Recognition of temporal sequences of patterns using state-dependent synapses

1994 J. Phys. A: Math. Gen. 27 5879

(http://iopscience.iop.org/0305-4470/27/17/020)

J. Phys. A: Math. Gen. 27 (1994) 5879–5887. Printed in the UK

Recognition of temporal sequences of patterns using state-dependent synapses*

F Zertuche†§, R López-Peña‡ and H Waelbroeck‡

† Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, Sección Cuernavaca, AP 139-B 62191 Cuernavaca, Morelos, Mexico

‡ Instituto de Ciencias Nucleares, UNAM Apdo. Postal 70-543, México DF, 04510 Mexico

Received 17 March 1994, in final form 5 July 1994

Abstract. Using an asymmetric associative network with synchronous updating, it is possible to recall a sequence of patterns. To obtain a stable sequence generation with a large storage capacity, we introduce a threshold that eliminates the contribution of weakly correlated patterns. For this system we find a set of evolution equations for the overlaps of the states with the patterns to be recognized. We solve these equations in the limit of the stationary cycle, and obtain the critical value of the capacity as a function of the threshold and temperature. Finally, a numerical simulation is made, confirming the theoretical results.

1. Introduction

The application of neural networks to time-sequence recognition has a wide range of potential applications, from associative memory to speech recognition to the prediction and control of complex dynamical systems. Networks with symmetric synapses, such as the Hopfield [1] and Little [2] models are unable to reproduce complex sequences since they minimize an energy function [3, 4]. The equilibrium properties of these models have been studied analytically by constructing the partition function associated with the energy. In [5] the asynchronous case (the Hopfield model) and the synchronous one (the Little model) have been studied far from saturation. That is to say, for $\alpha = 0$ with $\alpha = p/N$, where p is the number of patterns and N is the number of neurons as N goes to infinity. The near saturation case ($\alpha \neq 0$) was studied by Amit *et al* for the asynchronous case [6] and by Fontanari *et al* for the synchronous one [7] by using replica symmetry techniques. Networks with asymmetric synapses are able to reproduce sequences of patterns, however, in general, the equilibrium statistical mechanics cannot be applied. Instead of using an energy approach Coolen and Ruijgrok [8] started from the master equation for the microscopic states of the network and derived an evolution equation for the probability density for the overlaps of the states with the patterns in an asynchronous asymmetric network. Using this approach other authors have studied asynchronous asymmetric nets far from saturation [9, 10]. The synchronous case was also analysed far from saturation by Bernier [11].

Near saturation networks have been studied recently with success using statistical methods based on the central limit theorem (CLT). In [12] Nishimori and Ozeki investigated the relaxation to equilibrium in the Little model [2] (synchronous dynamics) near saturation

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

§ E-mail address: zertuche@ce.ifisicam.unam.mx

using a mean-field treatment and solving the equations by means of the CLT. In the equilibrium limit they obtained a good qualitative agreement with the results by Amit *et al* [6]. An energy approach and the CLT was used by Shukla [13] in the Little model to obtain a phase diagram that coincided well with the numerical results. In [14] we studied a symmetric, asynchronous, state-dependent network with a threshold using a CLT method developed by Geszti and Peretto [15] and find that its capacity increases with η (the threshold) as $\alpha \sim \eta^{-1}e^{\eta^2/2}$. Moreover, for $\eta = 0$ we recovered the Hopfield model and found the same results as with replica symmetry techniques [6].

In this paper, we will consider an asymmetric neural network synchronously updated with a state-dependent synapses and calculate its capacity to follow a sequence of patterns. For this we first derive the evolution equations for the overlaps using the transition probability for the neuron states and a mean-field-type approximation. We then solve them by the statistical method of Geszti and Peretto. We will show that this network has a maximal storage capacity $\alpha_c = p/N = 0.278$ in the case $\eta = 0$, which is the limit in which the network is state-independent. Beyond this value, the number of weakly correlated patterns becomes so large that their contribution dominates over that of the highly correlated pattern, which should force the transition to the next state in the sequence, and the system diverges from the learned sequence. We will show how the storage capacity increases if one limits the contribution of weakly correlated patterns, by taking $\eta \neq 0$: only those patterns whose correlation with the state of the system is greater than or equal to the threshold are left to give a contribution to the synapses. We will give the dynamical phase diagram of this network and compute the critical capacity as a function of the temperature and the threshold: the learned sequence is stable up to the critical capacity, which increases with the threshold. These results are confirmed in the deterministic limit $T = 0$, with the help of a numerical simulation.

Although biological synapses are likely to be asymmetric, and this fact is probably an important factor in the complexity observed in biological neural systems, the application of a state-dependent threshold and the use of a central clock for the parallel updating of the neurons implies that the results of this paper are not likely to be relevant in the physiological domain. On the other hand, the use of synchronous parallel updating allows for an efficient use of modern parallel-processing computers. We will discuss possible applications of our work to the prediction of chaotic time series in the conclusion.

In section 2, we review the increase in capacity with the threshold parameter in the Hopfield model, as well as the solutions of the mean-field equations for symmetric synapses [14]. In section 3, we consider the asymmetric rule with a threshold, and parallel updating of the neurons; the evolution equations for the overlaps are derived and solved using the statistical method developed in [15], and the generalized phase diagram for the stationary limit of the dynamics is constructed. In section 4, we give the results of the numerical simulation and in secton 5 we give the conclusions.

2. State-dependent synapses and associative memory

In [14] we considered a network with symmetric state-dependent synapses for associative memory tasks. The main idea of this type of synapses is to cut off the contribution of those patterns whose correlation is less than a given value (the threshold). A set of N neurons with states $s_i^n = \pm 1$ at time $n = 1, 2, \ldots$ ($i = 1, \ldots, N$) interact through a state-dependent synapses matrix W_{ij}^n. The network evolves by updating one neuron at a time in random order through the rule

$$s_i^{n+1} = \pm 1 \qquad \text{with probability} \quad \frac{1}{1 + e^{\mp 2\beta h_i^n}} \tag{1}$$

where

$$h_i^n = \sum_{j=1}^{N} W_{ij}^n s_j^n \tag{2}$$

and where $\beta^{-1} = T$ is the temperature parameter. Let $\{\xi_i^\mu = \pm 1\}$ $(\mu = 1, \dots, p)$ be a set of p patterns generated randomly, such that

$$\langle\langle \xi_i^\mu \xi_j^\nu \rangle\rangle = \delta^{\mu\nu} \delta_{ij} \tag{3}$$

where $\langle\langle \cdots \rangle\rangle$ denotes the average over the distribution of patterns. The synapses matrix was taken to be

$$W_{ij}^n = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \; \Theta\left(\left(m_\mu\left(s^n\right)\right)^2 - \frac{\eta^2}{N} \right) \tag{4}$$

where

$$m_\mu\left(s^n\right) \equiv \frac{1}{N} \sum_{j=1}^{N} \xi_j^\mu s_j^n \tag{5}$$

is the correlation of the state s_i^n with the pattern ξ_i^μ, and the step function $\Theta(x)$ ensures that the pattern ξ^μ is set to zero if $\left(m_\mu\left(s^n\right)\right)^2 < \eta^2/N$. η is called the threshold and for $\eta = 0$ Hebb's rule for the synapses is recovered [3, 11]. From (1) and (5) one obtains the equation of motion for the instantaneous average activity of the neurons

$$m_\mu^{n+1} = \frac{1}{N} \sum_j \xi_j^\mu \tanh\left\{ \beta \sum_\nu \xi_j^\nu m_\nu\left(s^n\right) \; \Theta\left(\left(m_\nu\left(s^n\right)\right)^2 - \frac{\eta^2}{N} \right) \right\} \tag{6}$$

where

$$m_\mu^n \equiv \langle m_\mu\left(s^n\right) \rangle = \frac{1}{N} \sum_{j=1}^{N} \xi_j^\mu \langle s_j^n \rangle \tag{7}$$

with $\langle \cdots \rangle$ denoting the thermal average over the distribution (1). Since $m_\mu^n - m_\mu\left(s^n\right) \simeq O(1/\sqrt{N})$ one can substitute $m_\mu\left(s^n\right)$ by m_μ^n in (6) making an error of

$$O\left(\sqrt{\alpha} \text{Prob}\left[\left(m_\mu^n\right)^2 - \frac{\eta^2}{N} \geqslant 0 \right] \right) \tag{8}$$

where $\alpha = p/N$ and Prob $[x \geqslant 0]$ is the probability that x be greater or equal to zero. In the limit $\alpha \to 0$ this substitution is exact, while for $\alpha \neq 0$ the approximation remains good for large values of η since then the probability that a weakly correlated pattern pass the threshold becomes small. With this approximation, one obtains the mean-field equations

$$m_\mu^{n+1} = \frac{1}{N} \sum_j \xi_j^\mu \tanh\left\{ \beta \sum_\nu \xi_j^\nu m_\nu^n \Theta\left(\left(m_\nu^n\right)^2 - \frac{\eta^2}{N} \right) \right\}. \tag{9}$$

A study of these equations by means of the statistical techniques developed in [15] showed that the storage capacity $\alpha_c = p/N$ can be increased up to any preassigned value (for $N \to \infty$) by taking a sufficiently large value of the threshold. Those results were confirmed by means of a numerical simulation.

3. Sequence recognition with state-dependent synapses

Let us consider a neural network with N neurons and a time series of q random patterns $\{\xi_i^\mu\}$ which obeys (3). We want the network to be capable of following the sequence of patterns in order from $\mu = 1$ to $\mu = p$, where $p < q$. The capacity of the network is again defined by $\alpha = p/N$. The system evolves now in parallel by the transition probability from the state $s' = (s_i')$ to the state $s = (s_i)$

$$\phi\left(s' \to s\right) \equiv \prod_{i=1}^{N} \tfrac{1}{2}\left[1 + s_i \tanh\left(\beta h_i'\right)\right] \tag{10}$$

where h_i' is given by (2) and the synapses now have

$$W_{ij}^n = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{\mu+1}\xi_j^\mu \Theta\left(\left(m_\mu\left(s^n\right)\right)^2 - \frac{\eta^2}{N}\right). \tag{11}$$

The evolution of the probability of the network $P_n(s)$ to be in state s at time n is given by [10, 11]

$$P_{n+1}(s) = \sum_{s'} \phi\left(s' \to s\right) P_n\left(s'\right). \tag{12}$$

The average overlap at step $n+1$ is then

$$m^{n+1} = \frac{1}{N}\left\langle \sum_i \xi_i s_i^{n+1} \right\rangle_{n+1} \tag{13}$$

where

$$\langle A \rangle_n \equiv \sum_s P_n(s)\, A(s). \tag{14}$$

After substituting (10) and (12) in the expression (13) and then summing over s, one finds

$$m^{n+1} = \frac{1}{N}\left\langle \sum_i \xi_i \tanh\left\{\beta \sum_\mu \xi_i^{\mu+1} m_\mu \Theta\left(\left(m_\mu\right)^2 - \frac{\eta^2}{N}\right)\right\}\right\rangle_n. \tag{15}$$

We can substitute the overlaps in the right-hand side of (15) by its thermal average making an error of order given by (8), obtaining

$$m_\mu^{n+1} = \frac{1}{N}\sum_j \xi_j^\mu \tanh\left\{\beta \sum_\nu \xi_j^{\nu+1} m_\nu^n \Theta\left(\left(m_\nu^n\right)^2 - \frac{\eta^2}{N}\right)\right\}. \tag{16}$$

We are interested in finding solutions of (16) which have a large overlap only with the pattern $\mu = n$:

$$\langle\langle m_\mu^n \rangle\rangle = \delta_\mu^n m_n. \tag{17}$$

Let us consider in (16) $\mu = \nu + 1 \neq n + 1$ and expand the term proportional to $m^n_\mu \Theta\big((m^n_\mu)^2 - \eta^2/N\big)$ to first order. One finds

$$m^{n+1}_{\nu+1} = \frac{1}{N}\sum_i \xi_i^{\nu+1}\xi_i^{n+1}\tanh\left[\beta\left(m_n + \Upsilon_i^{n,\nu}\right)\right] + \frac{\beta}{N} m^n_\nu \Theta\left(\left(m^n_\nu\right)^2 - \frac{\eta^2}{N}\right)\sum_i\left[1 - \tanh^2\beta\left(m_n + \Upsilon_i^{n,\nu}\right)\right] \tag{18}$$

where

$$\Upsilon_i^{n,\nu} \equiv \sum_{\mu\neq n,\nu} \xi_i^{\mu+1}\xi_i^{n+1} m^n_\mu \Theta\left(\left(m^n_\mu\right)^2 - \frac{\eta^2}{N}\right). \tag{19}$$

Because m^n_μ, $\mu \neq n$ are the sum of a large number of random variables $\xi_i^\mu \langle s_i^n\rangle$ we may assume that they have normal distributions centred at zero, with variance σ_n^2/N, by the CLT. For $\eta \neq 0$, only the patterns which pass the threshold contribute to (19): the distribution of m^n_μ over the reduced set of patterns is Gaussian for $(m^n_\mu)^2 > \eta^2/N$ and zero in the strip $(m^n_\mu)^2 < \eta^2/N$. We will assume that the number of patterns which pass the threshold is large, so that $\Upsilon_i^{n,\nu}$ has a normal distribution also for $\eta \neq 0$, with variance αr_n and average zero (note, however, that the m^n_μ, $\mu \neq n$ are not strictly independent variables, since they are related through the (16)). Then

$$\alpha r_n = \langle\langle\left(\Upsilon_i^{n,\nu}\right)^2\rangle\rangle \tag{20}$$

and

$$q_n \equiv \frac{1}{N}\sum_i \tanh^2\beta\left(m_n + \Upsilon_i^{n,\nu}\right) = \int\frac{\mathrm{d}z}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\tanh^2\beta\left(m_n + \sqrt{\alpha r_n}z\right). \tag{21}$$

From equation (18) we get

$$m^{n+1}_{\nu+1} - \beta m^n_\nu \Theta\left(\left(m^n_\nu\right) - \frac{\eta^2}{N}\right)(1 - q_n) = \frac{1}{N}\sum_i \xi_i^{\nu+1}\xi_i^{n+1}\tanh\beta\left(m_n + \Upsilon_i^{n,\nu}\right). \tag{22}$$

Squaring this expression, averaging over the distribution of patterns and using (19), (20) and (22) we get for $p \gg 1$

$$\sigma^2_{n+1} - \beta^2(1 - q_n)^2 r_n = q_n. \tag{23}$$

Another relation can be obtained directly from (19) and (20), giving

$$r_n = \frac{2}{\sqrt{\pi}}\sigma_n^2\Gamma\left(\frac{3}{2}, \frac{\eta^2}{2\sigma_n^2}\right) \tag{24}$$

where Γ is the incomplete gamma function [16]. Now taking the mean-field equations for $\mu = n + 1$ we obtain, with similar calculations

$$m_{n+1} = \int\frac{\mathrm{d}z}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\tanh\beta\left(m_n + \sqrt{\alpha r_n}z\right). \tag{25}$$

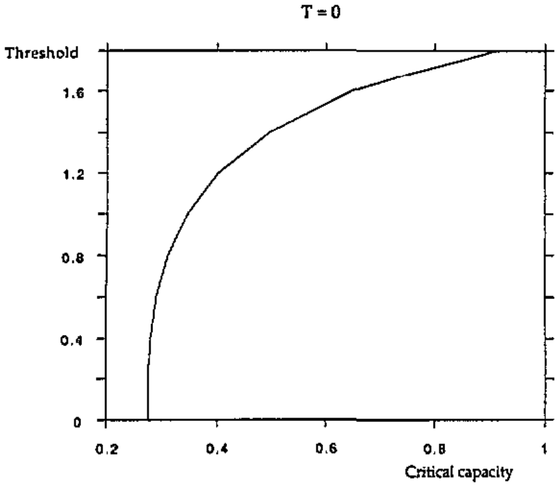

Figure 1. The increase in the storage capacity of the neural network with the threshold is represented, in the deterministic limit $T = 0$. The critical capacity with threshold equal to zero is $\alpha_c = 0.278$. With a threshold $\eta = 1$ the critical capacity increases to 0.36, and at $\eta = 2$ it reaches $\alpha_c = 1.1$

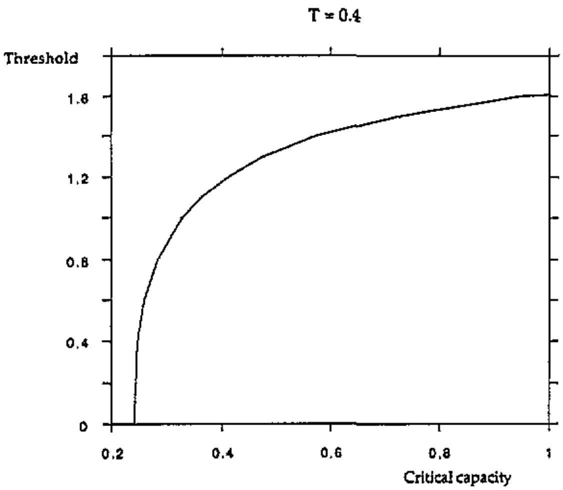

Figure 2. The maximum storage capacity is given as a function of the threshold at $T = 0.4$. For low values of the threshold the capacity is less than in the deterministic case $T = 0$, as thermal fluctuations inhibit a clean recall of the sequence. However, the improvement in the storage capacity with the threshold is stronger than at zero temperature, and for thresholds above $\eta = 1.15$ the capacity of the $T = 0.4$ network surpasses that of the deterministic network.

Equations (21), (23), (24) and (25) are simultaneous equations for the unknowns m_n, q_n, σ_n and r_n. We can solve them in the stationary limit in which the order parameters are constant in time: $m_n = m$, $q_n = q$, $\sigma_n = \sigma$ and $r_n = r$. In figure 1, we describe the space of solutions at $T = 0$: for $\eta = 0$ there exist solutions for $\alpha \leqslant 0.278$, while for $\eta \neq 0$ this value increases as is shown by the critical line $\alpha_c(\eta)$; $m \neq 0$. Above this line one has temporal sequence solutions of the form (11), while below it $m = 0$. For $\eta \to \infty$, $\alpha_c \to \infty$

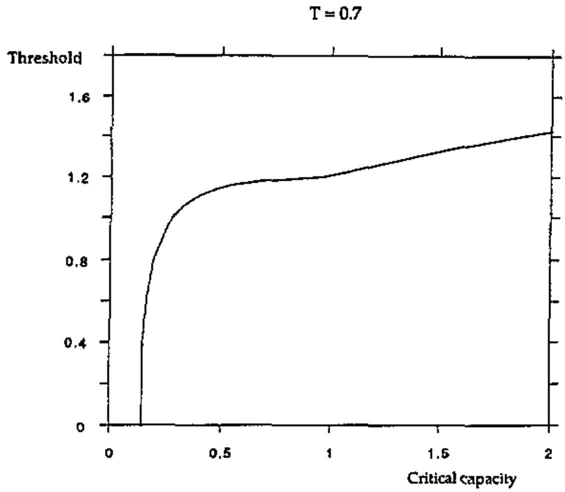

Figure 3. The storage capacity is given as a function of the threshold at $T = 0.7$.

as

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

In figures 2 and 3 we give the critical capacity as a function of the threshold for finite temperatures.

4. Numerical simulation

A numerical simulation of a network of $N = 144$ neurons was carried out (see figure 4). The network was initiated near the first pattern in each set, with an error in one of the neurons, and allowed to evolve according to the deterministic synchronous evolution rule with the synapses (11). The overlap with the last pattern in the sequence was determined, and this result was averaged over 200 sets of randomly generated patterns and over 25 different choices of the erroneous initial neuron. For low values of the threshold the overlap begins to decline sharply at a critical value of the capacity close to the theoretical estimate. For $\eta \geqslant 2$ the decline is smoother and it is difficult to determine accurately the critical capacity. Following the first referee's suggestion, we increased N to 1681 neurons, using the 'trick' described by Penna and Oliveira [17] (1681 is the square of 41; the odd number avoids memory drift conflicts in the vectorized code). The low-η graphs were unchanged, but the new $\eta = 2$ graph displays a sharp drop near the theoretical critical capacity estimate, namely at $\alpha_c \approx 1.1$ (figure 5).

The need for a greater number of neurons for large-η simulations reflects the fact that the number of patterns which pass the threshold must be large, in order to apply the CLT in (16). Since m_μ has a normal distribution with variance $\sigma^2/N \sim 1/N$, this number is of the order $\alpha N(1 - \mathrm{erf}(\eta/\sqrt{2}))$, where erf() is the error function [16]. For $N = 144$, $\eta = 2$ and $\alpha = 1$, an average of only 6.55 patterns pass the threshold, while for $N = 1681$ this number increases to 76.5, enough to invoke the CLT. As this interpretation suggests, there should be no need to increase further the number of neurons; this fact was confirmed by a run at $N = 6561$, which reproduced the same graph as figure 5.

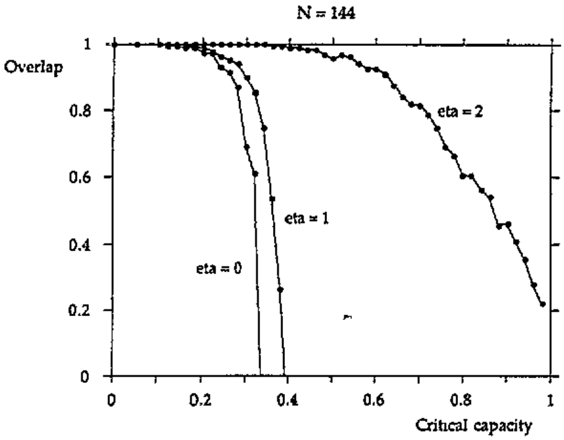

Figure 4. The results of the numerical simulation of the network are represented. The average overlap with the learned sequence is given as a function of the storage capacity $\alpha = p/N$, for various values of the threshold. A sharp decline in the overlap is detected near the theoretical value $\alpha_c = 0.28$ for $\eta = 0$. The increase in capacity from $\eta = 0$ to $\eta = 1$ is roughly equal to 0.06, again close to the theoretical prediction. For $\eta = 2$ the capacity increases substantially: one has an accurate recall of the sequence up to $\alpha = 0.6$, and the overlap with the sequence goes to zero near the theoretical critical value $\alpha_c = 1.1$.

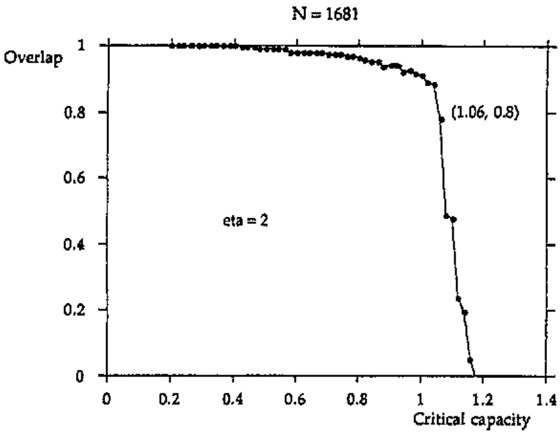

Figure 5. The numerical simulation with $N = 1681$ neurons reveals a critical capacity $\alpha_c \approx 1.1$, in agreement with the theoretical calculation.

5. Conclusions

We have proposed a neural network with state-dependent synapses and synchronous updating capable of storing long sequences of random patterns. In order to achieve this result, we introduced a threshold which cuts off the contribution of the weakly correlated patterns. The critical storage capacity was computed as a function of the threshold parameter and the temperature, and the results for the deterministic case $T = 0$ were confirmed by numerical simulation.

One of the motivations for this work was the possibility of applying this type of network to chaotic time series prediction. There it is necessary to 'teach' the network a large number of patterns, typically of the order of 10^4 for a low-dimensional attractor. Since the present network is capable of storing such a chaotic time series, it is possible that as a dynamical system this network would itself display a chaotic behaviour, with an attractor which approximates reasonably well the attractor of the physical system that generated the time series. Finally, it may be possible to adjust the value of the threshold so that the first Liapunov coefficient of the neural network be equal to that of the time series; this would probably give a subcritical value of the threshold, such that the learned sequence is unstable. This approach to nonlinear modelling suggests the exciting prospect of designing models which come close to topological equivalence with the physical system.

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*, and the Belgian Ambassador to Mexico, for their support of our project. The first author (FZ) has benefited from discussions with P González Casanova and computational advice from G Krötszch.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554; 1984 *Proc. Natl Acad. Sci. USA* **81** 3088
[2] Little W A 1974 *Math. Biosci.* **19** 101
[3] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City, CA: Addison-Wesley) ch 2
[4] Peretto P 1984 *Biol. Cybern.* **50** 51
[5] Amit D, Gutfreund H and Sompolinsky H 1985 *Phys. Rev.* A **32** 1007
[6] Amit D, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530; 1987 *Ann. Phys., NY* **173** 30
[7] Fontanari J F and Koberle R 1988 *J. Physique* **49** 13
[8] Coolen A C C and Ruijgrok Th W 1988 *Phys. Rev.* A **38** 4253
[9] Shiino M 1990 *J. Stat. Phys.* **59** 1051
Nishimori H, Nakamura T and Shiino M 1990 *Phys. Rev.* A **41** 3346
[10] Coolen A C C and Sherrington D 1992 *J. Phys. A: Math. Gen.* **25** 5493
[11] Bernier O 1991 *Europhys. Lett.* **16** 531
[12] Nishimori H and Ozeki T 1993 *J. Phys. A: Math. Gen.* **26** 859
[13] Shukla P 1993 *J. Stat. Phys.* **71** 705
[14] Zertuche F, López-Peña R and Waelbroeck H 1994 *J. Phys. A: Math. Gen.* **27** 1575
[15] Peretto P 1988 *J. Physique* **49** 711
Geszti T 1990 *Physical Models of Neural Networks* (Singapore: World Scientific) ch 4
[16] Arfken G 1970 *Mathematical Methods for Physicists* (New York: Academic) ch 10
[17] Penna T J P and Oliveira P M C 1989 *J. Phys. A: Math. Gen.* **22** L719