Berlin), p. 177.
Wetlaufer, D. B., 1984, Ed., *The Protein Folding Problem* (Westview Press, Boulder).
Wilson, A. C., 1985, Sci. Am. **253**, 148.
Young, A. P., A. J. Bray, and M. A. Moore, 1984, J. Phys. C **17**, L155.
Zuckerkandl, E., and L. Pauling, 1965, J. Theor. Biol. **8**, 357.