Revista Mexicana de Física **42**, *No. 1 (1996) 97–103*

Attraction domains in a Hebb-like neural network with unequally weighted patterns: temperature dependence

L. VIANA*, C. MARTÍNEZ AND E. COTA†
Laboratorio de Ensenada, Instituto de Física
Universidad Nacional Autónoma de México
Apartado postal 2681, 22800 Ensenada, B.C., México
Recibido el 13 de marzo de 1995; aceptado el 23 de agosto de 1995

ABSTRACT. A long range Ising Neural Network is considered, where a finite number p of uncorrelated patterns has been stored with different embedding strengths w_μ according to a modified Hebb rule. A systematic study of the size of the basins of attraction as a function of the weights $\{w_\mu\}$ and the noise level T is performed for $p \leq 4$. This is done by iterating numerically the flux equations for the overlaps between the stored patterns and the dynamical state of the system. The temperature above which all spurious minima disappear as a function of the embedding strengths is obtained. This temperature is found to take its highest value in the Hebb case.

RESUMEN. Se considera una red neuronal de Ising con interacciones de largo alcance en la cual un número finito p de patrones ha sido almacenado con pesos diferentes w_μ de acuerdo con una regla de aprendizaje tipo Hebb modificada. Para $p \leq 4$ se hace una evaluación sistemática del tamaño de las cuencas de atracción de este sistema, como función de los pesos $\{w_\mu\}$ de las memorias y del nivel de ruido T. Esto se hace mediante la iteración numérica de las ecuaciones de flujo para los traslapes entre los patrones almacenados y el estado dinámico del sistema. Se obtiene la temperatura T^*, como función de $\{w_\mu\}$, por encima de la cual desaparecen los estados espurios. Se encuentra que esta temperatura toma su valor más alto en el caso de Hebb.

PACS: 87.30; 75.10; 64.60.I

1. INTRODUCTION

The face of theoretical physics has changed dramatically during the last few years and a large amount of new ideas, concepts and vocabulary such as disordered systems, dynamical systems and chaos has emerged. One of the new paradigms that has come out is that of neural networks (NN), as complex, disordered systems which show properties as content addressable memories. In NN the information is not stored explicitly, what is stored is the affinity of the net to "be" in some particular states, which act as attractors of the dynamics of the system.

In living beings we find that noise is inherent to the functioning of their nervous systems. In the modelling of NN, this factor was first considered by Little [1], who introduced it as the temperature T analogue of spin systems. Since then, the beneficial role played by a moderate amount of noise T in the recoverability of stored information, in this

*E-mail: laura@ifuname.ifisicaen.unam.mx.
†E-mail: ernesto@ifuname.ifisicaen.unam.mx.

kind of systems, has been recognized: This noise allows the system to overcome small barriers and to get to deeper, and therefore more important minima, thus avoiding some possible minima related to non nominated patterns. Amit *et al.* [2] have found that for T above 0.461 no spurious memories exist at all in a NN with a Hebb learning prescription. However, it should be clear that in the presence of any amount of noise there are no fixed points, *i.e.* in the equilibrium state the system is fluctuating among microscopically different but macroscopically equivalent states.

In this paper we will evaluate the effect of noise on the size of the basins of attraction for a modified Hebb model of NN, as a function of the load parameter $\alpha = p/N$, where p is the number of stored patterns $\{\xi_i^\mu\}$, and N is the number of neurons in the system. We will also evaluate the importance of "spurious memories", which are attractors related to mixtures of stored patterns that jeopardize the recovery of meaningful information. For this purpose we consider a Hebb-like learning rule in which patterns are stored with different embedding strengths $\{w_\mu\}$ in order to reflect various degrees of training [3].

The size of the basins of attraction is an important parameter to evaluate because it is directly related to the recoverability of information. Gardner defined stability parameters [4] which were assumed to give an idea about the sizes of the domains of attraction; however, for this model it was demonstrated that these parameters are not a direct measure of the absolute sizes of the attraction domains [5]. Therefore, the use of a more direct method is necessary. For our purpose, we define the cumulative size of the attraction domains f_p as the total fraction of states leading towards any of the p stored memories, that is $f_p = 2\sum_{\mu=1}^{p} f_p(\mu)$ (the factor of 2 comes from the symmetry $\vec{\xi}_\mu \to -\vec{\xi}_\mu$), where $f_p(\mu)$ is the ratio between the number of initial states which eventually evolve towards the μ–th pattern and the total number of possible states, when p patterns are stored.

The evaluation of f_p can be done at the microscopic level, by carrying out the actual simulation of the dynamics in an N–element NN where p elements are stored [6]. Instead, we will make this evaluation at the macroscopic level, by iterating the flux equations for the p overlaps q_μ between the present microstate $\{S_i\}$ of the system and each of the p stored patterns $\{\xi_i^\mu\}$, starting from random initial states. This level of treatment is particularly convenient as microscopical details of the system are generally irrelevant.

A similar evaluation has already been done, in the noiseless case, for small values of p ($p \leq 4$), with some interesting results [7]: It was found that in some regions in the weights $\{w_\mu\}$ space, the flux equations for the overlaps have no attractive fixed points related to mixtures of patterns (these would be characterized by having simultaneously a macroscopic value for two or more overlaps q_μ). This means that any initial state leads the system towards one of the stored patterns. In the case $p = 3$, the region where all attractors correspond to any of the stored patterns, so no spurious memories exist, is defined by $w_2 + w_3 < w_1$, where the convention $0 < w_3 \leq w_2 \leq w_1 = 1$ was used.

2. The flux equations

We will consider a system composed of N neuron like Ising elements S_i, whose (symmetrical) interactions J_{ij} between pairs (ij) reflect the storage of a finite number p of random unbiased patterns $\{\xi_i^\mu\} = \pm 1$, with $\mu = 1, \ldots, p$, according to a modified Hebb rule [8], as given by

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} w_\mu \xi_i^\mu \xi_j^\mu (1 - \delta_{ij}), \tag{1}$$

where $w_\mu > 0$, is the weight associated to the μ–th pattern, and $w_\mu \geq w_\nu$ for $\nu > \mu$, and $w_1 = 1$. In the case where all embedding strengths are equal ($w_\mu = 1$ for all μ), Eq. (1) reduces to the Hebb rule. Following Coolen and Ruijgrok [9], we will derive the flux equations for the overlaps; for that end, we will assume that each spin has a probability $w(S_i \to -S_i)$ to flip, given by

$$w(S_i \to -S_i) = \tfrac{1}{2}\big[1 - \tanh(\beta S_j h_j)\big], \tag{2}$$

where β is a fictitious temperature inverse $\beta = 1/T$, introduced to account for the noise level, and h_j is the "field" at site j due to the presence of the rest of the neurons. Therefore, the time evolution of the probability $p(\vec{S}, t)$ of finding the system at time t in a given state $\vec{S} = (S_1, S_2, \ldots S_N)$ is given by

$$\frac{\partial p(\vec{S}, t)}{\partial t} = \sum_{j=1}^{N} w(-S_j \to S_j) p(F_j \vec{S}, t) - \sum_{j=1}^{N} p(\vec{S}, t) w(S_j \to -S_j), \tag{3}$$

where F_j is a spin flip operator acting on the j–th neuron. We intend to rewrite this expression in the overlaps space, by using the definition of the overlap vector $\vec{q}$ whose μ–th component is given by

$$q_\mu(\vec{S}) = \frac{1}{N} \sum_{j=1}^{N} \xi_j^\mu S_j, \tag{4}$$

together with the connection formulae for the microscopic probability $p(\vec{S}, t)$ and its macroscopic counterpart $P(\vec{q}, t)$, given by

$$P(\vec{q}, t) = \sum_{\vec{S}} p(\vec{S}, t)\, \delta(\vec{q} - \vec{q}_{\vec{S}}). \tag{5}$$

After some mathematical manipulation, and assuming that in this limit ($N \to \infty$, $\alpha \to 0$) strong averaging applies [10], the master equation for the system can be written as

$$\frac{\partial \vec{q}}{\partial t} = F(\vec{q}\,) - \vec{q}, \tag{6a}$$

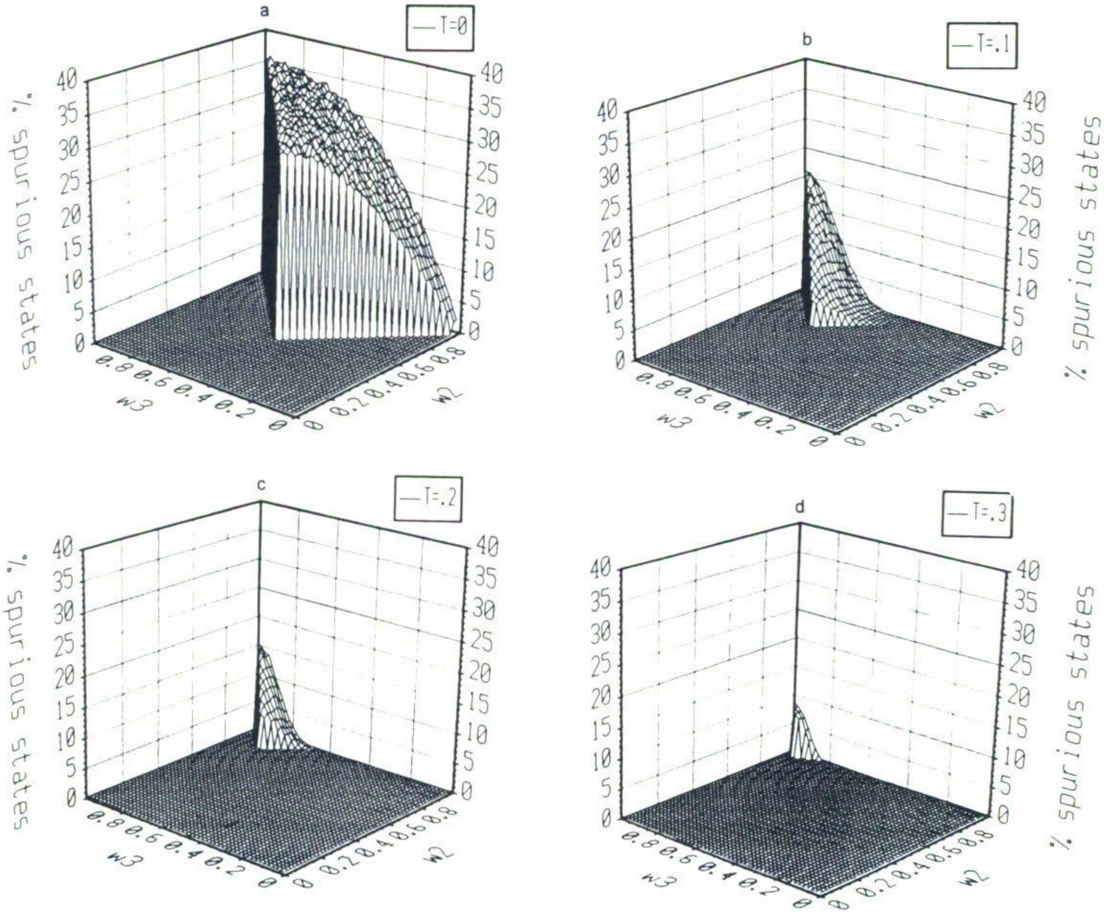

FIGURE 1. Cumulative size of the spurious basins of attraction $f_s = (1 - f_p)$ (multiplied by 100), in the parameters $\{w_\mu\}$ space for several values of T: a) $T = 0$, b) $T = 0.1$ c) $T = 0.2$, and d) $T = 0.3$. Notice that as we considered $0 < w_3 \leq w_2 \leq w_1 = 1$, only the region below the line $w_2 = w_3$ should be taken into account.

with $F(\vec{q})$ given by

$$F(\vec{q}) = \left\langle\!\!\left\langle \vec{\eta} \tanh\left\{\beta \sum_\gamma w_\gamma q_\gamma \eta_\gamma\right\}\right\rangle\!\!\right\rangle_{\vec{\eta}}, \tag{6b}$$

where the double bracket $\langle\!\langle \quad \rangle\!\rangle_{\vec{\eta}}$ indicates averaging over the 2^p corners $\vec{\eta} \in \{-1, 1\}^p$. For synchronous dynamics the corresponding equations are given by

$$\vec{q}(n+1) = F(\vec{q}(n)), \tag{7}$$

with $F(\vec{q})$ given by Eq. (6b). Therefore, numerical iteration of this equation, starting from any initial value for $\vec{q}$ will predict the time evolution of this parameter until a fixed point of the dynamics is reached; that is, until $F^n(\vec{q}) = F^{n-1}(\vec{q})$, for some n.

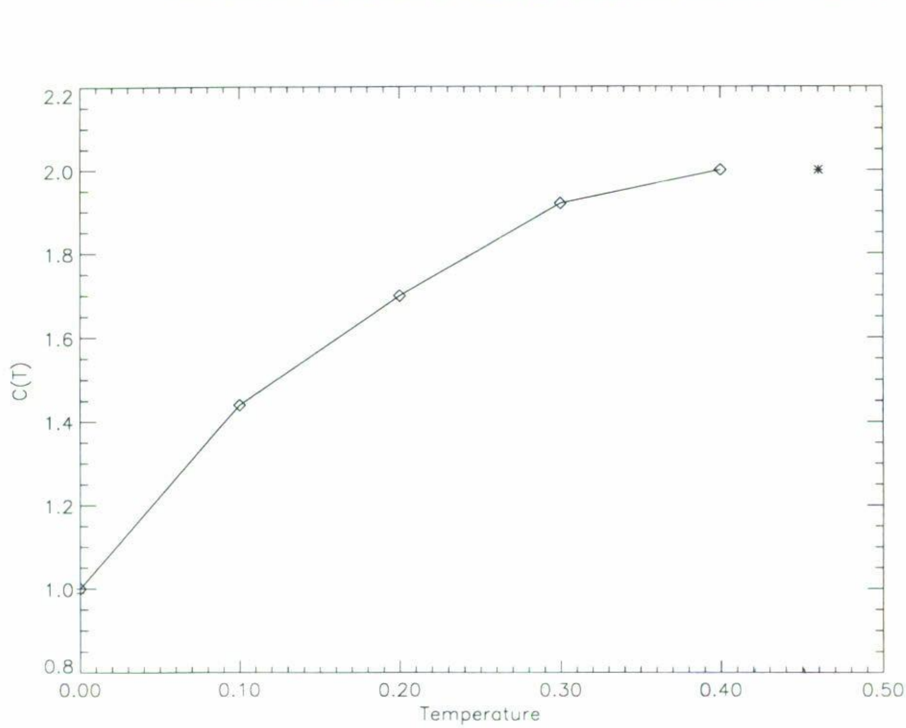

FIGURE 2. There is a line in the parameter space dividing the regions where mixed spurious memories do and do not exist; this line can be written as $w_2 + w_3 = C(T)$. Diamonds ($\diamond$) in Fig. 2 show $C(T)$ *vs.* T as obtained from the simulations, while the asterisk ($*$) indicates the analytical prediction of this temperature for the Hebb model ($w_1 = w_2 = w_3 = 1$) [2,11].

3. DYNAMICAL EVOLUTION

The regions delimited by the planes $\vec{\eta} \cdot \vec{q} = 0$ are convex for any p value, which means that if $\vec{q}_0$ is within one of them, $F(\vec{q}_0)$ is also there. Because of this, it is appropriate to define the set $\Omega_\mu \subset R^p$

$$\Omega_\mu \equiv \left\{ \vec{q}_0 \mid \left|q_\mu^0\right| > \frac{1}{w_\mu} \sum_{\lambda \neq \mu} w_\lambda \left|q_\lambda^0\right| \right\}, \tag{8}$$

Due to the convexity of this set, it has been used to calculate a lower bound to the fraction f_p at $T = 0$ [7]. It has been found that for $p = 1, 2$ and $T = 0$ all $\vec{q}_0 \in \Omega_\mu$ evolve in one single time step towards the μ–th pattern. In the case $p = 3$ and $T = 0$, the behaviour in the parameters' space $\{w_\mu\}$ is divided by the line $w_2 + w_3 = w_1 = 1$ into two different regimes: a) If $w_1 > w_2 + w_3$ then $F^2(\vec{q}) = F(\vec{q})$, for all $\vec{q}$. In this way, if $\vec{q}_0 \in \Omega_\mu$, then $F(\vec{q}) = (0, \ldots, \operatorname{sgn}[q_\mu], 0, \ldots)$, so its elements evolve towards the μ–th pattern in one single time step; on the other hand, if $\vec{q}_0 \notin \bigcup \Omega_\mu$ the system will evolve towards a spurious memory in a single time step. b) For $w_1 < w_2 + w_3$, no spurious minima exist, so the system will evolve towards any of the stored patterns in either one or more time steps depending on whether $\vec{q}_0$ is, or is not, inside a region Ω_μ. For $p = 4$ there are also small regions in the $\{w_\mu\}$ space, where no spurious minima exist; however these regions

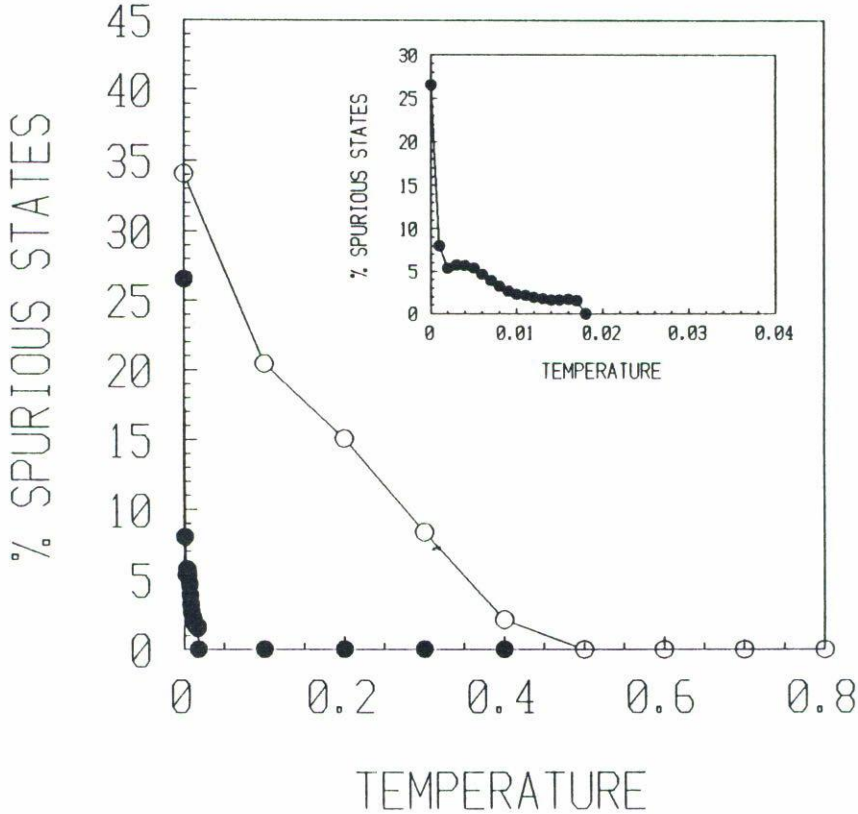

FIGURE 3. **Cumulative size of the basins related to spurious memories $f_s = (1 - f_p)$ (multiplied by** 100) as a function of T for a typical case with $w_2 = .7$, $w_3 = .4$ (filled circles $\bullet$) and for the Hebb model (open circles $\circ$). In this model, the Hebb case is the one who looses its spurious memories at a higher temperature. The inset shows an enlargement of the region with low T, for the modified Hebb model.

are not so clearly defined as in the case $p = 3$, and the transition between regions with and without spurious memories is soft (second order) [7].

We iterated Eq. (6b), for several noise levels T, in order to find f_p for several temperatures, with initial values for $\vec{q}_0$, given by a gaussian distribution with zero mean and width $\sigma = 10^{-5}$, which would correspond to a system of size $N \sim 10^{10}$. This is the only size dependence of our results, since the Eqs. (6b) are exact in the thermodynamical limit; however σ was chosen within the range where results were found to vary very little as a function of σ. As expected, for $p = 1, 2$ we found identical results to those in the noiseless case: there are no spurious minima and the basins of attraction are clearly defined by the regions Ω_μ.

Figure 1 shows the cumulative size of the spurious basins of attraction $f_s = (1 - f_p)$ (multiplied by 100) for $p = 3$, for various noise levels, in the parameters space $\{w_\mu\}$. As we can see, f_s has a behaviour similar to that found at $T = 0$: It presents an abrupt transition between two regions where spurious minima do and do not exist; the size of the region with spurious minima, in this space, diminishes as T grows. These two regions are separated by a line which can be written as $w_2 + w_3 = C(T)$; Fig. 2 shows $C(T)$ for several temperatures, where the diamonds' length indicates the uncertainty in the measurements. $C(T)$ is an important quantity because the mixed spurious states composed by 3 memories are the ones which disappear at higher temperatures; therefore, if we know the value for $w_2 + w_3$, we can estimate from this graph, the temperature above which no spurious memories exist for $p = 3$. Although the estimation of this temperature from $C(T)$ becomes less accurate as we get closer to the upper limit of our model, the Hebb model, the tendency towards the theoretical prediction for the latter, $T \approx 0.46$ [2,11], can be clearly observed, this is shown by an asterisk ($*$) in Fig. 2.

Figure 3 shows the percentage of states occupied by the cumulative basin of attraction of the spurious memories, as a function of the noise level T, for a particular case with $p = 3$, $w_2 = .7$, $w_3 = .4$ (filled circles $\bullet$), and for the Hebb learning prescription (open circles $\circ$). The domain size f_p of this model has its lowest value in the case $w_\mu = 1$, for all μ, that is, for the Hebb model (for any p). We also considered a few cases for $p = 4$, and found regions with no spurious states for all T; however,in this case, there is a second order transition between the two kinds of regions.

It can be concluded that this model presents the following characteristics as compared with the usual model with the Hebb learning prescription: First, it accounts for different degrees of training, since the size of the basins of attraction can be controlled by varying the relative sizes w_μ / w_ν. Second, it has a higher cumulative size f_p than that for the Hebb prescription, therefore the possibility of the system getting trapped by any of the spurious minima is much lower. Finally, the amount of noise necessary to improve retrieval is also lower.

ACKNOWLEDGEMENTS

One of the authors (LV) wishes to thank MC. Armando Reyes for his computing support. This work was partially financed by project DGAPA IN100294 of the National University of México.

REFERENCES

1. W.A. Little, *Math. Biosci.* **19** (1975) 101.
2. D.J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev.* **A 32** (1985) 1007; *Phys. Rev. Lett.* **55** (1985) 1530.
3. L. Viana, *J. Physique* **49** (1988) 167.
4. E. Gardner, *J. Phys.* **A21** (1988) 257.
5. A.C.C. Coolen, *Europhys. Lett.* **16** (1991) 73.
6. L. Viana, E. Cota and C. Martínez, in Statistical Mechanics of Neural Networks, L. Garrido, Ed. *Lecture Notes Phys.* Springer Verlag **368** (1990) 97; C. Martínez, E. Cota and L. Viana, *Rev. Mex. Fís.* **36** (1990) 415.
7. L. Viana and A. C. C. Coolen, *J. Phys. I France* **3** (1993) 777.
8. D.O. Hebb, *The Organization of Behaviour*, Wiley, New York (1949).
9. A.C.C. Coolen and Th. W. Ruijgrok, *Phys. Rev.* **A38** (1988) 4253.
10. J.L. Van Hemmen, D. Grensing, A. Huber and R. Kuhn, *Z. Phys.* **B 65** (1986) 53.
11. D.J. Amit, *Modeling Brain Function*, Cambridge University press (1989).