Storage capacity of a neural network with state-dependent synapses

1994 J. Phys. A: Math. Gen. 27 1575

(http://iopscience.iop.org/0305-4470/27/5/021)

J. Phys. A: Math. Gen. 27 (1994) 1575–1583. Printed in the UK

Storage capacity of a neural network with state-dependent synapses†

F Zertuche†, R López‡ and H Waelbroeck‡

† Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, Sección Cuernavaca, AP 139-B 62191 Cuernavaca, Morelos, México
‡ Instituto de Ciencias Nucleares, UNAM, Apdo Postal 70-543, México, DF, 04510 México

Received 6 July 1993, in final form 19 October 1993

Abstract. The storage capacity of the Hopfield network is limited to $p/N \leqslant \alpha_c = 0.138$ (p = number of patterns, N = number of neurons), beyond which the contribution of weakly correlated patterns surpasses that of the desired pattern. This contribution can be eliminated by introducing a threshold: a pattern correlation below this threshold is simply set to zero in the synapses. We solve the mean-field equations and derive the critical value of the threshold required to stabilize $p = \alpha N$ patterns with $\alpha > \alpha_c$.

1. Introduction

The maximal storage capacity of a pattern recognition neural network has been a subject of intense research, for its obvious importance in associative memory applications. The Hebbian learning scheme of the Hopfield model [1], leads to a maximal storage capacity $\alpha_c = p/N = 0.138$, for a system with p patterns and N neurons [2]. Beyond this value, the number of weakly correlated patterns becomes so large that their contribution dominates over that of the highly correlated pattern being recalled, and the so-called spurious states become the stable attractors. This apparent maximal storage capacity is not an intrinsic limit on neural networks, but rather reflects the limitations of the Hebbian learning rule. In fact for networks with a general quadratic interaction, replica symmetry techniques give the value $\alpha_c = 2$ for the maximal storage capacity [3]. Other variants of the Hopfield model have been studied, always giving $\alpha_c < 2$ [4]. Further attempts have been made to increase the storage capacity and performance of the Hopfield model by taking more complex synapses. Gardner has studied a generalization of the Hopfield model which has n-interactions among the neurons and has found that α_c grows along with n in such a way that $\alpha_c \sim n/(4 \ln N)$ for $n \to \infty$ [5]. In [6] a generalization of Hebb's rule with state-dependent synapses, which turns out to be a special case of Gardner's networks, has been studied. The authors show by means of numerical simulations that the number of spurious states is reduced and that the stability of the memorized states is improved, thus suggesting that α_c should increase. Beyond this, state-dependent synapses are of interest in describing some biological phenomena such as habituation and sensitization [7]. Our purpose in this article is to study the storage capacity of a state dependent synapses model where there is a parameter, the 'threshold', which determines which patterns contribute to the synapses.

† This work is supported in part by CONACyT grant 400349-5-1714E and by the Association Générale pour la Coopération et le Développement (Belgium).

0305-4470/94/051575+09$19.50

Only those patterns whose correlation with the state of the system is greater or equal to the threshold are left to give a Hebbian contribution to the synapses.

In section 2, the Hopfield model is reviewed, as well as the solutions of the mean-field equations at zero temperature [2, 8]. In section 3, Hebb's rule is modified by introducing the threshold, which makes the synapses state dependent. An energy function is also constructed, which is minimized by the deterministic evolution rule of the system (zero temperature dynamics). In section 4, the mean-field equations of the model are solved by using the statistical method developed in [9, 10], the phase space for the memorized states is constructed, and the increase in the capacity α_c with the threshold is confirmed with a numerical simulation. For large values of the threshold, the information theoretical maximum capacity $\alpha = 2^N/N$ is attained. In section 5, we give the conclusions.

2. The Hopfield model

The study of associative memory was pioneered by Little [11] and Hopfield [1]. In Hopfield's model, a set of N neurons S_i ($i = 1, \ldots, N$) which can take the values ± 1 interact through a synapses matrix W_{ij}. The network evolves by updating one neuron at a time in random order with the deterministic rule

$$S_i(t+1) = \mathrm{sgn}\left(\sum_{j=1}^{N} W_{ij} S_j(t)\right) \tag{1}$$

for the zero temperature dynamics ($T = 0$), or by Glauber dynamics [12] in the case of finite temperature

$$S_i(t+1) = \pm 1 \qquad \text{with} \quad \text{probability} \qquad \frac{1}{1 + \mathrm{e}^{\mp 2\beta h_i(t)}} \tag{2}$$

where

$$h_i(t) = \sum_j W_{ij} S_j(t) \tag{3}$$

and where $\beta^{-1} = T$ is the temperature parameter. The synapses matrix is given by Hebb's rule

$$W_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu . \tag{4}$$

where $\{\xi_i^\mu = \pm 1\}$ ($\mu = 1, \ldots, p$) is a set of p patterns generated randomly, such that $\langle\langle \xi_i^\mu \xi_j^\nu \rangle\rangle = \delta^{\mu\nu} \delta_{ij}$, where $\langle\langle \cdots \rangle\rangle$ denotes the average over the distribution of patterns.

Taking the thermal average of S_i by means of (2), and approximating h_i by its thermal average, one obtains the so called mean-field equations

$$\langle S_\nu \rangle = \frac{1}{N} \sum_j \xi_j^\nu \tanh\left(\beta \sum_\mu \xi_j^\mu \langle S_\mu \rangle\right) \tag{5}$$

where

$$S_\mu = \frac{1}{N} \sum_j \xi_j^\mu S_j \tag{6}$$

is the correlation of the state S_i with the pattern ξ_i^μ, and $\langle \cdots \rangle$ denotes the thermal average.

In [2] the mean-field equations (5) have been solved using replica symmetry techniques. For the memorized states, $\langle S_\mu \rangle = (m, 0, 0, \ldots)$, the authors obtain solutions for $\alpha \leqslant \alpha_c = 0.138$, in the zero temperature limit $\beta \to \infty$.

3. State dependent synapses

The limit on the storage capacity of the Hopfield model is due to the contribution in (5) of a large number of weakly correlated patterns, so one could attempt to eliminate the problem by introducing a threshold $\eta \geqslant 0$ in the synapses (4):

$$W_{ij} = \frac{1}{N} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \; \Theta \left(S_{\mu}^2 - \frac{\eta^2}{N} \right) \tag{7}$$

where the step function $\Theta(x)$ ensures that the pattern ξ^{μ} is set to zero if $S_{\mu}^2 < \eta^2/N$. The factor N^{-1} in η^2 is introduced because the weakly correlated patterns are statistically of order $1/\sqrt{N}$. Note that by setting $\eta = 0$, we recover the Hopfield model.

The energy function

$$H = -\frac{N}{2} \sum_{\mu} \Theta \left(S_{\mu}^2(t) - \frac{\eta^2}{N} \right) \left(S_{\mu}^2(t) - \frac{\eta^2}{N} \right) \tag{8}$$

tends to be minimized by the evolution rule (1). Indeed, if the i-neuron changes its state as a consequence of (1), then

$$S_i' \equiv S_i(t+1) = -S_i(t) = \operatorname{sgn}(h_i(t)) \tag{9}$$

and the change of H is given by

$$\Delta H = H' - H = \Delta H_{\Theta} + \Delta H_S$$

where

$$\Delta H_{\Theta} \equiv -\frac{N}{2} \sum_{\mu} \Delta\Theta_{\mu} \; \left(S_{\mu}'^2 - \frac{\eta^2}{N} \right) \tag{10}$$

$$\Delta H_S \equiv -\frac{N}{2} \sum_{\mu} \Theta \left(S_{\mu}'^2 - \frac{\eta^2}{N} \right) \; \Delta S_{\mu}^2 \tag{11}$$

with

$$\Delta\Theta_{\mu} \equiv \Theta \left(S_{\mu}'^2 - \frac{\eta^2}{N} \right) - \Theta \left(S_{\mu}^2 - \frac{\eta^2}{N} \right) . \tag{12}$$

It is easy to check that all the terms in the sum (10) are positive, so that $\Delta H_{\Theta} \leqslant 0$: if $\Delta\Theta_{\mu} < 0$, then $S_{\mu}'^2 - \eta^2/N < 0$ and $\Delta\theta_{\mu} \left(S_{\mu}' - \eta^2/N \right)$ is positive, and if $\Delta\Theta_{\mu} > 0$, then $S_{\mu}'^2 - \eta^2/N > 0$, and once again we have a positive term. We must also show that $\Delta H_S \leqslant 0$. From (9) we have

$$S_{\mu}' = \frac{1}{N} \sum_{j} \xi_j^{\mu} S_j' = \frac{1}{N} \left(-\xi_i^{\mu} S_i + \sum_{j \neq i} \xi_j^{\mu} S_j \right) = -\frac{2}{N} \xi_i^{\mu} S_i + S_{\mu}$$

so that

$$\Delta S_{\mu}^2 = S_{\mu}'^2 - S_{\mu}^2 = \frac{4}{N^2} - \frac{4}{N} \xi_i^{\mu} S_i S_{\mu} = \frac{4}{N^2} + \frac{4}{N} \xi_i^{\mu} S_{\mu} \operatorname{sgn}(h_i) .$$

Substituting in (11), one obtains

$$\Delta H_S = -\frac{N}{2}\sum_\mu \Theta\left(S_\mu^2 - \frac{\eta^2}{N}\right)\left(\frac{4}{N^2} + \frac{4}{N}\xi_i^\mu S_\mu \,\mathrm{sgn}(h_i)\right)$$
$$= -\frac{2}{N}\sum_\mu \Theta\left(S_\mu^2 - \frac{\eta^2}{N}\right) - 2|h_i| \leqslant 0\,.$$

Thus the attractors are local minima of the energy (8).

Averaging S_i using (2), and approximating $h_i(S_j)$ by $h_i(\langle S_j\rangle)$ one gets the equations

$$\langle S_\mu\rangle = \frac{1}{N}\sum_j \xi_j^\mu \tanh\left\{\beta\sum_\nu \xi_j^\nu \langle S_\nu\rangle \Theta\left(\langle S_\nu\rangle^2 - \frac{\eta^2}{N}\right)\right\}. \tag{13}$$

With this approximation we are going beyond a mean-field treatment in which h_i is just replaced by $\langle h_i\rangle$ [8]: instead we are also passing the thermal average inside the threshold function. To justify this approximation, we note that it is also possible to obtain the equations (13) for finite p by calculating the partition function associated to the energy function (8), in the saddle point approximation. Although this does not imply that the approximation remains valid for $p = \alpha N$ with $\alpha \neq 0$, we will see later that a numerical simulation leads to a critical capacity in agreement with the theoretical predictions.

4. The memorized states

The neural network is a useful associative memory device, as long as the mean-field equations have solutions of the form $\langle S_\mu\rangle = (m, 0, 0, \ldots)$ in the thermodynamic limit ($N \to \infty$). Following the statistical method developed by Geszti [9] and Peretto [10] (see also [8]), we take $\langle S_1\rangle = m \sim \mathrm{O}(1)$ and $\langle S_\nu\rangle \sim \mathrm{O}(1/\sqrt{N})$ for $\nu \neq 1$, so that $m^2 > \eta^2/N$. Taking $\mu = \nu \neq 1$ in (13), separating the term proportional to $\langle S_\nu\rangle\Theta\left(\langle S_\nu\rangle^2 - \eta^2/N\right)$ and expanding in Taylor series to first order, one obtains

$$\langle S_\nu\rangle = \frac{1}{N}\sum_j \xi_j^\nu \xi_j^1 \tanh\left[\beta\left(m + \eta_j^\nu\right)\right]$$
$$+ \frac{\beta}{N}\langle S_\nu\rangle \Theta\left(\langle S_\nu\rangle^2 - \frac{\eta^2}{N}\right)\sum_j \left[1 - \tanh^2\beta\left(m + \eta_j^\nu\right)\right] \tag{14}$$

where

$$\eta_j^\nu \equiv \sum_{\mu\neq 1,\nu} \xi_j^\mu \xi_j^1 \langle S_\mu\rangle \Theta\left(\langle S_\mu\rangle^2 - \frac{\eta^2}{N}\right) \tag{15}$$

is the sum of a large number ($p \gg 1$) of small terms. We now assume that the small correlations $\langle S_\mu\rangle$, $\mu \neq 1$ have identical normal distributions centred at zero, with variance σ^2/N. This is partially justified from (6) and by the central limit theorem (note, however, that the $\langle S_\mu\rangle$, $\mu \neq 1$ are not strictly independent variables, since they are related through

(13)). By the same arguments η_j^ν is assumed to have a Gaussian distribution with variance αr and average zero, so that

$$\alpha r = \langle\langle (\eta_j^\nu)^2 \rangle\rangle \tag{16}$$

and one obtains

$$q \equiv \frac{1}{N}\sum_j \tanh^2 \beta \left(m + \eta_j^\nu\right) = \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh^2 \beta \left(m + \sqrt{\alpha r} z\right) \tag{17}$$

which is the Edwards–Anderson [13] order parameter. Substituting back (17) in (14) we have

$$\left[1 - \beta(1-q)\,\Theta\left(\langle S_\nu\rangle^2 - \frac{\eta^2}{N}\right)\right]\langle S_\nu\rangle = \frac{1}{N}\sum_j \xi_j^\nu \xi_j^1 \ \tanh \beta \left(m + \eta_j^\nu\right) .$$

Squaring this expression, using (16) and averaging over the distribution of patterns (here and in the following we are taking $p - 2 \approx p$), we get

$$\sigma^2 + Mr = q \tag{18}$$

where

$$M \equiv \left[1 - \beta(1-q)\right]^2 - 1 . \tag{19}$$

Equation (18) gives a relation among r and σ^2. Another relation can be obtained from (16), using (15) and averaging over the distribution of patterns:

$$\alpha r = p\left\langle\left\langle \langle S_\nu\rangle^2 \Theta\left(\langle S_\nu\rangle^2 - \frac{\eta^2}{N}\right)\right\rangle\right\rangle = p\int \frac{dz}{\sqrt{2\pi\sigma^2/N}} \exp\left(-\frac{Nz^2}{2\sigma^2}\right) z^2 \Theta\left(z^2 - \frac{\eta^2}{N}\right)$$

so that

$$r = \frac{2}{\sqrt{\pi}}\sigma^2 \Gamma\left(\frac{3}{2}, \frac{\eta^2}{2\sigma^2}\right) \tag{20}$$

where Γ is the incomplete gamma function [14].

Another equation is obtained for m by taking $\mu = 1$ in (13) and using the same procedure as before,

$$m = \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh \beta \left(m + \sqrt{\alpha r} z\right) . \tag{21}$$

In conclusion we have obtained the four equations (17), (18), (20) and (21) for the four unknowns q, r, σ and m. These must be solved numerically. In the limit $\beta \to \infty$, in which one recovers the deterministic evolution law (1), $q \to 1$ in such a way that

$$C \equiv \beta(1-q) = \sqrt{\frac{2}{\pi \alpha r}}\, e^{-m^2/2\alpha r} . \tag{22}$$

The other equations become

$$r = \frac{1}{(1-C)^2 + k} \tag{23}$$

$$\sigma^2 = \frac{1+k}{(1-C)^2 + k} \tag{24}$$

$$m = \mathrm{erf}\left(\frac{m}{\sqrt{2\alpha r}}\right) \tag{25}$$

where

$$k \equiv \frac{\sqrt{\pi}}{2} \frac{1}{\Gamma\left(\frac{3}{2}, \eta^2/2\sigma^2\right)} - 1 \,. \tag{26}$$

For $\eta = 0$, one has $k = 0$ and from (22)–(25) one recovers the equations of the Hopfield model at $T = 0$ [2, 8]. On the other hand, for $\eta \neq 0$, the equations have a non-trivial solution with $\alpha > 0.138$. In fact, for $\eta \gg 1$, one can show that the solution always exists up to

$$\alpha_c \approx \sqrt{\frac{2}{\pi}} \eta^{-1} \, e^{\eta^2/2} \,.$$

Thus $\alpha_c \to \infty$ as $\eta \to \infty$. This is an analogous situation to that already found by Gardner in the case of a neural network which has n interactions among its neurons, when $n \to \infty$ [5]. For finite N, the maximum storage capacity is given by the information-theoretical limit $p = 2^N$.

In figure 1 we illustrate the whole phase space spanned by the parameters α, η and T; below the surface there are solutions for the memory states ($m \neq 0$), and above it the network is useless as an associative memory device ($m = 0$). In figure 2 we show the critical line $\alpha_c(\eta)$ at $T = 0$, with $\alpha_c = 0.138$ at $\eta = 0$. $m \neq 0$ above this line, and below it $m = 0$. One can see how the storage capacity of the network is improved as the value of η is increased. In figure 3 we give a slice of the phase space for η vs α at $T = 0.5$. The critical curve $\alpha_c(\eta)$ has its origin at 0.0587 and the memory solutions are in the region above the curve. Once again the capacity increases with η.

A numerical simulation was carried out using a neural network with $N = 100$ neurons, using 200 sets of αN patterns generated randomly. The initial state was set equal to the first pattern with an error in neuron i, and the network was updated asynchronously $8N$. At the end, the resulting state was compared with the first pattern; if there was coincidence the network was considered stable, and average of this stability was taken on the 200 sets of patterns, and on 25 choices of the erroneous neuron i. In figure 4, a plot of the average stability vs capacity is displayed for $\eta = 0, 1$. The two graphs differ roughly by a translation of 0.04. At the stability level of 97%, the increase in the capacity of the Hopfield model is from 0.14 to 0.17, in agreement with the theoretical calculation.

5. Conclusions

We have examined a modification of the Hopfield model, where only the patterns whose correlation with the state of the system is greater or equal to the threshold are left to give

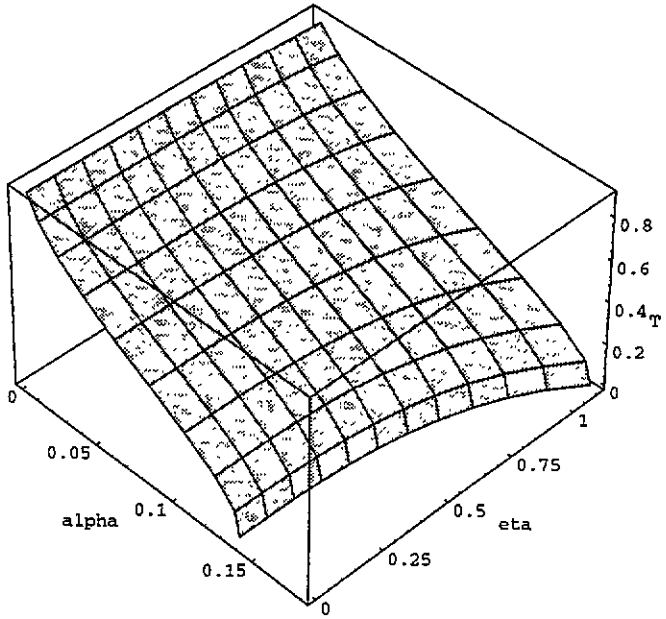

Figure 1. Phase space for the neural network with threshold. The memorized states are stable below the critical surface.

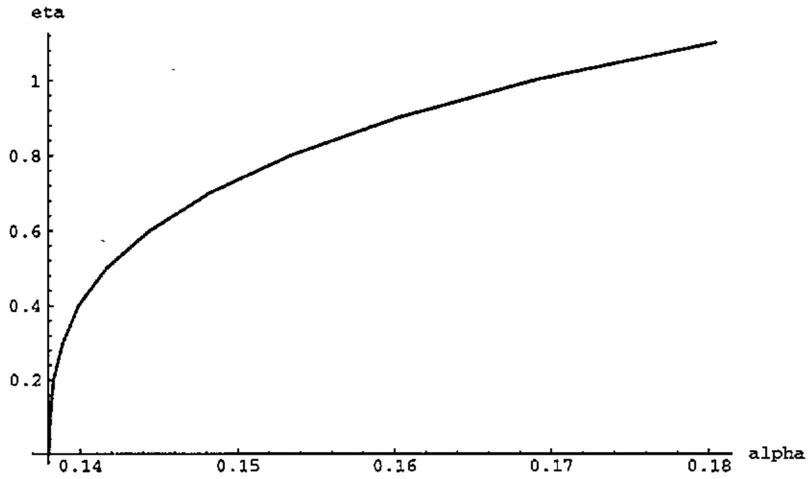

Figure 2. Slice of the phase space for the neural network with threshold at $T = 0$. The Hopfield value $\alpha_c = 0.138$ is found at $\eta = 0$; for a threshold equal to 1 the critical capacity increases to $\alpha_c(1) = 0.17$.

a Hebbian contribution to the synapses. We gave an energy function for this model, which tends to be minimized by the zero-temperature evolution rule, and reduced the mean-field equations to a set of four coupled algebraic equations, by means of a statistical calculation. The phase diagram was derived from a numerical solution of these equations; it shows an

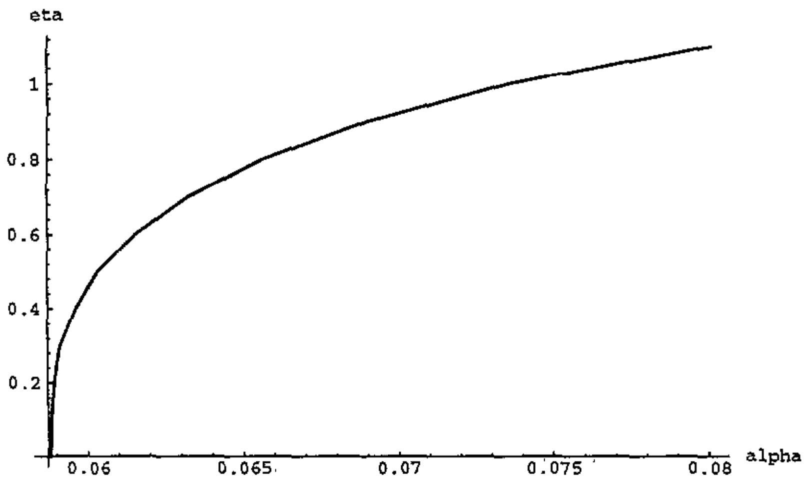

Figure 3. Slice of the phase space for the neural network with threshold at $T = 0.5$.

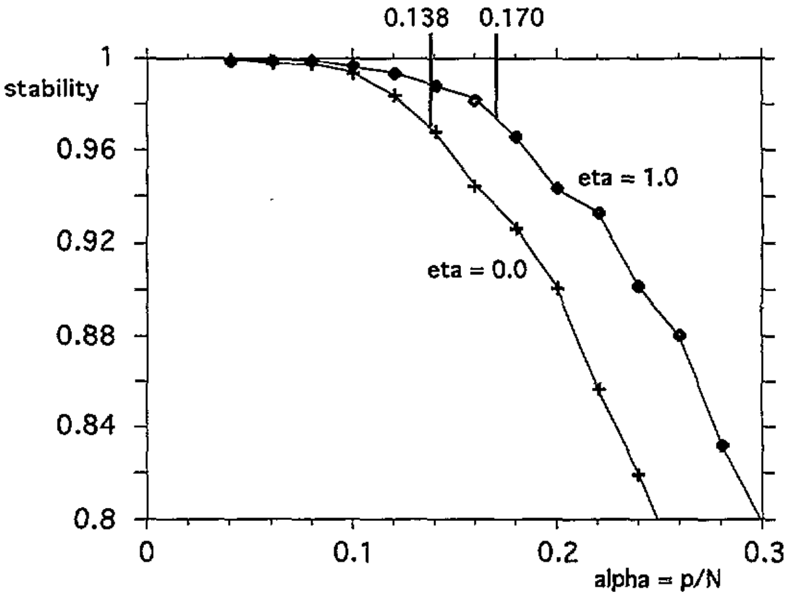

Figure 4. The results of numerical simulations with $N = 100$ neurons is represented, for the Hopfield model ($\eta = 0$) and for a threshold of 1. The average stability falls below 97% for $\alpha_c = 0.14$ without threshold, and $\alpha_c = 0.17$ with the threshold.

increasingly rapid improvement in the storage capacity as a function of the threshold, both for $T = 0$ and for $T > 0$.

The increase in storage capacity was confirmed by numerical simulation, for a threshold equal to 1, at $T = 0$. The quantitative agreement with the theoretical predictions indicates that the approximations, of passing the thermal average inside the non-linear threshold function and later considering the weakly correlated patterns as independent random variables, do not significantly affect the final results.

An interesting extension of this work is the application of a threshold to time-sequence recognition models. A preliminary study indicates that zero-temperature dynamics at large storage capacities is characterized by two phase transitions, from no memorized sequences at low thresholds, to an unstable sequence, to a stable attracting sequence at large values of the threshold. The intermediate phase, where the memorized sequence is unstable, may have interesting applications in non-linear dynamics and chaotic time series prediction.

Acknowledgments

We would like to express our gratitude to the Comité de Supercómputo of the UNAM, for access to the Cray YM-P4, and other computer facilities. One of us (HW) is also indebted to C Duqué, director of the Association pour le Développement par la Recherche et l'Action Integrées, L Haroldo Rodríguez, director of the Coordinación para el Desarrollo Rural Integral, and the Belgian ambassador to México, for their support of our project.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554; 1984 *Proc. Natl Acad. Sci. USA* **81** 3088
[2] Amit D, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530; 1987 *Ann. Phys.* **173** 30
[3] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[4] van Hemmen J L 1987 *Phys. Rev.* A **36** 1959
Krauth W and Mézard M 1989 *J. Physique* **50** 3057
Bouten M, Komoda A and Serneels 1990 *J. Phys. A: Math. Gen.* **23** 2605
[5] Gardner E 1987 *J. Phys. A: Math. Gen.* **20** 3453
[6] Matus I J and Perez P 1990 *Phys. Rev.* A **41** 7013
[7] Kandel E R and Schwartz J H 1991 *Principles of Neural Science* (Amsterdam: Elsevier/North-Holland) ch 65
[8] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Reading, MA: Addison-Wesley) ch 2
[9] Geszti T 1990 *Physical Models of Neural Networks* (Singapore: World Scientific) ch 4
[10] Peretto P 1988 *J. Physique* **49** 711
[11] Little W A 1974 *Math. Biosci.* **19** 101
[12] Glauber R J 1963 *J. Math. Phys.* **4** 294
[13] Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
[14] Arfken G 1970 *Mathematical Methods for Physicists* (New York: Academic) ch 10