UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE FÍSICA
CAIXA POSTAL 20516
01452-990 SÃO PAULO - SP
BRASIL

PUBLICAÇÕES

IFUSP/P-1080

DREAMING IN ANALOG ATTRACTOR NEURAL NETS

Silvia M. Kuva and **Nestor Caticha**
Instituto de Física, Universidade de São Paulo

Outubro/1993

DREAMING IN ANALOG ATTRACTOR NEURAL NETS

Silvia M. Kuva and Nestor Caticha
Instituto de Física, Universidade de São Paulo
CP 20516, 01452-990 São Paulo, SP, Brazil

Abstract

The removal of spurious minima in attractor neural nets by reverse Hebbian learning has been called "dreaming". This self-organized process leads to an increase in the storage capacity and has been previously studied numerically at "zero temperature" (infinite analog gain) (Hopfield *et al.* 1983, Van Hemmen *et al.* 1989). Approximate differential equations are proposed to describe the evolution of quantities related to the mean stabilities under the dreaming dynamics; their solutions are consistent with the results of numerical simulations of the analog attractor neural net. The phase diagram in the $(\alpha, \beta^{-1}, \tau)$ space (where α is the ratio of patterns to neurons, β is the analog gain, and τ the dream load), obtained numerically, shows a large increase in the storage capacity with respect to the original net with the simple Hebbian prescription, even at moderately high temperatures . A modification of the process, where the norm of the coupling matrix is kept fixed, is also studied. The retrieval region can thus be increased even further.

1 Introduction

The reason for dreaming is not quite well established. A possible approach, due to Crick and Mitchison [2], suggests that dreaming during the Rapid Eye Movement (REM) sleep might be a process used in weeding out certain undesired cortex modes in mammalia. The related, though independent, idea of improving the performance of attractor neural nets (ANN) [1, 5] by Hebbian unlearning, introduced by Hopfield, Feinstein and Palmer [6] led to the concept of "dreaming" in neural nets.

Although beautiful, this idea might not turn out to be correct in biological systems, and since technical difficulties to experimentally check its validity are immense, it may take a long time to actually accept it as an explanation for the role of dreams or to discard it. It will certainly lead to controversies far away from our interest here. Nonetheless, its application in the field of neural nets has been numerically established [3, 4, 6] and if the biological validity of the process is not vindicated, a change of nomenclature is all that will be needed to maintain the idea of dreaming in good standing as a technique for cleaning an ANN of undesired spurious minima.

The present understanding of the process rests entirely on somewhat extensive numerical simulations of the Hopfield model at zero temperature. Several characteristics, such as the retrieval of patterns of different activities and the enlarged storage capacity of the net [3, 4] after dreaming, strongly recomend this type of learning and thus further studies of its properties, either from a more analytical point of view or from simulations at finite temperature will be interesting. The main object of this paper is to perform a numerical study of the phase diagram of the analog attractor neural net [7 - 10] after the unlearning has taken place for finite "temperature", more specifically in the $(\alpha, \beta^{-1}, \tau)$ space, where as usual α is the ratio of patterns to neurons, and β is the analog gain. The dream load τ measures the amount of unlearning that the net has undergone. Secondly, in an attempt to understand analytically the dreaming dynamics in a continuous dreaming load approximation, we obtain very simple approximate differential equations to describe the evolution of mean stability-like quantites. A partial solution of these equations, using information from the simulations, helps to understand the evolution of the stabilities of the patterns in some regions of the phase diagram, in agreement with numerical results.

We also introduce a new version of the dreaming algorithm, where the norm of the synaptic matrix is kept fixed, and which is more robust than the simple Hebbian unlearning. The dreaming dynamics can be again partially understood in terms of simple differential equations. This leads to improved results, such as a larger retrieval phase, but it lacks the simplicity and appeal of the former method, as it is no longer local. This opens up the question of whether there are other dreaming algorithms which lead to better performances and what would be the optimal way to dream, that we have not tried to answer here.

2 The Dreaming Dynamics

Although the mechanism of the putative biological process is unknown, the dreaming procedure in neural nets is rather simple. We first describe the model [9] to fix our notation. Consider a set of continuous neural variables S_i, $i = 1, \ldots, N$, and a fast discrete neural dynamics given by

$$S_i(t+1) = f(h_i(t)), \tag{1}$$

where $h_i = \sum_j J_{ij} S_j$, is the local post-synaptic field and f is a sigmoid function. We will use the hyperbolic tangent: $f(h_i) = \tanh(\beta h_i)$. The inverse of the analog gain β^{-1} is related to the temperature T of an analogous stochastic dynamics.

The reason for looking into this model and not into the one with a continuous-time dynamics is simply that, although the structure of the fixed-points is the same, the discrete version is much simpler to simulate.

An appropriate order parameter is the normalized overlap

$$m^\nu = \frac{1}{\|\tilde{m}^\nu\|} \sum_{i=1}^{N} \xi_i^\nu S_i, \tag{2}$$

where $\|\tilde{m}^\nu\| = \sqrt{N \times \sum_{i=1}^{N} (S_i)^2}$ which is N at $T = 0$. Its behavior will be analized in section 3.

The synaptic matrix $\boldsymbol{J}$ is supposed to be fixed at the fast time scale of the dynamics and varies during the (un)learning process at a much slower time scale labeled by d. This means that the evolution of equation (1) is always for fixed $\boldsymbol{J}$, which itself evolves in a slower time scale under a set of $\boldsymbol{S}$ which have had enough time to relax. Through learning, the synaptic matrix is built such that equation (1) has fixed-points at or near desired configurations. These form the 'learning set' $\{\xi_i^\mu = \pm 1\}$ with $i = 1, \ldots, N$ and $\mu = 1, \ldots, P = \alpha N$. Initially the synaptic matrix will be taken to be Hebbian:

$$J_{ij}(d = 0) = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu, \tag{3}$$

which is known [8, 10] to work for $\alpha \lesssim 0.14$ at infinite gain, or less as the gain is decreased. Starting from a random configuration $\boldsymbol{\eta}(0)$, the fast dynamics will relax to

$$\eta_i(t \to \infty) \to \eta_i. \tag{4}$$

This configuration $\boldsymbol{\eta}$ will be called a "dream" and will be used in a process of reverse learning to modify the synaptic matrix [6]

$$J_{ij}(d+1) = J_{ij}(d) - \frac{\varepsilon}{N} sign(\eta_i) sign(\eta_j), \tag{5}$$

where the subtraction is made with the sign of $\boldsymbol{\eta}$, but could be made using $\boldsymbol{\eta}$, with this last choice just leading, under some conditions, to the same results only more slowly. ε is a small parameter which defines the time scale of the slow dreaming dynamics.

The reason that this procedure works efficiently in eliminating spurious minima is that for a net in a non-retrieval phase the configurations $\boldsymbol{\eta}$ are most likely to be spinglass metastable states, which are removed by the reverse Hebbian procedure. There is an exponentially ($\propto \exp(kP)$) large number of these minima, but as discussed by van Hemmen [4], if α is not very large each subtraction destabilizes an exponentially large number of near lying minima which are correlated to some pattern. Only a linear (in P) number of dreams is needed for the net to be rendered functional. For $\alpha(\beta)$ too large the spurious minima are not sufficiently correlated (localized in small clusters) and the process doesn't lead to a retrieval phase. The patterns are destabilized before this. These arguments are not understood from an analytical point of view.

Also we introduce a normalized version, where first the subtraction is performed

$$\tilde{J}_{ij}(d+1) = J_{ij}(d) - \frac{\varepsilon}{N} sign(\eta_i) sign(\eta_j), \tag{6}$$

and then the normalization is restored

$$J_{ij}(d+1) = y \tilde{J}_{ij}(d+1), \tag{7}$$

where

$$y = \sqrt{\frac{\sum_{ij} J_{ij}^2(d)}{\sum_{ij} \tilde{J}_{ij}^2(d)}}. \tag{8}$$

In the average, if the process is started from the simple Hebb prescription (2), the norm of $\boldsymbol{J}$ will be always approximately P , the number of patterns.

An important quantity is K^d

$$K_S^d = \frac{1}{2N} \sum_{ij} J_{ij}^d S_i S_j, \tag{9}$$

related to the mean stability per site of a generic configuration $\boldsymbol{S}$. The difference between (9) and the so-called stability is very small in the high-gain limit (if the $\{S_i\}$ are near to plus or minus one).

K_S^{d+1} is the stability of a fixed-point $\boldsymbol{S}$ after $d+1$ dreams and is given by

$$K_S^{d+1} = K_S^d - \frac{\varepsilon}{2} M_{\eta S}^2, \tag{10}$$

where $M_{\eta S}$ is the usual overlap between $\boldsymbol{S}$ and the $(d+1) - th$ dream, given by

$$M_{\eta S} = \frac{1}{N} \sum_{i=1}^{N} \eta_i S_i. \tag{11}$$

After d dreams, defining a variable $\tau = \varepsilon d / 2P$ and taking the thermodynamic limit, we can write a differential equation for the evolution of K averaged over the possible dreams

$$\frac{dK_S}{d\tau} = -\langle P M_{\eta S}^2 \rangle. \tag{12}$$

For the fixed-norm case we have

$$K_S^{d+1} = K_S^d - \varepsilon M_{\eta S}^2 + \frac{\varepsilon}{P} K_\eta^d K_S^d, \tag{13}$$

that leads, in the thermodynamic limit to

$$\frac{dK_S}{d\tau} = -\langle P M_{\eta S}^2 \rangle + \langle K_\eta \rangle K_S. \tag{14}$$

We have kept only terms up to first order in ε.

These rather simple derivations lead to some important insights of the slow dynamics. The solution of these equations is not possible without knowledge from the simulations. The measured overlaps can be used in the differential equations. In region 1 of the phase diagram, that leads to simple equations, which can then be solved for K and compared to the values obtained in the simulations. This will be discussed in section 4.

3 Numerical Results

The phase diagram in the (α, β^{-1}) plane is shown in fig. (1) for the free-norm case. Three regions can be distinguished. Region 1 is where no dreaming was needed, since it is the retrieval phase of the simple Hebb couplings. That means that the minimum number of dreams for retrieval is zero. Region 2 is where dreaming is able to destroy the spinglass phase turning it to ferromagnetic. The ferromagnetic state is achieved only after a minimum number of dreams d^{min} (or τ^{min}). As already noticed in [3, 4], in both regions there is a maximum number of dreams, d^{max} (τ^{max}) which if surpassed, leads the net to a nonretrieval phase. The evolutions of the magnetizations (overlaps with pattern 1) m^1, are shown in fig. (2).

Finally region 3 of the phase diagram is such that no amount of dreaming is sufficient to enter a retrieval phase. The borders of the phase diagram were determined for a given α by the largest β which can have a magnetization one (or near one, depending on the temperature), letting the net relax to a fixed-point configuration starting from an initial configuration equal to a memory pattern. This measures the onset of null-radius basins of attraction. Region 2 could be larger if we adopted a less stringent criterium. Note (fig. 2) that the order parameter is still quite large in region 3. As usual, these borders don't signal the occurrence of a real thermodynamic phase transition, but describe the nature of the metastable states.

The waterfall-like graphs in fig.s (3a, 4a) show the order parameter as a function of the number of dream iterations d and α (fig. 3a) or β^{-1} (fig. 4a) in region 2. The very sharp rises, when the minimum (maximum) number of dreams is reached (surpassed), suggest discontinuous transitions.

This is not so clear for higher inverse gains, and the transition seems to be continuous at $\beta = 1$. The plateaux are quite extensive, indicating a broad region where the net works properly as a memory.

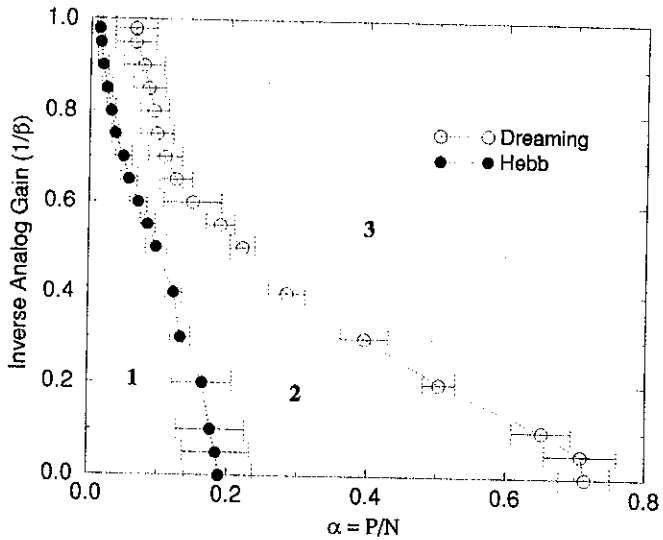


Figure 1: Phase Diagram (α-β^{-1}) for the free- norm case. Average over 5 simulations, with $N = 100$ spins for ∘ and $N = 400$ for • , $\varepsilon = 0.02$.

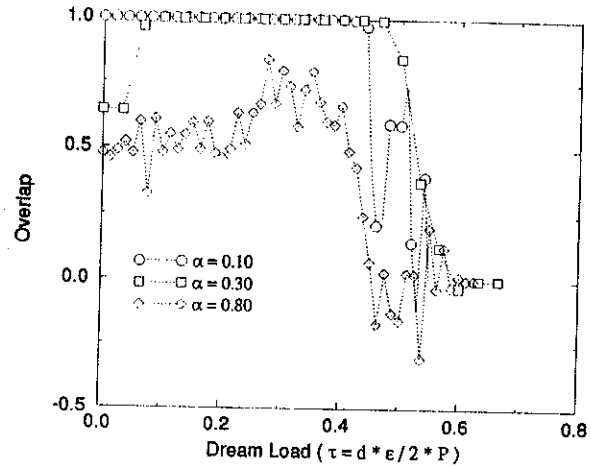
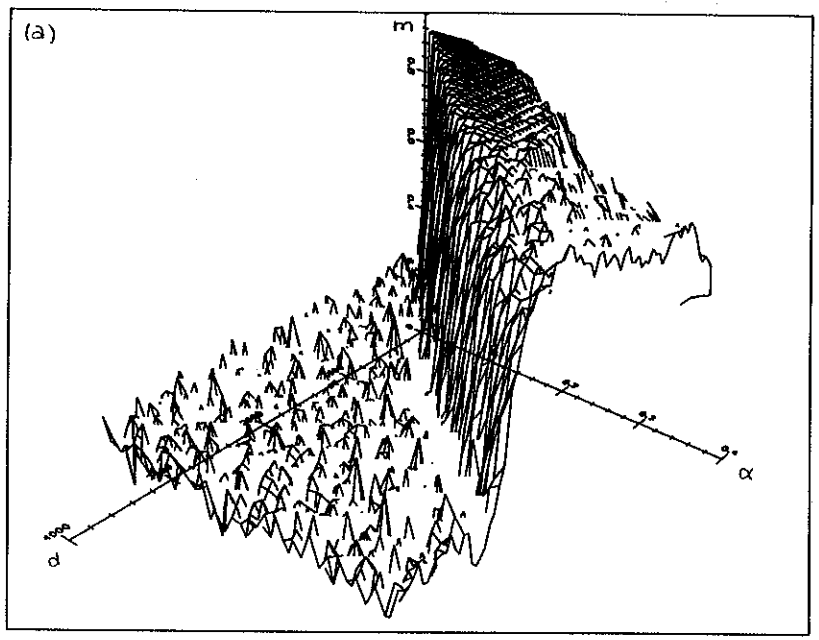


Figure 2: Evolution of magnetizations m^1 under dreaming, for $\beta^{-1} = 0.10$.

For different values of α, we show in fig. (3b) the plateaux of retrieval. The upper (lower) curve of each pair shows the maximum (minimum) number of dream iterations as a function of the inverse gain. It can be seen that for $\beta^{-1} \to 0$ the maximum dream load is independent of the pattern load α (see also fig. 2). The meeting of the two lines defines for a given α the maximum β^{-1} or which there is a plateau. The phase boundary of fig. 1 is determined by these meetings, that is, the vanishing of the order parameter's plateau. Figure (4b) shows the plateaux for different β^{-1}. The situation is less clear than in the fig. (3b), since the determination of the closing point is numerically more difficult. The phase diagram (fig. 1) is a cut of the three-dimensional space $(\alpha, \beta^{-1}, \tau)$ at nonconstant τ, projected onto the (α, β^{-1}) plane. For $T \to 0$ we have found that $\alpha_c \approx 0.71 \pm 0.04$.

In fig. (5) we show the evolution of m^1 under dreaming for the fixed-norm case. These curves may be compared to those of fig. (2), related to the free norm case. The capacity α_c is certainly increased, but since the plateaux are not so well defined as in the free-norm case and the dreaming dynamics is now much slower, we can only say that as $T \to 0$ we have $0.9 \lesssim \alpha_c \lesssim 1.0$.



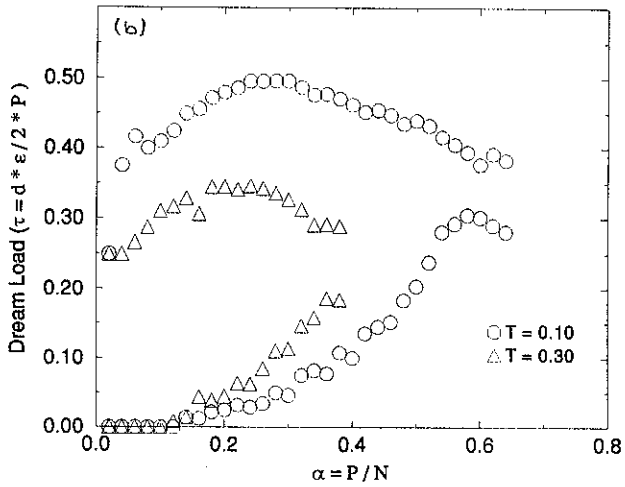
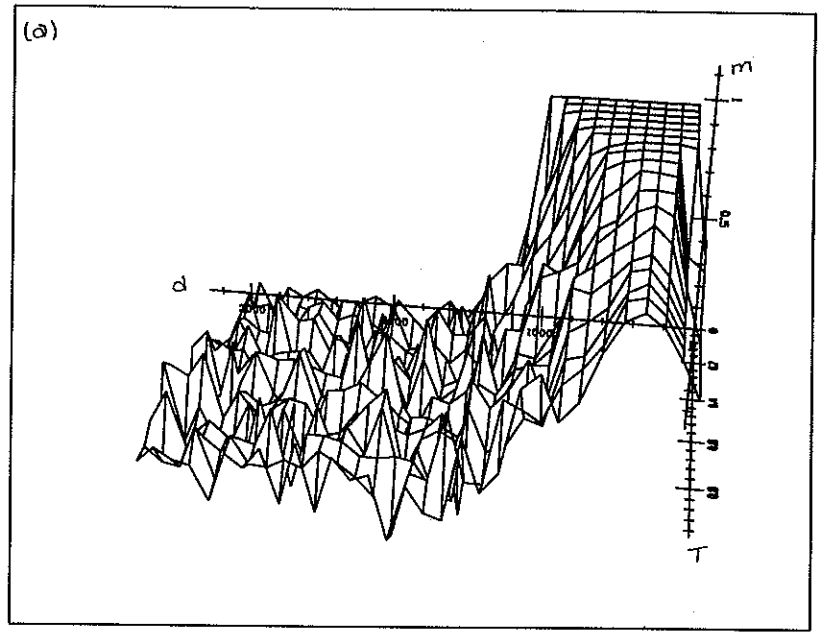


Figure 3: (a) Overlap as a function of the load α and the number of dream iterations d for $\beta^{-1} = 0.20$. The retrieval region is characterized by the plateau, where $m^1 \approx 1$; (b) phase diagrams (plateaux) $(\tau - \alpha)$ for $\beta^{-1} = 0.10$ and 0.30. Upper and lower curves represent maximum and minimum dream loads. The regions between them are retrieval phases.



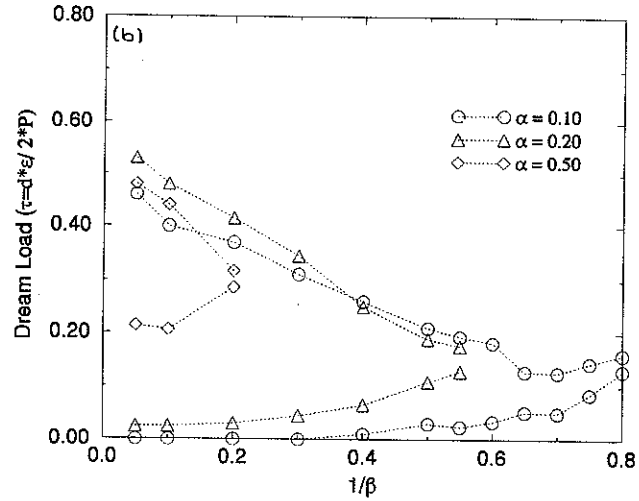


Figure 4: (a) Overlap as a function of the inverse gain $\beta^{-1} = T$ and the number of dream iterations d. The plateau corresponds to the retrieval region for $\alpha = 0.20$; (b) phase diagrams (plateaux) $(\tau$-$\beta^{-1})$ for $\alpha = 0.10, 0.20$ and 0.50. Upper and lower curves represent maximum and minimum dream loads and the regions between them are retrieval phases, as in the two previous fig.s.

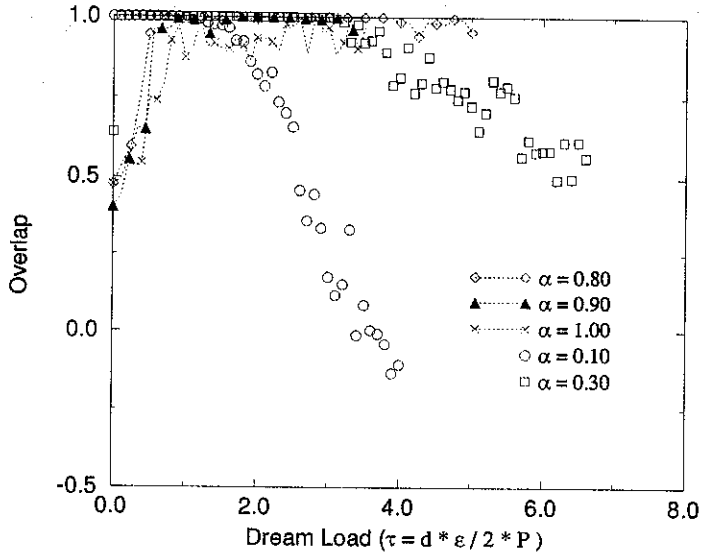


Figure 5: Evolution of the magnetizations m^1 under dreaming for the fixed-norm case.

4 Evolution of K

Here we discuss an approach to partially solve the differential equations (12) and (14). By partial it is meant that information from the simulation will be used. We restrict ourselves to region 1 of the phase diagram (fig. 1) for a while. The information to be used from the simulation is the average quadratic overlap of a dream with the fixed-point configuration $\boldsymbol{S}$. A dream configuration is obtained by letting a randomly chosen configuration to relax. In region 1, it will relax to a given pattern or antipattern with probability $1/2P$; this means that $\langle PM^2_{\eta S}\rangle \approx 1$ if the fixed-point $\boldsymbol{S}$ is a pattern, say $\boldsymbol{\xi}^1$. Thus for free-norm dreaming

$$\frac{dK_{\xi^1}}{d\tau} = -1, \tag{15}$$

if the number of dreams is less than the maximum number, and if larger then

$$\frac{dK_{\xi^1}}{d\tau} = 0. \tag{16}$$

So the K-stability should decrease with slope -1 up to the maximum number of dreams, and then level off. This accounts very well for the shape of the measured curve in fig. (6a), which has a slope, obtained by regression in the linear region, of -0.98 for $\beta^{-1} = 0.10$. Otherwise, we can also calculate the value of $\langle PM^2_{\eta\xi_1}\rangle$ for $\tau^{min} \leq \tau \leq \tau^{max}$, and obtain the slope directly. This later method includes finite-size effects (the dreams overlaps macroscopically with many

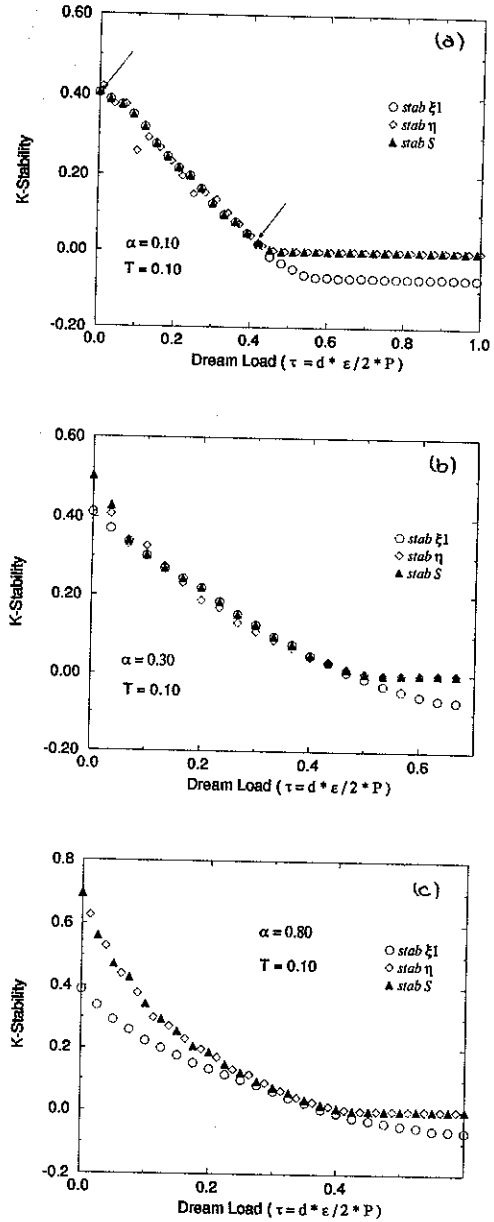


Figure 6: Evolution of the K-stabilities of the memory patterns ξ^1(circles), dreams η(diamonds) and fixed-points S (triangles) under dreaming in regions 1 (fig.a), 2 (fig.b) and 3 (fig.c). The maximum and minimum τ are indicated by arrows in (a).

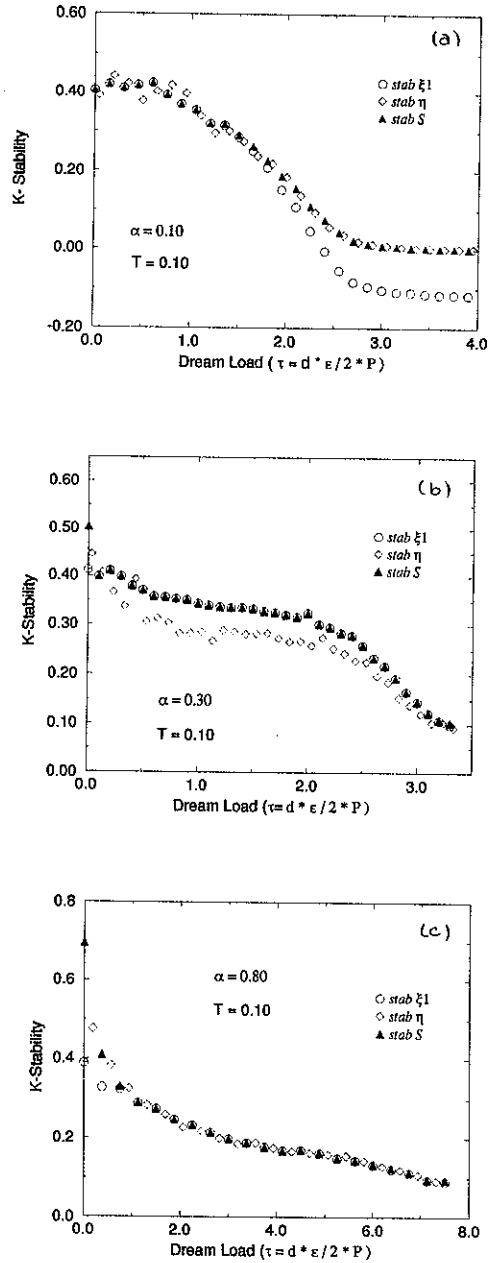


Figure 7: Same as fig. (7), but for the fixed-norm procedure.

of the patterns and mixtures), leading to larger values of the slope. We have obtained 1.11.

In region 2 the probability to hit a pattern or antipattern is smaller and so they must destabilize slower. The slopes measured by linear regression in this region are in agreement with this. We have obtained -0.86 and -0.79 for the slopes of the curves corresponding to $\alpha = 0.20$ and 0.30 for $\beta^{-1} = 0.10$. Direct calculation of $\langle PM^2_{\eta\xi^1}\rangle$ has given 1.09 and 1.17, respectively.

In region 3 we cannot evaluate the slope in this simple way. Only information given by the simulations is shown in fig. (6c).

For the fixed-norm case, as in the previous one, it is possible to evaluate $\langle PM^2_{\eta\xi^1}\rangle$ in region 1. Thus, noting that $K_{\xi^1} \approx K_\eta$ when $\langle PM^2_{\eta\xi^1}\rangle \approx 1$, we obtain that K as a function of τ is given approximately by

$$\frac{dK_{\xi^1}}{d\tau} = -1 + \langle K_{\xi^1}\rangle^2 \tag{17}$$

and so the K-stability should vary according to

$$K_{\xi^1}(\tau) = K_0 \tanh(\tau_0 - \tau)/\tanh(\tau_0) \tag{18}$$

where τ_0 and K_0 are constants. Fig.(8) shows a fitting of function (18) to numerical data in the region 1 of the phase diagram (1). It was made just for $\tau^{min} \leq \tau \leq \tau^{max}$, which corresponds to a retrieval phase (where triangles and circles superposes in fig. (7a)). The fitted value τ_0 is the dream load which leads to $K \to 0$, and K_0 is the initial stability for $\tau = 0$.

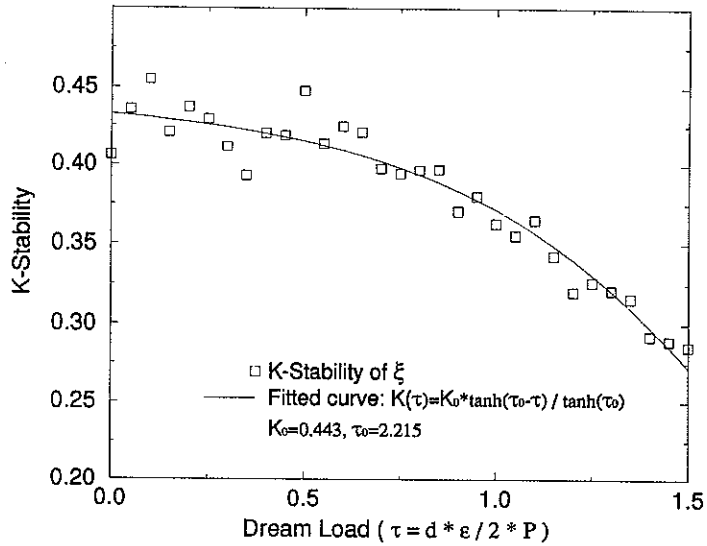


Figure 8: Fitting of function (18) to numerical data. $\beta^{-1} = 0.10$, $\alpha = 0.10$. Fitted values: $K_0 = 0.443$ and $\tau_0 = 2.215$.

5 Conclusion

Here we summarize our results. We have performed numerical simulations to measure the ability of an analog attractor neural net in recognizing non-correlated patterns learned according to Hebb's rule and submitted to dreaming.

The performance of the net in different regions of the phase diagram as a function of the dream load τ has been investigated. We have measured the increase of the capacity α_c as a function of the temperature β^{-1} due to dreaming.

A fixed-norm version for the dreaming algorithm was introduced. Numerical simulations have shown that this version is even more robust than the traditional one, leading to a greater increase in the retrieval region of the phase diagram.

We have also obtained simple differential equations which describe well the evolution of K-stabilities in the retrieval phase, for both versions of dreaming. By looking at the form of the differential equations, it is possible to understand qualitatively how the algorithms work. The subtraction of the dream configuration reduces the K-stability of a minimum by an amount which depends on the square of their overlaps. In the fixed-norm case, the restoration of the norm increases slightly the stability and this explains why the fixed-norm algorithm works on larger $\tau-$scales than the free-norm one.

Aknowledgments

SMK has received financial support from CNPq and NC has received partial support from CNPq and FAPESP.

References

[1] D. J. Amit: "Modelling Brain Function", Cambridge University Press, (1989).

[2] F. Crick and G. Mitchison: *Nature*, **304**, 111,(1983).

[3] J. L. van Hemmen, L. B. Ioffe, R. Kühn, M. Vaas: *Physica A*, **163**, 386, (1989).

[4] J. L. van Hemmen, in: Proc. STATPHYS 17 Workshop on Neural Nets and Spin Glasses. W. Theumann and R. Köberle, eds., Porto Alegre, (World Scientific, Singapore), (1990).

[5] J. Hertz, A. Krogh and R. G. Palmer: "Introduction to the Theory of Neural Computation", Addison-Wesley, (1991).

[6] J. J. Hopfield, D. I. Feinstein and R. G. Palmer: *Nature*, **304**, 158, (1983).

[7] J. J. Hopfield: *Proc. Natl. Acad. Sci. USA* **81**, 3088, (1984).

[8] R. Kühn: *Lecture notes in Physics* **368**, (1990).

[9] C. M. Marcus and R. M. Westervelt: *Phys. Rev. A*, **40**, 1, (1989).

[10] M. Shiino and T. Fukai: *J. Phys. A: Math. and Gen.* **23**, L1009, (1990).