



Relevant parameters for a class of sequence-retrieving neural networks

O. Lefèvre, J.-P. Nadal

► To cite this version:

O. Lefèvre, J.-P. Nadal. Relevant parameters for a class of sequence-retrieving neural networks. Journal de Physique I, 1993, 3 (6), pp.1303-1328. 10.1051/jp1:1993101 . jpa-00246798

HAL Id: jpa-00246798

<https://hal.science/jpa-00246798>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification
Physics Abstracts
05.20 — 87.30

Relevant parameters for a class of sequence-retrieving neural networks

O. Lefèvre ⁽¹⁾ and J.-P. Nadal ⁽²⁾

⁽¹⁾ II Università di Roma « Tor Vergata », Dipartimento di Fisica via Emanuele Carnevale, I-00173 Roma, Italy

⁽²⁾ Laboratoire de Physique Statistique (*), Ecole Normale Supérieure, 24 rue Lhomond, F-75231 Paris Cedex 05, France

(Received 21 November 1990, revised 21 December 1992, accepted 15 February 1993)

Abstract. — Reconsidering a recently introduced model of sequence-retrieving neural network, we introduce appropriate analogues of the well-known stabilities and show how these, together with two coupling parameters λ and ϑ , entirely control the dynamics in the case of strong dilution. The model is exactly solved and phase diagrams are drawn for two different choices of the synaptic matrices ; they reveal a rich structure. We then briefly speculate as to the role of these parameters within a more general framework.

1. Introduction.

After Hopfield's seminal papers [1] of 1982-1984 had revived interest in neural networks within the physics community and the basic model had been solved by Amit *et al.* [2], research in this field has proliferated along quite a wide range of lines (see Refs [23-26] for reviews). Here we will be concerned with sequence retrieval. Although that issue has been addressed in quite a number of papers [3-15], there are not that many models. Essentially, those proposed so far fall into two categories : while some, using either time-dependent synaptic strengths [3-5] or time-delayed signals [6-11], have somehow embedded the sequentiality into their very definition, others [12-15] rely solely on instantaneous interactions. In all cases the connection matrices are asymmetric and the model, by its very nature, exhibits some kind of frustration. Indeed, even while the network has to be stabilized on the desired patterns (by means of an appropriate choice of the connection matrix) as in usual, non-sequential models, the patterns

(*) Laboratoire associé au CNRS (URA 1306) et aux Universités Paris VI et VII.

here must not be too stable either, lest this should preclude the desired transitions from taking place. Given that the features of all these models are not otherwise crucially different and that they all seemed to perform quite well when numerically implemented, which particular scheme is to be preferred by a given investigator depends on the issue he wishes to address.

Our purpose, here, is to identify the parameters governing the dynamics of such a sequence-processing network. Ideally, it is hoped that all the complexity of the dynamics can be reduced to a list comprising only a few parameters types ; more exactly there may be as much as N , or even Np such parameters for each type : one per site as well as for each of the p patterns to be memorized but so long as the whole family can still be defined by a single formula we may still be content with it. To date and as far as usual neural networks are concerned (by which we mean those models functioning as content-addressable memories, thereafter abbreviated as CAM), the best candidates have been the so-called stabilities. They are defined by :

$$\Delta_{i, \mu} = \xi_{i, \mu} \sum_j J_{ij} \cdot \xi_{j, \mu} / \|\mathbf{J}_i\| \quad (1)$$

where J is the connection matrix and \mathbf{J}_i its i -th column-vector, whose euclidean norm $\|\mathbf{J}_i\|$ is :

$$\|\mathbf{J}_i\| = [\sum_j (J_{ij})^2]^{1/2} \quad (2)$$

The vectors ξ_μ are the p patterns to be learnt ; they are random patterns with (possibly) a bias a . It is well known by now, both numerically [18] and analytically [19, 20], that the depth and size of the basins of attraction of every pattern μ in fully-connected network models are well correlated with the N stabilities of index μ ; these also offer a convenient framework for the definition of the maximal storage capacity of the network. It is however an intrinsic feature of fully-connected models that no completely satisfying reduction their dynamics can be performed. In order to achieve that, one has to recourse to a diluted model where, because of the extreme sparsity of the connection matrix, the dynamics of the whole net reduces to a set of N single-site equations. Such a model was proposed for the first time in reference [17] and it was found that its dynamics could be expressed in terms of the stabilities. Models of that kind go under the name of strongly diluted (or DGZ) models ; it is usually the case that they can be solved exactly. It is also important to realize that their use imposes no limitation upon the choice of a learning rule : one simply applies it on the remaining connections (the converse procedure, i.e. to apply the rule first and then dilute, would be tantamount to loosing the information embedded in the rule and raises the issue of the robustness of that rule against random dilution, something we are not interested in here).

Considering our purpose, the time-delayed models are particularly appropriate for, as we shall see shortly, the stability/transition duality takes in such models a very neat and convenient form. Moreover their equations of evolution are fairly simple (hence few parameters) whereas those from the other group of models tend to be composed of a very high number of terms which would be unpractical here. More specifically, we are going to use the model of references [6-8] and, like the authors of that last paper, to use it in its strongly diluted version ; we have also used the same notations wherever possible. Essentially, we find that the dynamics can be expressed as a function of two sets of stabilities, one of which is the set already defined in equation (1), plus two coupling parameters and that the usual CAM model can be viewed as the zero-limit (with respect to those coupling parameters) of the sequence-processing network. Specific cases of Hebbian and Pseudoinverse [19] type rules will be studied and « phase » diagrams drawn for each. Note that (in most cases, at least) the « phases » of our diagrams will not be true thermodynamical phases but merely regions of the phase-space where the dynamics of the network exhibits certain qualitative features : we will

call them modes (or, equivalently, regimes) of the network, basic modes of such a model being the sequential and CAM modes. Because these two rules are very different and, in a sense, complementary, we expect them, taken together, to give a good picture of what can take place within such a model. Moreover, in our view at least, they need not be taken too literally; rather, they should be viewed as representatives of their respective class of learning rules.

It will turn out that the model has a richer structure than was suspected in the original references, which was due to an inappropriate parametrization. Here, by contrast, all « phases » of the model have been identified and analytical, often explicit equations will be given for all separatrices. Although it will be presented here as a by-product of our search for the (physically) most relevant parameters of the model, such an in-depth study of that particular model is certainly valuable in itself if one considers that only a handful of the many sequential network models to date have been solved exactly and that, therefore, little is known about their « phase » diagrams.

2. The model.

The prototypical neural network consists in a single layer of N Ising spins S_i ($S_i = \pm 1$; $i = 1, \dots, N$) sequentially updated according to the Glauber dynamics :

$$P(S_i(t+1) = \pm 1) = [1 + \exp(\mp 2\beta h_i(t))] \quad (3a)$$

more often expressed in the following form :

$$\langle S_i(t+1) \rangle = \tanh [\beta h_i(t)] \quad (3b)$$

where $\langle \dots \rangle$ denotes a thermal average, β is an inverse temperature and h_i the internal field defined by

$$h_i(t) = \mathbf{J}_i \cdot \mathbf{S}(t). \quad (4)$$

It is tacitly assumed that the relevant sum runs over $j \neq i$ or, in other words, that we set all self-couplings J_{ii} to zero. Then, with a proper choice of synaptic strengths, the states represented by a set of p preassigned patterns ξ_μ will coincide with equilibrium states of the network (or merely close to them, if one allows for errors) : in the usual terminology, we say that the net has memorized the patterns.

In order to obtain a sequence-processing model from that archetype, one merely supplements the internal field h_i as defined above with a time-delayed term and rewrites (4) as :

$$h_i(t) = \mathbf{J}_i^S \cdot \mathbf{S}(t) + \mathbf{J}_i^T \cdot \mathcal{T}(t) \quad (5)$$

where \mathbf{J}^S and \mathbf{J}^T are now two distinct connection matrices, and $\mathcal{T}(t)$ is the convolution of $\mathbf{S}(t)$ with some suitable delay function whose precise form is not essential. For the sake of simplicity, we adopt for $\mathcal{T}(t)$ the simplest and most commonly used form, namely :

$$\mathcal{T}(t) = \mathbf{S}(t - \tau). \quad (6)$$

The basic idea is to reach a steady-state regime in which the network is induced by the second, delayed term in (5) to undergo transitions from one pattern to the next while it is stabilized by the first, instantaneous term on each pattern in turn as it would be in the CAM mode (whence the superscripts T and S). The duration of the stay on each pattern is of course expected to be roughly equal to the time-delay τ , which has to be much longer than the time τ_i needed to complete a transition (about which, besides, we shall have very little to say). In that sense our

solution will be a long- τ approximation but this is no real limitation since τ is a tunable parameter.

Let us now introduce m_μ , the magnetization of the network over the μ -th pattern :

$$m_\mu(t) = N^{-1} \mathbf{S}(t) \cdot \xi_\mu. \quad (7)$$

We say that at time t_0 the network has stabilized on pattern ξ_μ (or that it is in that state) when all magnetizations take negligible values but for the μ -th one, which we denote by Q_μ , i.e.

$$m_\nu(t_0) \approx Q_\mu \delta_{\mu, \nu}. \quad (8)$$

As in usual models, a good recall of ξ_μ will be characterized by $Q_\mu \approx 1$. Then, assuming that a sequential regime has been or will be established, we monitor the state of the network by means of successive pairs of magnetizations $\{m_\mu(t), m_{\mu+1}(t)\}$; indeed, starting from state ξ_μ , the network will remain in that state for some τ time-steps and then, quite abruptly, evolve towards and stabilize onto the state $\xi_{\mu+1}$, so that the other magnetizations should play no role (except, possibly, during the brief time-span of the transition). In that way the phase-space of the dynamics will always be a 2D space at most. Note that within such a framework a starting point has to be provided for the network, i.e. we must assume that at the very start of the dynamics the network had already been stabilized on the first pattern for at least τ time-steps. Hence Q_1 , the initial overlap of the net with the first pattern in the series, is actually a parameter of the model and, hereafter, will be denoted simply by Q . The starting assumption of the model (i.e. that the network had been in state ξ_1 for at least τ time-steps) thus reads :

$$m_\mu(t) \approx Q \delta_{\mu, 1}, \quad t = 0, \dots, \tau. \quad (9)$$

At this point, the possibility of the patterns having a non-zero bias a makes it necessary to shift to a new set of variables. Assuming that at time t the network is in a state with macroscopic overlap onto patterns ξ_μ and $\xi_{\mu+1}$ we may introduce $x_\mu(t)$ and $x_{\mu+1}(t)$ defined by :

$$\langle \mathbf{S}(t) \rangle = x_\mu(t) \xi_\mu + x_{\mu+1}(t) \xi_{\mu+1} \quad (10)$$

where $\langle \rangle$ denotes as before a thermal average, and rewrite the magnetizations m_μ and $m_{\mu+1}$ as linear combinations of x_μ and $x_{\mu+1}$.

$$m_\mu(t) = x_\mu(t) + q_\mu x_{\mu+1}(t) \quad (11a)$$

$$m_{\mu+1}(t) = q_\mu x_\mu(t) + x_{\mu+1}(t) \quad (11b)$$

where q_μ is the overlap between the two successive patterns :

$$q_\mu = N^{-1} \xi_\mu \cdot \xi_{\mu+1}. \quad (12)$$

It will then turn out (see Appendix A) that we can get rid of any explicit dependance on the bias and thus that the dynamics of the system is most conveniently parametrized by using not the successive pairs $\{m_\mu, m_{\mu+1}\}$ but the new ones $\{x^+, x^-\}$ defined by :

$$x^\pm(t) = \pm x_\mu(t) + x_{\mu+1}(t) \quad (13)$$

the index μ running from 1 to p as every successive transition is completed.

With these definitions and assumptions, the short-term equations of evolution which describe the transition from some pattern μ to the next pattern $\mu+1$ in the strongly diluted

network are (see Appendix A) :

$$x^\pm(t+1) = \int \rho(\Delta_b^s, \Delta_f^s, \Delta^t, \lambda, J) d\Delta_b^s d\Delta_f^s d\Delta^t d\lambda dJ \int Dz \times \\ \times \tanh \beta J [\pm \Delta_b^s x_\mu(t) + \Delta_f^s x_{\mu+1}(t) + \lambda Q_\mu \Delta^t + z\sigma(x^\pm(t))] \\ \rho(\Delta_b^s, \Delta_f^s, \Delta^t, \lambda, J) = N^{-1} \sum_i \delta(\Delta_b^s - \Delta_{\mu,i}^s) \delta(\Delta_f^s - \Delta_{\mu+1,i}^s) \delta(\Delta^t - \Delta_{\mu,i}^T) \delta(\lambda - \lambda_i) \delta(J - J_i) \quad (14)$$

$$\sigma^2(x) = (1 - x^2) + \lambda^2(1 - Q^2)$$

where Dz is the centered, normalized Gaussian measure :

$$Dz = (2\pi)^{-1/2} \exp(-z^2/2) dz \quad (15)$$

and J_i denotes the norm of the vector \mathbf{J}_i^S :

$$J_i = \|\mathbf{J}_i^S\| = [\sum_j (J_{ij}^S)^2]^{1/2} \quad (16)$$

The Δ^S and Δ^T introduced in these equations are the two classes of stabilities called for by the present model ; for each neuron i and pattern μ they are defined by

$$\Delta_{\mu,i}^S = \|\mathbf{J}_i^S\|^{-1} \xi_{i,\mu} \mathbf{J}_i^S \cdot \xi_\mu \quad (17a)$$

$$\Delta_{\mu,i}^T = \|\mathbf{J}_i^T\|^{-1} \xi_{i,\mu} \mathbf{J}_i^T \cdot \xi_{\mu+1} \quad (17b)$$

Note that the $\Delta_{\mu,i}^S$ are just the standard stabilities of CAM models. On the other hand λ_i defined by

$$\lambda_i = \|\mathbf{J}_i^T\| / \|\mathbf{J}_i^S\| \quad (18)$$

is the first (and by far the most important) of the two coupling parameters of the model ; it is a measure of the relative strength of the couplings responsible for the stabilization and of those responsible for the transition. So far the joint probability distribution $\rho(\dots)$ in (14) depends on the learning rule ; in most practical cases however, the norms are site-independant and exhibit no fluctuations (in the thermodynamic limit). It is then possible to simplify (14) by rescaling the temperature from βJ to β , which amounts to setting $J = 1$. In the following we will always assume that such is the case and we can thus rewrite (14) as :

$$x^\pm(t+1) = \int \rho(\Delta_b^s, \Delta_f^s, \Delta^t, \lambda) d\Delta_b^s d\Delta_f^s d\Delta^t d\lambda \int Dz \times \\ \times \tanh \beta [\pm \Delta_b^s x_\mu(t) + \Delta_f^s x_{\mu+1}(t) + \lambda Q_\mu \Delta^t + z\sigma(x^\pm(t))] \quad (19)$$

Equation (19) is the main result of this section. For any given learning algorithm for which the distribution ρ can be computed, in principle the full phase diagram can then be derived. It would certainly be interesting to perform a classification of the possible distributions, in the line of what has been done for CAM models [21]. We will not address this question here, although it is very likely that very similar results will hold since the algorithms that can be used here in order to obtain the couplings are essentially the same as those already in use for CAM networks. As a first step towards such a classification, it is thus of interest to consider two extreme cases, one where the stabilities are widely distributed with positive and negative

values, and another where all stabilities are strictly positive and take most of their values within a short interval. As prototypes of these two extremes, we will study the following particular cases, named after the corresponding CAM rule :

$$(a) \quad \text{Hebb rule :} \quad J_{ij}^S = \sum_{\mu} \xi_{i, \mu} \xi_{j, \mu} ; \quad J_{ij}^T = \lambda \sum_{\mu} \xi_{i, \mu} \xi_{j, \mu+1} \quad (20a)$$

$$(b) \quad \text{« Pseudoinverse » rule :} \quad \Delta_{i, \mu}^S = \Delta^S > 0 ; \quad \Delta_{i, \mu}^T = \Delta^T > 0 . \quad (20b)$$

Note that (20) is valid as such only for unbiased patterns but that the Hebb rule might still be defined by a similar formula in the general case. On the other hand three remarks are in order here about our definition of the « Pseudoinverse ». First in our view (20b) only means that all the stability parameters of a given type take the same value ; in the case of CAM, this would indeed correspond to the choice of the standard Pseudoinverse algorithm [19, 20]. Here, however, there are at least two ways of generalizing the standard algorithm. The first is to compute *separately* J^S and J^T , thus effectively solving for (20b), but it is not the most relevant definition in this context and we adopted it for the sake of simplicity only. A better extension of the Pseudoinverse algorithm in this context [11] is to solve for

$$\sum_j \xi_{i, \mu+1} J_{ij}^S \xi_{j, \mu+1} + \sum_j \xi_{i, \mu+1} J_{ij}^T \xi_{j, \mu} = 1 \quad (21)$$

which amounts to imposing the constancy not of Δ^S and Δ^T separately but that of the total stability Δ defined by :

$$\Delta_{i, \mu} = [\Delta_{i, \mu+1}^S + \lambda \Delta_{i, \mu}^T] [1 + \lambda^2]^{-1/2} \quad (22)$$

where the factor $(1 + \lambda^2)^{-1/2}$ comes from a normalization by the global norm $[\sum_j (J_{ij}^S)^2 + \sum_j (J_{ij}^T)^2]^{1/2}$. The reason for that choice will appear later when we will show that such is the relevant definition of the stability for the long-term dynamics and hence that Δ (and not Δ^S or Δ^T) is the parameter we must optimize. The second point is that our definition (20b) of the « Pseudoinverse » implies $\lambda = 1$. That, however (even if we stick to (20b) *strictu sensu*), is a spurious limitation since λ may be recovered as a free parameter simply by incorporating it into the definition of J^T resulting from (20b), whatever it may be ; note, besides, that this is precisely the way in which it had to be introduced into the Hebb rule, i.e. as an *a priori* parameter. Such a procedure is legitimated by the rescaling from βJ_i to J_i , as a consequence of which it is only the ratio λ_i of $\|J_i^T\|$ to $\|J_i^S\|$ that matters. Finally, we wish to stress again the fact that the conclusions of this study need not be restricted to the particular picture which would be obtained by using the pseudo-inverse algorithm, no matter how we define it exactly. Rather, as indicated before, Δ^S and Δ^T can be understood here as the *typical* stability parameters that would be produced by any algorithm resulting in a sufficiently peaked distribution of the stabilities. Thus we want to explore the phase diagram as a function of these parameters as if we were free to chose them all, λ included.

The study of the Hebbian case was already the subject of the original paper [8]. Within our general framework we will recover their results and also obtain a more detailed description of the phase diagram. A notable difference, however, is that we are working directly at the thermodynamic limit (instead of first using finite- N expressions of combinatorial origin in the line of [17]), thereby emphasizing in equation (14) the role of the stabilities at an early stage of the calculation. On the other hand, here as with CAM models, the study of the Hebbian case is particularly instructive, especially in order to contrast it with constant stability models in the line of the Pseudoinverse.

Now, in order to resume from (14), all we need is the distribution $\rho(\Delta_b^S, \Delta_i^S, \Delta^T, \lambda)$. Let us consider first the Hebb rule. Then $\rho(\Delta_b^S, \Delta_i^S, \Delta^T) = D(\Delta_b^S) D(\Delta_i^S) D(\Delta^T)$, the symbol D denoting here a Gaussian distribution with mean $1/\sqrt{\alpha}$ and variance 1, and the mean values of $\|J_i^S\|^2$ and $\|J_i^T\|^2$ are just α and $\lambda^2 \alpha$ respectively, so that λ introduced *a priori* in (20a) does

indeed meet the definition (18), as it should. On the other hand, α denotes as usual the storage ratio p/N . As a result, regrouping all terms into a single Gaussian distribution and under a unique integral over z , we retrieve the remarkably simple equation of evolution already derived in [8] :

$$\begin{cases} x^\pm(t+1) = f_H[x^\pm(t)] \\ f_H(x) = \int Dz \tanh \beta [(x + \lambda Q_\mu) / \sqrt{\alpha} + z\sigma_H] \\ \sigma_H^2 = 1 + \lambda^2 \end{cases} \quad (23)$$

For the Pseudoinverse rule the distribution $\rho(\cdot)$ is simply a product of delta distributions and we get

$$\begin{cases} x^\pm(t+1) = f_P[x^\pm(t)] \\ f_P(x) = \int Dz \tanh \beta [\Delta^S x + \lambda \Delta^T Q_\mu + z\sigma_P(x)] \\ \sigma_P^2(x) = (1 - x^2) + \lambda^2(1 - Q_\mu^2) \end{cases} \quad (24)$$

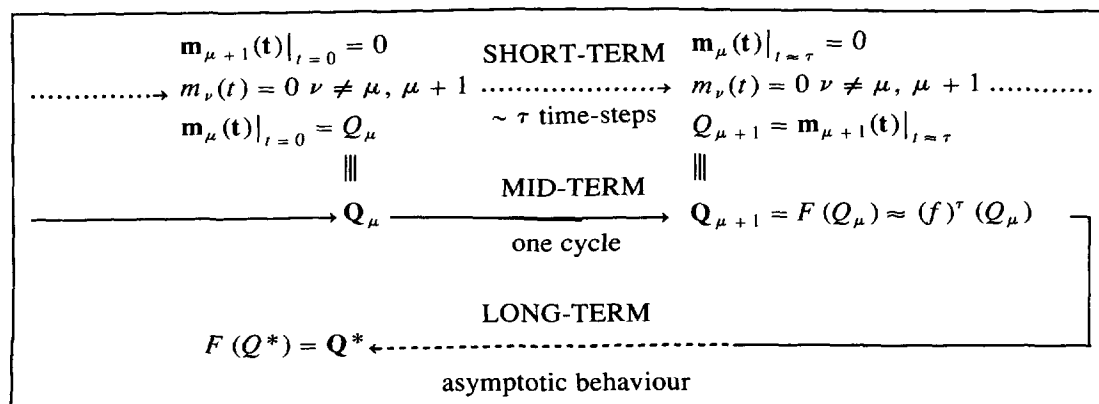
We now turn to a qualitative description of the various behaviours generated by the above equations of evolution. Its purpose is to keep the description of the phase diagrams in sections 4 to 6 as concise as possible, even while a whole zoo of phases will have to be introduced.

3. Overview of the dynamics.

In both cases the uncoupling of the equations of evolution of x^+ and x^- enables us to carry on an exhaustive analytical study of the model and an exact determination of its « phase » diagram ; we have failed to find other rules enjoying the same property. For simplicity only the zero-temperature case will be investigated. Note that since we are monitoring the state of the network in the space of successive pairs of magnetizations, for which (because of the dilution) the equations of evolution are of fixed-point type, the dynamics discussed below will be at worst a two-cycle ; of course with respect to individual spin trajectories more complicated (possibly chaotic) dynamics can not be ruled out.

3.1 DYNAMICS. — Sequential models such as this possess three time-scales ; they are summed up in the following table, where it is assumed that the network is locked in a sequential mode and where for each time-scale the relevant variables are printed in bold type :

Table I.



The meaning of this diagram is that the overall dynamics comes in cycles, each cycle being associated with the pattern on which the network has just stayed, for around τ time-steps. The succession of these shifts from pattern to pattern constitutes the mid-term dynamics, the most natural of the three time-scales in such a model. It is parametrized by the steady-state magnetizations Q_μ where the pattern index μ , running from 1 to p (where we set, if necessary, $\mu + p \equiv \mu$), acquires the meaning of a discrete time index. Underlying that global dynamics is a short-term, « microscopic » dynamics thanks to which the shift from Q_μ to $Q_{\mu+1}$ can be explicitated in terms of the instantaneous magnetizations $m_\nu(t)$, particular attention being paid to $m_\mu(t)$ and $m_{\mu+1}(t)$, the other ones fluctuating around zero. For all practical purposes the time variable t may be reset to zero at the onset of each transition since, past the first few ones, all shift are expected to be isomorphic (i.e. identical in every respect but for a translation of all pattern indices), which is precisely the rationale for a direct definition of the mid-term dynamics. If we denote by F its equation of evolution, then $F = (f)^\tau$ should approximatively hold; practically, F is derived from (19) in the form of a closed equation in Q_μ and $Q_{\mu+1}$. Finally if the mid-term dynamics is extrapolated to infinite values of the pattern index μ , we obtain the asymptotic magnetization Q^* as the fixed-point of F . We will refer to that asymptotic regime as the long-term dynamics.

3.2 DIAGRAM OF REGIONS (« PHASE » DIAGRAM). — We recall that, at the very start, the variables are x_1 and x_2 or, equivalently, x^+ and x^- . If we denote by the couple (x_1^*, x_2^*) the asymptotic solution of the short-term dynamics defined by equations (24) or (25) under initial assumption (9) (i.e., its fixed-points under the initial conditions $x^\pm(0) = \pm Q$), the sequential retrieval is then characterized by $x_1^* = 0$ and the network, initially in state ξ_1 , is now in state ξ_2 (in the sense of Eq. (8)). We denote by R the corresponding region of the phase-space and by C the remainder (i.e. the region where $x_1^* \neq 0$). Within the present framework, its physical content can be defined only negatively: rather than corresponding to any well-defined mode of the network, C is a sort of residue of our reasoning; more precisely, it is the region where the consistency of the initial assumption (9) can not be guaranteed. In other words, our reasoning is in essence a recurrence proof, which explains why the short-term separatrix (i.e. the R/C separatrix) has to be defined with respect to the behaviour of the network during the first τ time-steps (i.e. during the first cycle if we are in R). Hence, whereas in R consistency allows us to proceed further into the mid- and (possibly) long-term dynamics, the study of C is *per force* limited to the first cycle, so that the long-term behaviour of the system in that region can not be rigorously determined. Later we will see how some information can nonetheless be extracted and that the region C can be quite satisfyingly characterized.

Next, within the R region, the value of the fixed-point Q^* (implicitly $Q^*(Q, \lambda, \dots)$) of the implicit equation for $Q_\mu \cdot Q_{\mu+1} = Q_{\mu+1}(Q_\mu, Q_{\mu+1})$ characterizes the long-term behaviour of the network, whence a further subdivision in the phase diagram. Depending on whether Q^* is zero or finite, R is split into a « ferromagnetic » part R_{FG} ($Q^* \neq 0$) and a « paramagnetic » one R_{PG} ($Q^* = 0$).

Now, before we proceed further, let us forget for a while our previous definition (22) of the total stability Δ ; we are going to show that its existence has a necessary character and that the relations between the various regions defined so far may be clarified on a purely geometrical ground. On the first hand at null λ the sensible definition of a para- and a ferromagnetic « phase » is whether x_1^* is null or not (with $x_1^* \neq 0$ corresponding to the CAM mode); it is natural to extend that definition to the region C , which is then entirely FG by its very characterization (since $(0, 0)$ can not be a fixed-point if $\lambda \neq 0$). On the other hand, reasoning with the simpler Pseudoinverse case, we note that the parameters of the phase space are $\lambda, Q, \Delta^S, \Delta^T$ but that the list shrinks to Q, Δ^S at zero λ . Finally we may also notice that the

boundary between FG and P, again at zero λ , is simply $\Delta^S \leq \Delta_c$, where Δ_c is an as-yet-to-be-determined critical value of Δ . Hence, assuming some non-pathological topology of the « phase » diagram, there must exist a global parameter $\Delta(\Delta^S, \Delta^T, \lambda)$ (actually a kind of rectified Δ^S , i.e. obeying $\lim_{\lambda \rightarrow 0} \Delta = \Delta^S$) such that in the (λ, Δ) plane the relations between these

various regions and their definitions are simply :

$$FG = R_{FG} \cup C = \{(\lambda, \Delta) / \Delta \leq \Delta_c\} \quad (25a)$$

$$PG = R_{PG} = \{(\lambda, \Delta) / \Delta \geq \Delta_c\} . \quad (25b)$$

By the same reasoning a similar parameter $\Delta(\alpha, \lambda)$ should also exist for the Hebb rule. Moreover, since both models derive from the same equation (19), it should be possible to work out a unique formula for Δ irrespective of the rule. Such a parameter not only would neatly bridge the gap between both sets of definitions but also bring in almost one-to-one correspondence the long-term sequential mode of the network and the (trivially short-term in that context) CAM mode of the same network at zero λ . Of course Δ as defined in (22) is precisely such a parameter and we shall see shortly that (25) does indeed hold.

Finally, let us return to the mid-term dynamics, which we had skipped because its analysis requires knowledge of Q^* . It turns out that the dependance of the separatrices on Q can be the source of an interesting behaviour. If, indeed, once Q^* is known, it appears that the sequence of the Q_μ 's is a decreasing one while the R/C separatrix is also a (not necessarily uniformly) decreasing function of Q for the (λ, Δ) couple considered, then, according to an evident scenario, the sequential regime might end abruptly. We will denote by R^r the part of R (possibly) corresponding to such a transient sequential regime. That very possibility (which had been overlooked in [8]) is interesting because it may reveal something of the nature of the region C ; moreover we shall see that, in the case of the Pseudoinverse rule at least, a very considerable portion of R is concerned. On the other hand, if such a region exists at all, it will be interesting to see whether it belongs to R_{PG} or R_{FG} ; the latter case would indeed be worse since the sequential regime in R_{PG} is scheduled to fade out anyway (we might call it evanescent).

Before we conclude, let us recall that all the separatrices between the regions defined above are not necessarily of thermodynamical nature, which will then show up in their dependance on Q . In fact, only the FG/P separatrix is thermodynamical. The following table gathers the notations and definitions of all regions :

Table II.

R : the first Q_μ 's at least are non-zero (i.e. at the end of the first cycle $x_1^* = 0$ and $x_2^* \neq 0$)	$R_{FG} :$ $Q^* \neq 0$	the series $Q_\mu \mid_{\mu = 1, \dots, p}$ falls abruptly to zero : R_{FG}^{tr}
		all Q_μ 's are non-zero : R_{FG}^{st}
	$R_{PG} :$ $Q^* = 0$	the series $Q_\mu \mid_{\mu = 1, \dots, p}$ falls abruptly to zero : R_{PG}^{tr}
		all Q_μ 's are non-zero : R_{PG}^{st}
C : the sequential regime fails to set up (i.e. at the end of the first cycle $x_1^* \neq 0$ and $x_2^* = 0$) ; C is thus actually restricted to C_{FG}		

4. Short-term dynamics.

In order to obtain the equation of the R/C separatrix, the following reparametrization of the phase-space is useful :

$$(\lambda, \delta) \quad \delta = [\alpha (1 + \lambda^2)]^{-1/2} \text{ instead of } (\lambda, \alpha) : \quad \text{Hebb rule} \quad (26a)$$

$$(\lambda, \delta, \vartheta) \quad \delta = \Delta^S + \lambda \Delta^T \quad \text{instead of } (\lambda, \Delta^S, \Delta^T) : \text{Pseudoinverse rule} \quad (26b)$$

$$\vartheta = \Delta^T / \Delta^S$$

Note that the definitions of δ_{Hebb} in (26a) and δ_{Pseudo} in (26b) do not match ; we use them only in this section because they are the parameters naturally arising within the short-term equations. The diagrams drawn for the PS rule will also look better if these variables are used. We now let the temperature go to zero. In this way the functions of evolution become :

$$f_H(x) = \text{erf} \left[\frac{1}{\sqrt{2}} \delta (x + \lambda Q) \right] \quad (27a)$$

$$f_P(x) = \text{erf} \left[\frac{1}{\sqrt{2}} \frac{x + \lambda Q}{(1 + Q\lambda) \sigma(x)} \right]. \quad (27b)$$

A remark is in order here about ϑ : whereas the apparent limitations as to λ could be easily disposed of, it is here the very definition of ϑ that seems to imply that most learning rules will enjoy the property $\vartheta = 1$, i.e. that any imbalance between Δ^S and Δ^T is in fact a property of the coupling matrices J_{ij} and J_{ji}^T , which a calculation *à la Gardner* in the space of the couplings J_{ij} might confirm (it is, besides, the case of both rules investigated here if taken at face value). On the other hand, for reasons to appear shortly, we are particularly unwilling to limit our study to the case $\vartheta = 1$. As a compromise, only values of ϑ close to unity will be considered below. It is warranted by the fact that, as explained before, the rules investigated here are only representatives of their class and that departures from the value of ϑ implied by these rules may therefore be indulged if not too large.

Numerically the determination of λ_s , the R/C separatrix, is very easy : R is simply redefined as $\{(\lambda, \delta) / x_1^* < \varepsilon\}$, where (with our computer and *erf* routine), a good value of ε is 10^{-5} . Having obtained the numerical curve, we next proceed to its analytical determination and check that both coincide. All curves presented in the figures have been submitted to that double search procedure and in the following we will present only the analytical argument.

Analytically, we rewrite the natural definition of R : ($x_1^* = 0$) as $x^+ = x^-$, i.e. the same positive fixed-point must be reached no matter whether the iteration of f or g started from Q or $-Q$ (since $x^\pm(0) = \pm Q$). In other words, the interval $[-Q; Q]$ must be free of fixed-points ; it is on the latter, geometrical definition of R that we will rely in the following.

4.1 HEBB RULE. — The separatrices $\lambda_{s,Q}(\delta)$ are smooth (i.e. continuously differentiable) curves ; but for their endpoints $(\lambda, \delta) = (0, \sqrt{\pi/2})$ and $(\lambda, \delta) = (1, +\infty)$, they depend on Q . An interesting feature is that, in spite of their smoothness, they are composite curves (see Fig. 2 and Appendix B for their construction) whose proximal and terminal segment correspond to curves denoted by λ_I^- and λ_{II}^- , respectively. λ_{II}^- is expressed as a function of the important quantity $\Delta_c^H(Q)$ whose implicit equation is (see Fig. 1) :

$$Q = \text{erf} [Q \Delta_c^H / \sqrt{2}] \quad (28)$$

(which is none other than the equation for the fixed-point of f at zero λ and hence the Δ_c alluded to in Sect. 3.2). By far the most important property of the family of curves $\lambda_{s,Q}$ is the following :

$$\forall Q, Q' \quad Q < Q' \Rightarrow \forall \delta, \quad \lambda_{s,Q}(\delta) > \lambda_{s,Q'}(\delta) \quad (29)$$

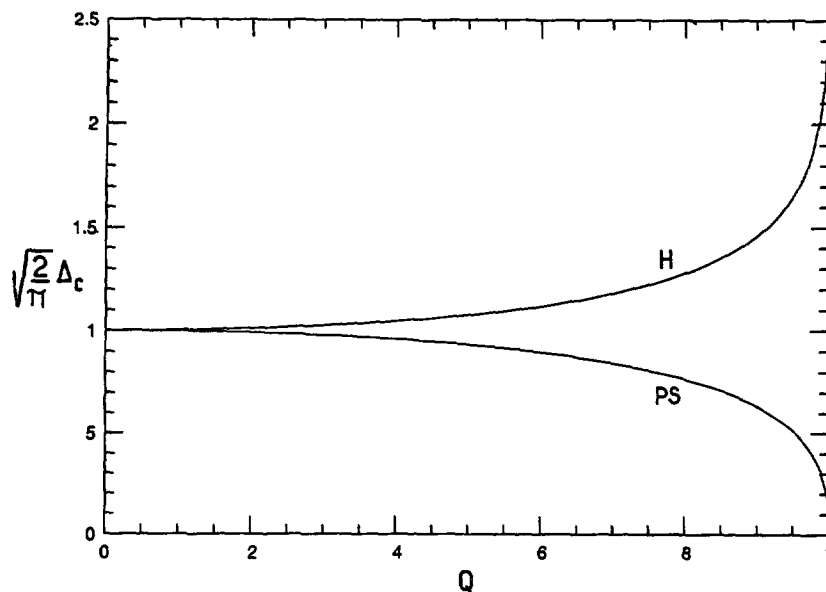


Fig. 1. — $\sqrt{2/\pi} \Delta_c^H$ and $\sqrt{2/\pi} \Delta_c^{PS}$ versus Q . Note that $\Delta_c^H(Q)$ has a vertical asymptote for $Q = 1$ while $\Delta_c^{PS}(1) = 0$.

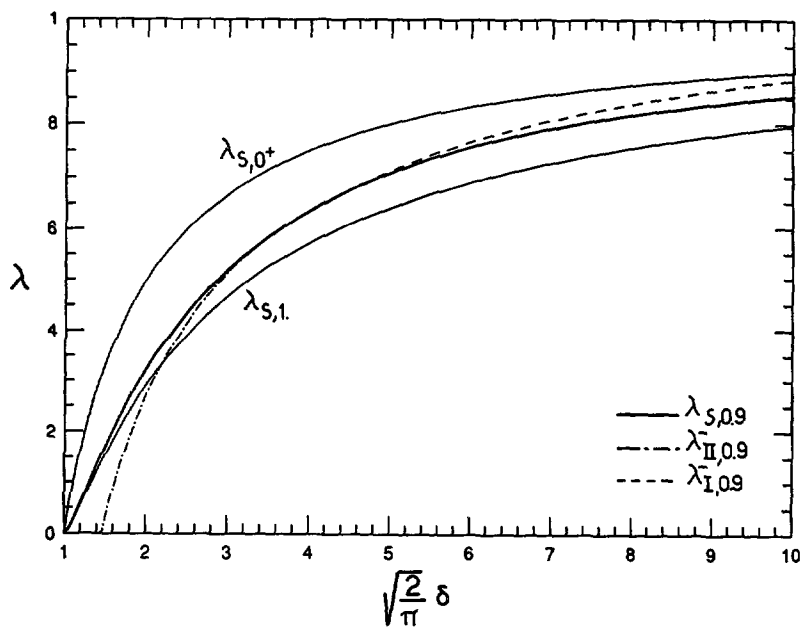


Fig. 2. — Hebb rule : 1) $\lambda_{I,Q}^-$ and $\lambda_{II,Q}^-$ versus $\sqrt{2/\pi} \delta$, building up $\lambda_{S,Q}$, $Q = 0.9$. See text for their respective pathways. 2) The upper and lower curves, on the other hand, are $\lambda_{S,1} = \lambda_{I,1}^-$ and $\lambda_{S,0^+} = \lambda_{II,0^+}^-$ versus $\sqrt{2/\pi} \delta$; thus all other curves $\lambda_{S,Q}$ are comprised between those two.

which can be proved by resorting to a blend of algebraic and geometrical reasoning. It is a kind of uniformity property; it implies that the $\lambda_{s,Q}$ s can be ordered according to their Q . In particular (see again Fig. 2) :

$$\forall Q, \lambda_{s,0^+} > \lambda_{s,Q} > \lambda_{s,1} \quad (30)$$

where by $\lambda_{s,0^+}$ we mean by definition $\lim_{Q \rightarrow 0, Q > 0} \lambda_{s,Q}$ (see Appendix B for more details).

Finally, what about the possible existence of a CAM regime? This (i.e. not only $x_1^* \neq 0$ but $x_2^* = 0$, too), by equation (27) implies a null λ , which is trivial. Numerically, however (Fig. 3), it turns out that a large part of C correspond to an « almost-CAM » regime characterized, moreover, by a high value of x_1^* . Thus the most probable fate of the network, in most if not all of C, is that it will be stuck in a local state characterized by ($x_2^* \approx 0$; $x_1^* > 0.9$) during the first cycle (which in C refers merely to the first τ time-steps) and then, by consistency, during the subsequent ones as well. To all practical purpose, C should thus be a good CAM regime.

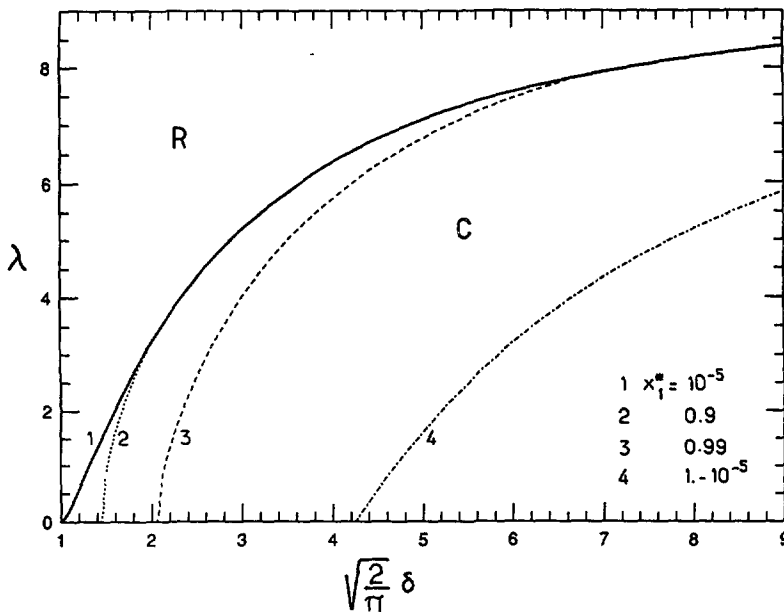


Fig. 3. — Hebb rule: $\lambda_{s,0.9}$ versus $\sqrt{2/\pi} \delta$; the curves drawn within C are curves of same x_1^* and the regions they delimit on their right are defined by: $\{x_1^* > x\}$, with $x = 1 - 10^{-5}$, 0.99, 0.9 and 10^{-5} from right to left, the latter region, i.e. $\{x_1^* > 10^{-5}\}$, being just C. On the other hand, to all practical ends $x_2^* = 0$ within C.

4.2 PSEUDOINVERSE RULE. — Depending on the value of Q , the separatrices $\lambda_{s,Q}$ are now made up of three segments if $Q < 0.9$ (approximatively), four otherwise; each segment is smooth but their intersections generate kinks in λ_s . Both cases are illustrated in figures 5 and 6, respectively. In all cases it is possible to recognize an ascending branch corresponding to the first segment and a descending one made up of the remainder (thus, « optional » segment, if any, included). Any direct comparison between λ_s^{Hebb} and $\lambda_s^{\text{Pseudo}}$ must be limited to the very

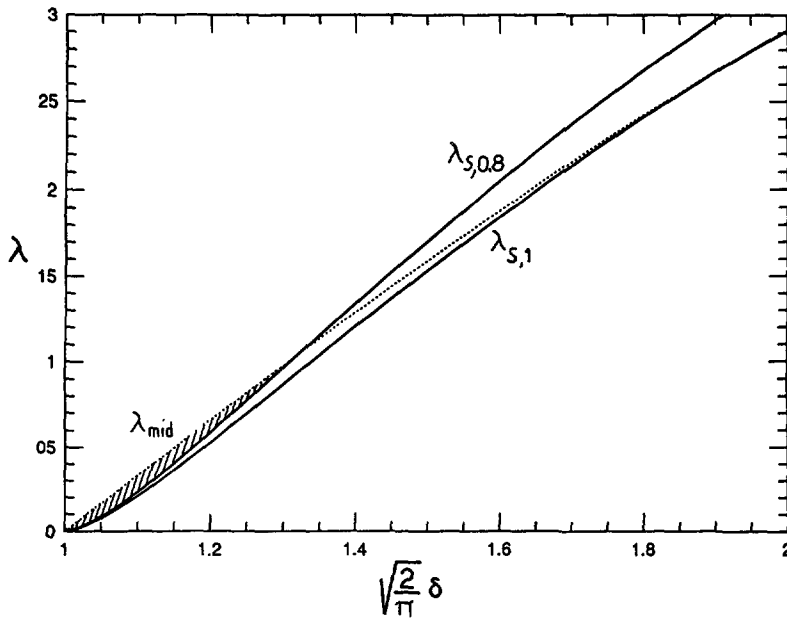


Fig. 4. — Hebb rule : $\lambda_{s,1}$, λ_{mid} and $\lambda_{s,0.8}$ versus $\sqrt{2/\pi} \delta$; the shaded area comprised between the latter pair of curves is therefore $R_{FG}^{tr}(0.8)$.

last segment for it alone corresponds to a contribution from the same underlying curve (actually λ_{II}^- , see Appendix B) ; other segments do not match. The one Q -independant endpoint of this model, namely $(\lambda, \delta) = (1/\vartheta, +\infty)$ (to be compared with $(\lambda, \delta) = (1, +\infty)$ for the Hebb rule) belongs precisely to the last segment ; the other endpoint is now a function of Q : $(\lambda, \delta) = (0, \Delta_C^{PS}(Q))$, where $\Delta_C^{PS}(Q)$ is defined in the same way as its counterpart $\Delta_C^H(Q)$ in (28), i.e. by :

$$Q = \text{erf} [Q \Delta_C^{PS} / \sqrt{2(1 - Q^2)}] . \quad (31)$$

At this point of the analysis, we may say that $\lambda_s^{\text{Pseudo}}$ looks like its counterpart λ_s^{Hebb} , except that its smooth start is somehow replaced by a « horn » to which contribute all the new segments.

Beyond this point it is more appropriate to revert to Δ defined in (22) (more precisely to its average value in case (a)), i.e. by :

$$\langle \Delta \rangle = \delta(1 + \lambda) : \text{Hebb} ; \quad \Delta = \delta / \sqrt{1 + \lambda^2} : \text{Pseudoinverse} \quad (32)$$

where $\langle \dots \rangle$ now denotes the ensemble average. The change is mandatory for any comparison of the two rules since Δ and not δ is the common parameter ; it also has the advantage that the λ_s , once their expressions have been obtained as a function of δ , are simpler to discuss (but also more cramped) if thus recast : only compare figures 5, 6 with figure 7.

1) The ordering of the $\lambda_{s,Q}$ s with respect to Q is now evident (Fig. 7) : the vertical part of λ_s (almost vertical when the « optional » segment contributes to its terminal end), whose equation becomes $\Delta = \Delta_C^{PS}$ (thus meeting (25)), recedes towards the origin as Q is increased ; the remaining part, on the other hand, obeys just the converse of (29), i.e. it is uniformly lowered as Q is decreased. As to the proximal and terminal segments, this simply follows from

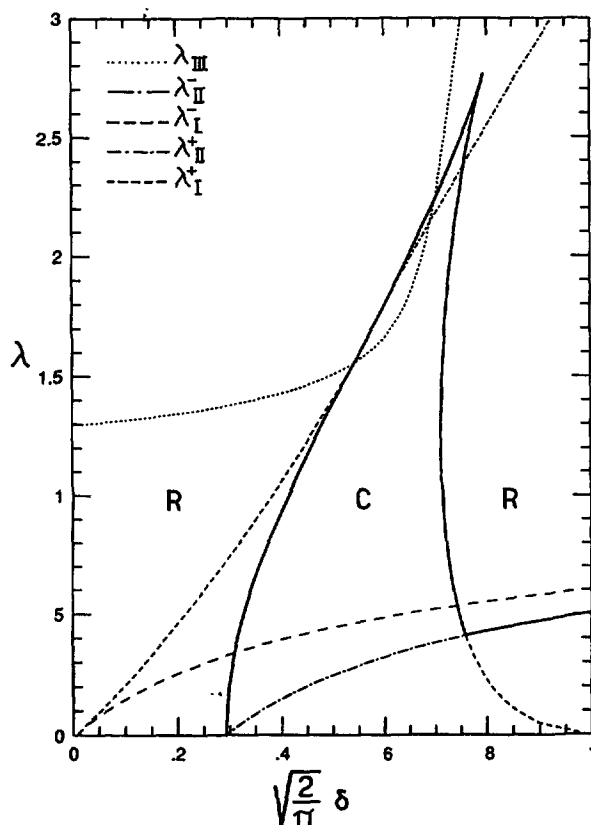


Fig. 5. — Pseudoinverse rule (with $\vartheta = 1$): here $Q = 0.9$. As in figure 2, $\lambda_{s,Q}$ is superimposed on the curves contributing to it, i.e. λ_{II}^+ , λ_{I}^+ and then λ_{II}^- , all are drawn versus $\sqrt{2/\pi} \delta$. Also displayed are λ_{I}^- and λ_{III} , though they do not contribute in this case. For lower values of Q , λ_{III} may not even be defined.

$\partial \Delta^{\text{PS}}(Q)/\partial Q < 0$; for the other, intermediate segment(s), we will be content with relying on numerical evidence. To sum up, even while we can not rigorously write any inequality in the line of (29) that would hold for the whole curve, the situation is nonetheless very much (but not quite) the opposite of what we had found in section (a) in that the relative area of C is now expected to increase with Q instead of decreasing.

2) A very neat ordering of the family of curves λ_s is also possible with respect to the parameter ϑ (see Fig. 8), with ϑ playing the role assigned to Q in equation (29). In figure 8 we have only displayed the graphs corresponding to $Q = 0.9$ but we have checked that the same conclusion holds for a large range of values of Q .

3) As to the comparison of $\lambda_{s,Q}$ for the Hebb and Pseudoinverse rule, it is not difficult to prove that $\lambda_{s,Q}^{\text{Hebb}}$ lies entirely below $\lambda_{s,Q}^{\text{Pseudo}}$ whatever the value of Q . The proof is exact and rests solely on the respective definitions of Δ_C and λ_{II}^- for the two rules. Figure 10 should give a feeling of how large the difference can be. The comparison has of course to be carried out for $\vartheta = 1$ since such is the value of ϑ with Hebb rule.

Finally, let us again examine the possibility of a CAM (or « quasi-CAM ») mode within region C (Fig. 9). As for the Hebb rule the largest part of C, namely the whole area lying below λ_{II}^- , corresponds to a quasi-CAM regime with very good characteristics; a pure CAM regime

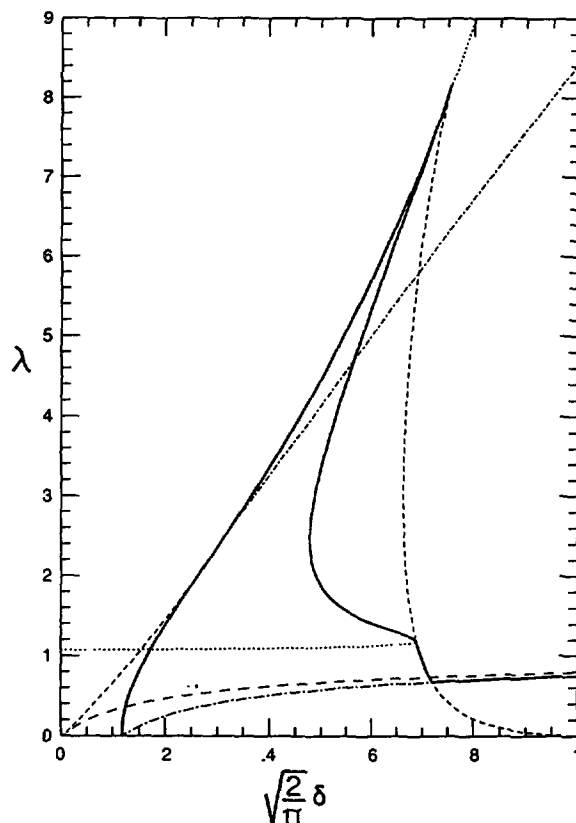


Fig. 6. — Pseudoinverse rule (with $\vartheta = 1$): same as before but $Q = 0.999$; note the lowering of λ_{III} , so that it now contributes to $\lambda_{s,Q}$.

is also possible, but only for $\lambda = 1$ or $Q = 1$. It is also noteworthy that if $Q = 1$ the model enjoys the property of perfect recall, i.e. $x_1^* = 1$ within C and $x_2^* = 1$ within R. The true novelty, however, is that the remainder of C (i.e. the portion lying above λ_{II}) corresponds to a regime of mixed attractors, where x_1^* and x_2^* are both non-zero and of comparable magnitude. The really meaningful definition of the « horn » is therefore

$$\text{«horn»} = C \cap \{(\lambda, \delta) / \lambda > \lambda_{II}(\delta)\} = \{(\lambda, \delta) / \lambda_{II}(\delta) < \lambda < \lambda_s(\delta)\}. \quad (33)$$

This suggests that the first sub-region of C_{pseudo} on the one hand and C_{Hebb} on the other are close analogues and that C_{pseudo} is supplemented with a horn rather than replaced by it, as we implied above in a first approximation. For (λ, δ) couples belonging to the horn the most likely fate of the network is a complete blurring as it overlaps with all patterns in turn without « forgetting » the former ones. During the first « cycle » (i.e. when x_1^* and x_2^* are still the only relevant parameters), the picture is simplified by the fact, to a very high accuracy, $x_1^* + x_2^* = 1$. In fact this is true also of other regions of the diagram (at least for both λ and Δ not too large), where it is trivial.

The following table presents a summary of the discussion (note that it does not take into account the long-term FG/PG distinction):

Table III.

Region or sub-region	Nature of 1 st cycle attractor	Regime	Long-term behaviour of the network
R	$x_1^* = 0, x_2^* > 0$	Sequential	A number of patterns ξ_μ are recovered in turn ; attractors ($x_\mu^* = 0, x_{\mu+1}^* > 0$), $\mu = 1$ to $\mu_c(\text{tr})$ or $\infty(\text{st})$
C ^{PS} (but horn), C ^H	$x_1^* > 0, x_2^* \approx 0$	Quasi-CAM (possibly pure)	The network gets stucked into ξ_1 ; the history of the system stops here
C ^{PS} (horn)	$x_1^* = 0, x_2^* > 0$ $x_1^* + x_2^* \approx 1$		Complete blurring occurs as a number of patterns are (partially) recovered and super-imposed

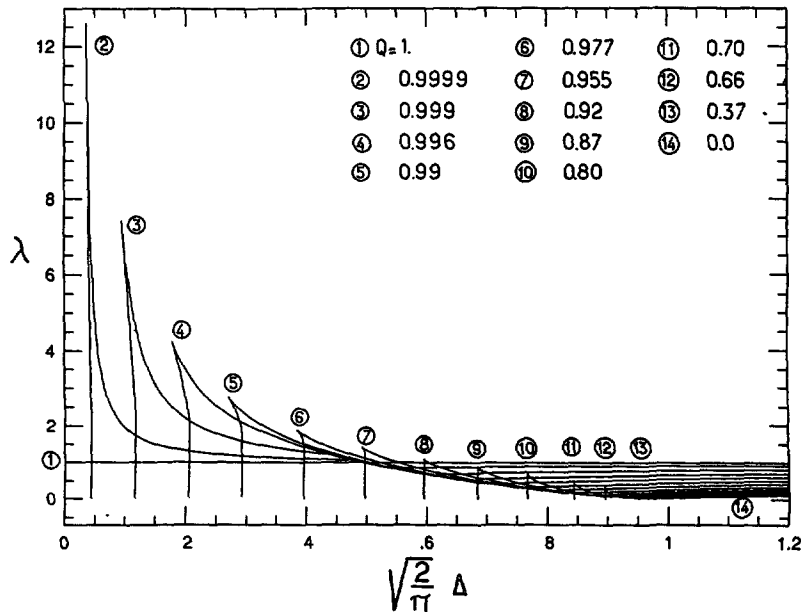


Fig. 7. — Pseudoinverse rule (with $\vartheta = 1$) : $\lambda_{\infty, Q}$ versus $\sqrt{2/\pi} \Delta$ for the various values of Q indicated within the figure. The curves to which λ_{III} contributes are recognizable in that their descending branch has two kinks ; roughly speaking it concerns all those with $Q > 0.99$.

5. Long-term dynamics ;

The easiest way to determine the phases of the long-term dynamics is to consider the implicit equation for Q^* , the asymptotic value of Q_μ . We no longer need, here, to restrict the analysis to any particular learning rule since the original set of equations (14) yields directly

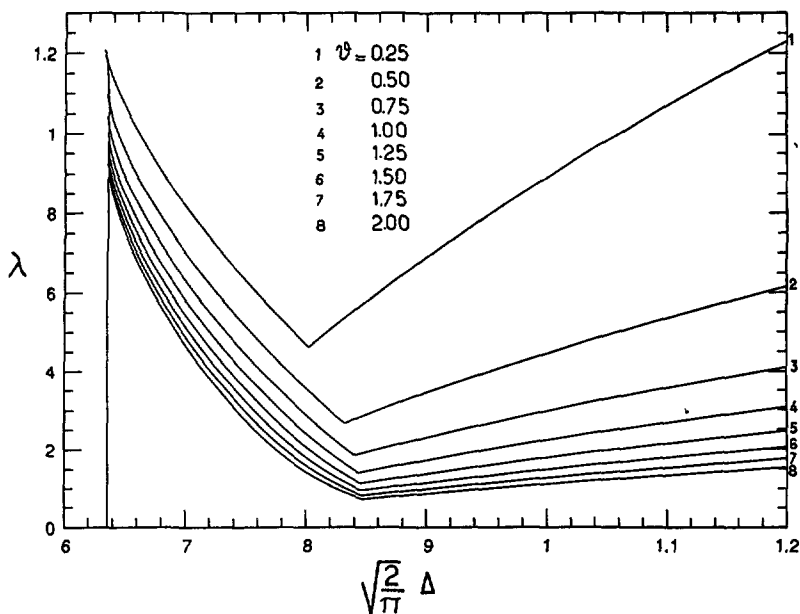


Fig. 8. — Pseudoinverse rule : $\lambda_{\cdot, 0.9}$ versus $\sqrt{2/\pi} \Delta$ for various values of ϑ .

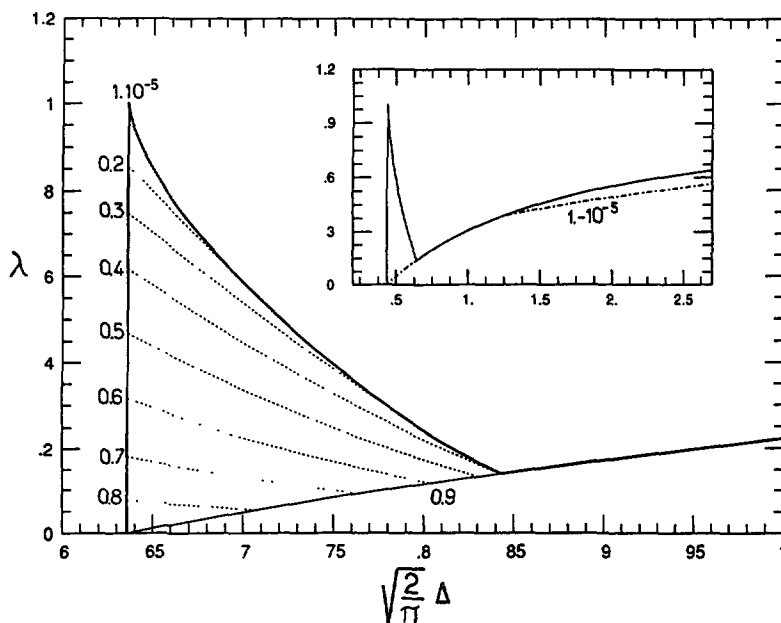


Fig. 9. — Pseudoinverse rule (with $\vartheta = 1$): same as figure 3. The full curve is $\lambda_{\cdot, 0.99}$ versus $\sqrt{2/\pi} \Delta$ and various « iso- x_1^* » are drawn within C, their values being those indicated within the figure. Within the so-called horn, those iso- x_1^* curves are iso- x_2^* curves, as well, because of the relation $x_1^* + x_2^* \approx 1$ (that relation holds everywhere up to $\sqrt{2/\pi} \Delta \approx 2$ or so but is trivial outside the horn). At some point they all fuse with the iso-0.9 curve, which is none other than λ_{\cdot} . On the other hand, the iso- x_1^* plotted within the insert coincides with it, too, for quite a substantial part of its path; this merely exemplifies the fact that, to all practical ends, $x_1^* \approx 1$ throughout C (with the exception of the horn).

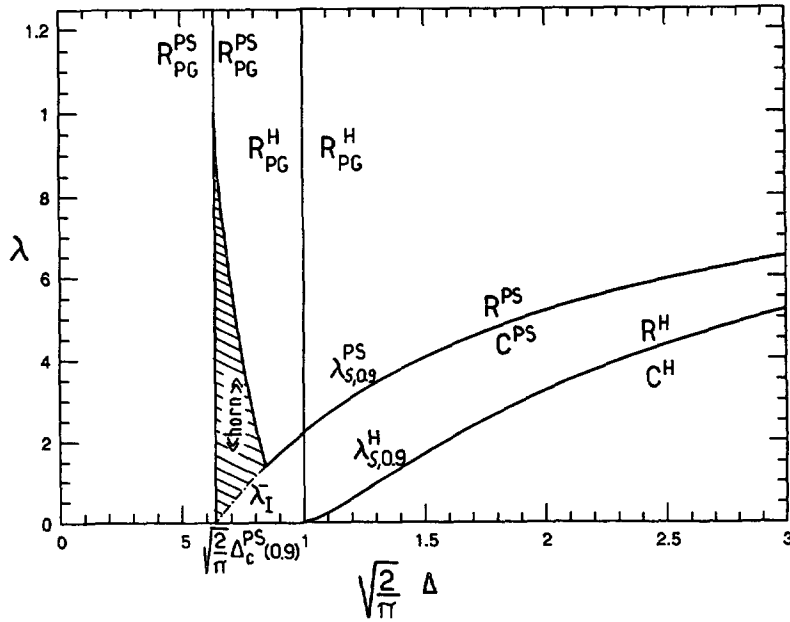


Fig. 10. — Comparison of the Hebb and Pseudoinverse rules : $\lambda_{s,0.9}^H$ and $\lambda_{s,0.9}^{PS}$ (for $\vartheta = 1$) versus $\sqrt{2/\pi} \Delta$, with the identity of the various regions of the phase-space indicated.

Q^* as the fixed-point of F , the mid-term equation of evolution. Q^* is thus defined by :

$$Q^* = F(Q^*) = \int \rho(\Delta) d\Delta \int Dz \tanh \beta [\Delta Q^* + z \sqrt{1 - (Q^*)^2}] \quad (34)$$

where we have used the same notations as above and assumed that $\rho(\Delta)$ depends neither on the spin nor on the pattern indices i and μ . We caution against the fact the sum in the definition (23) of Δ does not refer to $\Delta_{i,\mu}^S + \lambda \Delta_{i,\mu}^T$ but to $\Delta_{i,\mu+1}^S + \lambda \Delta_{i,\mu}^T$. Since both ϑ and λ disappear from the long-term equation, Δ is the only parameter and the situation is thus analogous to the zero- λ case ; a glance at (19) would even reveal that (34) is precisely the form that (19) would take if λ were set to zero. Hence, in agreement with our previous assertions, once expressed in terms of the proper parameter, the long-term dynamics becomes a faithful copy of the (necessarily short-term) dynamics of the corresponding zero- λ network. Expliciting (34) for each case at zero-temperature yields :

$$F_H(Q^*) = \text{erf} [\Delta Q^* / \sqrt{2} \sqrt{1 - (Q^*)^2}] \quad (35a)$$

$$F_P(Q^*) = \text{erf} [\langle \Delta \rangle Q^*]; \quad \langle \Delta \rangle = (1 + \lambda) / \sqrt{\alpha(1 + \lambda^2)} \quad (35b)$$

i.e. the same equations as those defining Δ_c in each case.

The equations of the long-term separatrix (telling R_{FG} from R_{PG}) are thus simply $\Delta = \Delta_c^H(0) = \sqrt{\pi/2}$ (whatever the value of Q) for the Hebb rule and $\Delta = \Delta_c^{PS}(Q)$ for the Pseudoinverse rule (see Appendix B for details). That particular Hebbian frontier enjoys the property of being the only true thermodynamical frontier and second-order separatrix in this paper : all other ones are of first-order type and the (sub-) regions they delimitate are not true thermodynamical phases.

6. Mid-term dynamics.

Let us recall that the purpose of this section is to check for a sudden disruption of the sequential regime within R. Because it is discussed here for the first time, full details are given in the text and there is no corresponding section in the Appendix.

6.1 HEBB RULE. — Here is the case alluded to in the overview : Q^* is the limit of a decreasing series of Q_μ 's while λ_s increases with decreasing Q 's ; there is then a risk that some portion of the original separatrix $\lambda_s(\Delta, Q)$ might fall below one of the successively arising $\lambda_s(\Delta, Q_\mu)$. Thus it will not be safe unless there is no intersection of $\lambda_s(\Delta, Q^*(\Delta))$ with $\lambda_s(\Delta, Q)$, which a little geometry would show can not happen but in the first part of λ_s , hence the curve :

$$\lambda_{\text{mind}} = \lambda_1^-(\Delta, Q^*(\Delta)) \quad (36)$$

(see Fig. 4). The condition for a persistent sequential regime reads : $\lambda > \lambda_{\text{mid}}(\Delta)$ and the variable Q has been eliminated as it should. The region of transient sequential regime is actually a very thin stripe running close to $\lambda_{s,1}$ and which all $\lambda_{s,Q}$ must cross by definition ; for any given Q , the small patch of R_{FG} comprised between $\lambda_{s,Q}$ and λ_{mid} must be discarded in order to obtain the region of persistent (or stable) sequential regime. The interesting point here is not so much the (negligible) extent of the loss than its very existence since, by an argument of continuity in the phase-plane, this fact rules out the existence of a corresponding stripe in C that would turn sequential in the long-run. *A contrario*, it suggests that the region C is somehow more stable, in the long run, than R_{FG} . This lends support to our former finding that C should correspond, to all practical ends, to a stable CAM mode, with the network locked into ξ_1 .

6.2 PSEUDOINVERSE RULE. — Here is the converse case. A graph of $Q^*(\Delta)$ (which has three branches) would show that the Q^* that is actually reached is always either 0 or 1 (there is also an intermediate fixed-point in some cases but it is unstable) ; hence within the R region the series Q_μ is an increasing series running from Q_1 to unity. A glance at figure 6 would then show that the region that must be subtracted from $R_{\text{FG}}(Q)$ is its intersection with $R_{\text{FG}}(1)$ or, in other words, the whole portion of $R_{\text{FG}}(Q)$ situated below $\lambda_{s,1}$ (whose equation is imply $\lambda = 1/\vartheta$), which is a much larger patch than formerly. As a result, the definitive equation of λ_s may be considerably simplified if $Q < 0.9$ (i.e. the value of Q below which, at $\vartheta = 1$, $\lambda_s(\Delta) < 1$ whatever the value of Δ) : in that case $C \cup R_{\text{FG}}^{\text{H}}$ is simply the $1/4$ plane $\{(\lambda, \Delta)/\lambda < 1/\vartheta ; \Delta < \Delta_C^{\text{PS}}(Q)\}$ and R^{st} the remainder. Practically, this restricts the role of the horn to the high- Q range, i.e. $\sim 0.9 \leq Q \leq 1$.

Evidently the situation is not so clear as regard R_{PG} . there might exist an intricate tessellation of transient and stable patches or, just as well, a very simple pattern in the line of what happens within R_{FG} . We do not think that this point really deserves further study : from a practical point of view, the interesting case is clearly the former one.

7. Discussion.

However strikingly different their geometrical outlook, the phase diagrams obtained for both rules are qualitatively the same, the Pseudoinverse being characterized only by a richer but also more clear-cut behaviour. Indeed, first, perfect recall is possible with that rule : it happens only for Q or $\lambda = 1$ within region C because it is then a short-term recall of the first pattern but the long-term recall within R (if any) is always perfect ; since the series Q_μ should converge rapidly, this is a significant feature of the rule. Also noteworthy is the dramatic dependance on

Q (both quantitative and qualitative) of the Pseudoinverse phase-diagram, in clear contrast with the mere 20 % or so increase of λ_s^H as Q is lowered from 1 to zero. The most important difference, however, concerns the extent of the various non-sequential regimes (broadly defined) but an assesment of the relative performances of the two rules in this field is not easy.

Everything considered the definitive C^{PS} , already much larger than the short-term C^H , does more than double its area when it is fused with the so-called transient and evanescent sequential regimes, which, from a practical point of view, are not truly sequential. This is the case, in particular, of the whole stripe $\{\lambda < 1/\vartheta\}$. A shift from the Hebb to the Pseudoinverse looks therefore detrimental ; the relative loss, however, is partially compensated by a gain in the leftmost part of the diagram. Indeed, the useless region on the left is now $\{0 < \Delta < \Delta_c\}$ which, given the definition (28) and (31) of Δ_c^{PS} and Δ_c^H , implies the (at least partial) gain of the whole stripe $\{\Delta_c^{PS}(Q) < \Delta < \sqrt{\pi/2}\}$. Now, this is far from being a negligible adjunction since it concerns the low- Δ range, i.e. precisely those values of Δ which are always most easy to obtain.

That, however, is not the whole story for another, equally relevant way of assessing R_{FG}^{st} is to reason in (α, λ) coordinates. Here there is no choice but to take our definitions at face value. Since the crude definition (20) of the Pseudoinverse implies $\Delta^S(\alpha) = \Delta^T(\alpha) = \sqrt{1/\alpha - 1}$, whence $\Delta^{PS}(\alpha) = [(1 + \lambda)/\sqrt{1 + \lambda^2}] \sqrt{1/\alpha - 1}$, high values of Δ (especially for $\lambda > 1$) translate into low values of α . Hence some compromise between high α 's and high Δ 's will be necessary and only the leftmost part of R_{FG}^{st} will display interesting features. The better definition (21), on the other hand, would yield $\Delta^{\langle \text{true} \rangle PS}(\alpha) = \sqrt{(2/\alpha) - 1}$, thus doubling the range of available values of α but leaving the antagonistic relation between α and Δ otherwise unchanged. By contrast the situation in the Hebbian case is a more mixed one : since $\langle \Delta^{Hebb}(\alpha) \rangle = \alpha^{-1/2} [(1 + \lambda)/\sqrt{1 + \lambda^2}]$, high values of Δ do not imply low ones for α but, on the other hand, no accurate control of Δ is possible (since it obeys a Gaussian distribution) and the resource of tuning Q is also denied to us since the frontier of the forbidden left stripe is then always $\Delta = \sqrt{\pi/2}$.

What finally emerges from that first survey is a conclusion already familiar from similar studies in other contexts. Indeed the relative superiority of our extended Pseudoinverse rule over the extended Hebb rule concern the core of the model, i.e. the true (long-term) sequential regime of R_{FG}^{st} . They are somehow bought at the expense of its performance in other regimes ; in particular a whole region of the diagram : the horn, is totally unsuited to either CAM or sequence-processing and is peculiar to the Pseudoinverse case. Hence, whenever the more elaborate rule of the two : the Pseudoinverse, performs better, it does so in a very « focused » way, so to say : in this context for high values of Q or low values of either α or Δ . By contrast the Hebb rule performs more homogeneously, i.e. is more robust, even while its « peak » performances are less promising. On the other hand, it should be noted that the two models differ little with respect to λ : in both cases the higher its value the better.

An additionnal lesson to be drawn from the study is the very special role played by $\lambda = 1/\vartheta$; in networks of ϑ also unity, this means that $\Delta^T/\Delta^S = \|\mathbf{J}_i^S\|/\|\mathbf{J}_i^T\| = 1$ is the gross equation of the R/C separatrix (which becomes exact in the limit of infinite Δ 's and/or — for the PS rule — of high Q 's or vanishingly small — for instance finite — storage capacities). In other words, the condition for the onset of a sequential regime would be, more or less, that the transitionnal stabilities take over their stabilizing counterparts, a rather intuitively evident conclusion within the present framework. That, besides, is the very reason why we determined ourselves in favour of a time-delayed model ; if we had used instead an instantaneous model, then by the very definition of those, a whole series of transitional stabilities « of

ν th order » :

$$\Delta_{i,\mu}^{\tau,\nu} = \xi_{i,\mu} \mathbf{J}_i^T \xi_{\mu+\nu} / \|\mathbf{J}_i^T\| \quad \mu, \nu = 1, \dots, p \quad (37)$$

would have been called for and, thus making intuitive-looking conclusions considerably more difficult to reach.

Regarding now more specifically the short-term dynamics, it comes rather as a surprise that, everything considered, the adoption of a Pseudoinverse-type rule should also enhance the performances of the network in the short-term regime. This was not expected since both (21) and (22) are extrapolations of a rule originally designed to enhance the CAM region of a zero- λ network (which besides it does here, *via* Δ_C^{PS}) which in this in this context, as explained before in section 3.2, corresponds to the long-term sequential retrieval properties. One must be careful here as it might an undesirable consequence of using the crude definition (20) instead of the better one (21). We do not think so, however, for numerical simulations on (21) and analysis of the resulting distributions of Δ^S and Δ^T showed that, for $\alpha < 1$ and to within 1 %, just $\lambda = \vartheta = 1$ did hold. For $\alpha > 1$, some of the Δ^S and Δ^T became of course negative but the former relation was still obeyed by most. Hence we feel rather confident that even the rather awkward definition we have used in order to obtain an exact phase diagram does not limit the validity of our approach and that equation (27b) is a tolerably good representative of its class.

That observation need not detain us from looking for yet better rules, designed more specifically in view of improving the long-term dynamics. Such rules (for instance (22)) will have to be based on the definition of the global stability Δ and not on those of its components Δ^T and Δ^S ; the problem, then, will be that very few rules 1) have workable stability distributions (remember that the short-term equations require knowledge of the individual distributions of Δ_μ^S and Δ_μ^T) and 2) make the system exactly solvable (which requires, amongst other things, the uncoupling of the equations of evolution for x^+ and x^-). For rules defined directly after Δ , the first point will undoubtedly be the more troublesome; on the other hand all rules leaving more than one integral in (11) will fail to meet the second point, even while they may meet the first, thereby making the short- and mid-term equations of evolution intractable and leaving only the long-term dynamics open to the investigation.

7. Conclusion.

The present study has revealed in the phase-space a richer structure than had previously been possible and has confirmed the merits of strongly-diluted models for the solution of even quite complicated network designs. It has also, as anticipated in the introduction, given a more explicit content to the short/long-term duality that is a peculiar feature of such models with the definition of a transitional counterpart (17b) to the usual stabilities (1) or (17a).

Another interesting result is that we have succeeded in deriving from a simple and exactly solved model a list of parameter governing the dynamics of a sequence-retrieving network. That list comprises Q , Δ , λ and ϑ , i.e. the initial magnetization on the first pattern, a global stability parameter (the analogue for a sequence-processing network of the usual stability of CAM models) and two coupling parameters measuring the relative strength of the transitional and stabilizing terms, the first (λ) in terms of the coupling matrices, the second (ϑ) in terms of the two classes of (normalized) stability parameters. As discussed in the text, what evidence we have obtained concerning ϑ is a little stretched but there is in fact numerical evidence [27], drawn moreover from a more realistically connected version of the present model, that the ratio $\vartheta = \Delta^T/\Delta^S$ does indeed play the role we assigned to it in this study. The above list of parameters is therefore a convenient framework within which to assess and discuss the performance of this and related models as various rules and, possibly, network

architectures, are tried. In that respect the essence of our results is that both ϑ and Δ should be maximised. More precisely :

1) Just as $\Delta^S(\alpha)$ must be maximized with respect to the storage capacity α in an ordinary CAM network, here it is $\Delta(\alpha)$ which should be maximized with respect to α .

2) High values of ϑ seem desirable for the short-term dynamic, probably because they facilitate the « microscopic » dynamics of the transitions.

3) Psychophysiological analogies strongly suggest that the ability to store sequence of patterns is bought at the expense of the capacity to retrieve them one by one (try to sing your favorite song starting by the middle : it will unfold well once a verse has been recalled but the individual verses are usually difficult to recall from scratch) ; hence we would like to see if Δ^S can not be maximized together with Δ or ϑ . Here for instance it is clear that Δ as defined in (22) (hence $\alpha_{\max} = 2$) can not be maximized for $1 < \alpha < 2$ without some of the Δ^S and Δ^T becoming negative.

As an alternative (or a preparatory step) to the search for particular, explicit learning rules, a most convenient way of assessing the feasibility of this program and of obtaining a measure of the relative cost of the maximisation of each parameter would be to undertake calculations in the space of couplings, in the line of previous work by Gardner. Additionnally, it should not be impossible in this way to understand why $\lambda = 1$ and $\vartheta = 1$ are such privileged values. We have just performed a number of such calculations, combining different sets of constraints, and will present the results shortly in what should be regarded as a companion paper [28] to the present one.

Acknowledgments.

Both authors acknowledge financial support of the EEC Initiative BRAIN, Contracts Nr. ST2J-0422-C(EDB) (J.-P. Nadal) and ST25-0321-C(EDB) (O. Lefèvre). The latter author also wishes to thank the Département de Physique of the Ecole Normale Supérieure for its hospitality during part of this work as well as Pr. Giorgio Parisi, through whose kind mediation he obtained his funds.

Appendix A.

General equations of the dynamics.

Note that this section makes use of non-normalized stabilities, i.e. of $\|J_i^X\| \Delta_{\mu,i}^X$ ($X = S, T$) instead of $\Delta_{\mu,i}^X$. If, in the text, the stabilities were introduced directly in normalized form, it was in order to keep the number of definitions down to a minimum. If we define m^+ and m^- in the same way as x^+ and x^- in the text then from the definition of the dynamics (3)-(6) we get :

$$\langle m^\pm(t) \rangle = 1/N \sum \langle (\pm \xi_{i,1} + \xi_{i,2}) S_i(t+1) \rangle \quad (A1)$$

$$= \pm 1/N \sum_i \langle \tanh [\beta \xi_{i,1} h_i(t)] \rangle + 1/N \sum_i \langle \tanh [\beta \xi_{i,2} h_i(t)] \rangle \quad (A2)$$

$$= \pm \langle m_1(t) \rangle + \langle m_2(t) \rangle \quad (A3)$$

because $\xi_{i,\mu}$ and $S_i(t+1)$, but for their sign, are always unity and the hyperbolic tangent is uneven. As in the text, we have denoted by $\langle \rangle$ the ensemble average, not to be confused with the thermal average $\langle \rangle$. The last formula may then be rewritten as a probability

distribution provided we know the first and second moments of $\xi_{i,1} h_i(t)$ and $\xi_{i,2} h_i(t)$. As to the first moments we have :

$$\langle \xi_{i,1} h_i(t) \rangle = x_1(t) \Delta_{i,1}^S + \varepsilon x_2(t) \Delta_{i,2}^S + \varepsilon Q \Delta_{i,1}^T \quad (\text{A4})$$

$$\langle \xi_{i,2} h_i(t) \rangle = \varepsilon x_1(t) \Delta_{i,1}^S + x_2(t) \Delta_{i,2}^S + Q \Delta_{i,1}^T \quad (\text{A5})$$

where by ε_i we denote $\xi_{i,1} \xi_{i,2}$; we have dropped the spin indices since all ε_i (and all $\xi_{i,\mu}$ as well) have the same probability distribution :

$$P(x) = 1/2(1+k)\delta(x-1) + 1/2(1-k)\delta(x+1) \quad (\text{A6})$$

where the value of k is a if x stands for $\xi_{i,\mu}$ and $q = a^2$ if it stands for ε_i . Similarly and whatever μ ($\mu = 1, 2$) we obtain for the second moments :

$$\langle [\xi_{i,\mu} h_i(t)]^2 \rangle - \langle \xi_{i,\mu} h_i(t) \rangle^2 = \|\mathbf{J}_i^S\|^2 [1 - (x_1(t) + \varepsilon x_2(t))] + \|\mathbf{J}_i^T\| (1 - Q^2) \quad (\text{A7})$$

$$= [\sigma_\varepsilon(t)]^2 \quad (\text{A8})$$

where we have used the fact that, under the hypothesis of strong dilution :

$$\langle S_i(t - \tau) S_j(t) \rangle = \delta_{\tau,0} \delta_{i,j} \quad (\text{A9})$$

(remember that we use discrete time). The validity of (A9) for $C \ll \ln N$, where C is the connectivity of the network, is proved in reference [17] where the trick was introduced. Finally :

$$m_1(t) = \int \cdot \int \text{DZ} \sum_{\varepsilon=\pm 1} P(\varepsilon) \tanh \beta [\Delta_1^S x_1(t) + \varepsilon \Delta_2^S x_2(t) + \varepsilon Q \Delta_1^T + z \sigma_\varepsilon(t)] \quad (\text{A10})$$

$$m_2(t) = \int \cdot \int \text{DZ} \sum_{\varepsilon=\pm 1} P(\varepsilon) \tanh \beta [\varepsilon \Delta_1^S x_1(t) + \Delta_2^S x_2(t) + Q \Delta_1^T + z \sigma_\varepsilon(t)] \quad (\text{A11})$$

where :

- $\int \cdot$ stands for the other integrals (the same ones as in (9) ;
- we have summed over equivalent spins ;
- the $\xi_{i,\mu} h_i(t)$ are treated as Gaussian variables (which follows from (A6) and the law of large numbers.

Given that the respective definitions of $m^\pm(t)$ and $x^\pm(t)$ imply

$$m^\pm(t) = (1 \pm q) x^\pm(t) \quad (\text{A12})$$

equation (10) of the text follow in a straightforward way. The usefulness of the shift from m^\pm to x^\pm was to get rid of the irrelevant factor $(1 \pm q)$ in the equations of the dynamics.

Appendix B.

Detailed equations of the separatrices.

SHORT-TERM DYNAMICS.

(a) *Hebb rule.*

Let us recall that the problem is to find the couples (λ, δ) for which the intervall

$[Q, -Q]$ is free of fixed points. The equation of evolution, f , being an homeomorphism with positive first derivative on that interval, we need not inquire beyond the first-order fixed-points. Ideally (and such is the case here) there is a unique positive fixed-point (whence x^+), so that the critical $\lambda(\delta)$ corresponds either to the disappearance of the two negatives fixed-points, one (the lowest-lying) stable, the other unstable, or to the moment (as δ is increased) where the latter goes below $-Q$. Such is the definition of λ_s in the present case. We must therefore consider the curves defined by :

$$\lambda_{I,Q}(\delta) : \begin{cases} x^* < 0 \\ f(x^*) = x^* \\ f'(x^*) = 1 \end{cases} \quad (\text{B1})$$

$$\lambda_{II,Q}(\delta) : f(-Q) = -Q \quad (\text{B2})$$

whose equations are :

$$\lambda_{I,Q}(\delta) = [\sqrt{2 \text{Ln}(\sqrt{2/\pi} \delta)} \delta - \text{erf}[\sqrt{2 \text{Ln}(\sqrt{2/\pi} \delta)}]] , \quad \delta \geq \sqrt{\pi/2} \quad (\text{B3})$$

$$\lambda_{II,Q}(\delta) = 1 - \Delta_C^H(Q)/\delta , \quad \delta = \Delta_C^H(Q) \quad (\text{B4})$$

λ_{II} is always smaller than λ_I^- , equality occurring only at $\delta(Q)$: it has to be so since δ marks the point where the two negative fixed-points coalesce and disappear (as either δ or λ is increased); its physical meaning is thus to tell the stable from the unstable point with respect to $-Q$. On the other hand $\partial f/\partial \lambda$ is always positive, whatever x and δ . Finally (see Fig. 2) :

$$\begin{aligned} \lambda_{s,Q}(\delta) &= \lambda_{I,Q}(\delta) , & \delta < \delta(Q) \\ &= \lambda_{II,Q}(\delta) , & \delta > \delta(Q) . \end{aligned} \quad (\text{B5})$$

Each curve contributes for one of the Q -independent endpoints $(0, \sqrt{\pi/2})$ (belonging to λ_I^-) and $(1, +\infty)$ (belonging to λ_{II}^-). Since λ_I^- and λ_{II}^- are tangent at $\delta(Q)$ and are everywhere differentiable, so is λ_s , even while it is a composite curve (which numerically obtained graphs could therefore not reveal).

It is interesting to note that the extreme curves $\lambda_{s,0^+}$ and $\lambda_{s,1}$ (and these alone) are not composite for they are entirely given by $\lambda_{I,0^+}$ and $\lambda_{II,1}$, respectively, the balance shifting progressively from the former to the latter. Note also that the $Q=0$ limit reveals a (slight) pathology of the model, namely that $\lambda_{s,0^+} \neq \lambda_{s,0}$. Indeed, while f no longer shows any dependance on λ at zero Q (the very notion of a sequential retrieval is then meaningless, besides) the equations defining λ_s . (B3)-(B5), on the other hand, do have a limit when Q approaches zero, which is therefore the natural way to handle the pathology. The same will be true of the Pseudoinverse rule.

(b) *Pseudoinverse rule.*

We will follow the same line of reasoning but will have to take into account the following additional complications with respect to the former case :

1) There may be more than one positive fixed-point, whence the need to supplement the curves $\lambda_{I,Q}$ and $\lambda_{II,Q}$ (defined as above, with f replaced by g) with their positive counterparts $\lambda_{I,Q}^+$ and $\lambda_{II,Q}^+$ defined, for the former, as in (B1) (but where $x^* < 0$ has been replaced by $x^* > 0$ and f by g) and by $g(Q) = Q$ for the latter.

Both λ_{II}^- and λ_{II}^+ can be expressed, in the line of (B4), in such a way as to confine the dependance on Q to a quantity $\Delta_C^{\text{PS}}(Q)$ defined in the text. More precisely, their expressions

follow from the inversion of :

$$\delta_{II,Q}^+(\lambda) = \Delta_C^{PS}(Q) \sqrt{1 + \lambda^2} \quad (B6)$$

$$\delta_{II,Q}^-(\lambda) = \delta_{II,Q}^+(\lambda) (1 + \vartheta \lambda) / (1 - \vartheta \lambda). \quad (B7)$$

2) A sign reversal of the first derivatives of g , namely $\partial g / \partial \lambda$ and $\partial g / \partial x$, may no longer be ruled out, thus undermining our simple geometrical reasoning. A further consequence of the same fact will be that $g(x)$ will be able to undergo two sign reversals over $[-Q, Q]$ instead of one ; in particular the sign of $g(-Q)$ will now depend on λ (whereas $f(-Q)$ was always negative in (a)) and will sometimes be positive. As a consequence of that, we have to take into account the possibility of its being a (second-order) fixed-point, too, and append on our list a fifth and last curve $\lambda_{III,Q}$ defined by $g[g(-Q)] = -Q$.

Clearly, given the last fact and as many as five curves competing to build up λ_s , breaking the (λ, δ) plane into a dozen patches or so, a purely geometrical reasoning, though still possible, would be rather tedious and not without risks as well. A rigorous determination of the separatrices is nonetheless still possible : it rests on the fact that the behaviour of g with respect to its fixed-points and the points $(-Q, -Q)$, $(-Q, g(-Q))$ must be homogeneous inside each of these patches. It will be enough, therefore, in order to assign each of them to either R or C, to pick up a single point out of it and to test it numerically ; the completion of the process will reveal the frontier. Provided the qualitative regimes of the model are clear-cut (which we have shown is the case) and the test-point lie well inside their respective patches, the lack of rigour incurred by such a « mixed » method will be negligible and only some (rather irrelevant) geometrical understanding will be lost. The detailed anatomy of two qualitatively different cases is displayed in figures 5 and 6.

As δ is increased, λ_s turns out to espouse the paths of λ_{II}^+ , then (possibly) of λ_{III} , followed by λ_I^- and finally by λ_{II}^- , the (entirely) missing contribution from λ_I^- is somehow replaced by the « horn » made up with the first two or three segments (remember that in λ_s^{Hebb} , the first contribution was that of λ_I^- , followed by that of λ_{II}^-). That part of λ_s is therefore a peculiar feature of the Pseudoinverse rule. So, too, is the need for λ_{III} . Although its characteristics have been but superficially investigated, its very existence seems limited to the high- Q range and its actual contribution to λ_s to yet higher values ; practically (for $\vartheta = 1$) it sets in around 0.8, by which time it is still lying too high and far away from the other curves to contribute, which it does starting from circa $Q = 0.99$ (see Fig. 7 and caption).

We also note that, here as before, the rule that several curves contribute to the actual R/C separatrix finds exceptions in the two limit-curves $\lambda_{s,0^+}$ and $\lambda_{s,1}$, both of which are here of type λ_{II}^- throughout. $\lambda_{s,1}$, besides, has the exceptionally simple equation ($\lambda = 1/\vartheta$), which can be interpreted as a flattening out of λ_{II}^- (and λ_{III}) combined with the progressive irrelevance of λ_{II}^+ and λ_I^+ . The following fact is also worth mentioning : if we consider $R_{FG}^r \cup C$ (i.e. the really meaningful definition of C), we note that it is the union of the actually infinite but constant stripe $\{\lambda < 1/\vartheta\}$ and of a portion of the horn. That latter part has zero area for $Q < 0.9$ (approximatively) and for $Q = 1$; hence, if C denotes the area of $R_{FG}^r \cup C$, then $\partial C / \partial Q$ must change sign at some Q_C close to unity and, from positive, turn negative. Although this is certainly a minor point, it gives a feeling of how unnatural that rule can be in some respects.

LONG-TERM DYNAMICS. — Because with the Hebb rule F in (35) is a convex function of its argument, with only one positive non-zero fixed-point, the equation of the R_{FG}/R_{FG} separatrix is simply $\partial h^{\text{Hebb}} / \partial Q |_{Q=0} = 1$, i.e. $\Delta = \Delta_C^H(0) = \sqrt{\pi/2}$ (the more correct definition

$\partial Q_{\mu+1}(Q_{\mu})/\partial Q_{\mu} |_{Q_{\mu}=0}$ would have yielded the same result). The situation is very different with the Pseudoinverse rule : $Q^* = 1$ is always a fixed-point but there is a second, unstable positive fixed-point if $\Delta < \sqrt{\pi/2}$. This implies that in the range $[0, \sqrt{\pi/2}]$ the « good » fixed-point $Q^* = 1$ can still be accessed provided that Q is larger than the intermediate one and the desired equation is therefore $\Delta = \Delta_C^{PS}(Q)$. It is because of this dependance on Q that it is not a thermodynamical frontie.

References

- [1] HOPFIELD J. J., *Proc. Nat. Acad. USA* **79** (1982) 2554 and **81** (1984) 3088.
- [2] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. Lett.* **55** (1985) 1530 ; *Ann. Phys. NY* **173** (1987) 30 ;
AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. Lett.* **55** (1985) 1530.
- [3] PERETTO P. and NIEZ J. J., *Disordered Systems and Biological Organization*, E. Bienenstock Ed. (Springer Verlag, Berlin, 1987).
- [4] DEHAENE S., CHANGEUX J.-P. and NADAL J.-P., *Proc. Nat. Acad. USA* **84** (1987) 2727.
- [5] METZGER Y. and LEHMANN D., *Network* **1** (1990) 169.
- [6] SOMPOLINSKY H. and KANTER I., *Phys. Rev. Lett.* **59** (1988) 2861.
- [7] KLEINFELD D., *Proc. Nat. Acad. USA* **83** (1986) 9469.
- [8] GUTFREUND H. and MÉZARD M., *Phys. Rev. Lett.* **61** (1988) 235.
- [9] HERZ A., SULZER S. B., KUHN R. and VAN HEMMEN J. L., *Europhys. Lett.* **7** (1988) 663 ;
RIEDEL U., KUHN R. and VAN HEMMEN J. L., *Phys. Rev. A* **38** (1988) 1105.
- [10] AMIT D. J., *Proc. Nat. Acad. USA* **85** (1988) 2141.
- [11] GUYON I., PERSONNAZ L., NADAL J.-P., DREYFUS G., *Phys. Rev. A* **38** (1988) 6872.
- [12] BUHMAN J. and SCHULTEN K., *Europhys. Lett.* **4** (1987) 1205 ; preprint Technische Universität München (1990).
- [13] COOLEN A. C. C. and RUIJGROK Th. W., *Phys. Rev. A* **38** (1988) 4253.
- [14] RIEGER H., SCHRECKENBERG M. and ZITTARTZ J., *Z. Phys. B* **72** (1988) 523.
- [15] SHIINO M., NISHIMORI H. and ONO M., *J. Phys. Soc. Jpn* **58** (1989) 763 ; preprint Tokyo Institute of Technology (1990).
- [16] AMARI S., *IEEE Trans. C* **21** (1972) 1197.
- [17] DERRIDA B., GARDNER R. and ZIPPELIUS A., *Europhys. Lett.* **4** (1987) 167.
- [18] FORREST B. M., *J. Phys. A* **21** (1988) 245.
- [19] KRAUTH W., MÉZARD M. and NADAL J.-P., *Complex Systems* **2** (1988) 387.
- [20] KEPLER T. B. and ABBOTT L. F., *J. Phys. France* **49** (1988) 1657.
- [21] ABBOTT L. F. and KEPLER T. B., *J. Phys. A* **22** (1989) 2031.
- [22] PERSONNAZ L., GUYON I. and DREYFUS G., *J. Phys. Lett. France* **46** (1985) 1359.
- [23] GESZTI T., *Physical Models of Neural Networks* (World Scientific Publishing Co., Singapore and London, 1989).
- [24] AMIT D. J., *Modelling Brain Function : The World of Attractor Neural Networks* (Cambridge University Press, 1990).
- [25] HERTZ J., KRUGH A. and PALMER R. G., *Introduction to the Theory of Neural Computation* (Addison-Wesley, Cambridge (Mass.), 1990).
- [26] PERETTO P., *Modelling Neural Networks* (Les Editions de Physique, Les Ulis, 1990).
- [27] SOMPOLINSKY H., unpublished data.
- [28] LEFÈVRE O. and NADAL J.-P., to be submitted to *J. Phys. A*.