BASINS OF ATTRACTION AND SPURIOUS STATES IN NEURAL NETWORKS
by L. Viana, E. Cota and C. Martínez,
Lab. de Ensenada, Instituto de Física, UNAM, A.Postal 2681, 22800 Ensenada B.C., México.

Abstract. A long range Ising Neural Network is considered, where patterns have been stored according to either the Hebb rule or a modification of it which stores patterns with different weights. By performing computer simulations, the size of the basins of attraction of pure and spurious memories is evaluated in the absence of noise, as a function of the load parameter of the system. It is found that the use of the modified Hebb prescription, decreases considerably the percentage of configuration space occupied by spurious memories, which translates into an improvement in the retrieval capabilities of the network.

1. Introduction.

In the last few years, there has been a growing interest in studying the global behaviour of Neural Networks (NN) due to their features as associative fault-tolerant memories. In this way, we find professionals of various fields such as neurobiologists, computer scientists, psychologists, physicist, etc, working on common grounds and trying to find results from their various points of view. The works presented in this conference reflect some of the interests within this wide field of current research.

Statistical physicists became interested in this problem after the pioneering work of Hopfield [1], who made a mathematical analogy between an assembly of neurons and certain disordered magnetic materials called Spin Glasses [2]. This connection allowed the use of methods developed in Statistical Mechanics to study some general properties of NN in the thermodynamic limit. In particular, when the interactions between elements are symmetric, it is possible to assign an energy function to the system.

The retrieval capabilities of NN are a natural consequence of their dynamics, as by reducing their energy, they evolve spontaneously towards a minimum of the free energy which has a large 'overlap' [3] with the initial state. Therefore the final state of the network will
depend on both the initial state and the energy 'landscape' which is determined by the connection coefficients \( \{J_{ij}\} \). In this way, learning is related to a suitable choice of the specific set \( \{J_{ij}\} \) which favours particular configurations by making them minima of the Hamiltonian. However, it has been found that in the process of storing information, other undesired minima appear; these minima also act as attractors and therefore have a negative influence on the retrieval capabilities of the system. We will call these states 'spurious' as opposed to the 'pure' states we stored on purpose.

For symmetrical NN \( (J_{ij} = J_{ji}) \), the analytical methods of statistical mechanics give us important information about the existence of equilibrium states (both pure and spurious). In this way, we know that spurious states are related to mixtures of pure memories and that their number grows very rapidly with the number of stored patterns. However, more relevant than the number of spurious minima is the total percentage of configuration space occupied by their domains of attraction, as this parameter will directly affect the retrieval capabilities of the network. If we are interested in obtaining a more complete picture of the configuration space, it is necessary to use other complementary techniques such as computer simulations.

2. Procedure

In this paper we carry out computer simulations for a long range Ising NN composed by \( N \) neurons-like elements, whose dynamics, in the absence of noise, is governed by an energy function given by:

\[
H = -\left(\frac{1}{2}\right) \sum_{i<j} J_{ij} S_i S_j,
\]

where \( S_i \) denotes the state of the \( i \)th neuron with \( S_i = +1 (-1) \) for a firing (quiescent) neuron; and \( J_{ij} = J_{ji} \) represents the 'synaptic strength' between neurons \( i \) and \( j \) given by a modified Hebb rule, designed to reflect training [4]. This rule assumes that \( J_{ij} \) will be given by the sum of \( p \) random patterns \( \{ \xi_i^\mu \} \), according to

\[
J_{ij} = \left(\frac{1}{N}\right) \sum_{\mu} J_{ij} \xi_i^\mu \xi_j^\mu. \tag{2}
\]