CONVERGENCE TIME ON THE RS MODEL FOR NEURAL NETWORKS

T. J. P. PENNA and P. M. C. de OLIVEIRA

Instituto de Física, Universidade Federal Fluminense,
C. P. 100296, 24020 - Niterói, RJ, Brazil

J. J. ARENZON, R. M. C. de ALMEIDA and J. R. IGLESIAS

Instituto de Física, Universidade Federal do Rio Grande do Sul,
C. P. 15051, 91500 - Porto Alegre, RS, Brazil

Convergence times and the corresponding dispersions have been studied numerically as parameters to measure the efficiency of neural network models. These quantities are also supposed to be related to the number of spurious states for each configuration of stored patterns. In this work we measure these quantities for a recent multineuron interaction model presenting an enhanced performance compared to other traditional schemes.

Neural networks are systems with associative memeory characteristics. The performance of a neural network as an associative memory device can be measured by some parameters like the quality of retrieval, robustness to destruction of connections, the convergence time, the size of basins of attraction and the storage capacity. The most extensively studied model of neural networks is the one originally proposed by Hopfield in 1982.[1] This model works as a good content-addressable memory device for uncorrelated patterns if $\alpha < 0.14$ ($P = \alpha N$), where P is the number of stored patterns and N is the number of neurons. In a recent work,[2] de Almeida and Iglesias introduced a new model (the RS model) where all the stored patterns and only them are global minima of the Hamiltonian, regardless of the loading of the net. This feature leads to an enhanced performance compared to Hopfield-like models. The quality of retrieval and the size of the basins of attraction for this new model are studied elsewhere.[3]

Whereas the retrieval and storage capacity has been studied by analytical methods, mainly for fully-connected neural networks,[4] numerical simulations[3,5-10] appear as the main technique to study dynamical quantities like, for instance, the convergence time. Kohring[5] used a multispin coding algorithm[6] to simulate the largest neural networks ($N = 10^5$) known until now and he has shown that, in the Hopfield model, larger networks have different behaviour than smaller ones

$(N \approx 10^3)$ (Ref. 7) concerning dynamical aspects. Kohring also showed that the average convergence time $\langle T \rangle$ (number of steps to reach a stationary state) is not a simple function of $\ln(N)$ as proposed by Kanter.[7] Also the dependence of the dispersion in convergence times, $\sigma = \langle T^2 \rangle - \langle T \rangle^2$, on the size of the net N, seems to be related to the structure of the phase space around stored memories:[8] if there is a great number of metastable states around the true information, there are also many different paths leading the stimulus to information. Besides the increase in the average time of convergence, this diversity of paths also implies in larger dispersions σ. In his paper, Kanter[7] proposed that σ would grow as $\ln(N)$, which is actually confirmed for nets with less than 10^3 neurons. For larger nets $(N > 4000)$, as shown by Kohring, this quantity decreases, presenting a different behaviour compared to small nets and highlighting the extreme nonlinear character of those systems.

In this paper, we will present the results of simulations on the RS models[2] which presents an enhanced storing capacity compared to other models.[3] We show also that the performance concerning dynamical aspects is improved. Briefly introducing the model: the Hamiltonian can be written as

$$H = N \prod_{\mu=1}^{P} \left(\frac{1}{2N} \sum_{i=1}^{N} (\xi_i^\mu - S_i)^2 \right) , \tag{1}$$

where ξ_i^μ means the state (± 1) of the ith-neuron in stored pattern μ and S_i means the current state of the ith-neuron in the network. This is a multi-neuron interaction model, i.e., the synpatic connections present products between different memories, and terms linking more than two neurons appear. Defining the overlaps between the patterns and the current state of the network as usual:

$$m_\mu = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu S_i , \tag{2}$$

thus we can rewrite Hamiltonian (1) as

$$H = N \prod_{\mu=1}^{P} (1 - m_\mu). \tag{3}$$

Due to the symmetry of the Hopfield model Hamiltonian, the antimemories (the negative image of each memory) also correspond to energy minima if the memories do. It can be seen from Eq. (1) that, in this new model, the antimemories are no more minima but maxima of Hamiltonian. However, if one wishes also to store explicitly the antimemories, the Hamiltonian can be written as

$$H = N \prod_{\mu=1}^{P} (1 - m_\mu^2). \tag{4}$$

Here we treat only the case where both memories and antimemories are stored, because the net exhibits a better quality of retrieval in this case.[3]

In order to save computer memory, we used a version of multineuron coding[6] to evaluate the overlaps m_μ. One neuron is stored as one bit in an integer computer word, and the overlaps are evaluated by means of logical operations as described below:

$$Nm_\mu(t) = N - 2 \sum_{j=1}^{N/B} \text{POPCOUNT}(\zeta_j^\mu \otimes \sigma_j(t)) \qquad (5)$$

where ζ_j^μ is the binary representation of ξ containing the states of B neurons ($B = 16$, 32 or 64 is the computer word length), $\sigma_j(t)$ is the binary representation of the current state of the net $S(t)$ and POPCOUNT is the function which counts the bits set to one in a given computer word. The symbol $\otimes$ represents the XOR (eXclusive OR) bitwise operation. Within this model the advantage of this technique is essentially memory saving, whereas in the Hopfield model it also decreases computing time by a factor up to 35 for a single processor.[9] The time saving is ruled out in this model by the product in Hamiltonian (4). Due to the large range of numbers required by this operation, floating point numbers are used for large networks or small initial overlaps, i.e., for large energy values.

We simulated networks with up to 4096 neurons, measuring the time spent for convergence $\langle T \rangle$ and the dispersion σ for different initial overlaps, load parameters α and activities of stored patterns (that can be understood as correlations), for several sets of networks. The convergence time is obtained also for the non-retrieval phase $(m(t = 0) < m_c)$, where m_c is the minimum initial overlap for which the retrieval is successful, i.e., the final overlap with the chosen memory is one for all trails. This definition is different from that used by Forrest[10] and in Ref. 3, where a finite size scaling approach for determination of critical overlap in the thermodynamical limit is shown. We used randomly generated patterns and different initial overlaps in order to compare our results with those extracted by Kohring[5] for the Hopfield model. The critical overlap for the new model is, in this case, $m_c \approx 0.25$. We do not need to simulate larger networks due to the behaviour shown in Fig. 1 where, for $m > m_c$, $\langle T \rangle$ does not increase with N, in contrast with the Hopfield model. We averaged different inputs for each set of stored information, i.e., different set of synaptic connections. The number of sets and inputs averaged depends on the network size and the α value, such that for larger networks and higher α, i.e., the situations where the computational effort is larger, we used more different sets, up to 50 for $N = 4096$, and less different inputs, up to 10 per set. For the smaller nets we used 50 inputs to 8 sets.

As shown in Fig. 1, in the retrieval phase only few steps were needed to reach the stable state, and this does not seem to depend on the network size in opposition to the Hopfield model, in which the larger the network the larger the number of steps. This is an important feature for practical applications and implies an enhancement in the dynamical performance of the net in comparison with other models because,
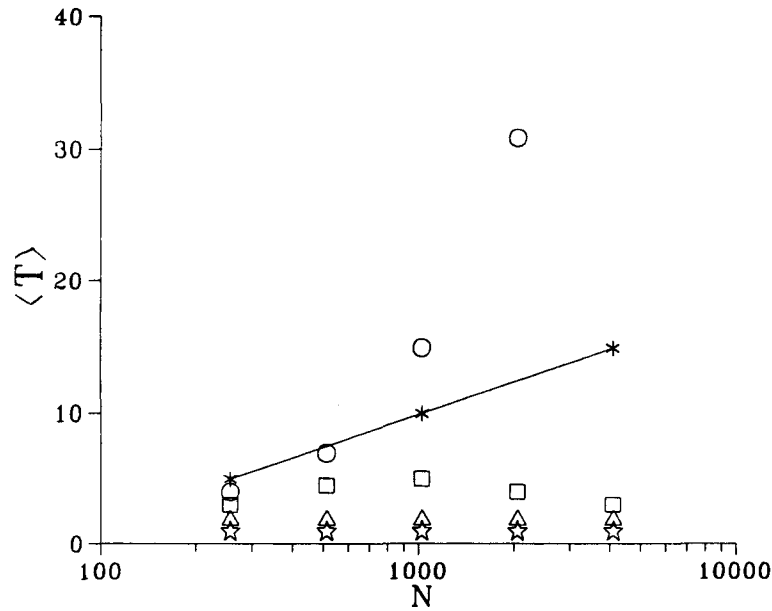
Fig. 1. Convergence time (T) as a function of N at $\alpha = 0.1$ for different intial overlaps. The data represented by asterisks are those extracted by Kohring for the Hopfield model $(m(t = 0) = 0.35)$. The critical overlap varies from 0.27 for $N = 256$ to 0.20 for $N = 4096$.
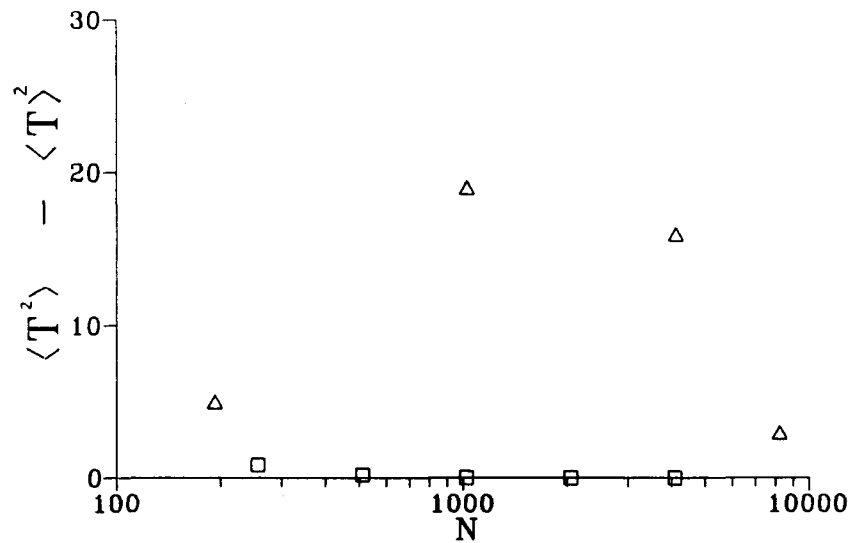

Fig. 2. Variance of the distribution of (T) as a function of N for the RS and Hopfield models. We consider $m(t=0)=0.4$ and $\alpha=0.1$.

as recognition by human brains is a fast process, corresponding to few updates of the neurons, the response of an ideal model network to a stimulus should be as fast as possible.

The plot of the dispersion σ versus the size of the net N, presented in Fig. 2, leads to similar conclusions: the vanishing dispersions reflect the smoothness and isotropy of the phase space around the minima because the possible different paths to reach the equilibrium do not imply in very different times of convergence.

In Fig. 3, we show the convergence time for different sizes and activities (the activity a is the fraction of active neurons in each pattern). When correlations are introduced (i.e., when $a \neq 50\%$) the time for convergence is less than for the uncorrelated case, in non-retrieval phase. One possible explanation for this fact is that, for the correlated case, attractors corresponding to neighbouring basins are closer to each other and the dynamics leads quickly the network to one of them, instead of crossing the phase space. On the other hand, in this specific case, the size of the basins of attraction goes to zero in the limit $N \approx \infty$ (Ref. 3).
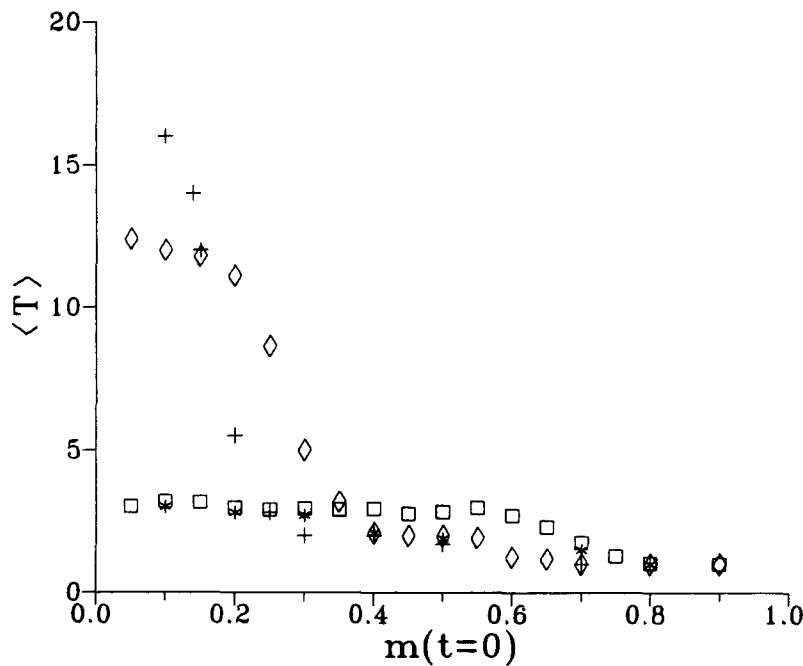

Fig. 3. Convergence times as a function of initial overlap for two different sizes ($N = 512, 1024$). Note that the convergence times are smaller in non-retrieval phase if correlations are introduced. The load parameter is $\alpha = 0.2$.

Finally one can see in Fig. 4, the behaviour of the convergence time for different values of α; again the model presents a good performance in the retrieval phase, i.e., few steps are needed to reach equilibrium. However in non-retrieval phase the
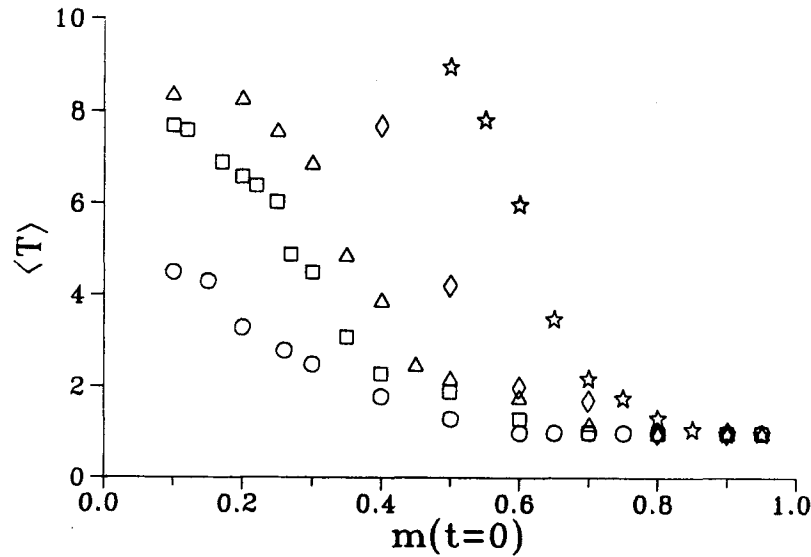
Fig. 4. Convergence times as a function of the initial overlap for different values of the load parameter α. The retrieval phase can be identified as the region where one or two steps are needed to reach a stable state. We used $N = 256$ neurons.

number of steps needed to reach the equilibrium grows with the number of neurons.

In summary we presented results concerning the dynamics of the RS neural network model. The average convergence time is very small, if the retrieval is successful, and depends only on the initial overlap and on the parameter α, independent of the size of the net. The dispersion of these convergence times is also small, indicating a homogeneity on the phase space.

Acknowledgments

References

1. J. J. Hopfield, *Proc. Acad. Sci. (USA)* **79**, 25 (1982).
2. R. M. C. de Almeida and J. R. Iglesias, *Phys. Lett.* **A146**, 239 (1990).
3. J. J. Arenzon, R. M. C. de Almeida, J. R. Iglesias, T. J. P. Penna, and P. M. C. de Oliveira, preprint, 1990.
4. D. J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. Lett.* **55**, 1530 (1985); *Phys. Rev.* **A32**, 1007 (1985); *Ann. Phys. (N.Y.)* **173**, 30 (1987); *Phys. Rev.* **A35**, 2293 (1987).
5. G. Kohring, *J. Phys.* **A23**, 2237 (1990).
6. T. J. P. Penna and P. M. C. de Oliveira, *J. Phys.* **A22**, L719 (1989); P. M. C. de Oliveira, in *Computing Boolean Statistical Models* (World Scientific, 1991).
7. I. Kanter, *Phys. Rev.* **A40**, 2611 (1990).

8. G. Koring, *J. Stat. Phys.* **59**, 1077 (1990); J. Komlós and R. Paturi, *Neural Networks* **1**, 239 (1988); E. Gardner, *J. Phys.* **A19**, L1047 (1986).
9. G. Kohring, *Int. J. Mod. Phys.* **C1**, 259 (1990).
10. B. M. Forrest, *J. Phys.* **A21**, 245 (1988).