Chapter 2

Biological

&

Artificial Neural Networks,

The Physical Approach Development

2-1 Biological Neural Networks

2-1-1 The Structure of the Neuron

The detailed investigation of the internal structure of neural cells has taken a long time to reveal its composition. Especially after the invention of the electron microscope some 50 years ago, the internal structure of the neural cells has revealed that all neurons are constructed from the same basic parts, independent of their size and shape (see Figure (1)). The bulbous central part is called the cell body or *soma*; from it, project several root-like extensions, the *dendrites*, as well as a single tubular fibre, the *axon*, which ramifies at its end into a number of small branches.

The size of the soma of a typical neuron is about 0-80 μm, while dendrites and *axons* have a diameter of a few μm. While the dendrites serve as receptors for signals from adjacent neurons, the axon's purpose is the transmission of the generated neural activity to other nerve cells or to muscle fibres. In the first case, the term *interneuron* is often used, whereas the neuron is called a *motor neuron* in the latter. A third type of neurons, which receive information from muscles or sensory organs, such as the eye or ear, is called a *receptor neuron.*

The joint between the end of an axonic branch, which assumes a plate-like shape, and another neuron or muscle, is called a *synapse*. At the synapse the two cells are separated by a tiny gap only about 200 *nm* wide (the synaptic gap or *cleft*), barely visible to R. Y. Cajal, but easily revealed by modern techniques. Structures are spoken of in relation to the synapse as *presynaptic* and *postsynaptic*, e.g. postsynaptic neuron. The synapses may be located either directly at the cell body, or at the dendrites of the subsequent neuron, and their strength generally diminishing with increasing distance from the cell body.
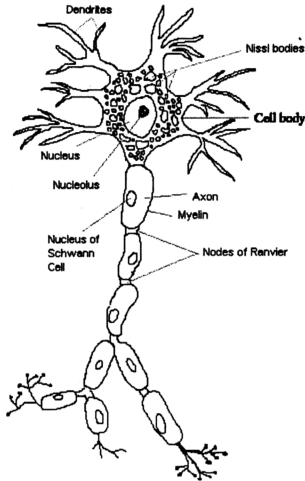
Figure (1). Structure of a typical neuron (schematic)

The total length of neurons shows great variations: from 0.01 *mm* for interneuron in the human brain, up to 1 *m* for the neurons in the limb.

2-1-2 The Inter-Neuronal Communication Mechanism

Nervous signals are transmitted either electrically or chemically. Electrical transmission prevails in the interior of a neuron, whereas chemical mechanisms operate between different neurons, i.e. at the synapses. Electrical transmission is based on an electrical discharge, which starts at the cell body and then travels down the axon to the various synaptic connections. In the state of inactivity, the interior of the neuron, the *protoplasm.*
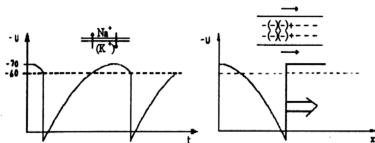
Figure (2). The mechanism of the propagation of the excitation wave.

The interval between two electrical spikes can take any value (longer than the regeneration period) and the combination of analogue and digital signal processing is utilised to obtain optimal quality, security, and simplicity of data transmission.

The speed of propagation of the discharge signal along the nerve fibre also varies greatly. In the cells of the human brain, the signal travels with a velocity of about 0.5 to 2.0 meter per second. Consequently, this would allow any two cells of the brain to communicate within 20-40 ms, which is something like a *temporal quantum* in the operation of the human central nervous system. However, it would cause unacceptably long reaction times for peripheral neurons connecting brain and limbs. It means that a person would hit the ground before even knowing that he had stumbled. To increase the speed of propagation, the axon for such neurons are composed of individual segments that are covered by an electrically insulating myelin sheath, which is interrupted from time to time at the so-called *Ranvier nodes* (see Figure (1)). The presence of an insulating cover causes the signal to propagate along the axon as in a wave guide from one Ranvier node to the next, triggering almost instantaneous discharge within the whole myelinated segment. This mode of propagation, called *salutatory conduction,* allows for transmission velocities of up to 100 *m/s.*

is negatively charged against the surrounding neural liquid. This resting potential of about -70 mV is supported by the action of the cell membrane, which is impenetrable for Na^+ ions, causing a deficiency of positive ions in the protoplasm.

Signals arriving from the synaptic connections result in a transient weakening, or *depolarisation,* of the resting potential. When this is reduced below -60 mV, the membrane suddenly loses its impermeability against Na^+ ions, which enter the protoplasm and neutralise the potential difference, as illustrated in the left part of the Figure (2). This discharge may be so violent that the interior of the neuron even acquires a slightly positive potential against its surroundings. The membrane then gradually recovers its original properties and regenerates its original resting potential over a period of several milliseconds. During this recovery period, the neuron remains incapable of further excitation. When the recovery is completed, the neuron is in its resting state and can "fire" again. The discharge, which initially occurs in the cell body, then propagates along the axon to the synapses. Because the depolarised parts of the neuron are in a state of recovery and cannot immediately become active again, the pulse of electrical activity always propagates in one direction: away from the cell body.

Since the discharge of each new segment of the axon is always complete, the intensity of the transmitted signal does not decay as it propagates along the nerve fibre. One might be tempted to conclude that the signal transmission in the nervous system is of *digital nature*: the neuron is either fully active or it is inactive. However, this conclusion would be wrong, because the intensity of a nervous signal is coded in the frequency of succession of the invariant pulses of activity, which can range from 1 to 100 per second.

2-1-3 The Synapses

The discharge signal travelling along the axon comes to a halt at the synapses, because there is no conducting bridge to the next neuron or muscle fibre. Transmission of the signal across the synaptic gap is mostly effected by chemical mechanisms. Direct electrical mechanism is also known to occur in rare cases, but is of less interest here. The reason behind this disinterestedness is the much lower degree of adjustability of this type of synapse. In chemical transmission, when the spike signal arrives at the presynaptic nerve terminal, special substances called *neurotransmitters* are liberated in tiny amounts from vesicles contained in the endplate (e.g. about 10^{-17} mol acetylcholin per impulse). The transmitter release appears to be triggered by the influx of Ca^{++} ions into the presynaptic axon during the depolarisation caused by the flow of Na^{+} ions. The neurotransmitter travels across the synaptic cleft, as shown in Figure (3), reaching the postsynaptic neuron (or muscle fibre) within about 0.5 milliseconds. Upon their arrival at special receptors, these substances modify the conductance of the postsynaptic membrane for particular ions (Na^{+}, K^{+}, CL^{-}, etc.). Then this modification flows in or out of the neuron, causing a polarisation or depolarisation of the local postsynaptic potential. After their action, the transmitter molecules are quickly broken up by enzymes into pieces that are less potent in changing the ionic conductance of the membrane.

If the induced polarisation potential δU is positive i.e. if the total strength of the resting potential is reduced, the synapse is termed *excitatory*, because the influence of the synapse tends to activate the postsynaptic neuron. If δU is negative, the synapse is called *inhibitory*, since it counteracts excitation of the neuron. Inhibitory synapses often terminate at the presynaptic plates of other axon, inhibiting their ability to send

neurotransmitters across the synaptic gap. In this case, one speaks of presynaptic inhibition. There is evidence that all the synaptic endings of an axon are either of an excitatory or an inhibitory nature (*Dale's law*). Besides, there are significant structural differences between those two types of synapses (e.g., the conductance for Na^+ and K^+ changes at excitatory synapses and that for Cl^- at inhibitory synapses).
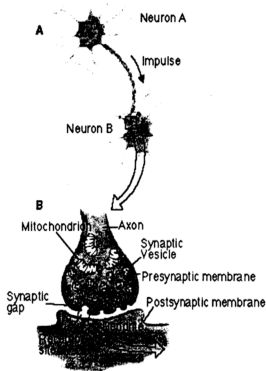

Figure (3). Synaptic-junctions Transmission Mechanism of the Neuron Excitation Wave.

However, an insisting question arises about the activity of the postsynaptic neuron. Under which condition is the postsynaptic neuron stimulated to become active? Although, in principle, a single synapse can inspire a neuron to fire, this is rarely so, especially if the synapse is located at the outer end of a dendrite. Just as each axon sends synapses to the dendrites and bodies of a number of downstream neurons, so each neuron connected to many upstream neurons, which transmit their signal to it. The body

21

of a neuron acts as a kind of *summing* device, which adds the depolarising effects of its various input signals. These effects decay with a characteristic time of 5 to 10 milliseconds, but if several signals arrive at the same synapse over such period, their excitatory effects accumulate. A high rate of repetition of firing of a neuron therefore expresses a large intensity of the signal. When the total magnitude of the depolarisation potential in the cell body exceeds the critical threshold (about 10 mV), the neuron fires.

The influence of a given synapse therefore depends on several aspects: the inherent strength of its depolarising effect, its location with respect to the cell body, and the repetition rate of the arriving signals. There is strong evidence that the inherent strength of a synapse is not fixed and that it will not remain at that state. As originally postulated by D. Hebb [1], the strength of a synaptic connection can be adjusted, if its level of activity changes. An active synapse, which repeatedly triggers the activation of its postsynaptic neuron, will grow in strength, while the others will gradually weaken. The mechanism of synaptic *plasticity* in the structure of neural connectivity, known as Hebb's rule, appears to play a dominant role in the complex process of learning.

The release of neurotransmitters as well as their action at the receptor sites on the postsynaptic membrane can be chemically inhibited by substances such as atropine or curare etc. Similar but less drastic changes of the synaptic efficacy are likely to occur naturally in the body, giving all synapses a certain degree of *plasticity*. Another point worth mentioning is that transmitter substance is randomly emitted at every nerve ending in quanta of a few 1000 molecules at a low rate. The rate of release is increased enormously upon arrival of an impulse, a single action potential causing the emission of 100 to 300 quanta within a very short time. However, even the random low-activity level

results in small depolarisation potentials, which may cause, from time to time, the spontaneous activation of the postsynaptic neuron.

2-2 Physics and Artificial Neural Networks

2-2-1 Historical Evolution

Historically, Crag and Temperley were the first physicists who mentioned the analogy between the activity of neural network and the collective states of coupled magnetic dipoles, in a paper which did not get much attention [2]. Twenty years later, Little [3] made another analogy with the magnetic system, where he related between the synaptic noise and temperature. Furthermore, he suggested that persistent firing states of neural network dynamics appear just like the ordered phases of the magnetic system. He argued for the relevance and applicability of the statistical mechanics concepts to the theory of neural networks. Other, but more important analogy was made by Hopfield ([4], [5]) who completed successfully the series of analogy between the magnetic system and neural networks by introducing the concept of *computational energy*. In his contribution, he emulated the computational energy of the neural networks to the Hamiltonian of the magnetic system. He studied the network with symmetric couplings, and therefore described the system by a Hamiltonian, and showed the equivalence between the asymptotic dynamical behaviour of such networks and equilibrium thermodynamical properties of random magnetic systems similar to spin glasses. The model (which had taken his name) was subsequently solved analytically using the powerful tools of statistical mechanics by Gutfreund et al [6]. This was the first successful and non-trivial application of statistical mechanics in this field, giving rise to a variety of new and surprising results. The last contribution was the work of Gardner [7], who proposed a

new approach, applying statistical mechanics in the space of possible networks. The Gardner approach allows however, the calculation of certain limits on the storage capacity and computational capability.

2-2-2 Statistical Mechanics Analysis Tool

During the last century, statistical mechanics has improved its tools perfectly in studying the emergence of collective behaviour of large interacted microscopic elements. Its usage has been successful to explain the macroscopic properties of the solids and gases using their basic units (atoms and molecules) properties. Consequently, statistical mechanics could be applied to study the emergence of the collective behaviour of intelligent elements such as neurons. Furthermore, as neural networks are processing information and as the relation between information and thermodynamics was established long time ago, ever since the work of Von Neumann and Shannon, then the use of statistical mechanics in neural networks is justified. Despite these arguments, the application of statistical mechanics to study neural networks is not straightforward. This is because of the non-symmetry of the interaction, between the biological neurons, transmitted through the synaptic strengths, whereas, statistical mechanics is based in its formalism on the symmetry of the interactions warranted by Newton 's third law.

During last few decades, statistical mechanics has explored heterogeneous systems, in contrast with simple homogeneous systems. Particularly, the study of spins glasses, which are spin systems with random interactions, has achieved tremendous technical and conceptual progress. The similarity of the neuron threshold to the automation behaviour of two-state Ising spin has created a tidy relation between neural networks and spin glasses. As a mater of fact, some models used in spin glass theory can be related just to

neural networks without having a real relation with the spin glass materials (neighbourhood interaction). At the same time, these models are not purely biological, which may decrease the power of the artificial model due to the huge information lost by denying some biological properties (interaction non-symmetry) of the neural networks. In brief, the study of neural networks by statistical physicists has acquired dynamics of its own. It will go on, even if the connection with neurobiological experiments remains loose for some time. However, despite this relative autonomy, physicists will be chronically looking for guidance from biology. The reason is simply because that the space of models that one can invent and study is too vast. Obviously, much of the excitement on neural networks, i.e., networks made of neurons interacting via synapses, comes from the fact that (biologically) the intelligent behaviour could be obtained with collections of such elements. It is a safe bet to predict that many further important steps and results can be achieved by such similar kinds of hints.

2-2-3 Introduction of Statistical Mechanics into Artificial Neural Networks

2-2-3-1 Neuron Modelling

It is assumed that communication among the biological neurons is mostly encoded in the pulse frequencies interchanged among them. Consequently, the neuron mathematical modelling is based on this assumption of frequency code i.e. all information is carried by the firing frequencies of the network neurons. In this model, the average frequencies of pulses incoming through excitatory and inhibitory synapses determine how often the receiving neuron gets into the state of firing a pulse. The output frequency y_i of the neuron i depends on the input frequencies x_j of the neurons connected to this neuron in a two-step manner. The first is that the local time-averaged potential z_i of the neuron

(soma) i accumulates in a smooth manner the input excitations arriving from the different synapses in a linear function:

$$z_i = z_0 + \sum_j J_{ij} x_j \quad (1)$$

The neuron rest potential is represented by z_0. The second step is the neuron non-linear response to the average potential z_i in determining the output firing as follows:

$$y_i = f\left(\sum_j J_{ij} x_j - U\right). \quad (2)$$

The transfer function of equation (2) is the sigmoid function similar to a step-wise function centred at the neuron firing frequency threshold. The reason of the first assumption is that (1) represents the first order of the electro-chemical response of the average potential of the soma to the incoming pulse. However, the second processing step is explained by the response of the biological neuron, which does not respond at all for low frequencies, and does not generate pulses of more than the refractory time.

In the simplified, model of McCulloch-Pitts the sigmoid function is approximated to the infinitely steep hard threshold, the Heaviside or sign function.

$$f_\infty(X) = \Theta(X) = 1 \text{ If } X > 0 \text{ or } 0 \text{ if } X \leq 0 \quad (3)$$

In this case, the output of the any neuron is 1 or 0 indicating, or not, firing state. This simple representation, of great consequences, is the cornerstone that founds the mathematical modelling of the artificial neural networks.

2-2-3-2 The Adopted Processing and Architecture

For reasons that will be clarified later, the fully connected architecture in which each neuron is connected to all other neurons, is adopted. In reality to apply equation (2) to the chosen network, the concept of the time step has to be applied to reflect the

dynamics of neurons firing in the network. The dynamics of the network is understood through the dynamical change of the network pattern vector composed of all neurons firing state. Therefore, the discrete-time approximation can well describe this activity dynamics through the following equation:

$$V(t+1) = f\left(\sum_{j=1}^{N} J_{ij} V_j(t) - U\right) \quad i = (1,2,.....N) \tag{4}$$

In order to use the statistical mechanics analysis tools to study the firing dynamic of artificial neural network, the concept of the stochasticity is to be introduced. Little [3] has realised this introduction by replacing the deterministic law (3) by a stochastic law where he interpreted the transfer function f as the probability that $V_i(t+1)$ can take one of the values '0' or '1'. In reality, the stochasticity introduction has also a biological reason that a certain degree of stochasticity is observed in biological neural networks. Neurons may become spontaneously active, without any external stimulus, or if the synaptic excitation does not exceed the activation threshold. This phenomenon does not, however, appear to be a simple thermal effect, rather it is a consequence of random emission of neuron transmitters at the synapses.

Secondly, the analogy between the neuron firing states and the upward and downward states of the spin could be shown by the transformation:

$$S_i = 2V_i - 1 \quad (i = 1,2...N) \quad \text{where } S_i = \pm 1 \text{ as } V_i = 0,1 \tag{4}$$

This transformation is to make the analogy between a network of long-range interaction spin system called spin glass network and the artificial neural network in order to analyse it by statistical mechanics. The up or down directions limitation of the spins in the spin glass network is just the Ising model. When the transformation (4) is replaced in

the argument of the transfer function in equation (3), it results after scaling the two

following terms: $\sum_j J_{ij} S_j(t) + \left(\sum_j J_{ij} - U \right)$. The first term represents the *effective field*

acting on the spin i by its interaction with other spins whereas the second represents the

external field resulting from the conversion (4) and thresholds. Therefore, instead of the

motion equation (3), using Little concept, the following equation (5) is deduced.

$$S_i(t+1) = +1 \text{ with probability } f(h_i(t))$$

$$\text{where } \quad f(h_i(t)) = \sum_j J_{ij} S_j(t) + h_i^{ext} \text{ and } h_i^{ext} = \sum_j J_{ij} - U \qquad (5)$$

Finally, calculating the potential energy of the neuron influenced by its local effective

magnetic field, which is $\varepsilon_i = -S_i h_i$, one can use the Boltzmann-factor $e^{-\beta \varepsilon_i} = e^{+\beta h_i S_i}$ to

get the normalised probability f. Finally, rewriting equation (5) that presents Little

model, results clearly and simply by the following:

$$S_i(t+1) = \pm 1 \text{ with probability } f(h_i(t)) = \frac{1}{1 + e^{\mp 2\beta h_i(t)}} \qquad (6)$$

In equation (6), an important variable, which characterises the temperature of physical

systems in statistical mechanics, was introduced by β. While β represents the inverse

temperature of the physical system (spins) and because of its linear dependence to its

entropy (at the equilibrium), it describes the noise of the neural network system.

Therefore, the probability that the spin will be upward or downward in the first case will

be equally equal to ½. However, in the second case the probability that the spin takes the

same direction as the local field, will be greater than the probability that it takes in the

opposite direction. For the extreme case where β tends to infinity the equation of

motion of the neuron tends to the following deterministic equation which is the cornerstone of almost any neural networks models dynamics.

$$S_i(t+1) = \text{sgn}\left(\sum_{j=1}^{N} J_{ij} S_j(t) + h_i\right). \qquad (7)$$

2-2-3-3 Hebbian Learning Rule

The Hebbian learning rule is neither a mathematical based rule, nor is a pure biological reality. Hebb observation of the biological neural stimulates him to deduce that the only parameter that could be the reason behind the learning and recalling capabilities of neurons is their synaptic strengths. Therefore, he suggests that if an input pattern of the neural network was applied to its neurons for learning the interconnection synapses for some time then the network should remain at this firing configuration after removing it. However, he assumed that the reason of the stability of the network at the learned configuration is that each neuron influences the others through their updated synaptic strength connections to preserve the learned configuration. In other words, if the given neuron was firing when the network has learned the memory configuration and another connected to it was also firing then the synaptic strength connecting them will get positively stronger. Therefore, each time that one of them is firing, it will influence the other to fire too and vice-versa and the same could be said when both of them were quiescent. The first case of these assumptions was observed in the biological dynamics of the synaptic strength later by Kandel [8]. The remaining state of the synaptic dynamics, which was suggested by Hebb, is that when the two interacting neurons were in different states. In this last case, the state of each of the interacting neurons is inhibited through their interconnecting synaptic strengths so that each of the neurons

·induce its reverse state in the other neuron. However, based on these assumptions Hebb formulated the following equation describing the dynamical changes of the synaptic strengths during the learning phase of the artificial neural network:

$$J_{ij} \longrightarrow J_{ij} + \lambda \xi_i \xi_j \qquad (8)$$

The value of λ is the learning amplitude of the association between the interacting neurons i and j. In the general case where many configurations have to be stored in the neural network, the learning amplitude is usually taken to be equal except in some cases to enhance the stability of the stored pattern (see Ref [12] of Chapter 4). The concept of stability will be developed later when dealing with the Hopfield model. However, in the learning phase of many configurations the synaptic strengths accumulate the effects resulting from their learning, as it is shown by the following equation:

$$J_{ij} = \lambda \sum_{\mu=1}^{P} \xi_i \xi_j \qquad \text{For } i \neq j \text{ and } p \text{ the number of memories} \qquad (9)$$

In the above equation, the network has been learned using the Hebbian rule from *tabula rasa* where all the initial synaptic strengths are nulls. The assumption of Hebb could be however verified in the simple case of one taught memory in the limit of vanished noise (low temperature) by replacing (9) in (7) one can get:

$$S_i(t+1) = \text{sgn}\left(\lambda \sum_{j=1}^{N} \xi_i \xi_j S_j(t)\right) \qquad (10)$$

If $S_j(t)$ is replaced by ξ_j and $\lambda > 0$, then the next state of the neuron i will be ξ_i and therefore the stored memory remains the stable state of the network. In the case of many stored memories and a fully interconnected architecture, as it was adopted earlier in the above equations, the Hopfield model is the best candidate to study it. The conditions that

the Hopfield model requires to be used are the fully interconnectivity of the network and the symmetry of the interaction among the neurons.

The first condition is explained by long-range interaction of spins, which is hard to be found in the real spins system whereas the symmetry of the interaction is reflected by Hebbian learning rule.

Although the symmetry of the interaction is in accordance with the Newton's third law, it is contrary to the biological neural network although it has some biological reality in Kandel's work [8].

Consequently, the Hopfield model is neither purely biologically inspired model nor a physical reality but a mixed model to analyse the dynamics of the artificial neural network. In the next section, the Hopfield model and its revolutionary idea of the computational energy is presented by analogy with the long-range interaction spin system.

2-2-3-4 The Hopfield Model

The Hopfield model is a revolutionary combination of the physical model of the spin system with the biological model of neural network adopting Little modelling of the neuron and a full inter-connective network. Hebbian learning rule is adopted because of the symmetry of the synapses despite its limited biological existence [8].

The revolutionary part of the Hopfield model is the computational energy [4] of neural network system, which is the total energy of its equivalent spin system defined by:

$$E\{S\} = -\frac{1}{2} \sum_{i,j(i \neq j)} J_{ij} S_i S_j - \sum_i h_i^{ext} S_i \qquad (11)$$

The computation property of the system is realised, as Hopfield proposed, by the minimisation of the energy during the neural processing. As the energy of the isolated system gets its minimum at a given configuration, it remains at that system configuration forever. However, the computation is performed by storing the solution(s) implicitly or explicitly in the synapses representing the *energy surface* in the configuration space, so that network energy is minimal at that solution (s). Note that this problem is the inverse problem to the general goal in the physics of collective phenomena, which is to find the equilibrium or asymptotic behaviour when the interactions are known. The solution of the problem as well as the stimulus is given by the firing pattern of the network. The earlier represents a centred minimum of the energy surface while the latter represents the initial point of the energy trajectory toward the closest (hopefully the best) solution in the configurations space. This Hopfield network can perform the task of associative recall because it adopts the Hebbian rule as the learning rule. Moreover, whenever a stimulus configuration (input pattern), which is closest to one of the learned memories, is presented to the network, then the later configuration is recalled during the system relaxation. This is well described by the fall of the system configuration, in its energy surface, to the nearest local minima, which correspond to the nearest stored memory. Furthermore, the network is also able, with some modifications of the Hebbian rule, to solve the combinatorial task of the Travelling Salesman Problem (TSP) [10].

It can be shown from equation (11), by a simple signal to noise analysis, [11] that as long as the number of stored patterns 'p' does not exceed $N/(2lnN)$, these memories are fixed points of the dynamics at null noise (T=0). In other words, the stored uncorrelated and unbiased patterns $\{\xi_i^\mu\}$, which have been selected randomly are, with fixed $\lambda > 0$,

the minimum of the energy surface in the configurations space. As $'p'$ increases, the random correlations between the memories create noise, which tends to destabilise them from being minima and to reduce their basin of attraction. Therefore, what happened when $'p'$ becomes proportional to $'N'$ (number of neurons) as $p = \alpha N$?

The question was answered by Amit et al., [12] who showed that the free energy of the Hopfield network could be calculated and analysed using methods, concepts and analogies of statistical mechanics of disordered systems [13]. They carried out mean field calculation, because of the long-range interaction of the neurons, using the replica trick [14] familiar from the theory of spin-glass. The results of the analysis are presented in the following and the reader is referred to Geszti [15] for more details.

1. At $T=0$ and finite storage level α, the memories are not fixed points, but new fixed points appear in their close neighbourhood, as long as α is bellow the critical value $\alpha_c = 0.144$.

2. The overlap of these fixed points with the memories determines the fraction of retrieval errors $f = 1/2(1 - m_\mu)$ where m_μ is the overlap between the current configuration and the memory ξ^μ defined by $m_\mu = \frac{1}{N}\sum_i \xi_i^\mu S_i$. At α_c, one gets $f \approx 0.01$. The percentage of errors increases with α and jumps discontinuously, at α_c, to 50%, corresponding to zero overlap with the learned pattern, indicating a complete loss of memory (see Figure (4)). This *forgetting catastrophe* is clearly undesirable in a realistic memory model.

3. As T increases, at a fixed value of α, the overlap with memories decreases until it drops discontinuously at T_M to zero. The retrieval phase below the line is

33

characterised by valleys with macroscopic energy barriers surrounding the memories. It coexists with the spin-glass phase, characterised by exponentially many local minima. This phase persists up to line T_G.

4. The attractors, corresponding to the memories, become global energy minima only below the line T_C at which a true thermodynamic phase transition takes place.

The above observations are presented in Figure (4). The Hopfield model turned out however, to be profitable both for statistical physics and neural networks theory. From the point of view of statistical mechanics, the Hopfield model has several interesting features:

1. It is another solvable and rich model of a random long-range system.

2. It spans a whole range of intermediate modes of behaviour, between the infinite range Ising ferromagnet when $p = 1$ (the Mattis Model [16]) and the Sherrington-Kirkpatrick (SK) spin-glass [17] when $\alpha \longrightarrow \infty$.

3. The existence of macroscopic free-energy valleys, which are not ground states of the system, at least between the lines T_C and T_M, is a novel property of this model.
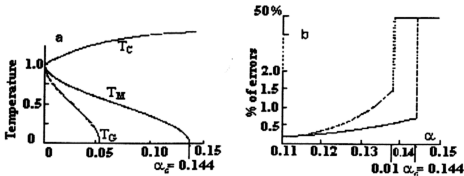

Figure (4). a) Phase diagram of the Hopfield model. b) Errors percentage in the retrieval phase. Broken line: The replica symmetric result; full line: effects of one-step replica symmetry breaking.

The major contribution of the Hopfield model to neural networks theory was to open a large gallery of concepts, techniques and analogies, and to direct the effort in a way, which extracts the computation from physical systems. The analysis of the model leads to results and insights, which go beyond the constraints of synaptic symmetry. It turned out to be a useful starting point for a variety of modifications, which removed some of the constraints and drawbacks of the original formulation of the model. It is worth mentioning that a layered-network model, with synaptic couplings between successive layers given by (9) can be analysed by similar methods of statistical mechanics [18].

2-2-3-5 Gardner Capacity of the Network

Memory stability is an important parameter, which characterises neural network recalling performance. The normalised stability of the stored memory or basin of attraction is defined by the following quantity:

$$\Delta_i^\mu \equiv \frac{\xi_i^\mu \sum_{j \neq i} J_{ij} \xi_j^\mu}{\sqrt{\sum_{j \neq i} J_{ij}^2}} \qquad (11)$$

where j vary from 1 to N and μ is the stored patterns index. In order to simplify the analysis calculation it is well suitable to have a common norm of the amplitude of the synaptic vector for all neurons and so that $\sum_{j \neq i} J_{ij}^2 = N$. Then, let define the stability degree by a given positive parameter, which describes well the basin of attraction of a stored memory defined as the lower limit of the stability parameter as follows:

$$\Delta_i^\mu \equiv \frac{\xi_i^\mu \sum_{j \neq i} J_{ij} \xi_j^\mu}{\sqrt{N}} > \kappa > 0 \qquad (12)$$

E. Gardner developed a method to calculate the storing capacity of a feedback dynamical network, independently of any particular learning algorithm. The starting point of Gardner's approach is to focus on the phase space of the interactions i.e. the space spanned by the connection strengths J_{ij}, all independent without any required symmetry. In this space, one calculates the volume, within which a given set of 'p' patterns are given a finite stability; their Δ_i''-s as defined in (11) satisfying the condition (12) with a finite stability κ. The volume within which the task is solved still depends on the actual patterns. Averaging over a distribution of patterns that can be independent or correlated, one obtains the volume as a function of a few parameters $V(p, N, \kappa)$. As the load of memory grows, V shrinks until it vanishes at a given critical $p_c(N, \kappa) = N\alpha_c(\kappa)$. This value defines then the capacity of the network. Without dealing with the details for which, the reader is referred to Geszti [15] and Gardner [19] for more clarification, the major results of the Gardner analysis are presented hereafter. For unbiased patterns the critical storage capacity α_c is given by:

$$\alpha_c(\kappa) = \left(\int_{-\kappa}^{\infty} Dt(t+\kappa)^2 \right)^{-1} \quad , \quad Dt \equiv \left(dt / \sqrt{2\pi} \right) e^{-t^2/2} \quad (13)$$

For the correlated patterns having the same magnetisation, which is just the average overlap of the current configuration with the stored memories, the result is

$$\alpha_c(m, \kappa) = \left[\frac{1+m}{2} \int_{(\vartheta m - \kappa)/\sqrt{1-m^2}}^{\infty} Dt \left(\frac{\kappa - \vartheta m}{\sqrt{1-m^2}} + t \right)^2 + \frac{1-m}{2} \int_{-(\vartheta m - \kappa)/\sqrt{1-m^2}}^{\infty} Dt \left(\frac{\kappa + \vartheta m}{1-m^2} + t \right)^2 \right]^{-1} \quad (14)$$

and

$$\vartheta = \frac{1+m}{2} \int_{(\vartheta m - \kappa)/\sqrt{1-m^2}}^{\infty} Dt \left(\frac{\kappa - \vartheta m}{\sqrt{1-m^2}} + t \right) = \frac{1-m}{2} \int_{(-\vartheta m - \kappa)/\sqrt{1-m^2}}^{\infty} Dt \left(\frac{\kappa + \vartheta m}{\sqrt{1-m^2}} + t \right) \tag{15}$$

The graph shown in Figure (5) gives an overview of the variation of the critical capacity versus the stability parameter κ.

The above analysis could not be done if the approximated analogy with the physical model of the spin system has not been formulated. Therefore, a physicist question could be formulated in the way of suggesting other physical models exhibiting computation like the spin-glass network model.
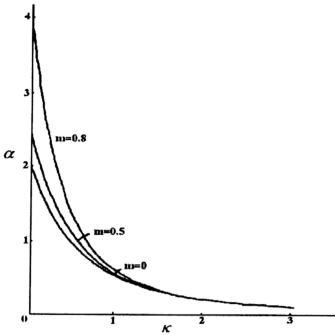

Figure (5). The critical storage ratio, α_c, as a function of the stability κ

In our point of view, the proposed physical model usefulness is related to its collective parallel-computational processing rather than to its theoretical background. In this way, the use the real physical model to emulate the parallel collective behaviour of neural networks is suggested as an alternative computation processor. We think that the use of

the *quantum-dot* arrays might be useful in this way, as prototype structures for nanoscopic parallel computers extracting their computation from the quantum-mechanical processes (diffusion, relaxation...etc) [20].

2-3 Artificial Neural Networks

The emergence of artificial neural networks is a result the imitation of the biological neural networks such as the brain and the nervous system. Right now, the biology of the nervous system does not reveal a clear picture about its processing mechanisms. Consequently, models have been formulated in order to understand its intelligent mechanism. Artificial neural network is the model in which an artificial neuron is proposed to imitate the biological counterpart representing the basic cellular unit of the nervous system. Different models of the biological neuron of different architectures have been proposed to perform special tasks such as associative memory, perception, self-organisation...etc.

2-3-1 Artificial Neural Networks Development

In 1943 W. McCulloch and W. Pitts [21] proposed general theory of information processing based on networks of binary switching or decision elements, which are somewhat euphemistically called "neurons", although they are far simpler than their real biological counterparts. Each one of these elements $i = 1,2,...n$ can only take the output values $n_i = 0,1$, where $n_i = 0$ represents the resting state and $n_i = 1$ the active state (firing) of the elementary unit. In order to simulate the finite regenerative period of real neurons, changes in the state of the network are supposed to occur in discrete time steps

$t = 0,1,2\ldots$etc. The new state of a certain neural unit is determined by the influence of all other neurons, as expressed by a *linear* combination of their output values:

$$h_i(t) = \sum_j w_{ij} n_j(t) \qquad (1)$$

Here the matrix w_{ij} represents the synaptic coupling strengths between neurons j and the neuron i, while $h_i(t)$ models the total postsynaptic polarisation potential at neuron i caused by the action of all other neurons. In equation (1), h_i can be considered as the input into the neural computing unit, and n_i as its output. The properties of neural networks are completely determined by the functional relation between $h_i(t)$ and $n_i(t+1)$. In the simplest case, the neuron is assumed to become active if its input exceeds a certain threshold ϑ_i, which may well differ from one unit to the next. The evolution of the network is then governed by the following rule:

$$n_i(t+1) = \theta(h_i(t) - \vartheta_i) \qquad (2)$$

where $\theta(x)$ is the unit step function, i.e. $\theta(x < 1) = 0$ and $\theta(x > 1) = 1$.

McCulloch and Pitts showed that in principle such networks could, carry out any imaginable computation, similar to a programmable, digital computer or its mathematical abstraction, the *Turing Machine*. In a certain sense, since the network contains a "program code", which governs the computational process, namely the coupling matrix w_{ij}, the network differs from the traditional computer in the execution way. This means that the steps of the program are not executed sequentially, but in parallel within each elementary unit. One might say that the program is compensated by the substitution of a vast number of processing elements (10^{11} in the human brain!) for a single processing unit of a conventional, sequential electronic computer. The designer of

a McCulloch-Pitts type neural network now faces the problem of how to choose the couplings w_{ij} so that a specific *cognitive task* is performed by the machine. Here the word cognitive task is used in a generalised sense; it can mean any task requiring digital or analogue information processing, such as the recognition of specific optical or acoustical pattern. This question was addressed in 1961 by E. Caianiello [22], who gave a *learning* algorithm that would allow the determination of the synaptic strengths of a neural network. This algorithm, called the *mnemonic* equation by Caianiello, incorporates in a simple way the basic principle of Hebb's learning rule [1]. Around 1960, F. Rosenblatt and his collaborators [23] extensively studied a specific type of neural network, which they called a *perceptron*. They considered it as a simplified model of the biological mechanisms for the processing of sensory information, i.e. perception. In its simplest form, a perceptron consists of two separate layers of neurons representing input and output layer, respectively, as illustrated in Figure (6). The neurons in the output layer receive synaptic signals from those of the input layer, but not vice versa, and the neurons within one layer do not communicate with each other. The flow of information is thus strictly directional; hence, one speaks of a *feed-forward* network.

Rosenblatt' s group introduced an iterative algorithm for constructing the synaptic couplings w_{ij} such that a specific input pattern is transformed into the desired output pattern, and even succeeding in proving its convergence [24]. However, M. Minsky and S. Papert [25] pointed out few years later that this proof applies only for to those problems, which can, in principle be solved by a perceptron.
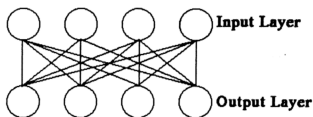
Figure (6). The simplest form, of the perceptron architecture.

The fact, which made matter worse, was that they showed the existence of very simple problems, which cannot be solved by any such two-layered perceptron. The most notorious of these is the *exclusive–OR* (XOR) logic gate. The XOR logic gate requires two input neurons to be connected with a single output neuron in such a way that the output unit is activated, if and only if, one of the input units is active. The XOR logic gate is a standard problem easily solved by computer simulation and thus the result of Minsky and Papert represented a severe blow to the perceptron concept. This problem was latter solved by the introduction of a practical algorithm (backpropagation rule) for the construction of the synaptic strengths (interconnections) w_{ij} of such generalised perceptrons [26].

Another very fruitful development began when W. Little [3] pointed out the similarity between a neural network of the type proposed by McCulloch and Pitts and systems of elementary magnetic moments or *spins* (see Figure (7)).

In these systems, called *Ising models*, the spin s_i at each lattice site i can take only two different orientations, up or down, denoted by $s_i = +1$ (up) and $s_i = -1$ (down). The analogy to a neural network is realised by identifying each spin with a neuron and

associating the upward orientation $s_i = +1$ with the active state $n_i = 1$ and the downward orientation $s_i = -1$ with the resting state $n_i = 0$.
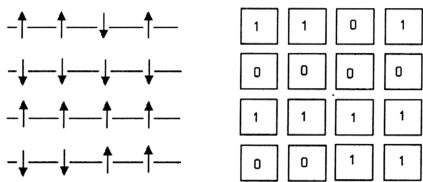

Figure (7). The Little analogy between the spins system and the digital system

These ideas were further developed by Little and G. Shaow [27] and later by J. Hopfield [29] who studied how such a neural network or spin system can store and retrieve information. The Little and Hopfield models differ in the manner in which the state of the system is updated. In Little' s model all neurons (spins) are updated synchronously according to the law (1), whereas the neurons are updated sequentially one at a time, either in certain fixed order or randomly, in the Hopfield model. Sequential updating has a considerable advantage when the network is simulated on a conventional digital computer and for building the theoretical analysis of the properties of the network. On the other hand, it holds the essential conceptual disadvantage that the basic feature of neural networks, namely the simultaneous operation of a large number of parallel units, is given up. Neurons in the human brain surely do not operate sequentially, this being precisely the reason for the brain' s superiority in complex tasks to even the fastest exciting electronic computer.

The analogy with the spin systems became especially fruitful owing to the advances in the understanding of the thermodynamic properties of disordered systems of spins, the so-called *spin glasses*, achieved over the past decade. In order to apply these results to neural networks it is necessary to replace the deterministic law (2) by a *stochastic* law. In the suggested stochastic law, the value of n_i $(t+1)$ is assigned according to a probabilistic function depending on the intensity of the synaptic input h_i. This probability function contains a parameter T that plays the role of a "temperature". However, T does not represent the physical temperature of a biological neural network, but a formal concept designed to introduce stochasticity. It reflects the disorder of the system through its relation to the entropy of the system by the relation $S = \ln(Z)/T$ where Z is the partition function of the system of spin and S its entropy. Thus, the temperature is introduced into the network in order to allow the application of the powerful techniques of statistical thermodynamics. In the limit of the vanishing temperature, the deterministic McCulloch-Pitts model is recovered. Studies have shown that stochastic evolution can render the network less susceptible to dynamic instabilities and may thus improve the overall quality of its operation. In the next chapter, the idea of how the introduction of physics through statistical mechanics has played this fruitful role in developing and understanding the behaviour of artificial neural networks system mainly the feedback networks, will be developed further. In recent years, the interest in layered, feed-forward networks (perceptrons) has been revived because of the need to apply this architecture in the image and sound processing as it was found being applied by the biological processing. This development was initiated by the rediscovery of an efficient algorithm for the determination of the synaptic coupling strengths for multi-

layered networks with *hidden layers* (see Figure (8)). The power of his method, initially suggested by Werbos [26] and now known as *error back-propagation*, was recognised around 1985 by several groups of scientists ([28], [29]). This learning algorithm is based on a simple but very effective principle: the synaptic strengths w_{ij} are modified iteratively such that the output signal differs as little as possible from the desired one. This is achieved by the application of gradient method, which yields the required modifications δw_{ij}. Since the operation of the network corresponds to a highly non-linear mapping between the input and the output (the step function in (1) is non-linear) the method must be applied many times until convergence (required error) is reached.
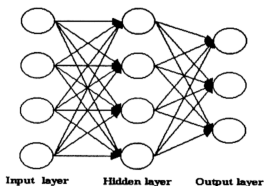

Input layer   Hidden layer   Output layer

Figure (8). A feed-forward network with one internal "*hidden*" layer of neurons

Error back-propagation is a particular example of a much larger class of learning algorithms, which are classified as *supervised learning*, because at each step the network is adjusted by comparing the actual output with the desired output. Not only are such algorithms most probably not implemented in biological neural networks, but also they suffer because they are applicable only when the desired output is known in detail. For this reason also other concepts of learning (the synaptic dynamics), are intensively studied, such as strategies based on reward and penalty, evolution and selection...etc.

2-4 Bidirectional Associative Memory

Bidirectional Associative Memory BAM is a neural network proposed for the implementation of the Parallel Learning-Processing PLP strategy. The origin of this ANN is presented as well as its general characteristic and its relation with the Hopfield network. The PLP neuron based BAM is suggested to solve the communication problems between any two digital systems.

2-4-1 BAM Definition

The BAM network belongs to the family of *heteroassociative* dynamic associative memory HDAM proposed first time by Okajima [30] in 1987. A descriptive block diagram of this kind of associative memory is shown in Figure (9).

It consists of two processing paths that form a closed loop. The first processing path computes the output vector $y \in \{-1,+1\}^L$ from an input vector $x \in \{-1,+1\}^n$ according to the parallel updating rule

$$y = F[W_1 x] \qquad (1)$$

or its serial (asynchronous) version, where one and only one unit updates its state at a given time. Here, F is usually the sign activation operator. Similarly, the second processing path computes a vector x according to the following equation:

$$x = F[W_2 y] \qquad (2)$$

or its serial version.

The vector y in equation (2) is the same vector generated by equation (1). The HDAM can be operated in either parallel or serial retrieval modes.
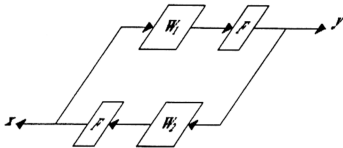
Figure (9). Representative Diagram of Heteroassociative Memory Dynamics

In the parallel mode, the HDAM starts from initial state x (0) (y (0)), computes its state y (x) (x (y)) according to equation (1) (equation (2)). Then, it updates the layer x (y) (y (x)) state according to equation (2) (equation (1)). This process is iterated until state x (and equivalently y) ceases to change. On the other hand, in the serial-update mode, only one randomly chosen component of the state x or y is updated at a given time.

Various methods have been proposed for storing a set of *heteroassociations* $\left\{ x^k, y^k \right\}$, $k = 1, 2, ..., m$, in the HDAM. In most of these methods, the interconnection matrices W_1 and W_2 are computed independently by requiring that all one-pass associations $x^k \longrightarrow y^k$ and $y^k \longrightarrow x^k$, respectively, are perfectly stored. It assumed that the set of associations to be stored forms a one-to-one mapping; otherwise, perfect storage becomes impossible. Examples of such HDAM recording methods include the use of projection recording ([31], [32]) and Householder transformation-based recording [33]. These methods require the linear independence of the vectors x^k (also y^k) for which a capacity of $m = \min(n, L)$ is achievable. One drawback of these techniques, though, is that they do not guarantee the stability of the HDAM; i.e., convergence to spurious

cycles is possible. Empirical results [32] show that parallel updating of the projection-recorded HDAMs exhibit significant oscillatory behaviour only at memory loading levels close to the HDAM capacity.

Kosko ([34], [35]) independently proposed a heteroassociative memory with the architecture of the HDAM but with the restriction $W_2^T = W_1 = W$. This kind of memory is known as the BAM. The interesting feature of the BAM is its stability for any choice of the real-valued interconnecting matrix W and for both serial and parallel retrieval modes. Starting from the BAM's bounded Liapunov energy function it can be shown the following property:

$$E(x,y) = -\frac{1}{2}x^T W y - \frac{1}{2}y^T W x = -\frac{1}{2}x^T W y \geq -\sum_i \sum_j |w_{ij}| \qquad (3)$$

This interesting dynamics is independent of the updating method. In the serial or parallel state updating the energy, $E(x, y)$ tends to decreases (Appendix A). The BAM stability can be proved observing that a BAM can be converted to a discrete autoassociative memory DAM (*Discrete Hopfield DAM*) with the state vectors $x' = \begin{bmatrix} x^T y^T \end{bmatrix}$ and interconnection matrix W' given by:

$$W' = \begin{bmatrix} 0 & W \\ W^T & 0 \end{bmatrix} \qquad (4)$$

Now, since W' is a symmetric zero-diagonal matrix, the autoassociative DAM is stable if serial updating is adopted [36]. Therefore, the serially updated BAM is stable.

This equivalence property could be used to show the stability of the parallel-updated BAM. Consequently, BAM always converges to a local minimum of its energy function defined in equation (3). It can be shown [37] that these local minima include all those

which correspond to associations $\{x^k, y^k\}$ which are successfully loaded into the BAM (i.e., associations which are equilibrium of the BAM dynamics.)

The most simple storage recipe for storing the associations as BAM equilibrium points is the correlation-recording recipe of the Hebbian rule. This recipe guarantees the BAM requirement that the forward path and backward path interconnection matrices W_1 and W_2 are the transpose of each other, since

$$W_1 = \sum_{k=1}^{m} y^k (x^k)^T \quad (5) \qquad \text{and} \qquad W_2 = \sum_{k=1}^{m} (x^k)^T y^k \quad (6)$$

However, some serious drawbacks of using the correlation recording recipe are low capacity and poor associative retrievals; when m random associations are stored in a correlation-recorded BAM, the condition $m \ll \min(n, L)$ must be satisfied if good associative performance is desired [38]. Heuristics for improving the performance of correlation-recorded BAMs have been demonstrated by Wang et al [39].

2-4-2 Comparison Between BAM and Hopfield Model

From the above BAM network description and recalling the Hopfield network, the major differences between the two networks could be summarised in the following points.

At the first sight, BAM needs a smaller correlation matrix to the Hopfield model because of the limited interconnections between the neurons. For N neurons implemented in the Hopfield network N^2 of synapses are needed. If this number is distributed to P neurons connected to $(N-P)$ neurons in the BAM network only $N \times (N-P)$ interconnections are needed, which are less [1] than N^2. Another advantage is that BAM has no oscillatory states because of its asynchronous version of the synchronous Hopfield model (one

---

[1] this is trivial because $N^2 > (N-P) \times P \Leftrightarrow ((N-P)^2 + N \times P) > 0$ and N, P are a positive numbers

updated layer at a time). This could be verified by the fact that $\Delta E < 0$ if just one-bit flips (see Appendix A). Finally, BAM is relatively easy to implement comparing with the Hopfield network, in both of the theoretical and hardware implementation in which Kosko [36] has suggested an optical method for its implementation.

Beside these advantages, BAM also suffers from many disadvantages such as its restricted capability to implement only one-to-one mapping. Another, but a common disadvantage of the Hebbian rule based networks, is its low capacity. Despite this property, BAM has larger capacity compared to the Hopfield network. It depends on the number of neurons of the minimum layer for the BAM while it is 14% of the total number of neurons for the Hopfield network (see Appendix A).

2-4-3 Other Models of the BAM

There have been several models of the BAM based on the different aspect of the artificial neuron, such as the transfer function, synaptic strength and the threshold. However, the continuous BAM suggested first time by Kosko ([36], [37]), suppose that the transfer function is a sigmoidal function. Besides, Kosko [36] proposed other model called the adaptive BAM by assuming the dynamic changes of the synapses. In this model, the synaptic change also during the processing.