Systems with Learning and Memory Abilities
J. Delacour and J.C.S. Levy (eds.)

CELLULAR AUTOMATA AND MODELS OF MEMORY

P. Ruján†

Institut für Festkörperforschung
der Kernforschunsanlage Jülich
Postfach 1913, D-5170 Jülich
Federal Republic of Germany

We review the status of physical models of memory and explore different propositions of using cellular automata and related models for pattern recognition tasks, for associative, content addressable memory and for the classification and the parallel processing of structured data sets.

1. Introduction

1.1. What is memory?

There are many living systems, like the central nervous system, the immune system, economico-social systems, etc., which show some common features as

a. They consist of a very large number of highly specialized components,
b. They are able to respond to an enormous variety of stimuli, some never experienced before;
c. They are cooperative systems in the sense that their responses seem to be quite insensitive to local perturbations or damages;
d. They are able to learn, store, and manipulate a large amount of information which is not "genetically" determined and is lost with the death of each individual,
e. Each part of such a system has the ability to distinguish between self and the outside world.

If physics is to be considered more as a way of thinking rather than being the science of the dead matter, then the assumption that these complex systems have some *universal* underlying functional principles leads naturally to the application of statistical physics and field theory. We shall discuss here only one of the features exposed above, the question of *memory*. As a general definition we might say that memory is the ability of a system to react to a given situation (instance) on the basis of its past experience. This definition emphasizes the active rôle of memory and goes beyond the notion that memory is a passive response to the environment. Nevertheless, in most physical models of associative memory one simply retrieves stored key patterns, without taking into account the general context within which they are interpreted. Accordingly, we shall work with more specific definitions appropriate to the particular aspects of memory being discussed.

---

†On leave from Institute for Theoretical Physics, Eötvös University, Budapest

There is no space - and no need - to give in this review a complete enumeration of the fast expanding number of models with memory abilities, since good books [Kohonen, 1987] and reviews [Sompolinsky, 1987; Amit, 1987; Hogg and Huberman 1988; Clark *et al*, 1985] are already available. Our aim is to present the main idea around much of the current activity is revolving, to point out the common points of the different approaches, and to guess their chances against natural selection. Given my physics background I have to apologize in advance to the numerous authors whose relevant works will not be mentioned here. As Mark Twain put it, the key to success is confidence and ignorance.

1.2. Associativity, Learning, and Adaptation

Let us call a Content Addressable Memory (CAM) a system able to *retrieve* a *learned* pattern when presented with some *fuzzy* input. In a mathematical formulation retrieval means that the dynamics will lead to a stored key pattern. It has been also suggested that recognition happens via a *grandmother* neuron, which represents a small group of neurons, firing only when a lot of specialized detectors have identified a class of features, say your grandmother. In this case it is not the key pattern which is retrieved, but the attractor corresponding to the key pattern. In this sense this type of memory mechanism is better called a *classifier* system.

Learning is strongly associated to perception: usually it means a slow training process during which the interactions between neurons are changed. The particular mechanism of these changes are largely unknown, for even the widely accepted synaptic plasticity hypothesis does not have a firm experimental proof [Rose, 1988]. The training of units is usually done before the actual tests are performed. In some cases modifications of interactions are also allowed whenever a pattern is presented and/or recognized. Some models even require learning (and forgetting) to be part of the retrieval process.

A fuzzy pattern is a slightly distorted key pattern. The degree of fuzziness can be given as the distance between the two patterns. The most common metric is the Hamming distance, the fraction of unmatching neuronal units. *Adaptable* systems are expected to build their own metric when repeatedly exposed to a given environment: then a completely new - but typical - entry shall be correctly classified.

1.3. Continuous vs. discrete models

Much is actually known about the structure of single neurons [Changeux, 1982]. The effect of postsynaptic excitation is to induce the firing of an impulse going down the axon and is possibly modulated by other synapses attached to the axon. The neuron needs a refractory time to regain again its firing ability. Synapses excite or inhibit the neuron via neuroreceptors, some synapses even produce both effects at the same time. A neuron seems to have a high level of synaptic noise and it is assumed that they behave as simple additional threshold units. A simpler description is obtained by assuming that from a functional point of view a neuronal unit might fire ($n_i = 1$) or not ($n_i = 0$) [McCulloch and Pitts, 1943]. Throughout this article we shall be concerned with units described in terms of discrete variables, rather than continuous ones. This need not be a strong limitation: lattice models display a very rich behaviour and appropriate coarse graining procedures lead us back to continuous variables. Probably a single neuron is a much more complex object than an additive threshold unit or a discrete variable. It is not clear today how this local complexity affects the macroscopic properties of a neuronal net.

The article is organized as following: in the Section 2. we provide some preliminary examples on the methods used in statistical physics for the calculation of ground states of spin models by using the linear programing and the trans-

fer matrix method. We also mention a list of models with very complex phase diagrams. In the following Section we discuss cellular automata and their relation to statistical physics. The transfer matrix approach provide a common methodological and interpretational tool for both fields. Different properties of homogeneous and random cellular automata are briefly summarized and results obtained with a cellular automaton preserving (and recognizing) the topology of an initial configuration are presented. In Section 4. different variants of neuronal network models are discussed and recent results summarized. These models include the Little - Hopfield models, projection methods and general asymmetric models storing random patterns. Section 5. deals with cellular automata like feed-forward layered structures. For sequential dynamics we give a general scheme of constructing cellular automata rules providing a set of desired attractors and discuss some implications of this approach. The effect of hidden units and internal degrees of freedom are briefly discussed. Section 6. closes this review by some comments on the present problems and the possible avenues of solving them in future.

2. Statistical Physical Models

2.1. Calculating ground states in lattice spin models

The linear programing method

In statistical physics one defines an energy functional which describes the interaction between some lattice variables. Consider Ising-type variables, $s_i = \pm 1$ interacting with an external field and with nearest neighbours. For example, in one-dimension, the energy is

$$E = -J \sum_n s_n s_{n+1} - H \sum_n s_n \tag{2.1}$$

Define by $p_+ = \lim_{N\to\infty} \frac{N_+}{N}$ the probability that a spin points up, where N_+ is the number of spins pointing up, N is the total number of spins. Likewise, the probability of having a nearest neighbour pair of spins pointing up is defined as $p_{++} = \lim_{N\to\infty} \frac{N_{++}}{N}$, etc. In terms of these probabilities, the energy per spin will have the following form in the thermodynamic limit $N \to \infty$:

$$\varepsilon \equiv \frac{E}{N} = -J(p_{++} + p_{--} - p_{+-} - p_{-+}) - H(p_+ - p_-) \tag{2.2}$$

From elementary probability theory one has

$$\sum_s p(s) = 1, \qquad \sum_{s_1} p(s_1, s_2) = p(s_2) \tag{2.3}$$

and since in addition $p_{+-} = p_{-+}$ only two parameters, say p_+ and p_{--} are independent. All parameters are probabilities $\in [0, 1]$, so

$$0 \le p_+, p_{--} \le 1 \tag{2.4a}$$

and

$$0 \le 2p_+ - 1 + p_{--} \le 1 \tag{2.4b}$$

$$0 \le 1 - p_+ - p_{--} \le 1 \tag{2.4c}$$

The minimization of the energy (2.2) subject to the conditions (2.4a-c) is a typical optimization task which can be solved using the method of linear programming [Papadimitriou and Steiglitz, 1982]. The extrema of the energy ε are

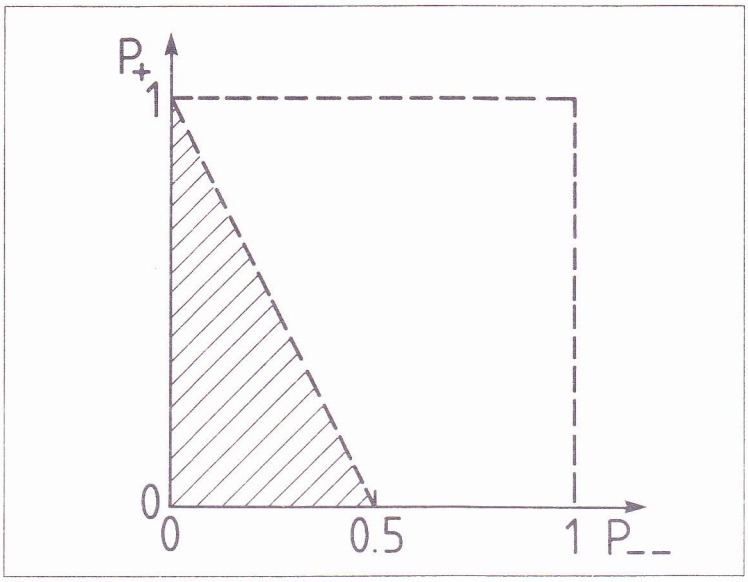

Fig. 1: The triangle corresponding to the allowed values of the probabilities p_+ and p_{--}

given by the vertices of the *convex polytop* defined by the constraints (2.4a-c) in the space of the independent variables (p_+, p_{--}) -see Fig. 1.

Only the three states

1. $p_+ = 0, \quad p_{--} = 1 \qquad (\downarrow\downarrow\downarrow\downarrow, \ldots)$
2. $p_+ = 1, \quad p_{--} = 0 \qquad (\uparrow\uparrow\uparrow\uparrow, \ldots)$
3. $p_+ = \frac{1}{2}, \quad p_{--} = 0 \qquad (\uparrow\downarrow\uparrow\downarrow, \ldots)$

can be ground states. Note that until now only the range of interaction, the symmetries of the energy, and trivial probability identities have been used! The actual phase diagram is obtained by matching the energy of different phases. Phases corresponding to the corners of the unit hypercube do not have common boundaries. At points where the energy of different phases are equal ("multicritical points") one has a multitude of configurations with the same energy. The linear programing method is quite powerful and can be extended to low temperatures, as shown in the contribution of W. Selke [Selke, 1988]. In some non-Bravais lattices the linear programing method "fails" - as shown by Kanamori [Kanamori, 1984] and using a simpler method by Villain [Villain, 1986] - walls separating different domains might form lattices (a triangular lattice, for example) which can be dilated to an arbitrary extent by the change of the external field.

Transfer matrix method

Another method for the determination of ground states is the transfer matrix method [e.g. Derrida, Pomeau and Vannimenus, 1978]. In statistical physics one has to calculate the partition function, defined as

$$Z = \sum_{\{\vec{s}\}} e^{-\beta E(\{\vec{s}\})} \tag{2.5}$$

If the interactions are short ranged, at least in some particular direction, i, so that $E = \sum_i E(i, i+1)$ the calculation of Z can be reduced to the calculation of

the largest eigenvalue of the transfer matrix $\mathbf{T}_{\vec{s}(i),\vec{s}(i+1)} = e^{E(i,i+1)}$:

$$Z = \prod_i \sum_{\{\vec{s}_i\}} e^{-\beta E(i,i+1)} = \mathrm{Tr}\mathbf{T}^N \simeq t_0^N \tag{2.6}$$

where t_0 is the largest eigenvalue of $\mathbf{T}$. The transfer matrix is taking into account the interaction between layer i and layer $i+1$, $E(i,i+1)$ and the product in (2.6) can be interpreted as adding interactions layer by layer. The $\beta = \frac{1}{k_B T} \to \infty$ ($T \to 0$) corresponds to the following algorithm: consider the first layer and calculate its ground state. The second layer is added by taking into account only the leading elements (Boltzmann-factors) of the transfer matrix in the limit $\beta \to \infty$ and so on. This allow for the construction of the ground state energy in a recursive way. To make the matter more clear, let us take again the simple example of Eq. (2.1): $E(i,i+1) = Ks_is_{i+1} + \frac{1}{2}H(s_i + s_{i+1})$. At finite β the transfer matrix is

$$\mathbf{T} = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix} \tag{2.7}$$

where $K = \beta J$, $h = \beta H$. If $J > H > 0$ $\mathbf{T}$ is proportional to the unity matrix. If the first spin is up, all successive spins are up. If $-J > H > 0$ $\mathbf{T}$ is the skew-diagonal unit matrix (the σ^x Pauli-matrix), so starting with the first spin up the next one is down, up, down, etc., creating the antiferromagnetic ground state. The transfer matrix method has been applied recently to quasiperiodic one dimensional models [Luck, 1987; Godrèche, Luck and Orland, 1986] which exhibit an infinity of ground states and also to continuous models with complicated phase diagrams [Chou and Griffiths, 1986; Marchand, Hood and Caillé, 1987].

At this point it seems worth mentioning that it is quite possible to use the form of the energy (2.1) in order to make some "mean field" type approximations, for example, by assuming that the mean field "energy" depends only on one-particle probabilities, like p_+. A systematic improvement of this approach is possible, with some truncation scheme for the long range or multispin interactions.

2.2. Physical models with many phases

In this paragraph we would like to mention a few spin models with many phases and to shortly explain the mechanism for creating such complicated structures. One class of models involves long range convex antiferromagnetic (repulsive) interactions and short range ferromagnetic (attractive) interactions (see in this volume [Levy and Mercier, 1988]). A general proof that such models have an infinity of commensurate phases has been given by Aubry [Aubry, 1978], a simplified treatment was given by Bak and Bruinsma [Bak and Bruinsma, 1982]. Consider the following energy functional

$$E = J \sum_{i,j \neq i} n_i A(|i-j|) n_j - H \sum_i n_i, \qquad J > 0 \tag{2.8}$$

where the variables n_i are binary variables $n_i = 0, 1$ and the repulsive interaction $A(x)$ is convex: $A(x+1) + A(x-1) > 2A(x)$. An example is a power like function, $A(x) = x^{-\alpha}$. If the field is small, all variables will assume the value 0 (empty lattice). As the field is increased, some periodic structures emerge: the simplest of them have unit cells with $n-1$ empty sites followed by one occupied site. Denote such structures by $< n >$. More complicated structures occur through a process of "dimerization", for example, between the phase $< n >$ and $< n+1 >$ a new structure, $< n, n+1 >$, is formed by the "concatanation" (dimerization) of the two elementary cells. This argument can be made more precise and the process of dimerization can be shown to continue *ad infinitum*. The change of the total coverage on such a phase diagram shows an infinite number of discrete steps and is pictorially called a "devil's staircase". Typical examples of models with

such interesting behaviour are Frenkel - Kontorova models and other models with continuous variables and competing interactions [Aubry, 1978; Burkov, 1985].

A second class of models is formed by spin models with axially competing interactions. Such are the Axial Next-Nearest Neighbour Ising model [for a review see P. Bak, 1982 and W. Selke, 1988] chiral clock models [Ostlund, 1981; Huse, 1981] and continuous magneto-elastic models [Marchand, Hood and Caillé, 1987]. In these models the interactions are *short* range but not convex. The mechanism through which a plenitude of phases occurs is now due to the competition between the energy and entropy. Heating up the system from $T = 0$ introduces thermal noise which weights differently the equienergetical configurations allowed near a multicritical point. This entropy difference leads to the stabilization of different commensurate phases at low temperatures. As recently pointed out [Shaw, Silverman and Pearson, 1985], such a mechanism might well apply to neuronal networks. The idea that thermal (or other) noise might induce order (stabilize commensurate phases) is very attractive from a biological point of view.

Finally, we mention a third large class of models with random, competing interactions. In these models (called spin-glass models) repulsive and attractive interactions are randomly distributed according to some given distribution. These models have many ground states (in generic spin glasses the number of ground states is an exponential of the numbers of spins) and even more metastable states. This type of models has been widely used recently [Hopfield, 1982] for CAM purposes.

To see how such a rich number of phases arise in disordered systems consider again the one dimensional Ising model of Eq.(2.1) but now suppose that the exchange coupling J is randomly distributed from $P(J) = p\delta(J-J_0)+(1-p)\delta(J+J_0)$. If the external field is zero we can repeat the minimization procedure using two types of transfer matrices (for J_0 and $-J_0$). One obtains a succession of clusters consisting of up spins and down spins, delimited by some "walls". When increasing the external field long clusters of down spins might turn up. The loss in exchange energy is proportional to the surface of the cluster: 2 ends = change of $4J_0$, while the gain is $2nH$, n is the length of the cluster. Thus, whenever $4J_0 = 2nH$, n integer, a macroscopic number of clusters turns around (phase transition). The model has again an infinity of phases characterized by the longest cluster of down pointing spins. In higher dimensions the mechanism is somewhat different but the fact remains that the introduction of competitive randomness leads to the creation of a hierarchical structure of clusters forming a heterogeneous phase called spin-glass. This mechanism is again attractive from a biological point of view, especially since the presence of "noise generators" is well established in the central nervous system [Changeux, 1982] The spin-glass problem is a very tough one and although the mean field theory is by now solved [Parisi, 1980a,b,c; Mézard *et al.*, 1984], the short range models are still not fully understood. One may try to go one step in between complete homogeneity and complete disorder. A quite popular class of models studied in physics are models with quasiperiodic energy functions. These models are better suited to analytic work than the fully disordered ones but still retain a lot of interesting properties.

In biology, computer science, etc. one is interested in models with many phases, which, however, are robust against thermal, chemical, etc. noise. In physics this is expressed as saying that the dimensionality of the phase must be above its lower critical dimension. The interesting behaviour occurring in the ANNNI model and in spin-glass models happens for dimensions at least equal to three. In many instances mean field techniques can be valuable tools, especially if one allows long range interactions.

3. Cellular Automata *alias* Systolic Arrays

Cellular automata consist of discrete dynamic variables (spins) defined on the lattice sites of discrete space-time manifolds. The time evolution of such systems is given in terms of dynamic rules which obey causality and involve - in cases of practical interest - a local neighbourhood of sites only. From the time of their invention by J. von Neumann and S. Ulam [Neumann, 1966; Codd, 1968; Burks, 1970] until now, cellular automata have provided a conceptually simple but flexible framework for modelling dynamic systems with many degrees of freedom in biology, mathematics, computer science, and physics.

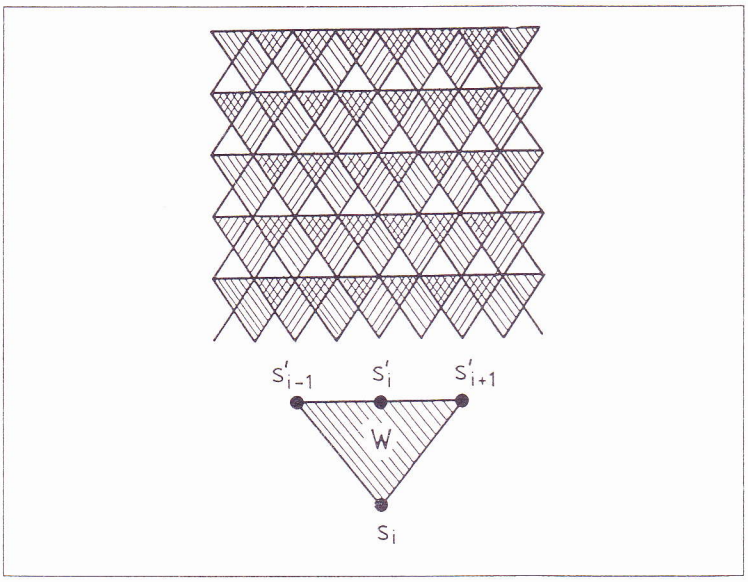

Fig. 2: Part of a square lattice in space and time. The range of the interaction corresponding to w_i is hashed

Consider for simplicity a square lattice (Fig. 2) where the $x-$direction is the space and the $y-$direction is the time. The configuration of the cellular automaton at time t is given by the state of the variables (spins) $s_i = 0, 1, 2, \ldots, q-1$ defined on the lattice sites of a row. Dynamic rules are imposed such that every spin ("child") $s_i(t+1)$ at time $t+1$ and site i is set *independently* of all others as a function involving only the previous time t values of the spins ("parents") s'_{i-1}, s'_i, s'_{i+1} at neighbouring sites $i-1$, i, $i+1$. These rules can be given in a concise way through the conditional probability w_i:

$$w_i(s_i|s'_{i-1}, s'_i, s'_{i+1}); \qquad \sum_{s_i=0}^{q-1} w(s_i|s'_{i-1}, s'_i, s'_{i+1}) = 1 \tag{3.1}$$

Given the distribution of the configurations at time t, $R(\{\vec{s'}\}, t)$, the distribution for time $t+1$ is given by

$$R(\{\vec{s}\}, t+1) = \mathbf{P}R(\{\vec{s}\}, t) = \sum_{\{\vec{s'}\}} (\prod_i w_i) R(\{\vec{s'}\}, t)) \tag{3.2}$$

The stochastic matrix ($\sum_i P_{i,j} = 1$) $\mathbf{P}$ is very similar to the transfer matrix: they are both non-negative, real matrices. The implications of this analogy have been

discussed elsewhere [Ruján, 1987]. An interesting point is that all cellular automata can be interpreted as statistical physical systems on a space-time lattice obeying constraints following from the normalization condition (3.1). Therefore, many "static" properties of statistical mechanical models can be translated in the language of cellular automata.

The theory of automata is considering mostly *deterministic* rules, meaning that w_i might assume only the 0,1 values. Therefore, every given configuration has an uniquely defined successor. However, some configurations may lead to the same successor, so the trajectories will merge on some special class of configurations, called *attractors* or *valleys*. A finite automaton must enter sooner or later a periodic cycle.

Many cellular automata algorithms are known for synchronization, pattern recognition and similar tasks [Vollmar, 1979]. Since Turing machines can be mapped into cellular automata, some cellular automata are able of universal computation [Smith III, 1971; Banks, 1971]. Von Neumann proved that among such cellular automata one can find in addition some which also "reproduce" themselves (universal constructors). Automata can easily recognize one-dimensional strings, draw pigmentation patterns, simulate cardiac dynamics, etc. [Wolfram, 1986; Physica D**22**, 1984]

New results have been obtained recently for the Kauffman model of cell differentiation [Kauffman, 1969] - or random boolean networks. In this model one has N spins, which are updated simultaneously according to

$$s_i(t+1) = F_k(s_{i_1}, s_{i_2}, \ldots, s_{i_k})_t \tag{3.3}$$

where every spin s_i has exactly k inputs from spins $(s_{i_1}, \ldots, s_{i_k})$ and F_k is a Boolean function ($F = \pm 1$). Initially the input spins of s_i are chosen at random and a particular function F_k is also chosen at random from all 2^{2^k} possibilities and thereafter are quenched. This model has two phases, a "frozen" phase, where the distance between two initial configurations converges ($k \leq 2$) and a"chaotic" phase, ($k > 2$), where this distance diverges in the $t \to \infty$ limit. This transition is also reflected in the average cycle length, which change from a power in N to an exponential in N. An interesting point made by Derrida and collaborators [Derrida, 1987] is that if the probability that a given site appears only once in a typical tree of ancestors is one, then the model can be treated as if the input sites and the corresponding rule would be independently chosen at each time step (annealed approximation). This idea was later applied to asymmetric, diluted neuronal networks [Derrida, Gardner and Zippelius, 1987].

The problem with - and the advantage of - random automata is that their behaviour can be analysed only in average: one does not have much control over a particular, small network. In some instances one might wish to have a handle on some properties of the system by changing the value of some "quantum numbers" left invariant by the dynamics. In the following example we discuss such an approach, a cellular automata which is able to perform some interesting tasks by design.

Example: A cellular automaton which knows topology

Consider a square lattice. The center of an elementary square is black if it is occupied and white when it is empty. Two sites are connected if they have a common edge, a cluster is formed by one, two, or more connected sites. The clusters may contain loops, the loops other loops, etc., as shown in Fig. 3.

In Fig. 4 we show how one may constructs the dynamic rules for an automaton recognizing the topological aspects of some initial configuration: one takes a 3×3 window: the center spin can be changed only if the connectivity of the visible 3x3 lattice is not changed. One has in total 2^8 possible probabilities w_i

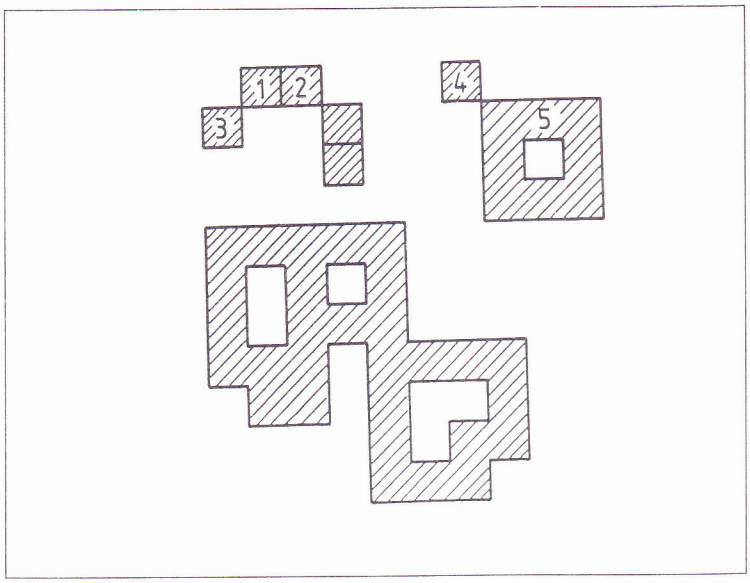

Fig. 3: A few examples of clusters and loops on the square lattice: 1 and 2 are connected, 1 and 3 not; 5 is a loop not connected to 4

- some of them are deterministic, some are at our disposal. Some examples are shown in Fig. 4.

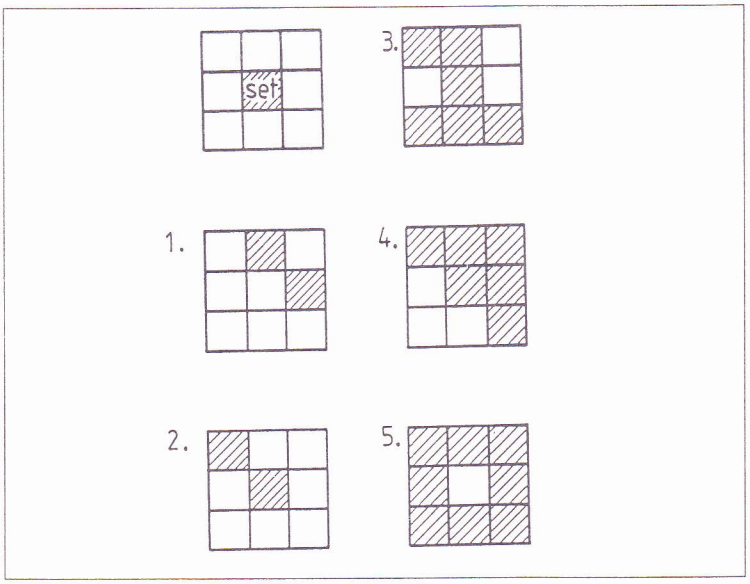

Fig. 4: The elementary window of the topological CA: 1. The center must remain 0. 3. The center must remain 1. 4. The center spin can be reset with some probability. 2. A cluster is identified. 5. A loop is identified

Call the rule "take away a variable whenever it is possible" the "take" rule, accordingly, one has an "add" rule. In Fig. 5a we show a configuration which

was generated at random: each site had the probability 0.55 to be set. One can see some clusters, loops, etc. - but it would be quite difficult for you to count their number (could it be that our brain dislikes randomness ?!).

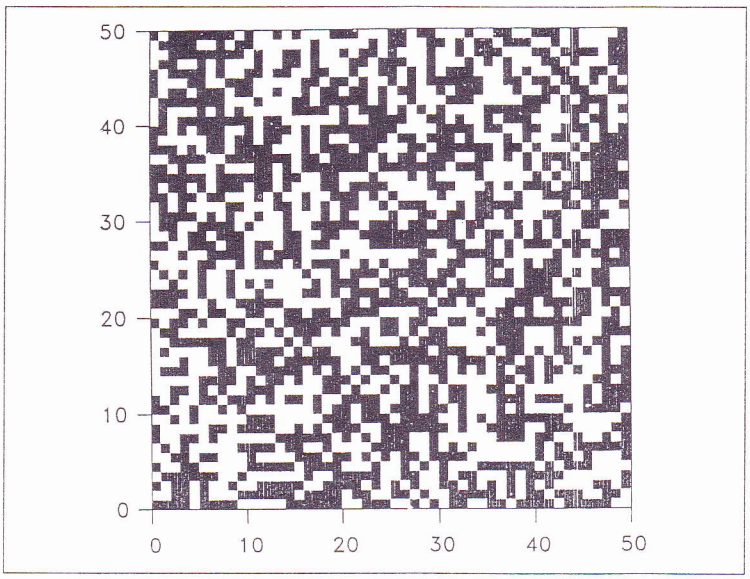

Fig. 5a: The initial random configuration p=0.55

Next apply the "take" rule: the loopless clusters are reduced to one site, the more complicated ones are also reduced, but do not fit within the 3×3 window. Nevertheless, we are now much better at recognizing what is connected to what - this simple filtering might be even thought as an abstraction process. The sequential algorithm we used needs 8 steps to get the Fig. 5b. (A nine-sublattice parallel algorithm also does it.)

Next, all loopless clusters are counted - there are 108 of them - and then eliminated. Note that by overlapping Fig. 5a and 5b one can also determine which were the original loopless clusters. The simplest object which can be now detected are most inner loops: Inverting from the "take" rule to the " add" rule, every possible site is filled and the centers of the most inner loops become a single white square (our picture Fig. 5c is a negative one).

Again, the loops are counted (there are 19 loops), filled up, the rules are inverted again, etc., until nothing remains. Not exactly: if our original square had periodic boundary conditions, then when the initial probability $p_0 > p_c \simeq 0.592$, suddenly a new kind of loop appears. This loop wraps around the torus geometry implied by the periodic boundary conditions and cannot be eliminated by the previous rules. It corresponds to the appearance of the so-called infinite cluster, spreading across the sample. This phase transition is called percolation transition and is a continuous one [Stauffer, 1985].

Obviously our simple algorithm is able to distinguish between different topologies. If writing would have been invented well after the letters had to be curved in wood or stone, we might have had the chance to define the letters to have different topologies: our machine would have been able to read any "fuzzy" handwriting except when connectivity rules are not obeyed!

Other interesting features appear when playing around with some optimization problems. Suppose some cities are placed at random in this square (one might

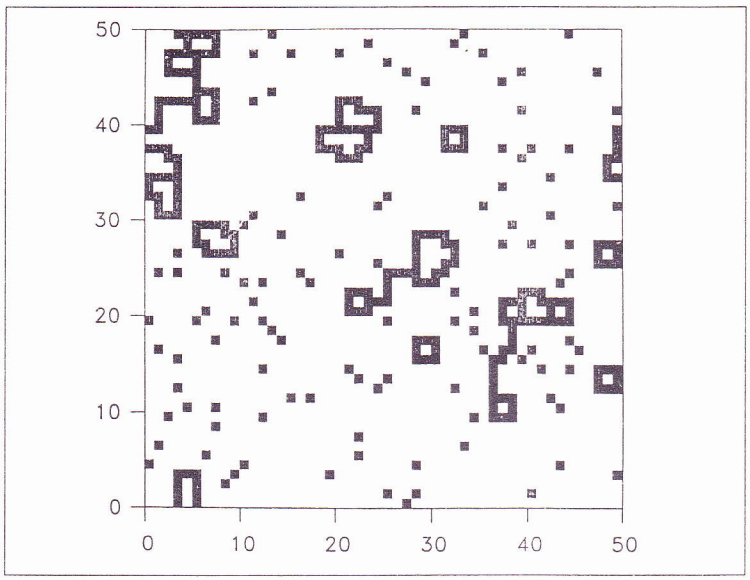

Fig. 5b: The configuration obtained after applying the "take" rule

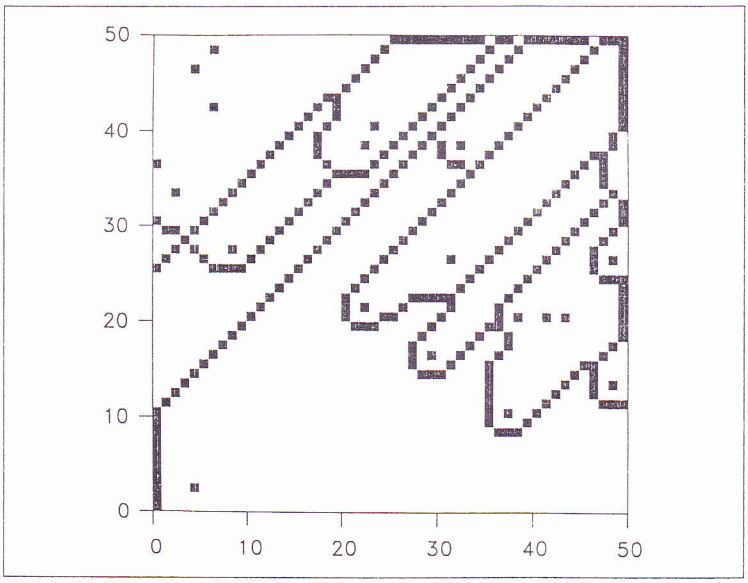

Fig. 5c: Negative of configuration after "add" rule. The little black squares are centers of most inner loops

call them also cells). We want to connect all the cities with a very short path. We take an initial configuration with all squares occupied and take out the center square - in this way we create a cluster with a loop. Now the "take" rule is applied gradually: the probability to take a site - except for the cities - is slowly increased. Some steps are shown in Fig. 6a-c. The final cluster is shown in Fig. 6d - the route is not really very short, but it seems acceptable.

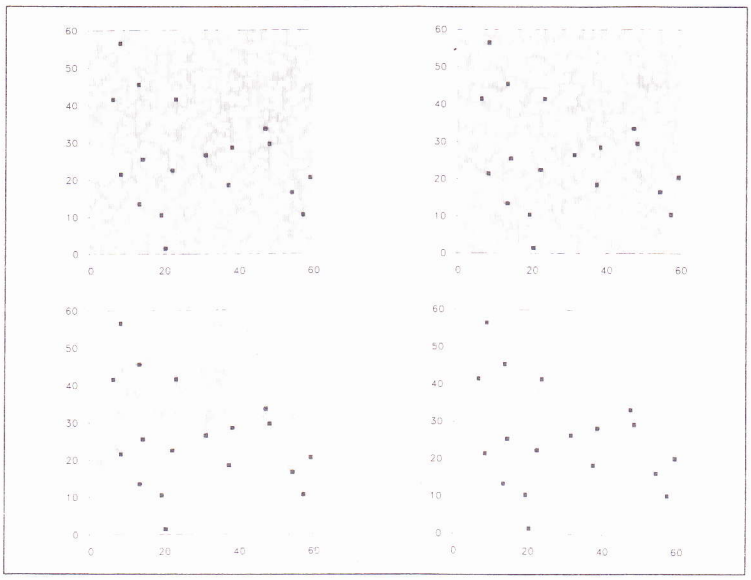

Fig. 6a-d: Slow cooling of a configuration with fixed "cities"

The main problem is related to the fact that the cluster has also some "legs", not only a "belly". Changes in the 512 rules might improve further the situation.

Finally, I would like to speculate on some possible applications of such algorithm to parallel computing. Suppose that one constructs a system which uses the two-dimensional lattice (or "retina") only as support - there are other units connected to each elementary square (for example registers, gates, etc.). To be concrete, suppose one would like to make in parallel the operation $(a+b)\times(c+d)$. Every ordered data set has a given hierarchy, usually represented as a tree graph, or equivalently, by Venn diagrams [Krishnamurthy, 1985]. In our case a and b are at the same level with c and d. One possibility is to use as many "retina" as different levels of hierarchy. Thus one starts at the lowest level with a two-dimensional topological CA and one inserts in a black see two elementary loops, whose registers contain a and b, the others' c and d. The programming is complete by setting the "operation modules" to addition on the first level and multiplication on the second (one can also imagine a structure where every level CA is able to perform at each site only a given operation). Now the "add" rule is initiated on the first (basic) layer: whenever a white site is set, its register content is added to one of its white site neighbours. When the computation reaches a fixed point, the remaining one-site loops are merged together into the second layer, where the multiplication is performed in the same way. At a given level one stores only simple loops, separating units of higher level than the ones present within the loops. The loops can eventually be inflated randomly, so that they would have very thick walls. In this state the memory would be quite insensitive to local noise, when bits flip independently with some probability p. Some kind of "dynamic maintenance" is also possible by using majority rules. However, if $p > p_c$, the critical percolation probability, the topological structure is destroyed. Much more thinking remains to be done if such schemes are to work in practice. Nevertheless, they have the right flavor: a parallel system with highly standardized parts but performing quite different tasks.

4. Neuronal networks as content addressable memory

In this Section we briefly describe different models for CAM, which use single layer units for retrieving key patterns. In all models we shall use N spin variables $s_i = \pm 1$. A model has three components: *structure,* (the kind of interactions allowed between spin-variables), *dynamics* , and *the learning rules* (how to choose the interactions in order to recall a given group of key patterns). The study of a given model is concerned with questions regarding the storage capacity (how many different key patterns can be stored in such a network), the information capacity , the effect of noise on the storage capacity (robustness), the size of the basins of attraction of different key patterns, the number of iterations needed to obtain a typical retrieval, etc. Different variants able to store time-ordered sequences [Kohonen, 1987; Peretto and Niez, 1986a; Sompolinsky and Kanter, 1986] patterns invariant under symmetries transformations [Lee *et al.*, 1986; Bienenstock and von der Marlsburg, 1987; Kree and Zippelius, 1987; Chen *et al.*, 1986] are also under active investigation. In what follows we discuss some basic models and mention some of their variants, as well as their main properties.

Denote by $\{\vec{\xi}^{\nu}\}_{\nu=1}^{p}$ the key patterns, where $\vec{\xi}^{\nu} = (\xi_1^{\nu}, \ldots, \xi_N^{\nu})$. Usually neuronal networks are modelled with continuous variables. In the high gain limit, however, one can go to occupation variables $n_i = 0, 1$ or to spin variables $s_i = \pm 1$. In most cases the spins are supposed to interact via two-spin interactions $J_{i,j}$. If the $J_{i,j} = J_{j,i}$ the model has *symmetric* interactions, otherwise the interactions are asymmetric. If key patterns are to be retrieved even if they are translated, rotated, etc., one has to introduce more complicated interactions, like three, four spin interactions, etc. [Lee *et al.* 1986], symmetric interactions one may define an *energy functional* of the form:

$$E = -\sum_{i,j} J_{i,j}^{(2)} s_i s_j - \sum_{i,j,k} J_{i,j,k}^{(3)} s_i s_j s_k + \ldots \tag{4.1}$$

where $J^{(n)}$ are n-spin interactions. Although not necessarily biologically sound, the existence of an energy has the advantage that standard statistical mechanical methods can now be used. The main idea is to associate key patterns with the ground states of E: construct E such, that its ground states are the desired patterns (and possibly no other ground state exists). We have seen in Section II. that the ground states are in principle determined through the linear programing method. The inverse procedure (called "(free)energy landscape gardening") is by no means trivial. Storing a single pattern is simple: $J_{i,j} = \xi_i \xi_j$, which is just another way of writing a ferromagnetic interaction by using the gauge symmetry of the energy. Hebb [Hebb, 1949] proposed a generalization of this rule:

$$J_{i,j} = \frac{J_0}{N} \sum_{\nu=1}^{p} \xi_i^{\nu} \xi_j^{\nu} \tag{4.2}$$

or in bra-ket notation

$$\mathbf{J} = \frac{J_0}{N} \sum_{\nu=1}^{p} |\vec{\xi}^{\nu} >< \vec{\xi}^{\nu}| \tag{4.3}$$

For orthogonal patterns obeying

$$< \vec{\xi}^{\nu} | \vec{\xi}^{\mu} >= \frac{1}{N} \sum_{i=1}^{N} \xi_i^{\nu} \xi_i^{\mu} = \delta_{\nu,\mu} \tag{4.4}$$

the Hebb rule ensures that every pattern $\vec{\xi}^{\nu}$ satisfies the condition

$$E(\vec{\xi}^{\nu}) = -N J_0, \quad \forall \nu \tag{4.5}$$

and possibly they are also ground states of the system. Hopfield [Hopfield, 1982] made the observation that the Hebb rule is useful also for random patterns: in this case the overlap between two patterns is not strictly zero, but proportional to $N^{-\frac{1}{2}}$. Thus the sum over all patterns different from $\vec{\xi}^{\nu}$ gives a correction of $(\frac{p-1}{N})^{\frac{1}{2}}$ for large enough p and N. This simple argument gives the correct scaling for the maximal number of stored random patterns as ($p_{max} = \alpha N$) if one allows for small distortions in retrieval. The number of *perfectly* storable patterns is $N/2\ln N$ for large N [Weisbuch and Fogelman-Soulié, 1985]. When applied to multispin interactions, it predicts for $k-$spin interactions the relation $p_{max} \sim \frac{N^{k-1}}{(k-1)!}$ [Peretto and Niez, 1986b; Gardner, 1987]. At low temperatures the patterns are well recognized if $\alpha < \alpha_c$ for relatively small disturbances. If $\alpha < \alpha_1$ the patterns are also ground states of the energy (4.1) ($\alpha_1 < \alpha_c$).

The Hopfield model has been generalized from different points of view, a general analysis for learning rules of the type:

$$J_{i,j} = \frac{\sqrt{p}}{N} A_{i,j} F(\sum_{\nu=1}^{p} \xi_i^{\nu} \xi_j^{\nu} / \sqrt{p}) + \zeta_{i,j} \tag{4.6}$$

where $A_{i,j}$ are random variables with unit average value, $\zeta_{i,j}$ is local white noise has been presented by Sompolinsky [Sompolinsky, 1987]. The formulation in terms of an energy functional has some obvious advantages:

- due to the additive character of the learning rules patterns can be stored one by one,
- the storage is robust against local noise and thermal fluctuations.

From Eq. (4.3) it is obvious that the Hebb-rules define the **J** interaction matrix as a projection matrix. In more general cases, when the key patterns are linear independent but not orthogonal, the **J** matrix can still be decomposed in terms of a biorthogonal basis $< \vec{\mu}^{\alpha} | \vec{\xi}^{\beta} >= \delta_{\alpha,\beta}$ as

$$\mathbf{J} = \sum_{\nu=1}^{p} \frac{|\vec{\xi}^{\nu} >< \vec{\mu}^{\nu}|}{< \vec{\mu}^{\nu} | \vec{\xi}^{\nu} >} \tag{4.7}$$

The vectors $\{\vec{\xi}^{\nu}\}$ are the key patterns. The calculation of the vectors $\{\vec{\mu}^{\nu}\}$ is done via the pseudo-inverse method [Kohonen, 1987; Personnaz, Guyon and Dreyfus, 1985; Diederich and Opper, 1987] or, in a slightly different form, by

$$\vec{\mu} = \mathbf{C}^{-1} \vec{\xi}, \qquad C_{\mu,\nu} =< \vec{\xi}^{\mu} | \vec{\xi}^{\nu} > \tag{4.8}$$

corresponding to the energy functional [Kanter and Sompolinsky, 1987]

$$E(\vec{s}) = -\frac{1}{2} < \vec{s}^{\perp} | \vec{s}^{\perp} > -\frac{N}{2}, \qquad \vec{s} = \sum_{\nu=1}^{p} a_{\nu} \vec{\xi}^{\nu} + \vec{s}^{\perp} \tag{4.9}$$

The interpretation of these equations is simple: every vector $\vec{s}$ has a component lying in the space spanned by the key patterns and has a component perpendicular to it. The energy (4.8) corresponds to the distance between $\vec{s}$ and the key pattern subspace : it is zero only if $\vec{s}$ is a key pattern (the probability of having a linear combination of key patterns also obeying the condition that its elements are ± 1 is negligible). Local learning rules are also known for this case [Denker, 1986].

Dynamics

In models where the energy functional is provided (symmetric interactions) the dynamics is defined as usual: one spin flip sequential Glauber dynamics obeying the detailed balance. Given enough time the iteration process will lead to the

equilibrium Boltzmann distribution, provided the low lying eigenvalues of the Liouville-operator are not exponentially degenerate. This later case occurs in random, frustrated models, for example in spin-glass models, where the ground state is not reached even by very long runs [Binder and Young, 1986]. Thus it might be of some interest to start directly with dynamic rules. Little [Little, 1974; Little and Shaw, 1978] has proposed cellular automata - type parallel dynamics, without an associated energy. The dynamics follows the general rules of the CA-dynamics discussed in Section III: all spins are set independently at each time step according to the transition probability $w_i(s_i|\vec{s'})$. At T=0 one can write

$$s_i(t+1) = \mathrm{sgn}(h_i - \theta_i), \qquad h_i = \sum_j J_{i,j} s_j(t) \tag{4.10}$$

where θ_i is some threshold (external field), and h_i is the "internal field" acting on s_i. At finite temperatures one has

$$s_i(t+1) = w(s_i|\vec{s'}(t)) = \frac{e^{s_i[h_i(\vec{s'}(t))-\theta_i]}}{2\cosh(h_i-\theta_i)} = e^{s_i(h_i-\theta_i)-f(\vec{s'})} \tag{4.11}$$

In a previous work [Ruján, 1987] we have shown that for a one time step parallel dynamics with only one (threshold) and symmetric two spin interactions between different time layers it is possible to define an energy functional:

$$E(\vec{s}) = \sum_i \ln\cosh(h_i(\vec{s}) - \theta_i) \tag{4.12}$$

Note that this expression for the energy involves all possible $2^N - 1$ interactions between the spins, from which only $\frac{N^2}{2}$ are independent. From the point of view of dynamics one can define stored patterns as being fixed points of the dynamics with a non-vanishing basin of attraction. For non-symmetric two-spin interactions obeying the spherical rule $\sum_i J_{i,j}^2 = N$. Gardner [Gardner, 1987] and Derrida and Gardner [Derrida and Gardner, 1987] have calculated recently the number of fixed points with a non-zero basin of attraction in the space of independent, asymmetric couplings $\{J_{i,j}\}$'s. The maximal number of stored patterns depends on the distribution of the patterns: for uncorrelated random pattern is at most $p = 2N$, giving the maximal information capacity as $S = \sum_i p_i \ln p_i = 2N^2 \ln 2$.

Other approaches, which take more closely into account the asynchron features of neuronal excitation mechanisms have been proposed [Carpenter and Grossberg, 1987; Horner, 1987]. The main observation is that the larger the internal field trying to flip a spin, the faster the spin flips. A practical dynamic rule taking this feature into account is to flip always the spin with largest negative internal field

5. Layered and automata-like structures

A different approach is advocated by some biologists [Hoffmann, 1986] - especially in connection to memory effects in the immune system - and computer scientists trying to generalize the idea embodied in Perceptrons [Minsky and Papert, 1969; Ackley, Hinton and Sejnowsky, 1985; Rumelhart and McClelland, 1986]. This architecture has layered structures with adaptive elements. The main conceptual difference is that now a stored pattern does not have to be recalled "bit-by-bit" : it is enough if a characteristic output is activated. These outputs are regarded as attractors* of the system. Learning means here that by changing some internal variables different inputs are canalized into the same desired outputs. If the system is large enough and is taught long enough, it is hoped that the basin of attraction of the outputs will somehow reflect common features of the learned inputs and will recognize them in any further input presented to the machine.

A layered version of the Hopfield model has been introduced and solved by Domany *et al.* [Domany, Meir and Kinzel, 1986; Domany and Meir, 1987] quite different approach has been proposed by Huberman and Hogg [Huberman and Hogg, 1984, Hogg and Huberman, 1985], Hinton *et al.* [Hinton, Sejnowski and Ackley, 1984] , Shaw *et al.* [Shaw, Silverman and Pearson, 1985], etc. For example, the architecture considered by Huberman and Hogg consists of automata with an internal degree of freedom, which works as a lateral inhibitor acting on the slope of neuronal response: see Fig. 7. The inputs are added and passed through a simple threshold unit, whose slope depends on the internal variables $M_{i,j}$ displaying lateral repulsion (inhibition): if the nearest neighbour outputs on the same row have larger (smaller) output values the value of M is decreased (increased). The system produces a large numbers of attractors and is used in speech recognition [Huberman and Hogg, 1985] and for the classification of outputs of a motion detector in presence of noise [Keirstead and Huberman, 1986].

The feed-forward layered structures consist of a layer of input units, a middle layer of so-called "hidden units", and a layer of output units. During training the connections between hidden units and the input-output layers is tuned in such a way, that when the input spins are set to some key pattern values they provide a desired output. A "back-propagation" algorithm is designed to change iteratively the connections as a function of the distance between the desired and the obtained outputs. Layered models are generalization of Perceptrons and have been shown to be able of computation, of recognizing symmetries, etc. [Sejnowski, Kienker and Hinton, 1986]. In Fig. 8 we present a simple layered machine operating in "grandmother neuron" mode. The number of hidden units must be at least equal to the number of key patterns. The first hidden unit couples to the i-th input unit with the strength ξ_i^1, the second one with strengths ξ_i^2, etc. Every hidden unit feels an internal field equal to the overlap between the configuration presented on the input layer and the key pattern it represents. If now one assumes lateral inhibition between the hidden units, corresponding to repulsive interactions - only the unit with the largest overlap will fire. With properly defined links between the hidden units and the output units an arbitrary output can be associated to any stored key pattern.

This type of "overlap" machine is very good at recognizing noisy patterns, but is rather sensitive to damages of the hidden units.

---

* The term attractor refers to an output obtained after running the system for a very long time. We use this notation loosely here, since the systems in question run only a few time steps.

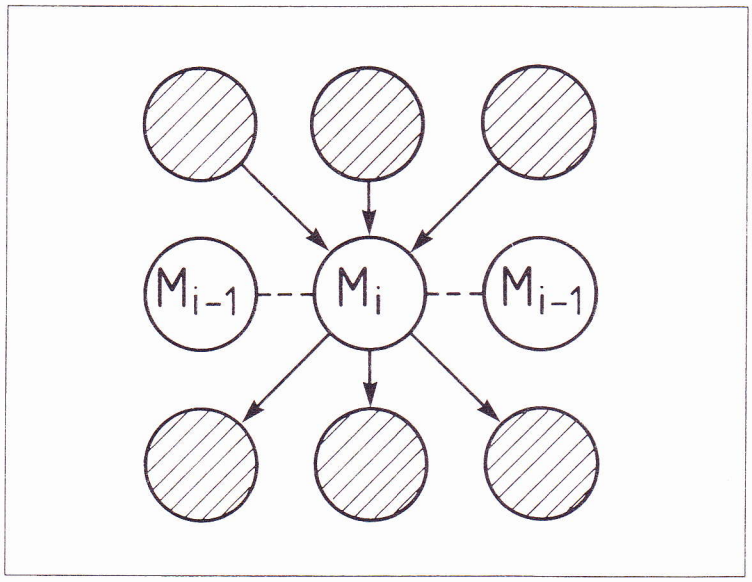

Fig. 7: The "Prism" architecture of Huberman and Hogg: inputs are taken from nearest neighbours: outputs are generated using the internal variables M as the slope of a sigmoid-type function of the sum of inputs. The M variables interact via lateral inhibition (broken lines)

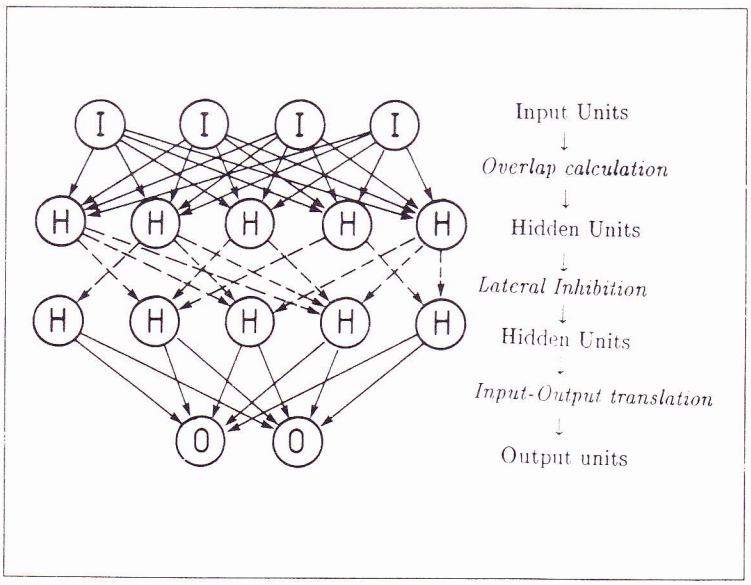

Fig. 8: A Perceptron-like feed-forward structure with laterally inhibiting hidden units. The top layer receives the input, the first layer of hidden units calculates the overlap with their own "key" patterns. The next layer of hidden units stores the overlaps surviving the inhibitory interactions. Finally, the outputs related to a given hidden unit are activated

The trion model of Shaw *et al.* [Shaw, 1988] represents a slightly different ap-

proach: it considers homogeneous cellular automata rules for an $s_i = 0, \pm 1$ three state model with two-time step rules. Because of competing interactions in the time direction, a large number of cycles and different temporal - spatial patterns are expected. By changing the parameters of the system the basins of attraction of different attractors are increased (decreased). When the basin of attraction of a particular pattern becomes very large, the pattern is said to be recognized. This approach to memory differs from the "free energy landscape gardening" one in many respects:

- The system is defined dynamically, in general no energy functional exists,
- Instead of having independent random connections, the system has originally homogeneous competing interactions leading to a large number of attractors.
- Learning is understood as a time dependent process during which the basin of attraction of a particular pattern is enlarged by changing the strength of particular couplings. It seems reasonable to think that the basic structure of our brain is genetically inherited: the space of what one could and what one could not learn is set from birth [Gould, 1982; Marler and Terrace, 1984] - learning and recognizing correspond to motion in this large "phase diagram" by changing a *few* parameters of the network. The model of the immune system of Hoffmann [Hoffmann, 1986] brings this philosophy to an extreme: here not even the couplings are allowed to change, the only control over the dynamical system is done through the initial conditions.

Let me finish this short list by mentioning another unorthodox approach. Bergman and Kerszberg [Kerszberg and Bergman, 1987] assume that learning can be dead serious: they suggest that an evolutionary process might be more effective than a strenuous training period: anyone involved in teaching would agree with this point of view. They start with populations of machines whose structure is given (the machines are automata with variable inhibitory - excitatory links between successive layers, all links being the same on a layer but changing from layer to layer) and submit them to tests like recognizing translational invariance, scale invariance, etc. Applying a computer assisted natural selection program, they were able to find (many) machines performing the given tasks - but were not able to understand how the machines are actually doing it - a paradigm for artificial intelligence going too far. The idea of working with interacting populations is a main ingredient of Jerne's network theory of the immune system [Jerne, 1973] and seems to play an important also in other approaches, like the classifier system [Holland, 1986] and for solutions of difficult optimization problems [Brady, 1985; Ruján, 1988]. In the next subsection we try to unify these idea in a compact mathematical description.

A dynamic approach to memory models

Consider a *sequential* time developing operator such that any initial configuration will fall into the key pattern with which it has the largest overlap (CAM). This operator has two components, a linear one and an extremely simple non-linear one. It is useful to think of the linear operator acting in the 2^N dimensional space of all possible configurations as consisting of a non-diagonal kinetic term and a diagonal potential term.

$$\hat{H} = \hat{K} + \hat{V} \tag{5.1}$$

Note that in the $|s_1 > |s_2 > ...|s_N >$ direct product basis all patterns are orthogonal to each other $< \vec{\xi}^{\nu} | \vec{\xi}^{\mu} >= \sum \prod < \xi_i^{\nu} | \xi_i^{\mu} >= \delta_{\mu,\nu}$. Unlike in the usual models, the key patterns are stored here as a *single* linear combination of slightly smeared

patterns:

$$\Psi_0 = \sum_{\nu=1}^{p} a_\nu e^{J_0 \sum_{i,k} \xi_i^\nu j_{i,k}^\nu \xi_k^\nu} \tag{5.2}$$

where J_0 is a large constant and $j_{i,k}^\nu$ are some couplings (± 1) fixing the ν-th pattern (for example *nearest neighbour couplings* $\xi_i^\nu \xi_k^\nu$).

The next step is to define the dynamics (the kinetic term): suppose one has single spin flips, so the kinetic term is

$$\hat{K} = \sum_i \sigma_i^x \qquad \text{where} \quad \sigma^x f(s) = f(-s) \tag{5.3}$$

Learning rules are required: the question is how to choose $\hat{V}(\vec{s})$ such that Ψ_0 is the ground state. A simple calculation gives the answer:

$$V(\vec{s}) = E_0 - \frac{\sum_\nu a_\nu e^{J_0(\sum_{i,k} s_i^\nu j_{i,k}^\nu s_k^\nu - 2 s_i \sum_k j_{i,k}^\nu s_k)}}{\sum_\nu a_\nu e^{J_0 \sum_{i,k} s_i^\nu j_{i,k}^\nu s_k^\nu}} \tag{5.4}$$

The Perron-Frobenius theorem ensures that for a finite automaton the ground state Ψ_0 is not degenerate - so all other energy levels must be less than E_0. Starting from an initial configuration a simple iteration should fall into the component of the ground state with which it has a maximal overlap, assuming that the overlap between the key patterns is not very large. It would be desirable to push down the energies of the excited states since a strongly degenerate excited state might slow down considerably the iteration process.

This general approach is not very practical: the function $V(\vec{s})$ is not local and in principle involves all possible 2^N interactions between the spins (interesting enough, the more symmetries are imbedded in the patterns, the more this number is reduced). In addition, the computational load is very high. From the discussion given above we expect that in order to store N (random) patterns one needs roughly N^2 free parameters. Instead of the exact solution for $\hat{V}$ one may start with a potential of form

$$\hat{V}^{trial}(\vec{s}) = \sum_i w_i \tag{5.5}$$

where w_i are some functions of the spin i and a few of its neighbours ($\sim 2 \ln_2 N$ of them). The w_i functions contain all possible interactions between these spins as free parameters to be obtained by minimizing the expression

$$\min_w \frac{< \Psi_0 | \hat{K} + \hat{V}^{trial} | \Psi_0 >}{< \Psi_0 | \hat{\Psi}_0 >} \tag{5.6}$$

and possibly also higher moments of $< \hat{H}^n >$, $n > 1$. For engineering applications one may constrain the functions w_i to boolean functions - then a larger number of neighbours is needed in w_i. The minimization procedure becomes then extremely difficult: checking whether a desired output has a non-empty basin of attraction corresponds to the famous satisfiability-problem, which is NP-complete [Garey and Johnson, 1979].

Suppose that one way or another the learning procedure has been completed. The system is presented with some input, which is processed by applying first the linear operator $\hat{H}$. As a result one obtains a linear combination of $N+1$ different patterns (the operator is very sparse). In usual Monte Carlo procedures one picks at random one of these configurations, and the procedure is repeated many times. We propose that instead of following the life history of a randomly chosen configuration, one keeps a typical population of patterns.

Then a non-linear transformation $\hat{S}$ can be performed by simply squaring the multiplicities of the different configurations after each iteration. Together with a proper normalization, this algorithm assures that the key pattern whose overlap with the initial configuration was the largest one will dominate after a few steps the population:

$$(SH)^n \psi^0 = (SH)^n \sum_n c_n \Psi_n = \sum_n E_n^{2^{n+1}-2} c_n^{2n} \Psi_n \tag{5.7}$$

Therefore the eventually large overlap with some excited state Ψ_n, $n > 0$ is suppressed by the difference in energy, while within Ψ_0 the difference in the initial overlap force the iteration into the right key pattern. A new - and necessary - feature of this algorithm is handling in parallel a population of patterns rather than following the time history of a single pattern.

The same arguments can be used when the purpose of the network is to associate given classes of inputs to given outputs: *classification.* Mathematically, one has to solve a generalized eigenvalue equation:

$$\hat{H}\Psi_0 = E_0 \hat{T} \Psi_0 \tag{5.8}$$

where now Ψ_0 is the input state and $\hat{T}\Psi_0 = \Psi^O$ is the output state. One can call $\hat{T}$ an "input-output translation operator". The CAM dynamics is recovered if the inputs and outputs are identical - $\hat{T}$ is the unity operator. The variational problem changes then into

$$\min_w \frac{< \Psi_0 | \hat{T}^{-1} \hat{H}^{trial} | \Psi_0 >}{< \Psi_0 | \hat{\Psi}_0 >} \tag{5.9}$$

6. Conclusions

We conclude with a few specific and some general remarks. There are three main area of scientific interest related to the study of memory:

a. The understanding of how the particular system - brain, immune system, economic system, etc., - functions;

b. The construction of "intelligent" parallel machines;

c. Most models of memory are quite intricate and interesting in their own right: quantitative results are usually hard to obtain and thus very valuable.

People who have actively contributed to this field have quite different backgrounds. The models and the methods used for their study are thus necessarily different and reflect their specific interest. Sometimes this leads to misunderstandings: biologists would give quite valid arguments against what they might perceive as intolerable simplifications of the reality, physicists will undoubtedly prefer models as simple as possible, while people interested in constructing chips would like to stick with logical gates and automata-like structures. The author is not convinced that biologically relevant models must be also easier to solve or even that they have greater technological potential.

Spin-glass like models of associative memory have been the focus of an intense activity recently. A common feature of these models is that the stored patterns are random patterns. Many techniques borrowed from the spin-glass physics, like replica-tricks, etc., have been proved valuable in obtaining a lot of interesting analytical results. One possible objection is that a highly uncorrelated environment does not help the evolution of the brain: otherwise it would be very

difficult to understand why dolphins did not outperform humans in this respect. In addition, it is not clear whether this type of models will be very useful for practical purposes, although the highly diluted models [Derrida, Gardner and Zippelius, 1987] seem quite reasonable. All this said, the Hopfield-type models already provide a good framework for asking concrete questions about learning, forgetting [Toulouse, Dehaene and Changeux, 1986], the rôle of sleep [Hopfield, Feinstein and Palmer, 1983], etc.

Projection methods and other associative memory models based on linear algebra seem to be more apt at storing correlated patterns. When local learning rules are also defined, these simple methods are amazingly powerful. Nevertheless, they are also questionable from a biological point of view.

As expected, for practical applications in computing Perceptron-like architectures with hidden units seem to perform best. However, it would be fair to say that our brain is probably functioning according to other - largely unknown - principles.

It is striking to notice that in many physical systems with short range competing interactions one may find a very large number of ground states: moving across such a rich phase diagram requires only the change of a *single* continuous parameter. Compare this with the extreme plasticity required in a Hebb-type model: one has to assume that after learning N^2 couplings are somehow stored unchanged for a long period of time. An almost regular structure is easier to generate genetically and the region of the parameter space to be searched can be easily influenced by the change of some global parameter. In this respect it would be important to explore from an experimental point of view what are the tasks our brain *cannot* perform: this would contain much needed information on the inherited overall functional structure of the brain.

A hierarchically organized structure has also many advantages, such models have been already proposed [Dotsenko, 1985; Feigel'man and Ioffe, 1987; Parga and Virasoro, 1986]. It would also allow for storing an exponential number of patterns. Another useful idea is the use of evolutionary strategies: why shouldn't nature use at microscale a strategy so successful at the macroscale? Isn't it possible that during perception we follow in parallel many similar but not identical interacting patterns, leading nevertheless to a very well defined result?

Our understanding of basic principles does not have to ultimately come from the slavish imitation of nature, as suggested by some approaches in which the problems are expected to disappear once one puts together a sufficiently large number of neuron-like units. Trying to *construct* a machine which *flies* might teach the survivor more about aerodynamics than studying the fine structure of bird feathers. The brain is the result of a long evolutionary process and as such must be understood also in light of its history. Let us be patient.

ACKNOWLEDGEMENTS

The author is grateful to J. Delacour and J. -C. Levy for the opportunity of presenting this contribution.

REFERENCES

Ackley D. A., Hinton G. E. and Sejnowski T. J. (1985) A learning algorithm for Boltzmann Machines, *Cognitive Science* **9**, pp. 147-169

Amit D. J. (1987), The properties of models of simple neural networks, in *Proceedings of the Heidelberg Colloquium on Glassy Dynamics and Optimization*, Lecture Notes in Physics **275**, pp. 430-484 Springer-Verlag, Heidelberg.

Aubry S. (1978), The new concept of transitions by breaking of analyticity in a crystallographic model, in *Solitons and Condensed Matter Physics*, A.R. Bishop and T. Schneider (Eds.), *Solid State Sci.* **8**, pp. 264-277

Bak P. and Bruinsma R. (1982), One-dimensional Ising model and the complete devil's staircase, *Phys. Rev. Lett.* **49**, pp. 249-251

Bak P. (1982) Commensurate phases, incommensurate phases and the devil's staircase, *Rep. Prog. Phys.* **45**, pp. 587-629

Banks E. R. (1971) *Information Processing and Transmission in Cellular Automata* Ph. D. Thesis, M.I.T.

Bienenstock E. and von der Marlsburg, C. (1987), A Neuronal Network for Invariant Pattern Recognition *Europhys. Lett.* **4**, pp. 121-126

Binder K. and Young A. P. (1986), Spin Glasses: Experimental facts, theoretical concepts, and open questions, *Rev. Mod. Phys.* **58**, 801-976

Brady R. M. (1985), Optimization strategies gleaned from biological evolution, *Nature* **317**, pp. 804-805

Burkov S. E. (1985), Fractal structure of the equilibrium crystal shape, *J. Physique (Paris)* **46**, pp. 317-328

Burks A. W. (1970) *Essays on Cellular Automata*, University of Illinois Press

Carpenter G. A. and Grossberg S. (1987) in *Computer Vision, Graphics, and Image Processing* to appear

Changeux J. P. (1982) *L'homme neuronal*, Librairie A. Fayard, Coll. Pluriel, Paris

Chen H. H. *et al.* (1986), High Order Correlation Models for Associative Memory in *Proceedings of the AIP Conference on Neuronal Networks for Computing*, Snowbird, Utah, J. Denker (ed.)

Chou, W. and Griffiths R. B. (1986), Ground states of one-dimensional systems using effective potentials, *Phys. Rev. B* **34**, pp. 6219-6234

Clark J. W., Rafelski J., and Winston J. V., 1985: Brain without mind: computer simulations of neuronal networks with modifiable neuronal interactions *Physics Reports* **123**, pp. 215-273

Codd E. F. (1968) *Cellular Automata*, Academic Press

Denker J. S. (1986) Neural network models of learning and adaptation *Physica D* **22**, pp. 216-232

Diederich S. and Opper M., 1987: Learning of correlated patterns in spin glass networks by local learning rules *Phys. Rev. Lett.* **58**, pp. 949-952

Derrida B. (1987) Dynamics of Automata, Spin Glasses, and Neuronal Network Models, Lectures given at the school *Nonlinear Evolution and Chaotic Phenomena, Noto (Sicily) 8-19 June 1987*

Derrida B., Gardner E. and Zippelius A. (1987) An exactly soluble asymmetric neural network model *Europhys. Lett.* **4**, pp. 167-173

Derrida B., Pomeau Y. and Vannimenus J. (1978) Simple frustrated systems: Chains, strips and squares *J. Phys. C* **11**, pp. 4749-4765

Domany E., Meir R. and Kinzel W. (1986) Storing and retreiving information in a layered spin system *Europhys. Lett.* **2**, pp. 175-185

Domany E. and Meir R. (1987), Exact solution of a layered neuronal network model, *Phys. Rev. Lett.* **59**, pp. 359-362

Dotsenko V. S. (1985), "Ordered" spin glass: a hierarchic memory machine, *J. Phys. C* **18** (1985), pp. L1017-L1022

Essam J. W. (1972), Percolation and cluster size, in *Phase Transitions and Critical Phenomena* Vol. 2 , C. Domb and M. S. Green (Eds.), pp. 197-270, Academic Press

Feigel'man M. V. and Ioffe L. B. (1987), The Augmented Models of Associative Memory *J. Mod. Phys. B* **1** pp. 51-68

Gardner E. (1987a), Multiconnected neural network models *J. Phys. A* **20**, pp. 3453-3460

Gardner E. (1987b) Maximum Storage Capacity in Neuronal Networks *Europhys. Lett.* **4**, pp. 481-485

Gardner E. (1987c), The phase space of interactions in neuronal network models *J. Phys. A* **21**, pp. 257-270

Gardner E. and Derrida B. (1987), Optimal storage properties of neuronal network models *J. Phys A.* **21**, pp. 271-284

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

Gould J. S. (1982), Ethology: The Mechanisms and Evolution of Behavior, in *The Biology of Learning*, Norton W. W. & Co., Marler P. and Terrace H. S. (Eds.), Springer-Verlag

Godrèche C., Luck J.-M. and Orland H. (1986), Magnetic phase structure on the Penrose lattice *J. Stat. Phys.* **45** , pp. 777-800

Hebb D. O. (1949), *The Organization of Behaviour* , Wiley, New York

Hinton G. E., Sejnowski T. J. and Ackley D. A. (1984), *Boltzmann Machines: Constraint Satisfaction Networks that Learn* Technical Rep. CMU-CS-84-119, May 1984, Carnegie-Mellon University.

Hoffmann G. W. (1986), A Neural Network Model Based on the Analogy with the Immune System, *J. theor. Biol.* **122**, pp. 33-40

Hogg T. and Huberman B. A. (1985), Parallel computing structure capable of flexible associations and recognition of fuzzy inputs, *J. Stat. Phys.* **41**, pp. 115-123

Hogg T. and Huberman B. A. (1988), Artificial Intelligence and Large Scale Computation: A Physics Perspective, *Physics Reports*, to appear

Holland J. H. (1986), A mathematical framework for studying learning in classifier systems *Physica D* **22**, pp. 307-317

Hopfield J. J. (1982), Neural networks and physical systems with emergent collective computational abilities *Proc. Nat. Acad. Sci. (USA)* **79**, pp. 2554-2557

Hopfield J. J., Feinstein D. I., and Palmer R. G., (1983), "Unlearning" has a stabilizing effect in collective memories, *Nature* **304**, pp. 158-159

Horner H. (1987), in *Proceedings of the Intern. Symp. on Synergetics*, Elmau 4-5. 5. 1987

Huberman B. A. and Hogg T. (1984), Adaptation and Self-Repair in Parallel Computing Structures *Phys. Rev. Lett.* **52**, pp. 1048-1051

Huberman B. A. and Hogg T. (1985), US Patent no. 4591980

Huse D. A. (1981), Simple three state model with infinitely many phases, *Phys. Rev. B* **24**, pp. 5180-5194

Jerne N. K. (1973) The immune system, *Sci. Am.* **229**, pp. 52-60

Kanamori J. (1984), Infinite series of ground states of the ising model on the honeycomb lattice, *J. Phys. Soc. Japan* **53**, pp. 250-260

Kanter I. and Sompolinsky H. (1987), Associative recall of memory without errors, *Phys. Rev. A* **35**, pp. 380-392

Kauffman S. A. (1969), *J. theor. Biol.* **22**, pp. 437

Keirstead W. P. and Huberman B. A. (1986), Collective Detection of Motion in the Presence of Noise, *Phys. Rev. Lett.* **56**, pp. 1094-1097

Kerszberg M. and Bergman A. (1987) in: The Evolution of Data Processing Abilities in Competing Automata in *Computer Simulation in Brain Science*, Cotterill R. J. M. (ed.), Cambridge University Press

Kohonen, T. (1987), *Self Organization and Associative Memory*, Springer Verlag, Berlin

Kree R. and Zippelius A. (1987), Recognition of topological features of graphs and images in neuronal networks, KFA preprint 1987

Krishnamurthy E. V. (1985), *Introductory Theory of Computer Science*, Springer-Verlag, New York

Lee Y. C. *et al.* (1986), Machine Learning Using a Higher Order Correlation Network *Physica D* **22**, pp. 276-306

Levy J. C. S. and Mercier D. (1988), A neuronal network with self organization, *in this Volume*

Little W. A. (1974), The Existence of Persistent States in the Brain, *Math. Biosci.* **19**, pp. 101-120

Little W. A. and Shaw G. L. (1978), Analytic Study of the Memory Capacity of a Neuronal Network, *Math. Biosci.* **39**, pp. 281-290

Luck J.-M. (1987), Frustration effects in quasicrystals: an exactly soluble example in one dimension, *J. Phys. A* **20** , pp. 1259-1268

Marchand M., Hood K. and Caillé A. (1987), Nonconvex interactions: A mechanism for the occurence of modulated order in condensed matter, *Phys. Rev. B* **37**, pp. 1898-1912

McCulloch W. S. and Pitts W. (1943), A logical calculus of the ideas immanent in nervous activity *Bull. Math. Biophys.* **5**, pp. 115-133

Mézard M. *at al* (1984), Nature of the spin-glass phase, *Phys. Rev. Lett.* **52**, pp. 1156-1159

Minsky M. and Papert S. (1969), *Perceptrons*, MIT Press

von Neumann J. (1966), *Theory of Self-Reproducing Automata*, (edited and completed by A.W. Burks), University of Illinois Press, Urbana - London

Ostlund, S. (1981), Commensurate and incommensurate phases in asymmetric clock models, *Phys. Rev. B* **24**, pp. 398-405

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

Parga N. and Virasoro M. A. (1986), The ultrametric organization of memories in a neural network, *J. Physique* **47**, pp. 1857-1864

Parisi G. (1980a), A sequence of approximated solutions to the S-K model of spin glasses *J. Phys. A* **13**, pp. L115-L121

Parisi G. (1980b), The order parameter for spin glasses: a function on the interval 0-1, *J. Phys. A* **13**, pp. 1101-1108

Parisi G. (1980c), Magnetic properties of spin glasses in a new mean field theory, *J. Phys. A* **13**, pp. 1887-1895

Peretto P. and Niez J. J. (1986a), in *Disordered Systems and Biological Organization*, E. Bienenstock (ed.), NATO ASI Series Vol F20, Springer-Verlag, Berlin

Peretto P. and Niez J. J. (1986b), Long term memory storage capacity of multiconnected neural networks, *Biol. Cybern.* **54**, pp. 53-63

Personnaz L., Guyon I., and Dreyfus G. (1985), Information storage and retrieval in spin glasss-like neural networks, *J. Physique Lett.(Paris)* **16**, pp. 359-365

Physica D (1984), the full Volume **10**

Rose S. P. R. (1976), *The conscious brain*, Pelican Books, London

Rose S. P. R. (1988), The biochemistry, physiology and morphology of engram formation in the chick brain, *In this Volume*

Ruján P. (1987), Cellular Automata and Statistical Physical Models *J. Stat. Phys.* **49**, pp. 139-222

Ruján P. (1988), Searching for optimal configurations by simulating tunneling, to be published

Rumelhart, D. E. and J. L. McClelland (Eds.) (1986), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1, *Foundations*, M. I. T. Press

Sejnowski T. J., Kienker P. and Hinton G. E. (1986), Learning Symmetry Groups with Hidden Units: Beyond the Perceptron, *Physica D* **22**, pp. 260-275

Selke W. (1988), A multitude of phases in the ANNNI model, *in this Volume*

Shaw G. L., Silverman D. J. and Pearson J. C. (1985), Model of cortical organization embodying a basis for a theory of information processing and memory recall, *Proc. Natl. Acad. Sci. USA* **82**, pp. 2364-2368

Shaw G. L. (1988), Memory storage in the Trion model of cortical organization, *in this Volume*

Sompolinsky H. (1987), The Theory of Neuronal Networks: The Hebb Rule and Beyond, in *Proceedings of the Heidelberg Colloquium on Glassy Dynamics and Optimization*, Lecture Notes in Physics **275**, pp. 483-527 Springer-Verlag, Heidelberg.

Sompolinsky H. and Kanter I. (1986), Temporal Association in Asymmetric Neronal Networks, *Phys. Rev. Lett.* **57**, pp. 2861-2864

Smith III A. R. (1971), Simple computation-universal cellular spaces, *Journal of ACM* **18**, pp. 339-353

Stauffer D. (1985), *Introduction to the percolation theory*, Pergamon Press

Toulouse G., Dehaene S. and Changeux J. P. (1986), Spin glass model of learning by selection, *Proc. Nat. Acad. Sci. (USA)* **83**, pp. 1695-1698

Villain J. (1986), On the ground states of Ising models, *Ferroelectrics* **66**, pp. 143-150

Vollmar R. (1979), *Algorithmen in Zellularautomaten*, B. G. Teubner Verlag, Stuttgart

Weisbuch C. and Fogelman-Soulié F. (1985), Scaling laws for the attractors of Hopfield networks, *J. Physique Lett. (Paris)* **46**, pp. L623-L630

Wolfram, S. (1986), *Theory and Applications of Cellular Automata*, World Scientific