Ising Spin Models of Partially Connected Neural Networks

Andrew M. Canning

Submitted for the degree of
Doctor of Philosophy

Department of Physics
University of Edinburgh
September 1988

To my Mother and Father,

and the memory of Elizabeth Gardner.

Declaration

All of the work in this thesis is my own except where otherwise indicated. Some of the work has been published in,

A. Canning and E. Gardner, 1988 *J. Phys. A: Math. Gen.* **21** 3275

Acknowledgements

My work at Edinburgh was supported by the Science and Engineering Research Council and A.M.T. Computers Ltd. Reading.

I am particularly indebted to Elizabeth Gardner who initiated the theoretical work in this thesis. Sections 2.2,2.3 and 3.1 were carried out in collaboration with her and throughout this work I have benefited from useful discussions with her. I am also grateful to David Wallace for his helpful suggestions, discussions and encouragement.

I would also like to thank Stewart Reddaway, Peter Flanders and David Hunt for making me welcome at A.M.T. and helping me write code for the DAP.

Finally I would like to thank all my colleagues in the Department of Physics for making my time in Edinburgh enjoyable.

Abstract

A partially connected version of the Hopfield neural network model is studied under the restriction that the number of connections per site becomes infinite as the size of the system, N becomes infinite with the connection structure at each site being the same. The connection architecture of the network is specified by a logical matrix $\mathbf{D}$ of dimensions $N \times N$, with $D_{ij} = 1, 0$ corresponding to sites being connected or disconnected. The replica-symmetric mean field theory equations for the order parameters are derived in terms of $\mathbf{D}$ and the external parameters of the system. The zero temperature forms of these equations are then solved exactly for a few different "local" connectivity architectures showing phase transitions at different critical storage ratios α_c. At α_c the states we are trying to store in the network become discontinuously unstable. We show that the information capacity per connection improves the more partial and random the connectivity of network becomes. We derive the full phase diagram for the particular case of the randomly connected model with of order $\sqrt{N}$ connections per site. The similarity between this model and the infinite range SK spin glass model is also discussed. The eigenvalue which controls the stability of the replica-symmetric solutions is also derived and then used to calculate the replica broken parts of the phase diagram for different connection architectures. Numerical simulations of finite size systems are also presented for a fully connected and one dimensionally connected network.

Contents

Chapter 1

An Introduction to Neural Networks and Spin Glasses

1.1 Introduction

The human brain is capable of massive processing tasks such as speech recognition, vision etc. which even the most advanced computers, as yet, cannot match. The brain achieves these tasks despite the fact that the processing and communication times between neurons are typically of the order of milliseconds compared to processing times on chips as low as 50 nanoseconds. Neural networks are thought to model some of the features of the brain that give it these remarkable powers, although a close comparison is difficult to make. This is partly because the biological working of the brain, while being quite well understood for an individual neuron, is little understood on the higher level of the network of interacting neurons. Whatever the similarity of neural network models to the brain they still stand or fall on their own merits as models of artificial intelligence.

Neural network models all share the common features of nodes (neurons) which can take different values representing the different levels of activity of a real neuron. They are connected together by synapses of different strengths representing the different synaptic resistances in the brain. The process of learning is associated with the modification of these synaptic strengths. The nodes update themselves by performing some kind of non-linear thresholding on the sum of their inputs which they receive from the other nodes through the synapses. In this way the process is very parallel with information being passed along many synapses simultaneously and possibly many neurons updating simultaneously, although the exact process depends on the model under study. It is this dynamical process of updating the neurons that is used to process information. The initial state of some of the neurons is chosen to match a specific pattern (the input). The network then evolves by the dynamical update scheme until a pattern is read from another set of neurons (the output). We could think of the network processing an image with each pixel of the image initially set up as a value on the input nodes. After processing the image the output from the network could be a specific firing pattern on a smaller output set of neurons corresponding to recognition of a certain object in the image. A learning procedure would have been carried out prior to this to choose the synaptic strengths. This, for example, could have involved presenting a sample set of noisy images to the network having known objects present and then choosing the synaptic strengths so that these objects were identified correctly.

The basic ideas behind most neural network models can be traced as far back as the 1940's to the seminal work of McCulloch and Pitts [1] and Hebb [2] but it is only in the last few years that there has been a great surge of interest from many different disciplines. One of the major motivations for interest in the physics community was the simple model proposed by Hopfield [3]. His model contained all the basic features of neural models and bore a strong similarity to models of disordered magnetic materials studied in physics. The analytical techniques used to study these magnetic systems have been applied to the Hopfield model and many variations of it by Amit, Gutfreund and Sompolinsky [4,5,6,7,8] (for review articles see [9,10]). It is the techniques developed in these papers that we will use in this thesis to study partially connected versions of the Hopfield model. We will therefore start this chapter by defining the Hopfield model in detail before going on to make a closer study of its similarity to the brain followed by a brief review of disordered magnetic systems. The similarity between certain models of magnets and the Hopfield model will then become clear. Chapters 2 and 3 will be devoted to the application of statistical physics techniques, reviewed in this chapter, to study a partially connected version of the Hopfield model. In

Chapter 4 we will present the results of some numerical simulations on finite sized versions of the Hopfield model.

1.2 The Hopfield Model

The basic Hopfield model consists of N neurons or nodes that are all connected to each other by synapses of different strengths. Each node receives inputs from all the other nodes along these synapses and determines its own state by summing all these inputs and thresholding them. The N neurons can only take two values 1 or -1 corresponding to the neuron firing or not firing. The state of the whole network can then be described by a vector of N values, $\{S_i, i = 1, N\}$. The input to neuron i at time t is then given by,

$$\Phi_i(t) = \sum_j T_{ij} S_j(t) \quad T_{ij} = T_{ji} \tag{1.1}$$

where T_{ij} is the synapse strength. We set $T_{ii} = 0$ and use a simple step like threshold function to define the new state of the neuron at time $t+1$ by,

$$S_i(t+1) = \mathrm{sgn}(\Phi_i(t) - U_i) \tag{1.2}$$

where U_i is the threshold which is chosen to have different values depending on the model under study. In the Hopfield model it is generally set to zero as it will be in all the calculations which follow. The question of the value of $S_i(t+1)$ when $\Phi_i(t) = U_i$ is not important in large systems as there is a very low probability of it occurring. For the simulations in chapter 4 the state of the node was chosen to be unchanged when $\Phi_i(t) = U_i$. The different types of updates which can be carried out are numerous, ranging from single random site update to synchronous update of all the neurons. Synchronous updating can lead to limit cycles whereas single spin update schemes will always lead to the system reaching a stable state. Simulations of a single random site update compared to synchronous updating of half of the neurons chosen at random by Bruce *et al* [12] showed little change in the properties of the system. Therefore for large systems we do not expect different update schemes to affect the results significantly providing they are not too synchronous. In general, repeated applications of the update scheme will lead to the net reaching a stable state. We can think of the initial state of the net as the input and the final stable state as the output with all the neurons being used for both input and output. This type of network can, for example, process N bit pixel images.

Under a serial or random single spin, dynamical update scheme defined by equations 1.1 and 1.2 the system will continually change its state until a stable state is reached. This corresponds to descending an energy landscape until a minimum is reached where the energy of a given state $\{S_i\}$ is given by the Hamiltonian,

$$H\{S_i\} = -\frac{1}{2}\sum_{ij} S_i T_{ij} S_j \tag{1.3}$$

In order to make this model useful we must control the stable states of the system and the energy surface. This is done by specifying the T_{ij}'s by some kind of algorithm usually called a learning algorithm. In the Hopfield model the simple Hebb rule [2] is used to store random binary patterns as stable states of the dynamics. If we wish to store p random patterns $\{\xi_i^\mu, \mu = 1, \ldots, p\}$ in the network the Hebb rule specifies the connection strengths to be,

$$T_{ij} = \begin{cases} \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu & i \neq j \\ 0 & i = j \end{cases} \tag{1.4}$$

This gives, providing the number of patterns is not too large, an energy space with basins of attraction associated with each of the states we are trying to store. This kind of storage is called distributed storage since the information in one state is stored throughout the whole system in the connection strengths rather than at a local site the way information is normally stored on a chip. For this reason the system is robust to synaptic death and can still accurately recall stored states when quite a high percentage of synapses have been cut. The system is considered to have content addressable memory since starting the network in a state close enough to the stable stored state to be in its basin of attraction will yield the stored state as output after performing the dynamical update scheme. An important parameter for studying the ability of the network to store patterns is,

$$\alpha = \frac{p}{N} \tag{1.5}$$

which is usually called the storage ratio of the system. As α is increased it turns out that we reach some value where the ability of the network to store patterns begins to break down. The maximum value of α which can be obtained before storage totally breaks down is called the critical storage ratio and is denoted α_c. For a state to be stable the sign of Φ_i, the input to site i from all the other sites, must be the same as the state at site i for every site. This means that for a state $\{\xi_i^\gamma\}$, which we wish to store, we must have,

$$\xi_i^\gamma \Phi_i = \xi_i^\gamma \sum_j T_{ij} \xi_j^\gamma > 0 \quad \forall i \tag{1.6}$$

for the state to be stored exactly. Putting in the expression for T_{ij} equation 1.4, this gives,

$$\Phi_i \xi_i^\gamma = \frac{N-1}{N} + \frac{1}{N} \sum_{j \neq i, \mu \neq \gamma} \xi_i^\mu \xi_j^\mu \xi_i^\gamma \xi_j^\gamma \tag{1.7}$$

The first term is a signal term which tends to make the state we wish to store stable while the second term is a noise term due to all the other states. If the noise term is too large it will destroy the storage. Another important parameter of the system which measures accuracy of storage is the overlap,

$$m^\mu = \frac{1}{N} \sum_i \xi_i^\mu S_i \tag{1.8}$$

This measures the fractional overlap of the state of the system with the state we wish to store. It is equal to one when the state is perfectly stored and zero when the state of the system is randomly related to the state we wish to store. If we consider a large system and assume the noise term in equation 1.7 is an independent Gaussian variable at each site we find m is only non-zero when $\alpha < \frac{2}{\pi}(= 0.637)$. This is far higher than the actual result $\alpha < 0.14$, [6] the difference being due to the correlations between the noise terms at each site. To cope with these correlations we have to use a more advanced technique known as replica symmetric mean field theory. In the next few sections we will develop the ideas behind this technique before applying it to partially connected networks in chapters 2 and 3.

1.3 Neural Network Models and the Brain

Neural network models fall into two main categories; feed back networks and feed forward networks. The Hopfield model is an example of the first type while multilayer perceptron models like those of Hinton are of the second type (for reviews of this type of network see [14]). In both models all the information is stored in the connections between the neurons and sometimes also in the threshold values, depending on the model. One of the major problems in neural network research, particularly in feed forward systems, is determining the connection strengths for a specific problem. Hopfield avoided this problem by having a simple one step learning algorithm. The main feature of the Hopfield model is its ability to operate as a content addressable storage system like human memory. The main property of feed forward networks is that they can process an input to give a totally different output. They can therefore be associated with human functions such as reflex of a finger to a very hot object. The first layer of neurons would receive the initial input from the finger's senses and then the intermediate layers process this information till the final layer sends the reflex message to the muscles in the finger.

The brain contains approximately 10^{10} neurons with about 10^{14} synaptic connections. It appears to have both feed back and feed forward networks depending on the region of the brain under study. In particular the cerebral cortex, which is associated with memory, has some feed back structure (see [15] and references therein). Another basic feature of the brain which appears to parallel neural network models is its robustness to synaptic and neuron death. It is when we consider the way in which neural networks learn and store information that we run into problems of direct comparison with the brain. The operation of an individual neuron in the brain is quite well understood and seems to broadly parallel the neural units in theoretical models. On the other hand the operation of a whole network of neurons and the role the synaptic resistance plays is not well understood at all. This gap in our knowledge about the brain is partly due to the problems of trying to monitor many neurons simultaneously and interpret their output. Neurons are extremely small and the brain very delicate so it is almost impossible to connect more than about twenty electrodes into the brain simultaneously without disrupting it greatly. A common analogy is that of trying to understand the operational details of a supercomputer using a small number of large electrodes and having almost no previous knowledge of computers. For these reasons the role of synaptic resistance in the brain is not well understood even though it is crucial for storage in all neural network models.

Another area where very little is known about the brain is that of learning. In neural network models learning is associated with altering the synaptic connection strengths. In the case of the Hopfield model we can reformulate the Hebb rule in terms of a more natural gradual learning process carried out on each pattern as it is presented to the observer,

$$\delta T_{ij} = \xi_i^\mu \xi_j^\mu \tag{1.9}$$

This means that synaptic connections between neurons that are stimulated increase in resistance and those between neurons which are not stimulated also increase in resistance. Inhibitory synapses are also present in this model between neurons that are stimulated and ones that are not. There is as yet almost no evidence to suggest that this is the type of learning process that occurs in the brain.

All learning processes developed for neural networks fall into two main categories; supervised and unsupervised learning. A supervised learning process requires an external controller to monitor the progress of learning and change the learning process depending on what stage has been reached. The learning algorithm in equation 1.9 is an unsupervised one since the same process is continued for every presentation of an image. To gain the maximum storage capacity from Hopfield type networks it is necessary to continually re-present images that are not stored accurately and apply the learning rule 1.9 [17,19]. This therefore requires a supervisor to monitor which images are not stored correctly and then continually represent them until they are. Most learning algorithms for feed forward networks require supervised learning.

Biologically speaking supervised learning seems slightly less plausible than unsupervised since most people are capable of learning things without supervision. However the true situation probably lies somewhere between the two with other humans playing the role of supervisors.

The basic Hopfield model outlined in the previous section clearly has many features which do not parallel the brain. The brain does not have full connectivity and at any one time only a small percentage of the neurons are firing. The relaxation of full connectivity is studied in this work and is not found to affect the basic features of the model. Amit *et al* [7] and Gardner [17] studied the properties of networks with low levels of activity and again found, with a slight alteration to the learning algorithm, that the basic features of the model were not affected. The condition of symmetric connections also seems unphysical and this has been studied by Derrida *et al* [46]. Relaxation of the connection symmetry means that a Hamiltonian no longer exists and the stable minima of the free energy can be replaced by wandering paths. This in turn leads to the possibility of cycles of patterns which is more physiologically plausible than terminating in stable states. The brain clearly operates in cycles in some way otherwise when a face was recognized the brain would drop into the basin of attraction of the state associated with the face and never leave it.

In summary the research carried out so far in neural networks has led to a lot of interesting models of artificial intelligence while their similarity to the brain is still very unclear. Many of the crucial ideas that these models are based on have not been substantiated by biological evidence.

1.4 Ising Spin Models of Ferromagnets and Spin Glasses

The Ising model [20] is a simple model which describes a highly idealized ferromagnet. Consider a periodic lattice in d dimensions with N magnetic ions situated at each site of the lattice. It is assumed that their magnetic moments (or spins) can only point in two directions; $S_i = +1$ for up and $S_i = -1$ for down. There is a quantum mechanical exchange interaction between the spins which at low temperatures makes it energetically favourable for them to be aligned. An external field h can also be introduced into the model with which the spins will tend to align. Thus the microstate of the system is specified by the set of spins $\{S_i, i = 1, \ldots, N\}$ just like a neural network. There are 2^N possible states of the system and each of these states can be thought of as a point in an N dimensional space called the phase space of the system. The energy of a specific state is given by the Hamiltonian,

$$H\{S_i\} = -\frac{1}{2}J \sum_{<ij>} S_iS_j - h\sum_i S_i \tag{1.10}$$

where J is the exchange interaction which is positive for a ferromagnet. The symbol $< ij >$ denotes the sum over spins and is usually restricted to nearest neighbours. In the case where the sum is taken over all sites and we let $N \rightarrow \infty$ the model is called infinite range and is exactly solvable [23]. It should be noted that the concepts of interaction range and lattice dimension are interchangeable since, for example, a nearest neighbour interaction model of infinite dimension is the same as a one dimensional model with infinite range interactions. In this case J has to be scaled with the system size in order to prevent the energy per site becoming infinite in the thermodynamic limit of an infinite size system.

Interactions in the Ising model play the same role as synapses in the Hopfield model. The only difference between the fully connected Hopfield model and the infinite range Ising model is that the interactions between sites in the Hopfield model do not all take the same value and actually take negative (anti-ferromagnetic) as well as positive (ferromagnetic) values.

Ising spin glasses are the same as the Ising model except they have random interactions rather than ferromagnetic interactions. Spin glasses are substances like AuFe which are formed by dissolving magnetic ions (Fe), in low concentrations in a non-magnetic host material (Au), at high temperatures . The substance is then cooled rapidly and the magnetic ions are frozen into position at random sites. This process is normally termed quenching, corresponding to freezing in disorder as opposed to annealing where a system is cooled slowly and order is allowed to build up. The conduction electrons in the spin glass become polarized by the magnetic spins which leads to an indirect exchange interaction between the spins described by the Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction [27]. This is a long range interaction which oscillates in sign with a period equal to the lattice spacing. Thus since the magnetic ions are at random sites the exchange interactions can be both positive or negative. We can therefore map the real lattice onto a new lattice with magnetic moments at each site but with random exchange interactions. The model which is thought to describe the essential features of this system is due to Edwards and Anderson [26] and its

Hamiltonian, with no external field, is given by,

$$H\{S_i\} = -\frac{1}{2}\sum_{<ij>} S_i J_{ij} S_j \tag{1.11}$$

where the interactions are defined by a Gaussian distribution,

$$P(J_{ij}) = \frac{1}{\sqrt{2\pi}J}\exp\frac{-(J_{ij}-J_0)^2}{2J^2} \tag{1.12}$$

Again the infinite range model (SK model), studied by Kirkpatrick and Sherrington [28] is closest to the Hopfield network having both positive and negative interactions. For the SK model the interactions have to be rescaled with the size of the system,

$$\begin{aligned} J_0 &\rightarrow \frac{J_0}{N} \\ J^2 &\rightarrow \frac{J^2}{N} \end{aligned} \tag{1.13}$$

which gives,

$$\begin{aligned} [J_{ij}]_{av} &= \frac{J_0}{N} \\ [J_{ij}^2]_{av} - [J_{ij}]_{av}^2 &= \frac{J^2}{N} \end{aligned} \tag{1.14}$$

where $[\;]_{av}$ denotes averaging over all the different interactions between sites. In the thermodynamic limit the interactions for the Hopfield neural network (see equation 1.4), also become Gaussian variables and we have for the synaptic strengths,

$$\begin{aligned} [T_{ij}]_{av} &= 0 \\ [T_{ij}^2]_{av} - [T_{ij}]_{av}^2 &= \frac{\alpha}{N} \end{aligned} \tag{1.15}$$

So at first sight the Hopfield model appears to be a spin glass with zero mean interaction. It is only when we look at the correlations between interactions at different sites that we see the essential difference between the Hopfield model and the SK spin glass. Consider for example the triple site correlation term,

$$[T_{ij}T_{jk}T_{ki}]_{av} \tag{1.16}$$

which contains terms of the form,

$$\xi_i^\mu \xi_j^\mu \xi_j^\mu \xi_k^\mu \xi_k^\mu \xi_i^\mu = 1 \tag{1.17}$$

as well as random terms with mean zero. Thus the expression in equation 1.16 is non-zero as are all the higher order bond correlations. In the case of the spin glass with $J_0 = 0$ all these interaction loops have zero mean. It is these correlation loops which distinguish a neural network from a spin glass and give it its characteristic storage properties. If we allow $\alpha \rightarrow \infty$ and rescale the interactions accordingly the loop averages tend to zero and the neural network behaves exactly like a spin glass. If the synapses are diluted giving a partially connected system some of these loops will be destroyed and we therefore expect the behaviour of the system to change. We may also expect the system to behave more like a spin glass. This is in fact exactly what happens and the full details of this change in behaviour are given in Chapters 2 and 3. Even with more complicated learning algorithms that keep T_{ij} Gaussian [17,19] we still expect these loops to play an important role in the behaviour of the system. We will now outline the mathematical formalism for studying Ising spin systems.

If we allow the system under study to equilibriate with a heat bath at temperature T then the probability of the system being in a particular state $\{S_i\}$ is given by the Boltzman distribution,

$$P\{S_i\} = \frac{\exp(-H\{S_i\}\beta)}{Z} \tag{1.18}$$

where β is the inverse temperature and Z is called the partition function and is given by,

$$Z = \sum_{\{S_i\}} \exp(-H\{S_i\}\beta) \tag{1.19}$$

where the sum is over possible realizations of the state of the system and is quite often written as $\underset{S_i}{\mathrm{Tr}}$. In general all the equilibrium thermodynamics of the system can be derived from the partition function or from the free energy which is closely related to it and given by,

$$F = -T \ln Z \tag{1.20}$$

where we have absorbed the Boltzman constant into the temperature. Thus, for example, the entropy of the system is given by,

$$S = -\frac{\partial F}{\partial T} \tag{1.21}$$

If we wish to know the value of some parameter of the system A, then we must calculate its thermodynamic average denoted $< A >$. This is given by,

$$< A >= \underset{S_i}{\mathrm{Tr}} A\{S_i\}P\{S_i\} \tag{1.22}$$

Equations 1.18 to 1.22 can be used to determine the thermodynamic properties of Ising type models at equilibrium.

An important class of parameters which are used to describe spin systems are order parameters. These characterize the ordering of the system at low temperature and have zero values at high temperatures. The regimes in which they are finite characterize different phases of the system. The system moves through these different phases when external parameters of the system such as temperature and magnetic field, are varied. The plot of these phases drawn in the space of all the important external parameters of the system is known as the phase diagram of the system. The single important order parameter for the Ising ferromagnetic model is the magnetization m, which is given by,

$$m = \frac{1}{N}\sum_i < S_i > \tag{1.23}$$

Below a certain critical temperature T_c this parameter takes on a non-zero value even in zero external field. This spontaneous magnetization corresponds to ferromagnetism in real magnets. The transition between two phases is termed first order if the derivative of the free energy changes discontinuously across the phase boundary and of order n if the lowest order of the derivative of the free energy which changes discontinuously across the phase boundary is n. In general for first order phase transitions the order parameters change discontinuously across the phase boundary and for second and higher order transitions the order parameters change continuously across the phase boundary. The phase diagram for the Ising model in greater than one dimension is given in fig 1.1. Spin glasses and neural networks require more than one order parameter to describe their low temperature behaviour and we shall discuss these parameters later when we have developed some more of the theoretical concepts for disordered systems.

Even though we have laid out a mathematical formalism in equations 1.18 to 1.22 for calculating the phase diagram of an Ising spin system it is often very difficult to calculate the spin sums in the partition function Z. The one dimensional

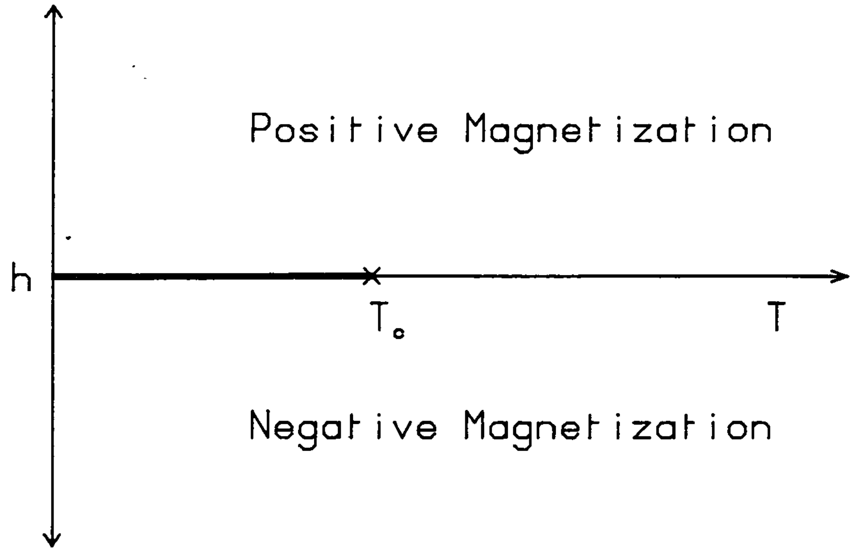

Figure 1.1: Phase diagram for the Ising model in greater than one dimension where h is the external magnetic field. At temperatures below T_c on the $h = 0$ axis the model enters a ferromagnetic phase (m $finite, h = 0$), while above T_c it enters a paramagnetic phase ($m = 0, h = 0$). At all finite values of h the model has a magnetic moment (m $finite$), in the same direction as h.

model is easily solved but does not exhibit a phase transition [20] while the two dimensional model has been solved by Onsager [21] but only for zero external field. The three dimensional model remains unsolved but many other techniques such as renormalization have described certain aspects of its phase transition (for a review see [22]). The two and three dimensional spin glass models remain unsolved and the infinite range model, normally found to very simple to solve, required a high degree of effort and mathematical complexity to solve (for a review see [28]). Many new techniques and theoretical concepts were developed in the study of spin glasses and it is these techniques that we shall apply to neural networks.

1.5 Saddle Point Mean Field Theory

Mean field theory is an approximate technique where the discrete interactions that one spin feels due to its neighbours are approximated by a continuous valued mean field which the spin sits in. A mean field calculation usually takes the form of replacing the discrete spins in the partition function by their thermodynamic averages and possibly a first order fluctuation part which is assumed to be small. The partition function can then be calculated by integrating over the thermodynamic averages which in the case of the Ising model is just the magnetization order parameter. This technique succeeds in predicting a phase transition for the Ising model but does not give the correct form of m across the phase boundary for systems of dimension three and lower. For dimensions higher than three the mean field theory model describes the behaviour of the magnetization across the phase boundary correctly. The threshold value of d, the dimension of the model above which the mean field approximation becomes qualitatively correct, is known as the upper critical dimension UCD. The UCD for the Ising model is therefore 4. The highest value of d for which the system does not exhibit a phase transition is known as the lower critical dimension and is one for the Ising model. These two values are not as yet known for the spin glass although there has been much speculation about their values [29].

In the case of infinite range models like neural networks mean field theory is exact. In these models each spin interacts with infinitely many other spins so the value of the total interaction each spin experiences becomes continuous and can be thought of as a field in which the spin sits. In this case we can evaluate the partition function exactly by means of what is termed the saddle point technique and hence the thermodynamic properties of the system can be deduced. This type of calculation usually takes the form of using a transformation to replace the sum over the discrete spins with an integration over the order parameters of the system. The infinite range Ising model represents a simple example of this type of model which can be generelised to more complicated systems such as spin glasses and neural networks.

Consider the partition function for the infinite range Ising model,

$$Z = \sum_{S_i} \exp(-\beta H\{S_i\}) \tag{1.24}$$

We will now split the spin sum into two sums; one which contains all the possible realizations of $\{S_i\}$ such that $\frac{1}{N}\sum_i S_i = m$ and another sum over all the possible values of m. We therefore have for the partition function,

$$Z = \sum_{m} \sum_{\{S_i\}_m} \exp(-\beta H\{S_i\}) \tag{1.25}$$

Now consider a free energy defined only on the given set of states which satisfy $\frac{1}{N}\sum_i S_i = m$ denoted $F(m)$. Then since from equation 1.20 we have $Z = \exp(-\beta F)$ we can rewrite the second sum as $\exp(-\beta F(m))$. In the limit of an infinite size system m can take continuous values so the first sum in the above equation becomes an integral, giving,

$$Z = \int \exp(-\beta F(m))dm \tag{1.26}$$

Thus we have replaced the sum over discrete spins by an integration over an order parameter. For infinite range models in the thermodynamic limit the N dependance of the free energy can be extracted from F to give,

$$Z = \int \exp(-\beta N f(m))dm \tag{1.27}$$

where f denotes the free energy per site which does not depend on N. As we take the thermodynamic limit the points corresponding to minima of f will dominate the integral. Therefore the values of m corresponding to the minima of f will be the only possible values of the order parameter for the system and they will vary with the external parameters of the system such as temperature and field. If there is more than one possible value of m then the value for the system will depend on its initial conditions (see section 1.9 for further discussion of this point). The values of m which dominate the integral are determined from,

$$\frac{\partial f(m)}{\partial m} = 0 \tag{1.28}$$

and the partition function is given by,

$$Z = \exp(-\beta N f(m)) \tag{1.29}$$

with the free energy per site of the system given by $f(m)$, the value of m being given by the solution of equation 1.28. This technique is called the saddle point method since it is the saddle point value of f which dominates the partition function. Calculations of this type are usually performed by evaluating Z and then determining the transformation which makes it of the form of equation 1.27. The free energy of the system then falls out from the exponent in the integral and the thermodynamic properties of the system can be determined from it. With other systems such as spin glasses or neural networks f is a function of more than one order parameter but the approach followed is very similar except that we end up with more order parameter equations. We will briefly illustrate the technique with the infinite range Ising model as in the mean field theory for neural networks many extra details associated with disordered systems mask the basic calculation. We will therefore formulate the model in the same way as our mean field calculations in Chapters 2 and 3 and it will be constantly referred to in these Chapters. The infinite range Ising model also illustrates the method by which second and higher order phase transitions can be determined analytically from transcendental order parameter equations. This technique will also be used in Chapter 3.

The Hamiltonian for the infinite range Ising model with no external field can be written as,

$$H\{S_i\} = -\frac{J}{2N}\sum_{i\neq j} S_i S_j = -\frac{J}{2N}\left(\left(\sum_i S_i\right)^2 - N\right) \tag{1.30}$$

and the Gaussian transformation,

$$\exp(a^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}y^2 + \sqrt{2}ay\right) dy \tag{1.31}$$

can be used to decouple the spins and introduce a new variable m, giving for the partition function, after some rescaling,

$$Z = \left(\frac{N\beta J}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta J}{2}\right) \underset{S_i}{\mathrm{Tr}} \int \exp \beta N J \left(-\frac{1}{2}m^2 + \frac{m}{N}\sum_i S_i\right) dm \tag{1.32}$$

The spins are now decoupled and can be summed giving,

$$Z = \left(\frac{N\beta J}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta J}{2}\right) \int \exp \beta N \left(-\frac{1}{2}Jm^2 + \frac{1}{\beta}\ln(2\cosh\beta J m)\right) dm \tag{1.33}$$

This is of the same form as equation 1.27 except for the constants in front of the integral. It is only constants of the form $\exp(Nc)$ which contribute to the free energy per site in the thermodynamic limit since the free energy per site is given by $\frac{1}{N}\ln(-\beta Z)$. They do however contribute to the total free energy of the system but this diverges in the thermodynamic limit. Therefore the exponent in the integral gives the free energy per site and the saddle point equation for f gives a transcendental equation in m,

$$m = \tanh \beta m J \tag{1.34}$$

So far in this calculation we have introduced m as a variable by means of a mathematical transformation. To find the physical meaning of m we must apply the saddle point method again to equation 1.32 which shows that m is the magnetization of the system, $\frac{1}{N}\sum_i < S_i >$.

Above a critical temperature T_c the only solution to equation 1.34 is $m = 0$ corresponding to the paramagnetic phase (see figure 1.1). Below T_c the equation has two solutions one positive and one negative which corresponds to the spins being aligned up or down in the ferromagnetic phase. Below T_c, $m = 0$ is also a solution but is a maximum rather than a minimum of the free energy and so does not give a stable state of the system. The order parameter m changes continuously across the phase boundary which turns out to be of second order. These results correspond to the $h = 0$ axis of figure 1.1.

Since the magnetization changes continuously across the phase boundary we can expand equation 1.34 about $m = 0$ which will correspond to being in close proximity to the phase boundary on the ferromagnetic side. This gives,

$$m = m\beta J - \frac{(m\beta J)^3}{3} + \cdots \tag{1.35}$$

If we now consider the limit $m \to 0$ then we are approaching the phase boundary from the ferromagnetic side. In this limit only low order terms in the sequence contribute and the first order terms give us the equation for the phase boundary,

$$m = m\beta_c J \tag{1.36}$$

which gives $T_c = J$. Solving the equation to second order tells us the value of m close to the phase boundary,

$$m \simeq \pm\sqrt{\frac{3}{(\beta J)^3}(\beta J - 1)} \tag{1.37}$$

If the first order equation for the phase boundary yields more than one solution then this second order equation can determine which is the valid one since, for a phase transition to be valid it must give a finite real solution at low temperatures and only a zero solution at high temperatures. The phase boundary must also be continuous for all real positive values of T. We have thus analytically determined the phase point on the $h = 0$ axis of the phase diagram for the infinite range Ising model (see figure 1.1).

In general for a spin glass or neural network many more transformations are required to decouple the spins and allow them to be summed. This in turn leads to many more order parameters in the expression for the free energy and hence the saddle point condition gives more than one order parameter equation. In general these equations will be of a transcendental form. The condition that all the eigenvalues of the matrix of second derivatives are positive is also required to guarantee that the saddle point is a minimum of the free energy. Expanding all the order parameter equations and then solving them to first order for a given order parameter equation yields potential candidates for a second or higher order phase change. Solving the equations to the next highest order gives the value of the order parameter close to the phase boundary and also determines which is the true phase change. In general if the phase boundary is first order or between two ordered phases numerical techniques have to be used to determine the phase boundary. This is because in these cases the order parameters are not all infintesimally small across the phase boundary and so we cannot expand the transcendental order parameter equations about zero. In some cases of second or higher order transitions between two ordered phases, the value of the order parameter which remains finite across the phase boundary is known and the phase boundary can then be calculated analytically. This does however depend on the form of the order parameter equations. We will in fact meet just such a case in Chapter 3 section 2.

The order parameters play a crucial role in the calculation of the free energy from the partition function in saddle point mean field theory. The introduction of the order parameters to the partition function by mathematical transformations allows the spin sums to be replaced with integrations over continuous variables. The correct choice of the order parameters, and the mathematical transformations required to introduce them, requires some intuitive skill and a knowledge of the distinctive features we expect the system to have at low temperatures. In

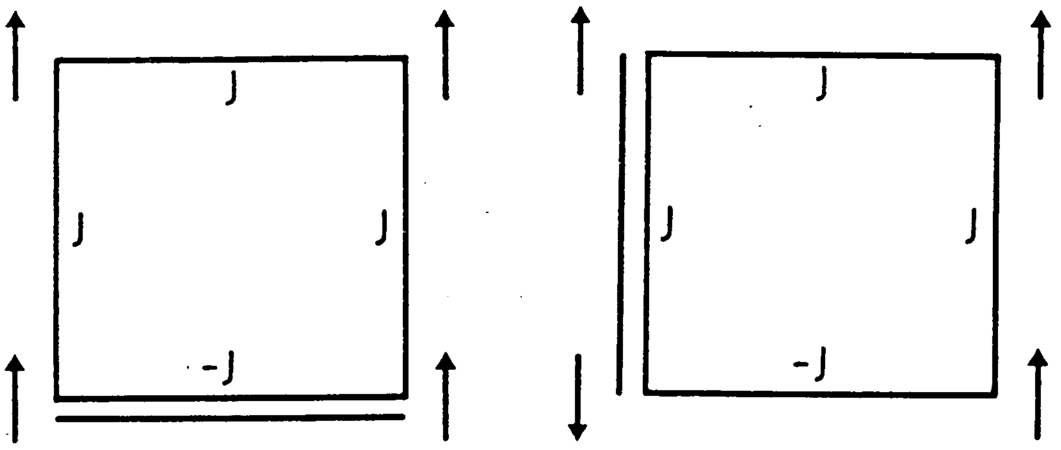

Figure 1.2: Plaquettes of spins in a neural network or spin glass. The bonds denoted by the double bar are not satisfied.

the next section we will discuss some of the common features of spin glasses and neural networks and the order parameters associated with these features. We have already seen in section 1.2 that the overlap m is an important parameter for the Hopfield model as it characterizes the correlation of the state of the system with the patterns nominated for storage. This will therefore be used as an order parameter in our mean field calculations in Chapter 2.

1.6 Frustration and Gauge Invariance

In the Ising model the interactions all have the same value J so at low temperatures the spins can align and satisfy the bonds. The free energy surface therefore has two minima corresponding to the spins being aligned upwards or downwards. In the case of neural networks or spin glasses the interactions can take both positive and negative values so the situation becomes more complex. Consider for example a plaquette with four spins sitting at each corner. For simplicity we will only consider interactions taking the values $\pm J$. Figure 1.2 shows typical plaquettes of random interactions and as can be seen in both arrangements of the spins one bond denoted by a double bar always remains unsatisfied. This inability to satisfy all the bonds is termed frustration and leads to a degeneracy of ground states with randomly orientated spins. If we now consider a system

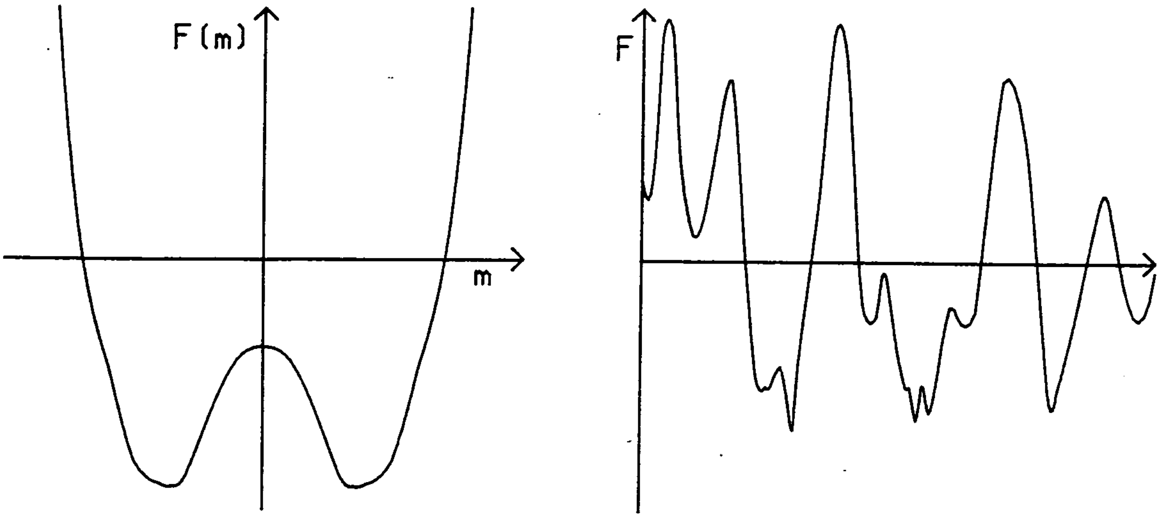

Figure 1.3: Left: Free energy surface of a ferromagnet below T_c as predicted by mean field theory, m is the magnetization order parameter.
Right: A section through the free energy surface of an equivalent spin glass or neural network system at low temperature plotted in the space of the order parameters.
In the thermodynamic limit the free energy barriers between the minima in both models become infinite.

of two plaquettes then, because there are two shared spins, some of the ground states of one plaquette may exclude the other plaquette from its ground state. We can thus have, not just degeneracy of the ground state of the two plaquette system but, also a degeneracy of higher energy metastable states. If we now consider a large system of many spins with a Gaussian bond distribution we would expect a large degeneracy of random metastable states at all different energy levels. This would cause the free energy surface to have a many valleyed structure at low temperatures. It is this many valleyed structure of the free energy surface that distinguishes a spin glass from the simple two valleyed structure of the ferromagnet (see figure 1.3). Neural networks also have this many valleyed structure with some of the valleys being associated with the states we are trying to store in the system and others being random spin glass states. Because of the randomness of the spin glass states we have, at low temperature, the possibility of the spins freezing into random positions. An order parameter (the EA order parameter), first proposed by Edwards and Anderson [26], which measures this random freezing is,

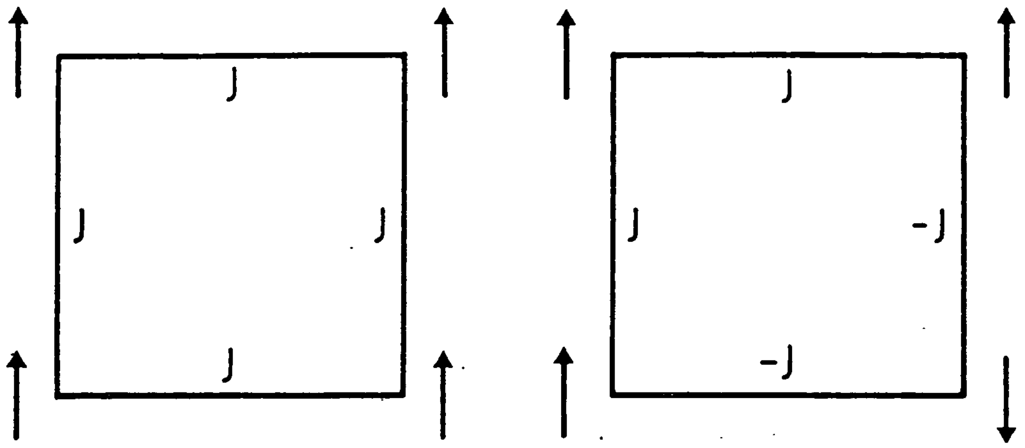

Figure 1.4: Left is a ferromagnetic plaquette while the right plaquette has ±J interactions but no frustrated bonds.

$$q = \frac{1}{N}\sum_i < S_i >^2 \tag{1.38}$$

This order parameter will tend to one as the temperature is lowered only reaching one in the zero temperature limit.

It is important to realize that it is not just the presence of $\pm J$ interactions which lead to frustration. Figure 1.4 shows how a plaquette can have $\pm J$ bonds but still have no frustration like the ferromagnetic case. It was Toulouse [30] who first introduced the parameter Φ which measures the frustration in a plaquette. Numbering the sites in a plaquette one to four we obtain,

$$\Phi = \mathrm{sgn}(J_{12}J_{23}J_{34}J_{41}) \tag{1.39}$$

If this expression is positive then all the bonds in the plaquette can be satisfied. In figure 1.2 we can see that $\Phi = -1$ but in figure 1.4 $\Phi = 1$ for both plaquettes.

There is a gauge symmetry present in the Hopfield model and spin glasses since the local transformation,

$$\begin{aligned} S_i' &= t_i S_i \quad t_i = \pm 1 \\ T_{ij}' &= t_i T_{ij} t_j \end{aligned} \tag{1.40}$$

leaves the partition function and hence the properties of the system unchanged. If we now consider a neural network where we are only trying to store one state

$\{\xi_i\}$ then,

$$T_{ij} = \frac{1}{N}\xi_i\xi_j \tag{1.41}$$

This is known as the Mattis model [31] and for any loop of sites Φ is always positive so there is no frustration in the system. Under the transformation given by equation 1.40 its disorder can be gauged away. Its Hamiltonian and partition function are therefore the same as the infinite range Ising model with two minima in the free energy at low temperature corresponding to the state $\{\xi_i\}$ and its image $\{-\xi_i\}$ (see figure 1.3). Therefore we have stored the nominated state in the network along with its image and there are no other minima in the free energy surface since there is no frustration in the system. It is only when we try and store more than one state in the system that we get frustration leading to extra unwanted minima in the free energy surface associated with spin glass states.

1.7 Quenched and Self-averaging

Consider a system of size N where we have chosen the set of interactions $\{J_{ij}\}$ from a Gaussian distribution. Now consider an order parameter A and let $A\{J_{ij}\}$ denote the value of that order parameter for a given choice of the interactions and let $\ll\ \gg$ denote averaging over all possible choices of the set $\{J_{ij}\}$. This type of averaging is called quenched averaging since we are averaging over the variables J_{ij} which are quenched; having no thermodynamic fluctuations. The parameter A is then said to self-average if,

$$A\{J_{ij}\} - \ll A\{J_{ij}\} \gg \rightarrow 0 \;\; as \;\; N \rightarrow \infty \tag{1.42}$$

We will now consider the Hopfield neural network model in more detail to decide which models self-average and which don't.

Consider one site i of a neural network of size N with all the interactions at that site defined by,

$$T_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{\mu}\xi_j^{\mu} \tag{1.43}$$

If p is finite and small the different T_{ij}'s at site i can only take a few different values and if $N \gg p$ these few values will be realized many times by the N

interactions at this site. In the limit $N \to \infty$ and p finite all the possible values of T_{ij} at site i will be realized infinitely many times. This means that the interaction at site i will not depend on the particular choice of $\{T_{ij}\}$ and so the system will self-average. We are in effect sampling a discrete distribution infinitely many times which gives an exact representation of that distribution. The finite p version of the Hopfield model was solved by Amit *et al* [5] as a precursor to the much more complicated infinite p model.

If we now consider p to be of the same order as N then a particular choice of the connection strengths T_{ij} does not properly sample the distribution. The interactions at site i will therefore depend on the particular choice of the set $\{T_{ij}\}$. This is also true in the thermodynamic limit where the distribution of T_{ij}'s becomes continuous. The system will not necessarily self-average and we may expect some of the properties of the system to change with the particular choice of the set $\{T_{ij}\}$. The lack of self-averaging of the interactions is only a sufficient condition for lack of self-averaging of individual parameters of the system. In the case of spin glasses an order parameter closely related to the EA order parameter q does not self-average (see [51]) while other parameters of the system do. The problem now arises as to how to introduce the quenched averaging into our calculation of the partition function. If the values of parameters vary with the specific choice of the interactions it is obviously the average of these values and possibly their standard deviation which we wish to calculate.

One might at first think that the easiest way to overcome this problem would be to perform the quenched averaging on the partition function and then carry on using standard techniques such as mean field theory to calculate the thermodynamic properties of the system. This turns out to be wrong which is well illustrated by this short example for finite size systems [29, page 838]. Consider an extensive observable of the system; we will use the free energy for simplicity since it is closely related to the partition function. In finite systems the lack of self-averaging of the interactions and their Gaussian nature produces a corresponding Gaussian distribution for the free energy [33]. We therefore have for the distribution of the free energy per site f,

$$P(f) \propto \exp\left(-\frac{N(f - f_{av})^2}{2(\delta f)^2}\right) \tag{1.44}$$

where the relationship between the free energy and the partition function is given by,

$$Nf = -\frac{\ln Z}{\beta} \tag{1.45}$$

If we now evaluate $-\beta \ln \ll Z \gg$ we do not get f_{av} but instead get $f_{av} + \beta(\delta f)^2$ which is clearly wrong. It is therefore necessary to perform the quenched averaging on real observables of the system rather than the partition function. The obvious observable to average is the free energy since most of the important parameters of the system can be derived from this. Unfortunately averaging the free energy turns out to be very difficult and a new technique know as the replica method [24,26] was developed to perform the quenched averaging. It is expected that the free energy self-averages in the thermodynamic limit for infinite range models although this has never been analytically proved. The self-averaging of the free energy for short range models has been proved by Khanin *et al* [50]. The numerical studies of the SK model by Kirkpatrick, Sherrington, [28] Palmer and Pond [34] suggest that the free energy self-averages in infinite range models, although a systematic study has never been carried out.

There is an important relationship between quenched averaging and configurational averaging for parameters that self-average. Consider the average magnetization,

$$m = \frac{1}{N}\sum_i \ll < S_i > \gg \tag{1.46}$$

The combination of quenched averaging and configurational averaging means that the spin-averages are translationaly invariant. We can therefore express the magnetization as,

$$m = \ll < S > \gg \tag{1.47}$$

where we have dropped the site index i so S can be the spin at any site. Now if m self-averages then from equation 1.42 we have,

$$\ll < S > \gg = \frac{1}{N}\sum_i < S_i > \tag{1.48}$$

Therefore, for parameters that self-average, the configurational average of a summation over all sites can be replaced with a quenched and configurational average at only one site. We will use this result in section 2.2 to derive the order parameter equations for a partially connected network.

1.8 The Replica Method

We wish to evaluate the free energy in order to determine the equilibrium thermodynamics of the Hopfield model. Even though we expect the free energy to self-average it is still necessary to make use of the quenched averaging to introduce some symmetry into the infinite sum of random interaction values which would otherwise be intractable. We therefore have to evaluate,

$$F = -\frac{1}{\beta} \ll \ln Z\{J_{ij}\} \gg \qquad (1.49)$$

The first step would be to evaluate $\ln Z\{J_{ij}\}$ but this is not possible since it depends on the infinite set $\{J_{ij}\}$ which, as we have mentioned, has no symmetry properties. Performing the quenched averaging on each term and then trying to sum them is not possible either since the log of a complicated function can generally not be integrated explicitly, especially if it involves some form of Gaussian averaging. So the log form of this expression prevents us from taking advantage of the quenched averaging to simplify the problem. The replica method [24,26] involves expressing f in such a way that the log is removed and the quenched averaging can be performed in such a way as to reduce the expression for the free energy to a tractable form. It is based on the expansion,

$$x^n = \exp(n \ln x) = 1 + n \ln x + \cdots \qquad (1.50)$$

Taking the limit $n \longrightarrow 0$ and rearranging this gives,

$$\ln x = \lim_{n \to 0} \frac{x^n - 1}{n} \qquad (1.51)$$

which putting $x = Z$ and performing the quenched averaging gives,

$$\ll \ln Z \gg = \lim_{n \to 0} \frac{\ll Z^n \gg -1}{n} \qquad (1.52)$$

Stricly speaking the thermodynamic limit $N \longrightarrow 0$ should be taken after the limit $n \longrightarrow 0$ but it is necessary to interchange the limits to make the calculation tractable. The interchange of these limits has been studied in detail [35,36] and it is now generally accepted that it does not lead to any problems. We thus have the following expression for the free energy,

$$F = \lim_{n \to 0} \lim_{N \to \infty} -\frac{\ll Z^n \gg -1}{\beta n} \qquad (1.53)$$

Therefore we have reduced the calculation of F to the calculation of the quenched average of the partition of n replicas of the system. In practice the replica calculation is carried out by performing saddle point mean field theory on n replicas of the system. The free energy of n replicas then falls out from the exponent of $\ll Z^n \gg$ with all the corresponding expressions for the order parameters of the n replica system. The limit $n \longrightarrow 0$ can then be taken on the free energy and order parameter equations to obtain the properties of a single system. The equations obtained in this way are called the replica symmetric equations of the system

The n systems in the calculation are non-interacting but are coupled to the same heat bath. Z^n thus contains n sets of spins $\{S_i^\alpha, \alpha = 1, \ldots, n\}$ and is invariant under permutations of indices α when n is an integer. It is not clear though, what happens when n is allowed to become continuous so that the limit $n \longrightarrow 0$ can be taken. The problem associated with the invalidity of this limit is called replica symmetry breaking and leads to invalid solutions in certain regions of the phase diagram. This problem will be discussed in much more detail in Chapters 2 and 3 once we have derived the replica and replica symmetric equations for partially connected Hopfield networks. It is interesting to note that the replica technique was originally introduced as a mathematical trick to allow the free energy to be evaluated. It eventually turned out that the n replicas of the system involved in the trick were essential to a complete description of the many valleyed structure of the phase space associated with spin glasses.

1.9 Ergodicity

A system is said to be ergodic if, during the period it is under study, it explores all regions of its phase space with equilibrium probabilities given by the Boltzman distribution (see equation 1.18). If we consider the Ising model with no external field then the probability of a given state remains unchanged under the flipping of all the spins. Therefore, the probability of the system being in a state with magnetization m is the same as that of being in a state with magnetization $-m$ and so the average magnetization is always zero if the system is ergodic.

If we look at the free energy surface of the Ising model at low temperature (see figure 1.3) then we can see that there is a free energy barrier separating the up spin states from the down spin states. If this energy barrier is large (which happens as the dimensionality of the system is increased) there will be a low probability of the system being able to move from one side to the other and the symmetry of the system will be broken. The typical time it takes before there is a big enough statistical fluctuation to push the system over the barrier is called the relaxation time of the model. On time scales less than the relaxation time the system is therefore non-ergodic and does not move through the whole phase space. To describe the system on time scales smaller than the relaxation time we therefore have to restrict the partition function to one side of the free energy phase diagram which correspondingly gives us a non-zero value of m. The property of ferromagnetism is therefore associated with extremely long relaxation times.

In the case of mean field calculations we are dealing with infinite size, infinite range models and the free energy barriers are of infinite height. These systems are therefore truly non-ergodic on all time scales and the system will stay in the free energy basin it is started in. This idea is shown in the mean field theory of the Ising model (see section 1.5 and figure 1.3) where the free energy surface has two minima, at low temperatures, situated at m and $-m$. In the case of spin glasses and neural networks the free energy surfaces have many minima, separated by barriers of infinite height, associated with different values of the many order parameters. Strictly speaking, in our mean field calculations we should always include a small symmetry breaking field h to pick out one of the minima and then take the limit $h \to 0$. In practice for all the calculations in this thesis all the results of mean field theory can be interpreted correctly without the need of this symmetry breaking field so we will not include it in the calculations.

Chapter 2

Replica and Replica-symmetric Mean Field Equations for a Partially Connected Hopfield Network

2.1 Introduction

In this chapter we will follow the techniques used by Kirkpatrick and Sherrington [28] to study the infinite range spin glass model and extended by Amit *et al* [6,4] to cover neural networks. The basic approach will be to use the replica method which defines the free energy F in terms of the partition function of n replicas of the system as,

$$F = \lim_{n \to 0} \lim_{N \to \infty} -\frac{\ll Z^n \gg -1}{\beta n} \tag{2.1}$$

As discussed in section 1.5 on mean field theory we will start by evaluating $\ll Z^n \gg$. Order parameters will then be introduced which will simplify the traces over the spins and in the case of the replica symmetry theory remove them all together. The free energy for n replicas of the system will fall out from the exponent of the integral over the order parameters in the expression for $\ll Z^n \gg$. The saddle point equations for the free energy will then give us the order parameter equations for the system. These along with the free energy will constitute what we shall refer to as the replica equations for the system which will be exact in the thermodynamic limit. We will then take the replica-symmetric limit $n \to 0$ on these equations to derive the replica-symmetric equations for the system. In the last section of this chapter we will look at the stability of the solutions of the replica-symmetric equations by considering replica fluctuations about them for different connection architectures. The actual replica-symmetric equations are rather complex to solve and we will leave it to Chapter 3 to present some solutions of them.

2.2 The Replica Equations

We define the connection strengths T_{ij} for a partially connected neural network as,

$$T_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} D_{ij}\xi_i^\mu \xi_j^\mu, \quad i \neq j, \;\; T_{ii} = 0 \tag{2.2}$$

D is a matrix with all elements equal to 1 or 0 corresponding to connections being present or not. In this way we can define any connection architecture we wish. In order for mean theory to be exact in the thermodynamic limit each site must interact with an infinite number of other sites. We will therefore only consider choices of **D** which satisfy this. We define the connectivity of the network w by,

$$w = \frac{\textit{total number of connections}}{N^2} \tag{2.3}$$

The Hamiltonian for a partially connected network is given by,

$$H = \frac{-1}{2N}\sum_{i,j,i\neq j,\mu} \xi_i^\mu S_i D_{ij} \xi_j^\mu S_j \tag{2.4}$$

We will first briefly consider the simple case when p, the number of states nominated for storage is finite. As we have already mentioned we are only considering cases where each site has an infinite number of connections to it so the system will self-average as discussed in section 1.4. We therefore expect that, providing the temperature is rescaled with w, the properties of all finite p, infinite-range partially connected networks to be the same as for a fully connected network. This means that all the states will be stored exactly by the network, and states which are a mixture of the stored states will also be stable. There is an infinite number of interactions at each site to store a finite amount of information so it is stored exactly. The details of the fully connected network for finite p are presented in reference [5]. We will now consider the case where p is of order N and the system does not self-average.

We define the storage ratio of the network as,

$$\alpha = \frac{p}{N} \tag{2.5}$$

and we have for the partition function of n replicas (labelled by $\rho = 1, 2, \ldots, n$),

$$\ll Z^n \gg = \ll \operatorname*{Tr}_{S^\rho} \exp\left[\frac{\beta}{2N}\sum_{ij\rho\mu}(\xi_i^\mu S_i^\rho) D_{ij} (\xi_j^\mu S_j^\rho) - \frac{1}{2}\beta pn\right] \gg \tag{2.6}$$

The $\frac{1}{2}\beta pn$ term comes from the $i = j$ term and we therefore set $D_{ii} = 1$. We will now use the general form of the Gaussian transformation to decouple the spins and introduce an order parameter. The calculation at this stage follows the same procedure as the mean field calculation for the infinite-range Ising model (see section 1.5). The Gaussian transformation for many variables is given by:

$$\exp(\frac{1}{2}\mathbf{s}^t\mathbf{Q}\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^k \mid \mathbf{Q} \mid}} \int d^k y \exp(-\frac{1}{2}\mathbf{y}^t\mathbf{Q}^{-1}\mathbf{y} + \mathbf{y}.\mathbf{s}) \tag{2.7}$$

where $\mathbf{Q}, \mathbf{s}$ and $\mathbf{y}$ are a matrix and two vectors of dimension k respectively. We set,

$$\mathbf{D} = \mathbf{Q}\beta N, \quad \beta\xi_i^\mu S_i^\rho = s_k \tag{2.8}$$

and introduce a new variable,

$$m_{\rho i}^\mu = y_k \tag{2.9}$$

which, as we shall see later, is an order parameter measuring the overlap of a nominated state with the state of the system at site i. We will look for solutions where only a finite number of patterns can condense out at low temperature though it is possible in the thermodynamic limit to have an infinite number condensing out. In the latter case standard mean field theory breaks down as we have an infinite number of order parameters and the free energy per site diverges as $N \to \infty$. These ideas are discussed in more detail in [37]. Eventually in deriving the mean field equations we will restrict ourselves to solutions with only one condensed pattern as we expect these to be the most important for storage.

To allow for a finite number of patterns condensing out at low temperature we will split the sum over the p patterns into two separate sums. One sum will correspond to a finite number s of patterns $(\sum_{\nu=1}^{s})$ which may condense out and the other sum will be over the remaining infinite number of $p-s$ patterns $(\sum_{\mu=s+1}^{p})$. Thus using the Gaussian transformation from equation 2.7 gives us,

$$\begin{aligned} \ll Z^n \gg \; = \; & \exp\left(-\frac{\beta pn}{2}\right) \ll \operatorname*{Tr}_{S_i^\rho} \left[\frac{(2\pi)^N}{N\beta} \mid \mathbf{D} \mid\right]^{-\frac{pn}{2}} \int \prod_{i,\mu,\rho} dm_{\rho i}^{\mu} \\ & . \exp \beta N \left(-\frac{1}{2} \sum_{ij\rho\nu} m_{\rho i}^{\nu} D_{ij}^{-1} m_{\rho j}^{\nu} + \frac{1}{N} \sum_{i\rho\nu} m_{\rho i}^{\nu} \xi_i^{\nu} S_i^{\rho}\right) \\ & . \exp \beta N \left(-\frac{1}{2} \sum_{ij\rho\mu} m_{\rho i}^{\mu} D_{ij}^{-1} m_{\rho j}^{\mu} + \frac{1}{N} \sum_{i\rho\mu} m_{\rho i}^{\mu} \xi_i^{\mu} S_i^{\rho}\right) \gg \end{aligned} \tag{2.10}$$

where $\prod_{i,\mu,\rho} dm_{\rho i}^{\mu}$ is defined for all $\mu = 1, 2, \ldots, p$.

When $\mathbf{D}$ is singular the expression for Z is undefined with the simplest example of this being when the network is fully connected ($D_{ij} = 1 \;\; \forall i, j$). The solution to this problem was suggested some time ago by Berlin and Kac [25] in their study of magnetic systems using order parameters. In our calculations so far we have chosen $D_{ii} = 1$ (see equation 2.6), but it can be chosen arbitrarily so that $D_{ii} = a$, (a real) removing the singularity of $\mathbf{D}$. In the thermodynamic limit, which corresponds to $\mathbf{D}$ being of infinite dimension, we expect the value of a required to remove the singularity to be finite or at least of negligible size compared to N. In the case of the fully connected model with $D_{ii} = 1$ the smallest eigenvalue of the matrix is zero therefore adding a small finite value to D_{ii} will remove the singularity. In all our subsequent calculations we will take $D_{ii} = 1$ for simplicity since, choosing $D_{ii} = a$ with a finite will not effect the form of the order parameter equations we shall derive.

The parameter $m_{\rho i}^{\mu}$ in equation 2.10 measures the condensation of a pattern so in the first sum over ν, $m_{\rho i}^{\nu}$ can be of order one but in the second, $m_{\rho i}^{\mu}$ is a random Gaussian variable with a standard deviation of order $\frac{1}{\sqrt{N}}$. It is therefore possible to carry out the quenched averaging over the $p-s$ high ξ's since $m_{\rho i}^{\mu}$ is uncorrelated with ξ_i^{μ}. We have for the last line of the previous equation after carrying out the quenched averaging:

$$\exp\left(-\frac{\beta N}{2} \sum_{ij\rho\mu} m_{\rho i}^{\mu} D_{ij}^{-1} m_{\rho j}^{\mu} + \sum_{i\mu} \ln \; \cosh \beta \sum_{\rho} m_{\rho i}^{\mu} S_i^{\rho}\right) \tag{2.11}$$

We rescale $m^{\mu}_{\rho i}$ to obtain a well defined limit as $N \rightarrow \infty$ in the integral,

$$m^{\mu}_{\rho i} \rightarrow \frac{m^{\mu}_{\rho i}}{\sqrt{N}} \tag{2.12}$$

and expanding in N we get as $N \rightarrow \infty$,

$$\exp \beta \left(-\frac{1}{2} \sum_{ij\rho\mu} m^{\mu}_{\rho i} D^{-1}_{ij} m^{\mu}_{\rho j} + \frac{\beta}{2N} \sum_{i\mu\rho\sigma} m^{\mu}_{\rho i} m^{\mu}_{\sigma i} S^{\rho}_{i} S^{\sigma}_{i} \right) \tag{2.13}$$

We are now in a position to integrate out the $m^{\mu}_{\rho i}$'s by using the general form of the Gaussian integral,

$$\left(\frac{1}{\sqrt{2\pi}} \right)^{m} \int \exp(-\frac{1}{2} \mathbf{x}^{t} \mathbf{A} \mathbf{x}) d\mathbf{x} = (\mid \mathbf{A} \mid)^{-\frac{1}{2}} \tag{2.14}$$

where m is the dimensionality of the matrix $\mathbf{A}$. Therefore defining,

$$\mathbf{K}_{ij\rho\sigma} = D^{-1}_{ij} \delta_{\rho\sigma} - \frac{\beta}{N} \delta_{ij} \delta_{\rho\sigma} - \frac{\beta}{N} \delta_{ij} q^{\rho\sigma}_{i} \tag{2.15}$$

where we have set,

$$q^{\rho\sigma}_{i} = S^{\rho}_{i} S^{\sigma}_{i} \tag{2.16}$$

we can now integrate over the $m^{\mu}_{\rho i}$'s which will give, ignoring constants and expressing the determinant of a matrix as the exponent of the trace of the log of the matrix,

$$\int \prod_{i,\rho>\sigma} dq^{\rho\sigma}_{i} \exp(-\frac{1}{2} \operatorname*{Tr}_{ij\rho\sigma} N\alpha \ln \beta \mathbf{K}) \prod_{i,\rho>\sigma} \delta(q^{\rho\sigma}_{i} - S^{\rho}_{i} S^{\sigma}_{i}) \tag{2.17}$$

where we have introduced $q^{\rho\sigma}_{i}$ via a delta function. We can introduce another parameter $r^{\rho\sigma}_{i}$ by means of the complex expression for the delta function,

$$\delta(a) = \int_{-i\infty}^{+i\infty} \frac{d\mu}{2\pi i} \exp(a\mu) \tag{2.18}$$

which brings the expression for the partition function back to the desired exponential form. The exponential form of the partition function is necessary for the application of the saddle point method. We therefore have for the delta function in equation 2.18.

$$\delta(q^{\rho\sigma}_{i} - S^{\rho}_{i} S^{\sigma}_{i}) = \frac{\alpha\beta^{2}}{2\pi i} \int_{-i\infty}^{+i\infty} dr^{\rho\sigma}_{i} \exp\left(-\frac{1}{2} \alpha\beta^{2} r^{\rho\sigma}_{i} q^{\rho\sigma}_{i} + \frac{1}{2} \alpha\beta^{2} S^{\rho}_{i} S^{\sigma}_{i} r^{\rho\sigma}_{i} \right) \tag{2.19}$$

where the integral is along a contour running in the direction of the imaginary axis. We can shift this contour so that $r^{\rho\sigma}_{i}$ can have a real as well as imaginary part. The introduction of the constants $\alpha\beta^2$ will, as we shall see later, make the final expressions for the order parameters much simpler. We now wish to put all this back into equation 2.10. First though we can simplify the expression by taking the constant term involving $\mid \mathbf{D} \mid$ from equation 2.10 inside the exponent giving,

$$\exp\left[-\frac{1}{2}N\alpha n \operatorname*{Tr}_{ij} \ln\left(\frac{\mathbf{D}}{N\beta}\right)\right] \tag{2.20}$$

We can then add this to the $\ln\beta\mathbf{K}$ factor and ignoring constants which occur in both terms we get,

$$\begin{aligned} & \operatorname*{Tr}_{ij\rho\sigma} \ln\beta\mathbf{K} + n\operatorname*{Tr}_{ij}\ln\left(\frac{\mathbf{D}}{N\beta}\right) && (2.21) \\ = \; & \operatorname*{Tr}_{ij\rho\sigma} \ln D_{ij}^{-1}N\beta\left(\delta_{\rho\sigma}\delta_{ij} - \frac{\beta}{N}D_{ij}\delta_{\rho\sigma} - \frac{\beta}{N}D_{ij}q_i^{\rho\sigma}\right) + \operatorname*{Tr}_{ij\rho\sigma} \ln\left(\delta_{\rho\sigma}\frac{D_{ij}}{N\beta}\right) \end{aligned}$$

which setting $q_i^{\rho\rho} = 1$ gives,

$$\operatorname*{Tr}_{ij\rho\sigma} \ln\left(\delta_{\rho\sigma}\delta_{ij} - \frac{\beta D_{ij}}{N}q_i^{\rho\sigma}\right) \tag{2.22}$$

Putting all this back in the expression for the partition function equation 2.10, gives us,

$$\begin{aligned} \ll Z^n \gg \; = \; & \left(\frac{\beta N}{2\pi}\right)^{\frac{1}{2}nsN}\left(\frac{\beta^2\alpha}{2\pi i}\right)^{\frac{1}{2}n(n-1)N} \ll \operatorname*{Tr}_{S_i^\rho}\int\prod_{i,\nu,\rho>\sigma} dm_{\rho i}^\nu dq_i^{\rho\sigma} dr_i^{\rho\sigma} \\ & \times\exp N\left[\frac{-\beta n\alpha}{2} - \frac{1}{2}\beta\sum_{ij\nu\rho} m_{\rho i}^\nu D_{ij}^{-1} m_{\rho j}^\nu\right. \\ & -\frac{1}{2}\alpha\operatorname*{Tr}_{ij\rho\sigma}\ln\left(\delta_{\rho\sigma}\delta_{ij} - \frac{\beta D_{ij}}{N}q_i^{\rho\sigma}\right) - \frac{1}{2}\alpha\beta^2\frac{1}{N}\sum_{i,\rho\neq\sigma} r_i^{\rho\sigma}q_i^{\rho\sigma} \qquad (2.23) \\ & \left. +\frac{1}{2}\alpha\beta^2\frac{1}{N}\sum_{i\rho\sigma} S_i^\rho S_i^\sigma r_i^{\sigma\rho} + \frac{\beta}{N}\sum_{i\nu\rho} m_{\rho i}^\nu \xi_i^\nu S_i^\rho\right] \gg \end{aligned}$$

Even though $r_i^{\rho\sigma}$ may be complex we will only be interested in real physically meaningful solutions although, when we are studying the stability of the order parameter equations it will be important to remember that $r_i^{\rho\sigma}$ must be stable to fluctuations along only the complex axis. It will also be necessary, when studying the stability of the solutions, to shift the fluctuation along the complex axis so that it passes through the saddle points of the other real order parameters. The stability conditions are studied in detail in the last section of this chapter.

Equation 2.24 is of the correct form for the saddle point method to be applied (see section 1.5), which gives, for the physical meaning of the parameters we have introduced,

$$\begin{aligned} m^{\nu}_{\rho i} &= \frac{1}{N} \ll \sum_j D_{ij}\xi^{\nu}_j < S^{\rho}_j > \gg \\ q^{\rho\sigma}_i &= \ll < S^{\rho}_i >< S^{\sigma}_i > \gg \\ r^{\rho\sigma}_i &= \frac{1}{\alpha} \sum_{\mu=s+1}^{\alpha N} \ll m^{\mu}_{\rho i} m^{\mu}_{\sigma i} \gg \end{aligned} \qquad (2.24)$$

At this point we cannot proceed any further unless we make some assumptions about the architecture of the system. The equations we have derived so far describe a system with an infinite number of order parameters, so we would end up with an infinite number of order parameter equations to be solved simultaneously. The standard mean field theory saddle point technique cannot cope with an infinite number of order parameters and this is reflected in the fact that the free energy per site becomes unbounded in the thermodynamic limit because of the N'th power in the constant terms in front of the integral (see equation 2.32). We must reduce the number of order parameters to a finite number in order to solve this problem. This can be done by restricting our choice of **D** to only translationally invariant matrices.

If we consider neurons to be sitting at sites on a hypercubic lattice then in the limit $N \rightarrow \infty$ we can think of the lattice as being continuous with every point representing a neuron. The lattice can then also be of finite size. For any neuron, the neurons connected to it will form a shape or shapes on this lattice. We will henceforth refer to this as the connection space of that neuron and only look at systems where the connection space is the same at every site. This concept of connection space is discussed in more detail in section 3.1 where we use it in setting up a numerical method of solving the order parameter equations. Choosing the connection space to be translationally invariant means that the matrix **D** will be translationally invariant. Therefore we can now define w, the connectivity ratio, as,

$$w = \frac{1}{N}\sum_i D_{ij}, \quad \forall j \qquad (2.25)$$

If we consider the values of $\ll \xi_i < S_i > \gg$ at each site they will form some kind of distribution curve. Then for a translationally invariant architecture the network contains a large number of macroscopic subsystems which are identical and have identical environments. We can think of these subsystems as building blocks from which the whole system can be constructed by performing translations on these blocks. Each of these macroscopic subsystems will have the same distribution curve for $\ll \xi_i < S_i > \gg$ and hence the same average value. The parameter $m^{\nu}_{\rho i}$ is an average over such a macroscopic subsystem and therefore its value will not depend on its index i. Therefore from equation 2.24 we can see that $r^{\rho\sigma}_i$ will also be independent of i, but we must be more careful with $q^{\rho\sigma}_i$ since its value depends on a single site rather than a macroscopic sum. If we now look at the log term containing $q^{\rho\sigma}_i$ in the partition function and expand in $\mathbf{D}/N$ we get,

$$\operatorname*{Tr}_{ij\rho\sigma} \ln\left(\delta_{\rho\sigma}\delta_{ij} - \frac{\beta D_{ij}}{N} q^{\rho\sigma}_i\right) = \frac{1}{N}\sum_{i\rho} D_{ii} q^{\rho\rho}_i + \frac{1}{N^2}\sum_{ij\rho\sigma} q^{\rho\sigma}_i D_{ij} q^{\rho\sigma}_j + \cdots \quad (2.26)$$

Therefore $q^{\rho\sigma}_i$ only occurs as a macroscopic sum in this term as it does in the other term containing it in the partiton function since $r^{\rho\sigma}_i$ is independent of i (see equation 2.24). We can now define a finite set of site independent order parameters,

$$\begin{aligned} m^{\nu}_{\rho} &= m^{\nu}_{\rho i} \\ r^{\rho\sigma} &= r^{\rho\sigma}_i \\ q^{\rho\sigma} &= \frac{1}{N}\sum_i D_{ij} q^{\rho\sigma}_i \end{aligned} \quad (2.27)$$

We now expect to be able to use mean field theory techniques and the saddle point method to derive solutions for the order parameters though the replicas complicate things considerably. It may be possible to solve a neural network model which is not translationally invariant but the connection matrix $\mathbf{D}$ would have to contain enough symmetry so that the number of order parameters required was finite. We can imagine a network where one set of sites has more connections to it than another set of sites and therefore the overlaps $m^{\nu}_{\rho i}$, associated with the second set, would be lower than the first.

At this point to make the form of the order parameter equations simpler and closer to the fully connected model we shall rescale some of the important parameters of the system,

$$m^{\nu}_{\rho} \rightarrow w m^{\nu}_{\rho}$$

$$\begin{aligned} q^{\rho\sigma} &\longrightarrow wq^{\rho\sigma} \\ r^{\rho\sigma} &\longrightarrow wr^{\rho\sigma} \\ T &\longrightarrow wT \\ \alpha &\longrightarrow w\alpha \end{aligned} \tag{2.28}$$

This will give us order parameter equations of a similar form to the fully connected model and the physical interpretations of the order parameters will lose their explicit w dependence. It is important to realize now that α is measure of the storage per connection and is given by,

$$\alpha = \frac{p}{wN} \tag{2.29}$$

We now have for the partition function,

$$\begin{aligned} \ll Z^n \gg \; = \; & \left(\frac{\beta N w}{2\pi}\right)^{\frac{1}{2}ns} \left(\frac{\beta^2 \alpha N w}{2\pi i}\right)^{\frac{1}{2}n(n-1)} \ll \operatorname*{Tr}_{S_i^\sigma} \int \prod_{\nu,\rho>\sigma} dm_\rho^\nu dq^{\rho\sigma} dr^{\rho\sigma} \\ & \times \exp Nn\beta w \left[\frac{1}{2}\alpha + \frac{w\alpha}{2\beta n} \operatorname*{Tr}_{ij\rho\sigma} \ln \left(\mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q} \times \mathbf{D}\right)\right. \\ & + \frac{1}{2n}\sum_{\rho\sigma}(m_\rho^\nu)^2 + \frac{\alpha\beta}{2n}\sum_{\rho\neq\sigma} r^{\rho\sigma} q^{\rho\sigma} \\ & \left. - \frac{1}{n\beta}\frac{1}{2}\alpha\beta^2 \frac{1}{N}\sum_{i,\rho\neq\sigma} S_i^\rho S_i^\sigma r^{\rho\sigma} + \frac{\beta}{N}\sum_{i\nu\rho} m_\rho^\nu \xi_i^\nu S_i^\rho \right] \gg \end{aligned} \tag{2.30}$$

where to simplify the notation we have introduced × which is the Kronecker product of two different dimensional matrices. The Kronecker product of two matrices is defined in Appendix A along with some important results for the product which will be used throughout this work. It has only been possible to introduce the Kronecker product at this stage because $q^{\rho\sigma}$ is independent of i. Therefore $\mathbf{q}$ is the matrix of $q^{\rho\sigma}$'s with $q^{\rho\rho} = 1$ and $\mathbf{I}_{nN}$ is the nN dimensional identity matrix. The constant terms in front of the integral do not contribute to the fee energy per site which is now bounded in the thermodynamic limit. The physical meanings of these new order parameters are then obtained from equation 2.31 by applying the saddle point method again giving,

$$\begin{aligned} m_\rho^\nu &= \frac{1}{N} \ll \sum_i \xi_i^\nu < S_i^\rho > \gg \\ q^{\rho\sigma} &= \frac{1}{N} \ll \sum_i < S_i^\rho >< S_i^\sigma > \gg \end{aligned}$$

$$\begin{aligned} r^{\rho\sigma} &= \frac{1}{\alpha}\sum_{\mu=s+1}^{p} \ll \frac{1}{N}\sum_i \xi_i^\mu < S_i^\rho > \frac{1}{N}\sum_i \xi_i^\mu < S_i^\sigma > \gg \\ &= \frac{1}{\alpha}\sum_{\mu=s+1}^{p} \ll m_\rho^\mu m_\sigma^\mu \gg \end{aligned} \tag{2.31}$$

which are the same order parameters Amit *et al* [6] used to describe a fully connected network. So m_ρ^ν is a measure of the overlap of the state of the system with the finite set of patterns nominated for condensation. The parameter $q^{\rho\sigma}$ is a measure of the alignment of spins at each site in different replicas at low temperature. $r^{\rho\sigma}$ is a measure of the overlap of the state of the system with the infinite set of $p - s$ patterns not nominated for condensation. Our aim in introducing these order parameters was to allow the spin sums to be removed from the expression for the partition function so that it could be formulated only in terms of order parameters. We will now see how the spin sums can only be simplified in replica theory giving us order parameter equations which can be solved but are extremely complicated. It is only in replica-symmetric theory that the spin sums can be removed from the partition function completely giving us analytical equations for the order parameters. Looking at the expression for the partition function (see equation 2.31), the last term involves a sum over the finite set s of states nominated for condensation. As discussed in section 1.7, terms with quenched averaging over a finite number of states will self-average as will the whole expression for the free energy. Rather than drop the quenched averaging at this point it is more convenient to remember that, for parameters that self-average, we can replace the configurational averaging over all sites with a quenched average of the configurational average at only one site (see section 1.7). We can therefore drop the site indices on the spins but keep in the quenched averaging. The trace over the spins can now be taken inside the exponent (see equation 2.31), so that Z can be written in the form given in equation 1.27 (see section 1.5),

$$\ll Z^n \gg = \left(\frac{\beta N w}{2\pi}\right)^{\frac{1}{2}ns} \left(\frac{\beta^2 \alpha N w}{2\pi i}\right)^{\frac{1}{2}n(n-1)} \int \prod_{\nu,\rho>\sigma} dm_\rho^\nu dq^{\rho\sigma} dr^{\rho\sigma} \exp[-Nn\beta f(m_\rho^\nu, q^{\rho\sigma}, r^{\rho\sigma})] \tag{2.32}$$

where the free energy per site f is given by,

$$\begin{aligned}\frac{f}{w} &= \frac{1}{2}\alpha + \frac{w\alpha}{2\beta n}\mathop{\mathrm{Tr}}_{ij\rho\sigma} \ln\left(\mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q}\times\mathbf{D}\right) \\ &+\frac{1}{2n}\sum_{\rho\sigma}(m_\rho^\nu)^2 + \frac{\alpha\beta}{2n}\sum_{\rho\neq\sigma} r^{\rho\sigma}q^{\rho\sigma} \\ &-\frac{1}{n\beta} \ll \ln \mathop{\mathrm{Tr}}_{S^\rho} \exp\left(\frac{1}{2}\alpha\beta^2\sum_{\rho\neq\sigma} S^\rho S^\sigma r^{\rho\sigma} + \beta\sum_{\nu\rho} m_\rho^\nu \xi^\nu S^\rho\right) \gg \qquad (2.33)\end{aligned}$$

the saddle point equations are then given by,

$$\frac{\partial f}{\partial m_\rho^\nu} = \frac{\partial f}{\partial q^{\rho\sigma}} = \frac{\partial f}{\partial r^{\rho\sigma}} = 0 \qquad (2.34)$$

As we would expect, the free energy scales with the connectivity ratio w (see equation 2.33), since this is a measure of the number of interactions at each site. It is worth noting at this point that, since f scales with w, in taking the thermodynamic limit to derive the saddle point equations we have assumed $wN \to \infty$. This compares to the simpler case of the infinite-range Ising model (fully connected), where $N \to \infty$, gives us the saddle point equations. Nw is in fact the number of connections per site so we are explicitly seeing here the condition that, for mean field theory to be exact, each site must interact with an infinite number of other sites. This means, for example, that each site could interact with $\sqrt{N}$ other sites and the order parameter equations derived from mean field theory would still be exact. This would give,

$$w = \frac{\sqrt{N}}{N} \to 0 \quad as \quad N \to \infty \qquad (2.35)$$

with $\frac{f}{w}$ remaining finite and so the analytical continuation of w to zero will be valid in our order parameter equations. We could have also chosen the number of connections per site to be $\ln N$ and the limit $w \to 0$ would also be valid. We will refer to the limit $w \to 0$ for simplicity as the $w = 0$ model and in Chapter 3 we will derive the phase diagram of the randomly connected version of this model.

The saddle point equations for the free energy in equation 2.2 give us the following solutions for the order parameters of the replica system,

$$m_\rho^\nu = \ll \xi^\nu < S^\rho > \gg$$

$$\begin{aligned} q^{\rho\sigma} &= \ll< S^{\rho} S^{\sigma} >\gg \\ r^{\rho\sigma} &= \frac{1}{\beta} \mathop{\mathrm{Tr}}_{ij} \left(\frac{\frac{1}{N}\mathbf{D}\mathbf{Q}^{\rho\sigma}}{\left| \mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q} \times \mathbf{D} \right|_{\rho\sigma}} \right) \end{aligned} \tag{2.36}$$

$| \ |_{\rho\sigma}$ is the determinant of the matrix with respect to the ρ and σ indices only and is therefore a matrix itself of dimension N (see Appendix A).
$\mathbf{Q}^{\rho\sigma}$ is the cofactor of $\mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{D} \times \mathbf{q}$ with respect to the indices ρ and σ and is therefore also a matrix of dimension N (see Appendix A).

These represent the order parameter equations for a system of n replicas and they are exact in the thermodynamic limit. At first sight it seems as if we have failed in our objective to remove spin sums from the calculation. The above equations could be used as a starting point to find solutions of the so-called replica symmetry broken phases of the model which has been done for the spin glass by Parisi and others (see [29] for a review of different replica symmetry breaking schemes). The replica symmetry broken phases are the areas on the phase diagram where the replica-symmetric solutions are unstable. In this work only the replica-symmetric equations will be solved and the replica equations will be used to study their stability and so determine the areas of broken replica symmetry on the phase diagram (see next section). While the spin glass phase is very important in the study of spin glasses, it is the storage part of the phase diagram (m_{ρ}^{ν} finite) that we are interested in for neural networks. Spin glass phases in neural networks and spin glasses are always replica symmetry broken phases but the storage part of the phase diagram only has a certain area that has broken symmetry. In the case of a fully connected network [6] this area is very small but, as we shall show later, this area does increase in size as the network becomes more partially connected. Therefore for spin glasses it is very important to study the replica equations to determine the nature of the spin glass phase but for neural networks the main features of storage capacity can be determined within replica-symmetric theory. In the paramagnetic part of the phase diagram all the order parameters are zero and the full replica stability conditions can be studied. This will be done in section 3.6 to show the consistency of replica symmetric theory and replica theory in determining the spin glass phase boundary for the randomly connected, $w = 0$ model.

2.3 The Replica-symmetric Equations

We will now derive the replica-symmetric form of the free energy and order parameter equations. This is done by making the assumption that all the order parameters are independent of their replica indices and then taking the $n \rightarrow 0$ limit on the free energy per site and the order parameter equations. Therefore we first set,

$$\begin{aligned} m_\rho^\nu &= m^\nu \\ q^{\rho\sigma} &= q \quad \rho \neq \sigma \\ r^{\rho\sigma} &= r \quad \rho \neq \sigma \end{aligned} \tag{2.37}$$

and then take the limit,

$$f_{replica\ symmetric} = \lim_{n \to 0} f_{replica} \tag{2.38}$$

on the free energy per site. These two steps correspond to firstly assuming that the minima in $f(m_\rho^\nu, q^{\rho\sigma}, r^{\rho\sigma})$ lie along the replica-symmetric direction and secondly, reducing the problem to only one system from n systems. We will look at the free energy per site first (see equation 2.2), and discuss the effect of the replica symmetry assumption on it, term by term. The first term remains unchanged but the second term is,

$$\frac{w\alpha}{2\beta n} \operatorname*{Tr}_{ij\rho\sigma} \ln\left(\mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q} \times \mathbf{D}\right) \tag{2.39}$$

where $\mathbf{q}$ is now a matrix with 1's on the diagonal and q for all the off diagonal terms. This can be written as, ignoring the constant term in front,

$$\frac{1}{n} \operatorname*{Tr}_{ij\rho\sigma} \ln\left(\mathbf{I}_{Nn} - \frac{\beta}{wN}(1-q)\mathbf{I}_n \times \mathbf{D} - \frac{\beta q}{wN}\mathbf{1}_n \times \mathbf{D}\right) \tag{2.40}$$

$$\operatorname*{Tr}_{ij} \ln\left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right) + \frac{1}{n} \operatorname*{Tr}_{ij\rho\sigma} \ln\left(\mathbf{I}_{Nn} - \frac{\beta q}{wN}\mathbf{1}_n \times \mathbf{D}\left[\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right]^{-1}\right)$$

$\mathbf{1}_n$ is an n dimensional matrix with all components set to 1.

We can now expand the second term in the above equation in $\mathbf{1}_n$ since we wish to take the limit $n \rightarrow 0$. This gives,

$$\operatorname*{Tr}_{ij} \ln\left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right) - \frac{1}{n}\sum_{k=1}^{\infty} \operatorname*{Tr}_{\rho\sigma} \frac{(\mathbf{1}_n)^k}{k!} \operatorname*{Tr}_{ij} \left(\frac{\beta q}{wN}\mathbf{D}\left[\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right]^{-1}\right)^k \tag{2.41}$$

In the limit $n \rightarrow 0$ only the first term in the sum contributes since,

$$\operatorname*{Tr}_{\rho\sigma}(\mathbf{1}_n)^k = n^k \tag{2.42}$$

Therefore we get, for the second term in the free energy equation 2.2, in the replica-symmetric limit,

$$\operatorname*{Tr}_{ij} \ln \left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q) \right) - \operatorname*{Tr}_{ij} \left(\frac{\beta q}{wN}\mathbf{D} \left[\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q) \right]^{-1} \right) \tag{2.43}$$

The remaining terms are the same as for a fully connected network [6] and we therefore have, for the third and fourth terms,

$$\frac{1}{n}\sum_{\nu\rho}(m^\nu)^2 + \frac{1}{n}\sum_{\rho\neq\sigma} rq = \sum_{\nu}(m^\nu)^2 + (n-1)rq \tag{2.44}$$

which taking the limit $n \rightarrow 0$ gives,

$$\sum_{\nu}(m^\nu)^2 - rq \tag{2.45}$$

The last term gives, setting $r^{\rho\sigma} = r$,

$$\frac{1}{n} \ll \ln \operatorname*{Tr}_{S^\rho} \exp \left(\frac{\alpha\beta^2}{2} r \left(\sum_\rho S^\rho \right)^2 - \frac{1}{2} n\alpha\beta^2 r + \beta \sum_\nu m^\nu \xi^\nu \sum_\rho S^\rho \right) \gg \tag{2.46}$$

The term $\frac{1}{2}n\alpha\beta^2 r$ comes from the term $\rho = \sigma$ which is excluded from the replica sum and therefore must be subtracted from the replica-symmetric sum. We can now decouple the spins using the single variable form of the Gaussian transformation in a similar way to its use in the mean field theory of the infinite-range Ising model (see section 1.5). We therefore get, using equation 1.31,

$$\frac{1}{n} \ll \ln \frac{dz}{\sqrt{2\pi}} \int \operatorname*{Tr}_{S^\rho} \exp \left(-\frac{1}{2}z^2 + \beta \left[\sqrt{\alpha r} z + \mathbf{m}.\xi \right] \sum_\rho S^\rho \right) \gg -\frac{1}{2}\alpha\beta^2 r \tag{2.47}$$

where $\mathbf{m}$ and ξ are the vectors $\{m^\rho\}$ and $\{\xi^\rho\}$.

The last term containing r is factored by the same constants as the previous term containing r calculated in equation 2.45, which is of the form $-qr$ so this gives us, adding the two terms and ignoring constants, a term of form $r(1-q)$. The spins S^ρ are now decoupled in the main term in equation 2.47 so the trace can be evaluated giving,

$$\frac{1}{n} \ll \ln \int \frac{dz}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}z^2 + n \ln \left[2\cosh\beta(\sqrt{\alpha r} z + \mathbf{m}.\xi) \right] \right) \gg \tag{2.48}$$

Since we are going to take the limit $n \to 0$ we can expand the exponential in n and only keep the leading term giving,

$$\frac{1}{n} \ll \ln \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \left(1 + n \ln\left[2\cosh\beta(\sqrt{\alpha r} z + \mathbf{m}.\xi)\right]\right) \gg \tag{2.49}$$

The first of the two terms in the above integral is just the Gaussian integral and is therefore equal to one. This means that the main log is of the form $\ln(1+ny)$ and so we can expand it in n and only the leading term will contribute when we take the limit $n \to 0$. This leading term is,

$$\ll \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \ln\left[2\cosh\beta(\sqrt{\alpha r} z + \mathbf{m}.\xi)\right] \gg \tag{2.50}$$

Putting all these terms together, we get for the replica-symmetric free energy per site,

$$\begin{aligned} \frac{f}{w} &= \frac{1}{2}\alpha + \frac{1}{2}\sum_{\nu}(m^{\nu})^2 + \frac{\alpha\beta r}{2}(1-q) + \\ & \frac{w\alpha}{2\beta}\left\{ \mathop{\mathrm{Tr}}_{ij} \ln\left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right) - \mathop{\mathrm{Tr}}_{ij}\left[\frac{\beta q}{wN}\mathbf{D}\left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right)^{-1}\right]\right\} \\ & -\frac{1}{\beta}\int \frac{dz}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \ll \ln\left[2\cosh\beta(\sqrt{\alpha r} z + \mathbf{m}.\xi)\right] \gg \end{aligned} \tag{2.51}$$

So now we have the free energy in a form which does not contain spin sums and we can also derive order parameter equations which do not depend on spin sums. The order parameter equations are derived from the finite set of saddle point equations,

$$\frac{\partial f}{\partial m^{\nu}} = \frac{\partial f}{\partial q} = \frac{\partial f}{\partial r} = 0 \tag{2.52}$$

Which gives,

$$\begin{aligned} m^{\nu} &= \ll \int \frac{dz}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \xi^{\nu} \tanh\beta\left(\sqrt{\alpha r} z + \mathbf{m}.\xi\right) \gg \\ q &= \ll \int \frac{dz}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tanh^2\beta\left(\sqrt{\alpha r} z + \mathbf{m}.\xi\right) \gg \\ r &= \mathop{\mathrm{Tr}}_{ij}\left[\frac{q}{w}\left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right)^{-2}\left(\frac{\mathbf{D}}{N}\right)^2\right] \end{aligned} \tag{2.53}$$

These equations can also be derived from the order parameter equations 2.36 with the replica-symmetric assumptions. This is done in the case of order parameter r in Appendix B. The equations for m^{ν} and q are exactly of the same form as for a fully connected network [6], while only r explicitly contains the matrix $\mathbf{D}$ which specifies the connection architecture. The external parameters of the system α and T are defined slightly differently from the fully connected system as they are factored by $\frac{1}{w}$ (see equation 2.28). Simultaneous solutions of these three equations will yield the phase diagram for the network. The physical meaning of these order parameters is then given by applying the replica assumption to equations 2.31 which gives,

$$\begin{aligned} m^\nu &= \frac{1}{N}\sum_{i=1}^{N} \ll \xi_i^\nu < S_i > \gg \\ q &= \frac{1}{N}\sum_{i=1}^{N} \ll < S_i >^2 \gg \\ r &= \frac{1}{\alpha}\sum_{\mu=s+1}^{p} \ll (m^\mu)^2 \gg \end{aligned} \qquad (2.54)$$

Therefore m^ν is the overlap of the state of the system with the nominated patterns. It was the same parameter we studied in section 1.2 using a simple statistical analysis technique. q is the Edwards Anderson [26] order parameter which is a measure of the freezing of the system at low temperature (see section 1.6). The other parameter r is a measure of the overlap with the infinite set of $p - s$ patterns, which are not nominated for condensation. Since we will only be looking at macroscopic overlaps with one pattern r corresponds to the sum of the squared overlaps with all the other patterns divided by α.

At this stage it is convenient to express r in a different form. We first define,

$$C = \beta(1-q) \qquad (2.55)$$

In numerical and theoretical calculations C was always found to be less than one. We can therefore expand r in C which gives,

$$r = \frac{q}{w}\operatorname*{Tr}_{ij}\left(\frac{\mathbf{D}}{N}\right)^2\left(1 + 2\frac{C\mathbf{D}}{wN} + 3\left(\frac{C\mathbf{D}}{wN}\right)^2 + \cdots\right) \qquad (2.56)$$

This can be more conveniently written as,

$$r = q\sum_{k=0}^{\infty} C^k(k+1)a_k(w) \qquad (2.57)$$

where,

$$a_k(w) = w\operatorname*{Tr}_{ij}\left(\frac{\mathbf{D}}{wN}\right)^{k+2} \qquad (2.58)$$

The physical significance of $a_k(w)$ in terms of the connection space of a network will be discussed in section 3.1 where we will also evaluate it numerically for different architectures, by a bounded random walk. At this point it is worth noting that $\mathbf{D}^{-1}$ has dropped out of our calculations with only traces of $\mathbf{D}$ being left in. In section 2.2 we found that in some cases it was necessary to choose $D_{ii} = a$, (a real and finite) rather than $D_{ii} = 1$, to remove the singularity of $\mathbf{D}$. At this stage in our calculations we can also see that a choice of $D_{ii} = a$ will not effect the results of the traces. In the case of the fully connected network the traces of $\mathbf{D}$ enter in the form,

$$\operatorname*{Tr}_{ij} \left(\frac{\mathbf{D}}{N} \right)^n = \frac{N(a^n + \cdots) + N^n}{N^n} \tag{2.59}$$

so the terms involving a are negligible for a finite in the thermodynamic limit.

In this thesis we shall only be looking at states which have a macroscopic overlap with one of the nominated states. This means that in the preceding equations $s = 1$ and the finite sum $\sum_{\rho=1}^{s}$ corresponds to only one term. States which have a macroscopic overlap with more than one nominated state will exist at low temperature. The details of the mixture states in the fully connected network are given in [6] where they only exist at low values of α and therefore do not play an important role in defining the maximum storage capacity of the network. Only solutions which have an overlap with an odd number of nominated states are stable and all the possible permutations of the bits of the nominated states give stable states. It is therefore difficult to justify these states as contributing to the storage capacity of the network. It is only the storage states with a single overlap that exist at high α and therefore determine the storage capacity of the network. We expect that the relationship between mixture states in a fully connected network and those in a partially connected network will be very similar to the relationship between single overlap states in the two models.

We have for retrieval states with a single macroscopic overlap,

$$m^{\nu} = m\delta_{\nu\eta} \tag{2.60}$$

This gives for the order parameter equation for m,

$$m = \ll \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2} \right) \xi \tanh \beta(\sqrt{\alpha r}\, z + m\xi) \gg \tag{2.61}$$

where we have dropped the index on ξ since only one component of $\{\xi^\nu\}$ now appears on the left hand side of equation 2.53. The quenched averaging can now be easily performed since ξ only takes two values, 1 or -1 so we only have two terms to average over. We get for the two terms,

$$m = \frac{1}{2}\int \frac{dz}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\left[\tanh\beta(\sqrt{\alpha r}z + m) - \tanh\beta(\sqrt{\alpha r}z - m)\right] \quad (2.62)$$

The transformation $z \rightarrow -z$ on the second term makes it positive and of the same value as the first term. A similar process can be carried out for q and we finally get the set of replica symmetric order parameter equations for a state with a single macroscopic overlap with one of the nominated states,

$$\begin{aligned} m &= \int \frac{dz}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\tanh\beta(\sqrt{\alpha r}z + m) \\ q &= \int \frac{dz}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\tanh^2\beta(\sqrt{\alpha r}z + m) \\ r &= q\sum_{k=0}^{\infty} C^k(k+1)a_k(w) \\ a_k(w) &= w\,\mathop{\mathrm{Tr}}_{ij}\left(\frac{\mathbf{D}}{wN}\right)^{k+2} \end{aligned} \quad (2.63)$$

and the free energy per site is,

$$\begin{aligned} \frac{f}{w} &= \frac{1}{2}\alpha + \frac{1}{2}m^2 + \frac{C\alpha r}{2} \\ &+\frac{w\alpha}{2\beta}\left\{\mathop{\mathrm{Tr}}_{ij}\ln\left(\mathbf{I}_N - \frac{C}{wN}\mathbf{D}\right) - \mathop{\mathrm{Tr}}_{ij}\left[\frac{\beta q}{wN}\mathbf{D}\left(\mathbf{I}_N - \frac{C}{wN}\mathbf{D}\right)^{-1}\right]\right\} \\ &-\frac{1}{\beta}\int\frac{dz}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right)\ln\left[2\cosh\beta(\sqrt{\alpha r}z + m)\right] \end{aligned} \quad (2.64)$$

which, expanding in C, can also be written in terms of $a_k(w)$ as,

$$\begin{aligned} \frac{f}{w} &= \frac{1}{2}m^2 + \frac{C\alpha r}{2} - \frac{\alpha}{2}\sum_{k=0}^{\infty} C^{k+1}a_k(w)\left(\frac{1+q(1+k)}{k+2}\right) \\ &-\frac{1}{\beta}\int\frac{dz}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right)\ln\left[2\cosh\beta(\sqrt{\alpha r}z + m)\right] \end{aligned} \quad (2.65)$$

It should be noted at this point that since we have never explicitly chosen with which state the system will have a macroscopic overlap, it can be any of the p states and there will be a basin of attraction at low temperature associated with each of these states. In all these calculations we are still allowing for the possibility that the macroscopic overlap m is zero. We thus have the possibility of spin glass states ($m = 0, q$ finite) and paramagnetic states ($m = 0, q = 0$) being solutions of these equations.

Although we have now reduced the problem to the simultaneous solution of three equations we have not looked at the stability of the solutions and the validity of the replica-symmetric assumptions. This will be the subject of the next section in this chapter.

2.4 Stability of Replica-symmetric Solution

The equation for the free energy of the n replica system equation 2.2, is exact in the thermodynamic limit. It is only when we perform the replica-symmetric continuation of $n \to 0$ that the expression we gain for the free energy may be invalid. The values of the order parameters which are then calculated by the saddle point method may then be invalid as well. We also have to check that the solutions we obtain are minima of the free energy rather than just saddle points. Both these stability conditions can be checked by expanding the replica order parameters about their replica-symmetric values and looking at the eigenvalues of the matrix of second derivatives of the free energy. This will tell us the regions of the phase diagram where our replica-symmetric solutions are stable. In certain areas of the phase diagram we can already see that replica theory gives strange results. At very low temperatures replica theory gives negative values for the entropy which is not possible for a system of discrete Ising spins. Another strange result is that the free energy is always a maximum rather than a minimum of q in the spin glass phase. This arises from the factor $n(n-1)/2$ which is the number of distinct $q^{\rho\sigma}$'s in replica theory occurring in the free energy. When we take the limit $n \to 0$ we get the strange result that this term becomes negative. This means that under the replica assumptions we end up with, in some sense, a negative number of order parameters. These strange and inconsistent results can only be understood by studying the stability of the replica-symmetric solutions in terms of the replica theory. We will therefore start by expanding the replica solutions about the replica-symmetric values,

$$m_\rho^\nu = m^\nu + \delta m_\rho^\nu$$

$$\begin{aligned} q^{\rho\sigma} &= q + \delta q^{\rho\sigma} \\ r^{\rho\sigma} &= r + \delta r^{\rho\sigma} \end{aligned} \tag{2.66}$$

and then look at the second derivatives of the free energy. In the analysis that follows we will use a similar approach to Lautrup [37] to classify the different eigenvalues.

We will try as much as possible to write the second derivatives of the free energy in terms of order parameters and closely related quantities. This will make it much easier to take the replica-symmetric limit on the second derivatives since we already know the replica-symmetric forms of the order parameters. We therefore start by defining $\mathbf{r}^{\rho\sigma}$ as the matrix from which the trace over the site indices gives $r^{\rho\sigma}$. We therefore have from equation 2.36,

$$\mathbf{r}^{\rho\sigma} = \frac{1}{\beta}\left(\frac{\frac{1}{N}\mathbf{D}\mathbf{Q}^{\rho\sigma}}{\left|\mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q}\times\mathbf{D}\right|_{\rho\sigma}}\right) \quad \forall\rho,\sigma \tag{2.67}$$

where,

$$r^{\rho\sigma} = \operatorname*{Tr}_{ij}\mathbf{r}^{\rho\sigma} \tag{2.68}$$

We have defined $\mathbf{r}^{\rho\sigma}$ for all values of ρ and σ including $\rho = \sigma$ which is valid since we have previously defined $q^{\rho\rho} = 1$, so $\mathbf{Q}^{\rho\rho}$ is defined. $\mathbf{Q}^{\rho\sigma}$ is the cofactor, with respect to the replica indices, of the matrix whose determinant occurs in $\mathbf{r}^{\rho\sigma}$. With this definition of $\mathbf{r}^{\rho\sigma}$ we obtain for the second derivatives of the replica free energy, equation 2.2, with the common factors w and $\frac{1}{n}$ taken out,

$$\begin{aligned}
\frac{n}{w}\frac{\partial^2 f}{\partial m_\rho^\nu \partial m_\sigma^\mu} &= \delta_{\nu\mu}\delta_{\rho\sigma} - \beta \ll \xi^\nu\xi^\mu(< S^\rho S^\sigma > - < S^\rho >< S^\sigma >) \gg \\
\frac{n}{w}\frac{\partial^2 f}{\partial m_\rho^\nu \partial q^{\gamma\lambda}} &= 0 \quad (\gamma < \lambda) \\
\frac{n}{w}\frac{\partial^2 f}{\partial m_\rho^\nu \partial r^{\gamma\lambda}} &= -\beta^2 \ll \xi^\nu(S^\rho S^\gamma S^\lambda > - < S^\rho >< S^\gamma S^\lambda >) \gg \quad (\gamma < \lambda) \\
\frac{n}{w}\frac{\partial^2 f}{\partial q^{\rho\sigma} \partial q^{\gamma\lambda}} &= -\frac{\beta^3\alpha}{w}\operatorname*{Tr}_{ij}(\mathbf{r}^{\rho\gamma}\mathbf{r}^{\lambda\sigma} + \mathbf{r}^{\rho\lambda}\mathbf{r}^{\gamma\sigma}) \quad (\rho < \sigma, \gamma < \lambda) \\
\frac{n}{w}\frac{\partial^2 f}{\partial q^{\rho\sigma} \partial r^{\gamma\lambda}} &= \beta\alpha\delta_{\rho\gamma}\delta_{\sigma\lambda} \quad (\rho < \sigma, \gamma < \lambda) \\
\frac{n}{w}\frac{\partial^2 f}{\partial r^{\rho\sigma} \partial r^{\gamma\lambda}} &= -\beta^3\alpha^2 \ll (< S^\rho S^\sigma S^\gamma S^\lambda > - < S^\rho S^\sigma >< S^\gamma S^\lambda >) \gg (2.69) \\
& \quad (\rho < \sigma, \gamma < \lambda)
\end{aligned}$$

We now wish to calculate all these derivatives under the replica-symmetric assumptions. We will only be looking at the states which have a macroscopic overlap with one of the nominated patterns. We therefore already know what the double and single spin averages which occur in these derivatives are in replica-symmetric theory since they are just equal to the order parameters m and q equation 2.54,

$$\begin{aligned} \ll \xi^{\nu} < S^{\rho} > \gg &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh\beta(\sqrt{\alpha r}z + m) \\ \ll < S^{\rho} S^{\sigma} > \gg &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2\beta(\sqrt{\alpha r}z + m) \quad \rho \neq \sigma \end{aligned} \tag{2.70}$$

The other spin averages can be calculated from the partition function constructed from the free energy in equation 2.2, with the replica assumptions. We will now introduce a more compact notation for defining these spin averages. We define,

$$M = \tanh\beta(\sqrt{\alpha r}z + m) \tag{2.71}$$

and,

$$<>_z = \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{2.72}$$

This gives us for all the spin averages required to evaluate the second derivatives of the free energy,

$$\begin{aligned} \ll \xi^{\nu} < S^{\rho} > \gg &= < M >_z \\ \ll < S^{\rho} S^{\sigma} > \gg &= < M^2 >_z + \delta_{\rho\sigma} < 1 - M^2 >_z \\ \ll \xi^{\nu} < S^{\rho} S^{\sigma} S^{\gamma} > \gg &= < M^3 >_z + \delta_{\rho\sigma}\delta_{\rho\gamma} < M(1 - M^2) >_z \quad (\sigma < \gamma) \\ \ll < S^{\rho} S^{\sigma} S^{\gamma} S^{\lambda} > \gg &= < M^4 >_z + \delta_{\rho\gamma}\delta_{\rho\lambda}\delta_{\sigma\gamma}\delta_{\sigma\lambda} < M^2(1 - M^2) >_z \\ &\quad + \delta_{\rho\gamma}\delta_{\sigma\lambda} < (1 - M^2)^2 >_z \quad (\rho < \sigma, \gamma < \lambda) \end{aligned} \tag{2.73}$$

Following Lautrup's [37] approach, instead of diagonalising the matrix of second derivatives to find the stability eigenvalues, we will evaluate the full second order fluctuation $\partial^2 f$ and characterise the eigenvalues into three specific groups. With the replica assumption that all the order parameters are independent of their replica indices $\mathbf{r}^{\rho\sigma}$, equation 2.66 will only take two forms depending on whether it is a diagonal or off-diagonal term. In the second stage of the replica-symmetric assumption $n \longrightarrow 0$ we will get two distinct matrices: the off-diagonal matrix $\mathbf{r}$, whose trace gives us r and the diagonal term matrix which we will denote as $\mathbf{r}_d$.

Both these matrices are calculated in Appendix B. We therefore have for the second order fluctuation in the free energy f,

$$\begin{aligned}
\frac{n}{w}\delta^2 f \;=\;& (1-\beta < 1 - M^2 >_z)\sum_\rho (\delta m_\rho)^2 \\
& -\beta(< M^2 >_z - < M >_z^2)\left(\sum_\rho \delta m_\rho\right)^2 \\
& -\alpha\beta^2(< M^3 >_z - < M >_z < M^2 >_z)\sum_\rho \delta m_\rho \sum_{\rho\sigma} \delta r^{\rho\sigma} \\
& -2\alpha\beta^2 < M(1-M^2) >_z \sum_{\rho\sigma} \delta m_\rho \delta r^{\rho\sigma} \\
& -\frac{\alpha\beta^3}{2w}\operatorname*{Tr}_{ij}\left[\mathbf{r}^2\left(\sum_{\rho\sigma}\delta q^{\rho\sigma}\right)^2 + 2\mathbf{r}(\mathbf{r}-\mathbf{r}_d)\sum_\rho\left(\sum_\sigma \delta q^{\rho\sigma}\right)^2 + (\mathbf{r}-\mathbf{r}_d)^2\sum_{\rho\sigma}(\delta q^{\rho\sigma})^2\right] \\
& +\alpha\beta\sum_{\rho\sigma}\delta q^{\rho\sigma}\delta r^{\rho\sigma} \\
& \frac{\alpha^2\beta^3}{4}(< M^4 >_z - < M^2 >_z^2)\left(\sum_{\rho\sigma}\delta r^{\rho\sigma}\right)^2 \\
& -\alpha^2\beta^3 < M^2(1-M^2) >_z \sum_\rho\left(\sum_\sigma \delta r^{\rho\sigma}\right)^2 \\
& -\frac{\alpha^2\beta^3}{2} < (1-M^2)^2 >_z \sum_{\rho\sigma}(\delta r^{\rho\sigma})^2
\end{aligned} \qquad (2.74)$$

The fluctuations $\delta q^{\rho\rho}$ and $\delta r^{\rho\rho}$ are defined as being zero since these terms do not correspond to replica order parameters, so the sums over ρ and σ are unrestricted. In the previous equations we have only set diagonal order parameters like $q^{\rho\rho}$ equal to certain values to make our expressions more compact.

We will now consider three classes of fluctuations which span the complete space of fluctuations which has dimension n^2. The most important one is what we will call strongly asymmetric fluctuations which are defined to satisfy,

$$\delta m_\rho = \sum_\sigma \delta q^{\rho\sigma} = \sum_\sigma \delta r^{\rho\sigma} = 0 \qquad (2.75)$$

and span a subspace of n(n-3) dimensions. Putting these expressions back into the equation for the fluctuation in the free energy equation 2.73 we obtain the strongly asymmetric fluctuation in the free energy,

$$\frac{n}{w}\delta^2 f = \sum_{\rho\sigma}\left(-\frac{\alpha\beta^3}{2w}\operatorname*{Tr}_{ij}(\mathbf{r}-\mathbf{r}_d)^2(\delta q^{\rho\sigma})^2 + \alpha\beta\delta q^{\rho\sigma}\delta r^{\rho\sigma} - \frac{\alpha^2\beta^3}{2} < (1-M^2)^2 >_z (\delta r^{\rho\sigma})^2\right) \qquad (2.76)$$

This is a quadratic in the fluctuations in $q^{\rho\sigma}$ and $r^{\rho\sigma}$ which seems to be negative-definite, implying that the replica solutions are always unstable but we must remember that the fluctuations in $r^{\rho\sigma}$ occur along a contour running in the direction of the imaginary axis. We can shift this contour in order to make it run through the saddle points. If we rewrite the above expression in the form,

$$\frac{n}{w}\delta^2 f = \sum_{\rho\sigma}\left[\frac{1}{2\beta}\lambda_S(\delta q^{\rho\sigma})^2 - \frac{\beta^3\alpha^2}{2}\lambda_r\left(\delta r^{\rho\sigma} - \frac{\delta q^{\rho\sigma}}{\beta^2\alpha\lambda_r}\right)^2\right] \tag{2.77}$$

where,

$$\begin{aligned} \lambda_r &= < (1-M^2)^2 >_z \\ \lambda_S &= \frac{1}{\lambda_r} - \frac{\alpha\beta^4}{w}\underset{ij}{\mathrm{Tr}}(\mathbf{r}-\mathbf{r}_d)^2 \end{aligned} \tag{2.78}$$

we have isolated the shift necessary in $\delta r^{\rho\sigma}$ which is,

$$\delta r^{\rho\sigma} = \frac{\delta q^{\rho\sigma}}{\beta^2\alpha\lambda_r} + i\delta s^{\rho\sigma} \tag{2.79}$$

where $s^{\rho\sigma}$ is real. The two eigenvalues λ_S and λ_r which characterize these fluctuations do not depend on n, the number of replicas so there is no problem with the limit $n \rightarrow 0$. Since λ_r is clearly always positive, asymmetric fluctuations in $\delta r^{\rho\sigma}$ are always damped in the imaginary direction and instabilities are entirely controlled by λ_s. This eigenvalue can take both positive and negative values and characterizes the replica and replica-broken phases. The other fluctuations can be split into two more classes which we will call weakly asymmetric and symmetric fluctuations. The weakly asymmetric fluctuations are defined to be of the form,

$$\begin{aligned} \delta q^{\rho\sigma} &= \delta q^{\rho} + \delta q^{\sigma} \quad \rho \neq \sigma \\ \delta r^{\rho\sigma} &= \delta r^{\rho} + \delta r^{\sigma} \quad \rho \neq \sigma \end{aligned} \tag{2.80}$$

where,

$$\sum_{\rho}\delta m_{\rho} = \sum_{\rho}\delta q^{\rho} = \sum_{\rho}\delta r^{\rho} = 0 \tag{2.81}$$

and the symmetric fluctuations are defined as,

$$\begin{aligned} \delta m_{\rho} &= \delta m \\ \delta q^{\rho\sigma} &= \delta q \quad \rho \neq \sigma \\ \delta r^{\rho\sigma} &= \delta r \quad \rho \neq \sigma \end{aligned} \tag{2.82}$$

These two types of fluctuation only introduce two more eigenvalues which are similar to the fully connected model [37] and are always larger than λ_S in the ordered phases so we will not consider them any further. The condition for stability of the replica-symmetric solution is therefore given by,

$$\lambda_S \geq 0 \tag{2.83}$$

with equality giving the lines on the phase diagram separating the symmetric phases from the symmetry broken phases. The condition for stability can be written as, from equation 2.77,

$$\frac{\beta^4 \alpha}{w} \underset{ij}{\mathrm{Tr}} (\mathbf{r} - \mathbf{r}_d)^2 \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sech}^4 \beta(\sqrt{\alpha r} z + m) \leq 1 \tag{2.84}$$

In Appendix B we find that this trace gives,

$$\underset{ij}{\mathrm{Tr}} (\mathbf{r} - \mathbf{r}_d)^2 = \frac{wr}{q\beta^2} \tag{2.85}$$

therefore the stability condition becomes,

$$\frac{\beta^2 \alpha r}{q} \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sech}^4 \beta(\sqrt{\alpha r} z + m) \leq 1 \tag{2.86}$$

Therefore to determine the areas of broken replica symmetry we must solve the order parameter equations 2.62 and plug the values obtained for the order parameters into the above equation. These calculations will be performed for different connection architectures in Chapter 3 but we will discuss some of the basic results in this section.

The eigenvalue λ_S is always negative at temperatures close to zero, because of the β^2 term (see equation 2.85), which explains why the replica-symmetric solution gives an invalid result of negative entropy at these temperatures. The spin glass phase ($m = 0$, q finite), always has replica symmetry broken while the memory phase (m finite, q finite), is split into two areas: one at low temperature with replica symmetry broken and the other at higher temperatures being replica-symmetric. The line which separates these two areas of the memory phase is called the Almeida-Thouless line [38] after the two physicists who first calculated it for the ferromagnetic phase of the Sherrington-Kirkpatrick infinite-range spin glass [28]. Its precise position varies with the connection architecture of the network. As we shall see later the memory phase actually co-exists with part of the spin glass phase. The paramagnetic phase ($m = q = r = 0$) is replica symmetric as it corresponds to the phase ($m_\rho = q^{\rho\sigma} = r^{\rho\sigma} = 0$) in the replica model which is trivially symmetric in the replica indices. The symmetry broken spin glass phase has been studied in great detail for the SK spin glass with many schemes being proposed to break the symmetry [39,40,41] but it is now generally accepted that Parisi's results [42,43,44,45] are the most satisfactory. We expect the spin glass phase in partially connected neural networks to be very similar in character to that of the SK spin glass. Amit *et al* [6] have shown this to be the case for the fully connected Hopfield network. The broken replica memory phase will also have similar properties to the spin glass phase so it is worth discussing some of the features of the spin glass phase in more detail.

To begin to understand what is happening in the spin glass phase we must go back to our consideration of frustration in section 1.6. In that section we saw how frustration in spin glasses and the Hopfield model leads to a very complicated energy surface with a high degeneracy of minima at different energy values. In the thermodynamic limit all these minima will have infinite energy barriers between them so they are truly stable states and the system remains in the basin of attraction of the state it starts in. This is the type of system we are studying with mean field theory and it is only in finite systems that we will have metastable states. At low temperatures all these minima will probably have similar values of the Edwards-Anderson order parameter q, but for the replica system $q^{\rho\sigma}$ will have to take a range of values to cope with all the overlaps between all the different spin glass states. The order parameter q does not bring out all the detail of the many stable states and it seems essential to the understanding of the spin glass phase that we work with replicas of the system. The replica-symmetric theory is only correct if there is one spin glass state with a specific q value. In this case all the replicas will sit in this same state and they will then be symmetric. This is clearly not the case and it seems that replica mean field theory cannot cope with a system having this type of degeneracy of states. The limit $n \to 0$ does not produce a unique solution if the different replicas can sit in different spin glass states. This is why the eigenvalue which controls replica fluctuations is always negative in the spin glass phase. In fact a continuous range of $q^{\rho\sigma}$ values is required to describe the spin glass state [42,43,44,45]. In the case of the memory states there is a basket of minima associated with each of the nominated states all with the same m value but with the errors at different sites [16]. At temperatures close to zero this gives a basket of minima in the free energy, so replica symmetry is broken. At moderate temperatures, unlike the spin glass states, these minima merge together giving a single minimum in the free energy and so the system is replica-symmetric. Hence the memory phase is broken up into two regions by the Almeida-Thouless line, the higher temperature phase being stable to replica symmetry breaking.

Replica calculations for spin glasses and neural networks have shown that in the replica broken phases near the symmetry breaking line, replica theory still gives good results [6,44,45]. There is therefore a kind of continuous divergence from the replica solutions as we move into the replica symmetry broken phases. This means that in the case of $q^{\rho\sigma}$ as we move into the spin glass phase from the replica-symmetric paramagnetic phase it takes a very small range of values close to the value predicted by replica theory. This is also the case as we move across the replica symmetry breaking line in the memory phase. The maximum value of the storage capacity, denoted by α_c, happens at zero temperature which is in the replica symmetry broken phase. Numerical simulations by Amit *et al* [6] for the fully connected model gave $\alpha_c = 0.145 \pm 0.01$ (see also Chapter 4 section 4.3 of this thesis). Calculations by Crisanti *et al* [11] with replica symmetry broken once gave $\alpha_c = 0.144$. These results are both very close to the theoretical result $\alpha_c = 0.138$, [6] predicted from replica-symmetric mean field theory. Therefore the replica-symmetric theory seems to give a result which is very close to the actual result. This is to be expected since the replica symmetry broken area in the memory phase is very small, and so the point $(\alpha_c = 0.138, T = 0)$ is very close to the Almeida-Thouless line. The results [6,11] also show that the replica-symmetric result always underestimates the maximum storage capacity as well as the accuracy of storage m. As we shall see in the next chapter, replica symmetry breaking plays an increasingly more important role the lower the connectivity of the network is. We shall also see in section 3.6 the equivalence of the spin glass phase boundary predicted by replica and replica-symmetric theory. It is interesting to note that the basic Hopfield model described in section 1.2 is a zero temperature model and can only be truly described by a very complex replica symmetry broken solution. This has never as yet been calculated.

Chapter 3

Solutions of the Replica-symmetric Order Parameter Equations

3.1 Calculation of $a_k(w)$

In this chapter we will solve the order parameter equations for a single condensed pattern,

$$\begin{aligned}
m &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh\beta(\sqrt{\alpha r}z + m) \\
q &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2\beta(\sqrt{\alpha r}z + m) \\
r &= q\sum_{k=0}^{\infty} C^k(k+1)a_k(w) \\
a_k(w) &= w \operatorname*{Tr}_{ij} \left(\frac{\mathbf{D}}{wN}\right)^{k+2}
\end{aligned} \tag{3.1}$$

for different connection architectures in order to derive the phase diagrams and maximum storage capacity for each architecture. The first thing we require before we can solve these equations is a method of evaluating $a_k(w)$ for different values of k and different connection architectures. The number of $a_k(w)$'s which have to be evaluated to calculate r accurately will depend on the value of C as well as the value of $a_k(w)$ itself. As we have already discussed in section 2.2, the network can be thought of as a hypercubic lattice of sites with each neuron connected to an infinite number of neurons in its neighbourhood. This neighbourhood defines what we have called the connection space of that neuron which is the same for every neuron. In this chapter we will only consider connection architectures that have hypercubic connection spaces in dimensions $d = 1, 2, 3, 4$ and 8. The results for a fully connected network, $w = 1$, will also be presented for comparison. Randomly connected models will also be considered and these will be referred to as $d = \infty$ models since a hypercubic connection space of infinite dimensions is equivalent to random connectivity. Even though the order parameter equations we have derived are for the same connection space at each site, in the case of the randomly connected model we expect the results to be the same as those obtained by site independent random dilution of a fully connected network. This is because in the thermodynamic limit the fluctuations from site to site in the connection architecture of the randomly diluted model will average out. Sompolinsky [8] studied the randomly diluted model at zero temperature and obtained the same order parameter equations as we obtain (see next section). He also showed the equivalence of a randomly diluted network and a fully connected network with Gaussian synaptic noise.

We will now look at the form of $a_k(w)$ in more detail which from equation 3.1 is,

$$a_k(w) = N^{-(k+2)} w^{-(k+1)} \underbrace{\sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_{k+2}=1}^{N} D_{i_1 i_2} D_{i_2 i_3} \ldots D_{i_{k+2} i_1}}_{S} \tag{3.2}$$

The sum S contains N^{k+2} terms each of which can take the value one or zero. A term has value one if a neuron i, is connected back to itself through neurons i_2 to i_{k+2}. The sum S contains all possible ways of choosing this loop therefore, $\frac{S}{N^{k+2}}$ is the probability that $k+2$ neurons chosen at random are connected together in a single loop. As we shall see in the next section, the less likely a loop is complete the lower the value of $a_k(w)$ and the correspondingly lower the value of α_c, the maximum storage capacity. We are thus explicitly seeing the loops of correlations which distinguish a neural network from a spin glass, entering into our calculations (see section 1.4).

The first term in the sequence, $a_0(w)$ is easy to calculate since the connections are symmetric. For any connection architecture we have,

$$\underset{ij}{\mathrm{Tr}}\,\mathbf{D}^2 = wN^2 \quad \forall w \tag{3.3}$$

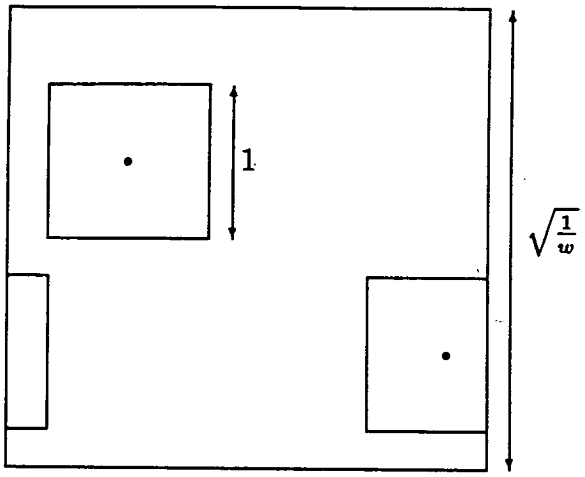

Figure 3.1: Square connection spaces of two neurons in a network with a two dimensional connectivity architecture.

which gives $a_0(w) = 1$. The next term $a_1(w)$ can also be calculated analytically for hypercubic connection architectures giving,

$$a_1(w) = \begin{cases} \left(\frac{3}{4}\right)^n & w < \left(\frac{2}{3}\right)^n \\ \left(\frac{3}{4} + \frac{1}{4}\left(3 - \frac{2}{\sqrt[n]{w}}\right)^2\right)^n & w > \left(\frac{2}{3}\right)^n \end{cases} \tag{3.4}$$

where n is the dimensionality of the hypercube. The details of this calculation of $a_1(w)$ is given in Appendix C. Beyond $k = 1$ it is extremely difficult to calculate $a_k(w)$ analytically and so we have to resort to a numerical method. This numerical method works for any type of connection space including hypercubic but for simplicity we will illustrate it for a two dimensional square connection space.

We can use a square lattice of finite size with cyclic boundary conditions to represent the network. Each point on this surface will represent a neuron and so in the limit of a continuous surface the system will be infinite. Consider therefore, a square of side $\sqrt{\frac{1}{w}}$ with cyclic boundary conditions as representing the network. The connection space of each neuron will then be a square of side one centred on each neuron (see figure 3.1). The first step in calculating $a_k(w)$ is to choose a point i_1, which can be any point on the square since all points are equivalent due to the cyclic boundary conditions. The next step is then to randomly choose another point i_2 in the connection space of the first point. We then continue this process for $k+1$ steps until we reach the point i_{k+2}. $a_k(w)$ is then the probability that the final point is in the connection space of the first point. At each step since we only choose a random point in the connection space of the previous point we are introducing a factor $\frac{1}{w}$ into the probability that sites are connected compared to just choosing points at random. This accounts for the factor $w^{-(k+1)}$ in equation 3.2 since we take $k+1$ steps in total to evaluate $a_k(w)$. The calculation of $a_k(w)$ is therefore reduced to the probability that a bounded random walk of $k+1$ steps ends in the connection space of the starting point. This method is much more efficient than simply choosing points at random and seeing if they are connected in a loop. At each step we are using knowledge of the connection space to avoid choosing points outside each other's connection space which are trivially unconnected. If we are carrying out this random walk on a computer the precision of the computer will limit the size of lattice we are working with. The lattice will therefore be finite with each neuron always being separated from its neighbour by an amount of the order of the precision of the computer whatever the shape of the connection space. Since the lattice size is $\frac{1}{w}$ units then if w is of order one, the size of the lattice will be of the order of, the inverse of the precision of the computer used. If w is smaller the lattice size and hence the size of the system will be larger since the neurons are always separated by a fixed amount.

The calculations of the $a_k(w)$'s was carried out on the ICL distributed array processor (DAP). This is a single instruction multiple data stream machine with 4096 bit processors forming a square lattice [1]. The DAP is very well suited to carrying out random walk calculations as 4096 different random walks can be carried out simultaneously. The DAP along with its programming languages are discussed in more detail in Appendix D. Single precision on the DAP gives about eight figure accuracy so the size of lattices we are working with are of the order 10^8. On the DAP about $\frac{2}{3}$ of a million random steps plus the calculation of $a_k(w)$ could be carried out per second. There are two possible sources of error in calculating $a_k(w)$ but both of them were found to be very small. The standard

[1] Some of the calculations were also carried out on the new DAP 510 which has 1024 processors but a clock cycle twice as fast as the 4096 DAP.

deviation of $a_k(w)$ due to random fluctuations was reduced to a negligible level by averaging over about two million random walks for each calculation of $a_k(w)$. The results we want for $a_k(w)$ should actually be for an infinite system but since the system we are working with is so large, of order 10^8 sites, we expect finite size effects to be negligible. For most choices of connection architecture the value of $a_k(w)$ is probably independent of the system size anyway. Comparison of the numerical results for $a_1(w)$ with the theoretical results (see equation 3.2) showed no significant difference.

An important aspect of the behaviour of $a_k(w)$ is that $a_k(w) \rightarrow w$ as k becomes large. This is because as the number of random steps increases the final position becomes less and less correlated with the initial position and in the limit of an infinite number of steps it is totally uncorrelated with the starting position. Since the connection space of the starting point occupies a w'th of the volume of the lattice, a random point has a probability of w of being in the connection space of a given point. Figure 3.2 shows some typical curves of $a_k(w)$ for different connection architectures and connectiveties. As the dimensionality of the connectivity increases and w increases $a_k(w) \rightarrow w$ more quickly as k increases.

In calculating r we always have to truncate the series in $a_k(w)$ at some point which is the main source of error in solving the order parameter equations. The fact that $a_k(w) \rightarrow w$ as k increases can help us to reduce this truncation error. To calculate r we must calculate the sum of the series,

$$\sum_{k=0}^{\infty} C^k(k+1)a_k(w) \tag{3.5}$$

If we calculate a finite number of terms numerically, say n terms, then the truncation error is given by,

$$Truncation\ \ error \ = \sum_{k=n+1}^{\infty} C^k(k+1)a_k(w) \tag{3.6}$$

If we had calculated enough terms so that $a_k(w) \simeq w$ for $k > n$ then the truncation error would be given by,

$$Truncation\ \ error \ \simeq w \sum_{k=n+1}^{\infty} (k+1)C^k \tag{3.7}$$

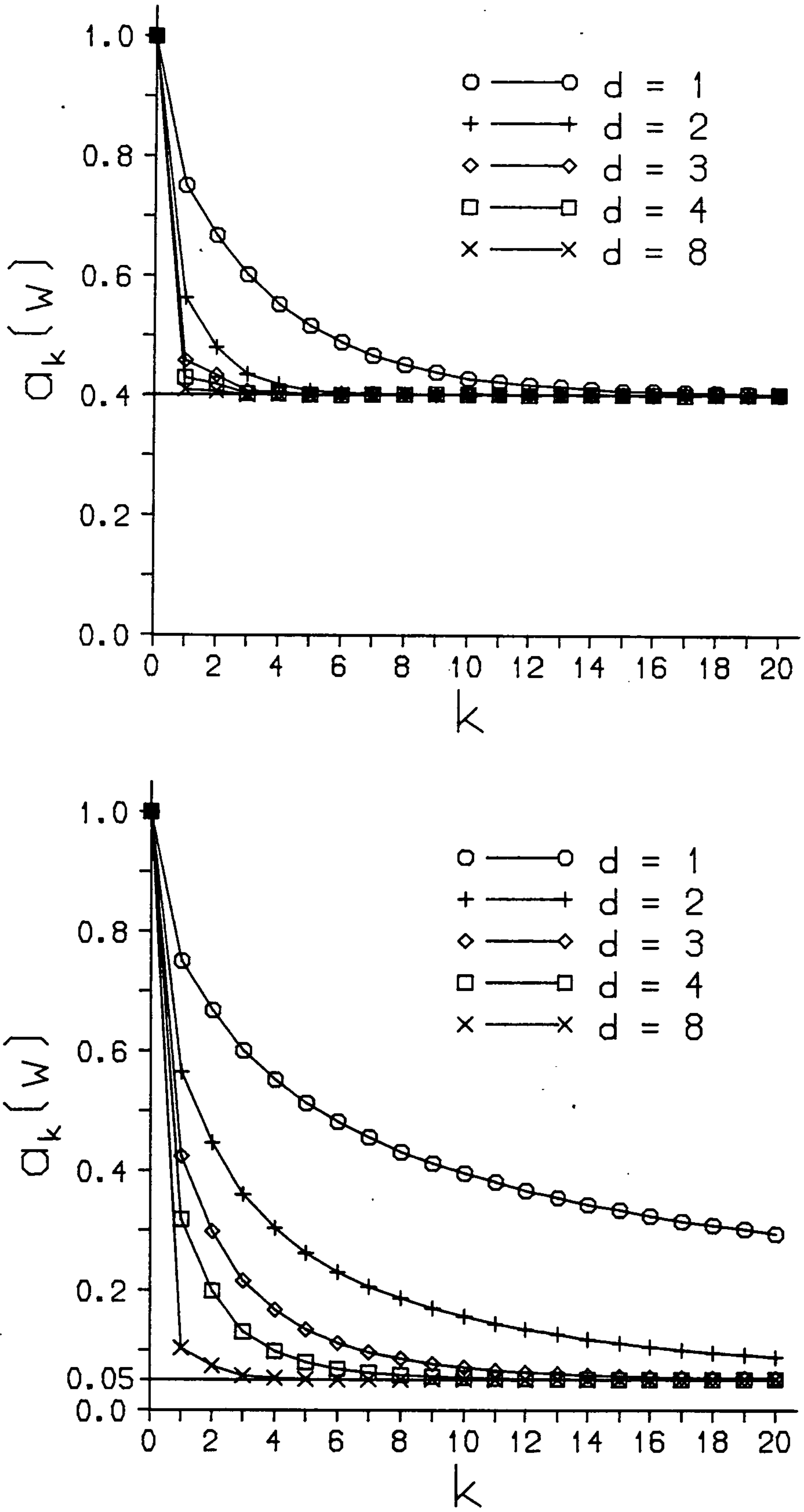

Figure 3.2: $a_k(w)$ **values for hypercubic connection architectures of dimensions** $d = 1, 2, 3, 4$ **and 8 with,** $k = 1, 20$**. Top figure** $w = 0.4$**, bottom figure** $w = 0.05$**.**

Series of this type can be evaluated by noting that,

$$(1+2x+3x^2+\cdots+nx^n)(1-x) = 1+x+x^2+\cdots+x^n-nx^{n+1} = \frac{1-x^{n+1}}{1-x}-nx^{n+1} \tag{3.8}$$

therefore,

$$\sum_{k=0}^{n}(k+1)C^k = \frac{1-C^{n+1}}{(1-C)^2} - \frac{nC^{n+1}}{1-C} \tag{3.9}$$

and the infinite sum is given by,

$$\sum_{k=0}^{\infty}(k+1)C^k = \frac{1}{(1-C)^2} \tag{3.10}$$

The truncation error is obtained by subtracting these two sums giving,

$$Truncation\ \ error \ \simeq \frac{wC^{n+1}(1+n(1-C))}{(1-C)^2} \tag{3.11}$$

for large n. This expression for the truncation error is very useful in evaluating r in situations where C is large and $a_k(w)$ tends rapidly to w as k increases. As can be seen from figure 3.2 and figure 3.5 in the next section, it turns out that these two situations tend to coincide since high dimensional connection architectures with low w give larger values of C. When w is close to one C tends to be smaller and only a few terms in the sequence are required for accurate evaluation of r. In the case of the fully connected network where the sum can be done analytically so the exact result is known, only six terms in the sequence are required to evaluate the maximum storage capacity $\alpha_c = 0.138$ to three significant figures. For all the connection architectures studied, using the truncation term, it was never found necessary to evaluate more than twenty terms numerically to obtain r very accurately. It was also found that the solutions to the order parameter equations are well behaved with small fluctuations in any of the order parameters causing only small fluctuations in the other order parameters. Therefore any small errors in r do not cause significantly larger errors in the other parameters of the system.

It is worth noting at this point that because of the monotonic decreasing nature of the sequence $a_k(w)C^k(k+1)$ it is the shorter correlation loops that count most in determining the thermodynamic properties of the network. This means for example, that low dimensional connectivity even with low values of w will have properties very similar to a fully connected network. This is because it is only the longer correlation loops that are lost in systems with low dimensional connectivity. Conversely, randomly connected systems with moderate values of w have many fewer short correlation loops so their properties will be significantly different from a fully connected network. These properties of partially connected systems will be borne out by the numerical and theoretical results in the next few sections of this chapter.

It is worth studying the case of random connectivity in more detail since r can be calculated analytically in this case. For the randomly connected network we have $a_0(w) = 1$ due to the symmetry of the connections and $a_k(w) = w$ for $k \geq 1$. This is because the connection space is a random set of points, so there is always the same probability w of being in the connection space of the starting point after any number of random steps greater than one. The expression for the sum of an infinite number of terms all with $a_k(w) = w$ can therefore be used to evaluate r, remembering that the first term in the sequence is 1 not w (see equation 3.10). This gives,

$$r = q\left[1 + w\left(\frac{1}{(1-C)^2} - 1\right)\right] \tag{3.12}$$

for a randomly connected network. In section 3 of this chapter we will look at the phase diagrams for randomly connected networks but in the next section we will only solve the zero temperature order parameter equations for different hypercubic connection architectures including the randomly connected model.

3.2 Zero Temperature Solutions of the Replica-symmetric Order Parameter Equations

In the limit $\beta \rightarrow \infty$, $\tanh\beta(\sqrt{\alpha r}z + m)$ can only take the values 1 or -1 and it will change between these two values at $m = -\sqrt{\alpha r}z$. We therefore obtain for m (see equation 3.1),

$$m = \int_{-\frac{m}{\sqrt{\alpha r}}}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) - \int_{-\infty}^{-\frac{m}{\sqrt{\alpha r}}} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{3.13}$$

Adding these two terms gives the zero temperature order parameter equation for m,

$$m = 2\,\mathrm{erf}\left(\frac{m}{\sqrt{\alpha r}}\right) \tag{3.14}$$

where,

$$\mathrm{erf}\,(x) = \frac{1}{\sqrt{2\pi}}\int_0^x \exp\left(\frac{-t^2}{2}\right) dt \tag{3.15}$$

At zero temperature q becomes one since all the spins freeze into position. C can be calculated in a similar way to m, and r remains unchanged since it does not explicitly contain β. Therefore the zero temperature order parameter equations are,

$$\begin{aligned} m &= 2\,\mathrm{erf}\,\frac{m}{\sqrt{\alpha r}} \\ q &= 1 \\ r &= 1 + \sum_{k=1}^{\infty} C^k(k+1)a_k(w) \\ C &= \sqrt{\frac{2}{\pi\alpha r}}\exp\left(\frac{-m^2}{2\alpha r}\right) \end{aligned} \tag{3.16}$$

We now wish to solve these equations to determine the extent of the memory phase (m, q, r all finite). The critical point at which m becomes zero as α is increased will give us the maximum storage capacity of the network α_c. The easiest way to solve the order parameter equations is to parameterize them by introducing,

$$t = \frac{m}{\sqrt{\alpha r}} \tag{3.17}$$

This gives for the order parameters,

$$\begin{aligned} m &= 2\,\mathrm{erf}\,(t) \\ C &= \frac{t\exp\left(\frac{-t^2}{2}\right)}{\sqrt{2\pi}\,\mathrm{erf}\,(t)} \\ r &= 1 + \sum_{k=1}^{\infty} C^k(k+1)a_k(w) \end{aligned} \tag{3.18}$$

An obvious, trivial solution to these equations is $m = t = 0$ with $q = 1$. This solution exists for all values of α and for all connection architectures and it corresponds to the spin glass phase. There also exists a non-trivial solution with $m \neq 0$ at low values of α which corresponds to the memory phase. Therefore the memory phase co-exists with the spin glass phase at low values of α. In the memory phase the two important parameters are α and m which determine how much information is stored and how accurately it is stored. We can write α in parametric form as well where r is given as a function of t through the parametric equation for C in equation 3.18,

$$\alpha(t) = \frac{4(\operatorname{erf}(t))^2}{t^2 r(t)} \tag{3.19}$$

The maximum value of $\alpha(t)$ will give us the maximum storage ratio α_c above which $m = 0$ is the only solution. A plot of $\alpha(t)$ against m will show how the accuracy of storage changes with the number of states stored. Plots of this type are shown in figure 3.3 for some different hypercubic connection architectures. The sections of the curves for $m < m_c$ are not shown as they are unstable solutions corresponding to maxima, rather than minima of the free energy with respect to fluctuations in the parameter m (see section 1.5). The end points of the curves (α_c, m_c), represent the phase transition point where we move from the co-existence phase having storage properties, to the pure spin glass phase with no storage. A very important result from these curves is that the higher the dimensionality of connectivity d, and the lower w is the better the maximum storage capacity per connection is. Also for a given error tolerance $(1-m)$ of the states we are storing, the more partially connected the network is the more states per connection it will store. Therefore a partially connected system will always out-perform a fully connected system with the same number of connections. The values of m_c and α_c are plotted in figure 3.4 for different connection architectures and all values of w.

The α_c and m_c family of curves is enveloped by the two curves $w = 1$, fully connected and $d = \infty$, randomly connected. The α_c and m_c curves for any connection architecture will lie between these two extremes. This result comes from the close correlation between the values of α_c , m_c and $a_k(w)$. In the case of α_c, the smaller the values of $a_k(w)$ the larger are the values of α_c. As mentioned in the previous section random connectivity gives the lowest values of $a_k(w)$ and hence the highest values of α_c. In the case of the fully connected network $a_k(w) = 1$ for all k which is the highest possible value of $a_k(w)$ and correspondingly the lowest possible value of α_c. In the case of m_c, the smaller the values of $a_k(w)$ the correspondingly smaller are the values of m_c. Another important result from these graphs is that the phase transition at zero temperature between the spin

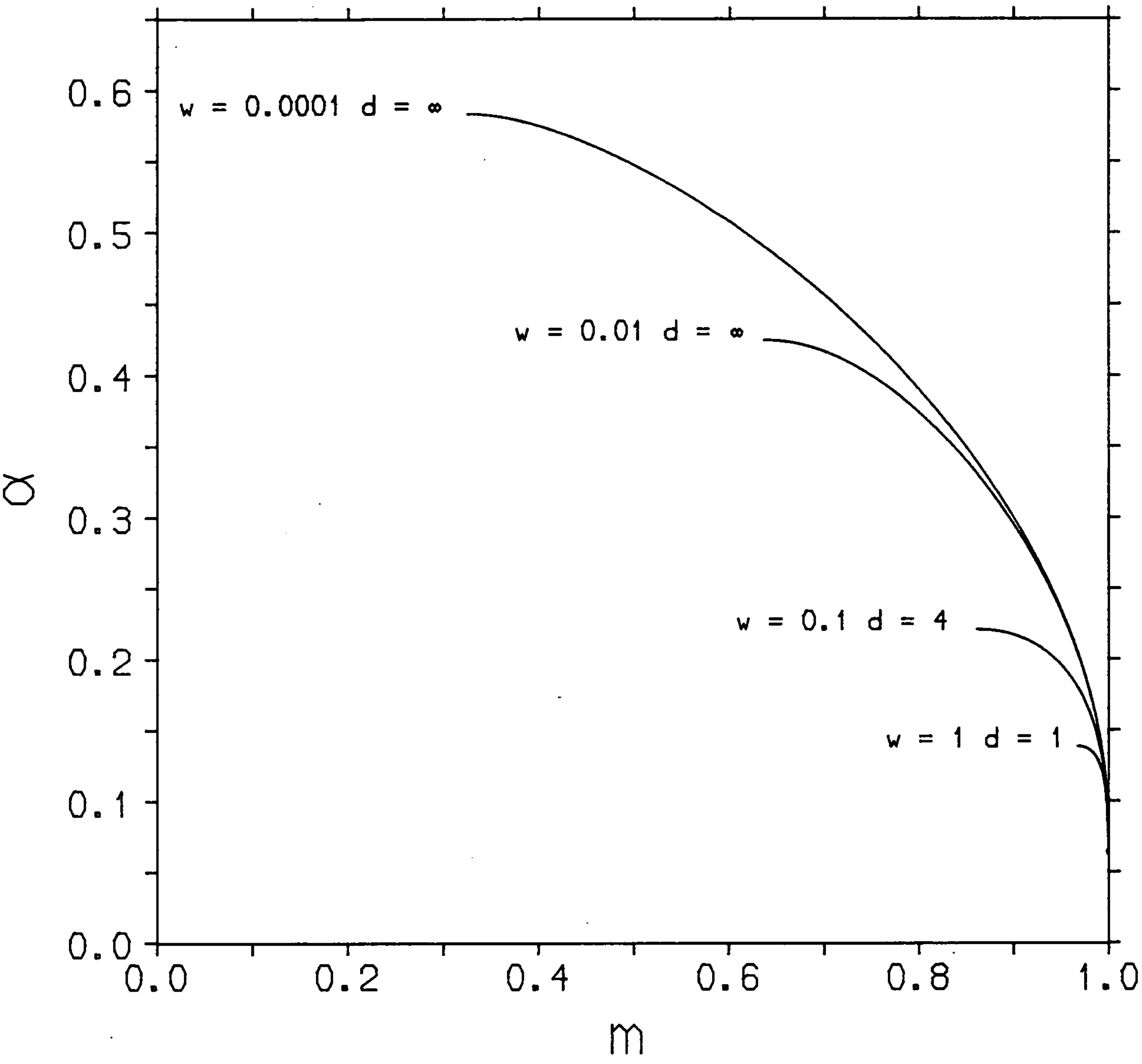

Figure 3.3: α against m for different hypercubic connection architectures in the memory phase.

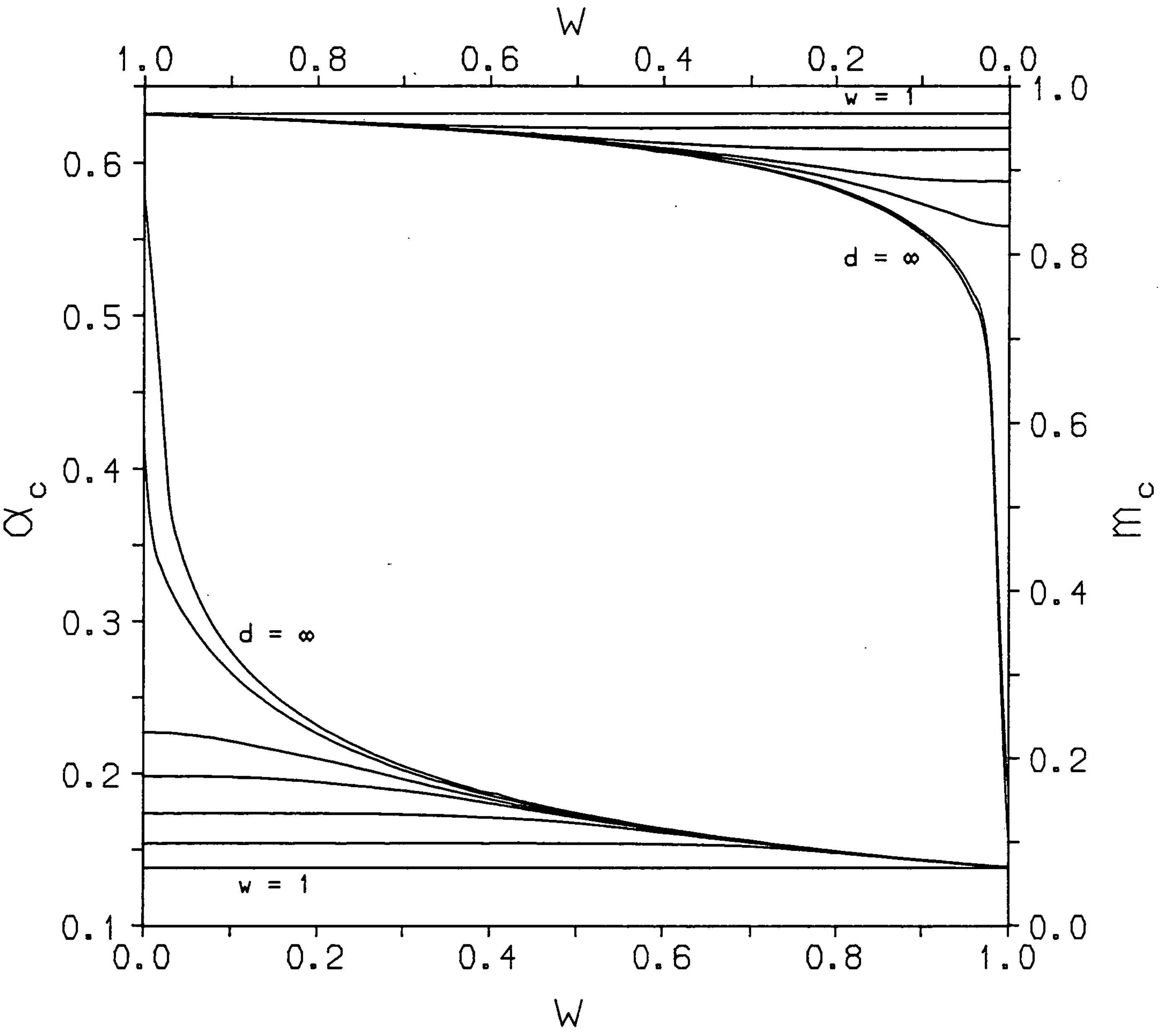

Figure 3.4: Critical values of the order parameters α and m are plotted against w. On the left and bottom axis are plotted α_c against w for different hypercubic connection architectures. The curves are, starting from the bottom, $w = 1, d = 1, 2, 3, 4, 8$ to $d = \infty$ at the top. On the right and upper axis is plotted m_c against w with the top curve being $w = 1$ through to $d = \infty$ on the bottom.

glass phase and the co-existence phase are all first order except in the limit of random connectivity and $w \to 0$ where the phase transition becomes second order. This limit of connectivity is discussed in more detail in section 3.4 of this chapter where the phase diagram for the randomly connected $w = 0$ model is calculated. The critical values of the other order parameter r_c and also C_c are shown in figures 3.5 and 3.6. As can be seen from the C_c curves the lower the connectivity and the higher the dimensionality of the connectivity, the closer to its maximum value of 1 C_c becomes. The r_c curves show oscillatory behaviour for intermediate values of d the dimensionality of the connectivity. This is due to the interplay of the terms C_c^{k+1} and $a_k(w)$ in the series for r. As w is lowered C_c increases and $a_k(w)$ decreases (see figure 3.2), and in some cases C_c can increase more rapidly than $a_k(w)$ so pushing the value of r_c up (see $d = 3$ curve). In other cases $a_k(w)$ dominates and pushes the value of r_c down as w is decreased. For some curves these two types of behaviour interchange at different values of w giving rise to oscillatory curves (see $d = 8$). This oscillatory behaviour in r_c is not reflected in α_c because the other terms in α_c always keep it increasing as w is decreased (see equation 3.19).

In the case of the randomly connected model we have an analytical expression for r and as $w \to 0$, $m \to 0$ also. The only order parameter which is not infinitesimally small across the phase boundary is q but we know that its value is 1. We can therefore analytically solve the order parameter equations by expanding them about m_c and w equal to zero. We have to be careful to what order we retain m_c,w and terms of the form $m_c w$ in our expansion as we do not know apriori, the relationship between m_c and w. We will in fact expand about $t = 0$ and derive the relationships between m_c, α_c and w from this. Integrating the Taylor expansion for the exponential function we obtain the Taylor series for the error function,

$$\operatorname{erf}(t) = \frac{1}{\sqrt{2\pi}}\left(t - \frac{t^3}{3!} + \frac{3t^5}{5!} - \cdots\right). \tag{3.20}$$

This gives for the order parameter equations,

$$\begin{aligned} m &= \sqrt{\frac{2}{\pi}}\left(t - \frac{t^3}{3!} + O(t^5)\right) \\ r &= 1 + w\left(\frac{9}{t^4}(1 + O(t^2))\right) \end{aligned} \tag{3.21}$$

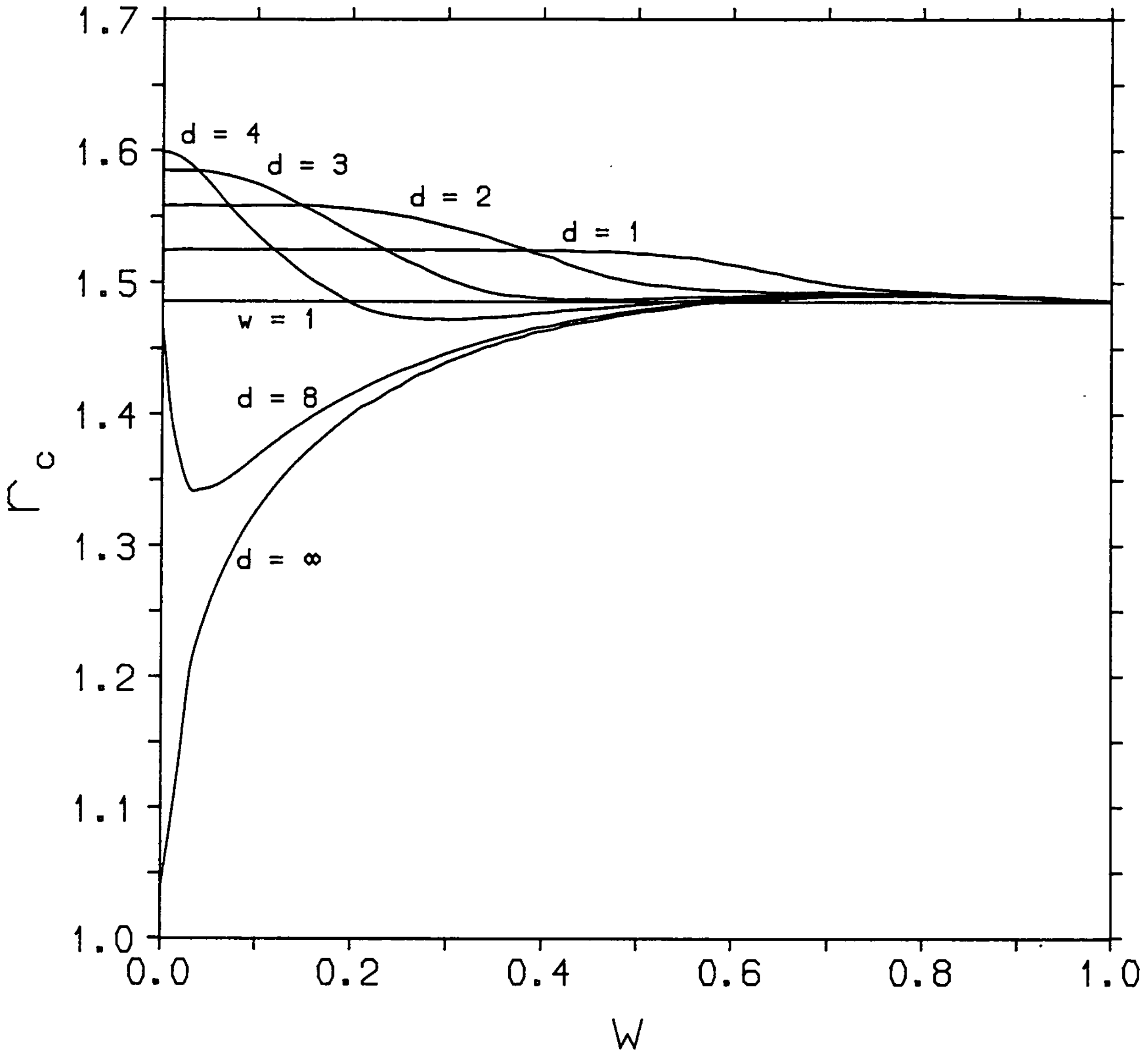

Figure 3.5: Critical values of r for different hypercubic connection architectures.

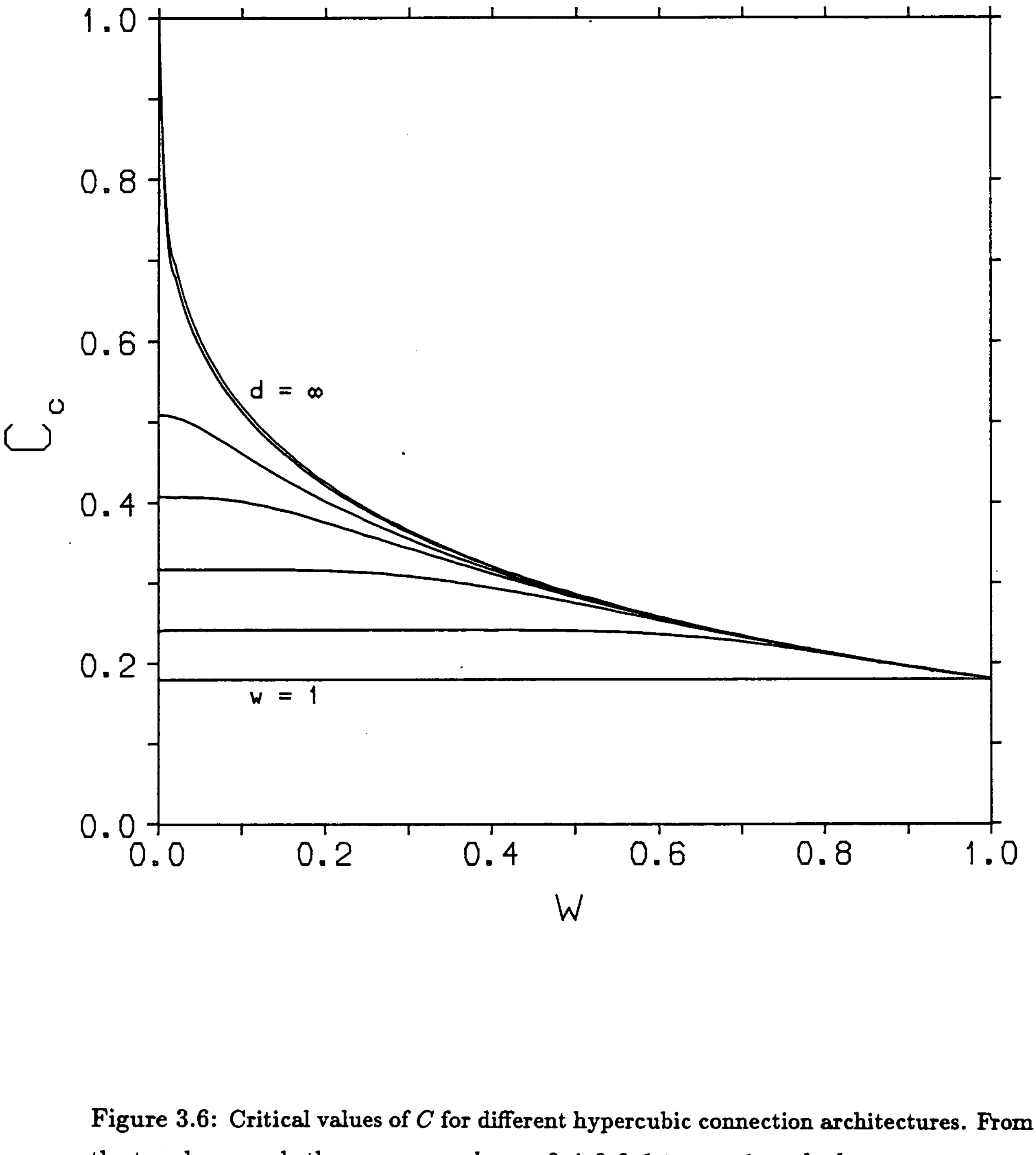

Figure 3.6: Critical values of C for different hypercubic connection architectures. From the top downwards the curves are $d = \infty, 8, 4, 3, 2, 1$ to $w = 1$ at the bottom.

and,

$$\alpha(t) = \frac{2}{\pi}\left(1 - \frac{t^2}{3} - \frac{9w}{t^4} + O(t^4) + O\left(\frac{w}{t^2}\right)\right) \tag{3.22}$$

For self consistency of these expansions we are assuming that w is of order t^6 or higher. We now have to maximize $\alpha(t)$ in order to determine t_c in terms of w and hence α_c, m_c and r_c. The first derivative of $\alpha(t)$ is given by,

$$\frac{\pi}{2}\frac{\partial\alpha(t)}{\partial t} = -\frac{2t}{3} + \frac{36w}{t^5} + O(t^3) + O\left(\frac{w}{t^3}\right) \tag{3.23}$$

Only keeping the first two terms in the above series gives α_c at t_c where,

$$w = \frac{1}{54}t_c^6 \tag{3.24}$$

We can now see the self consistency of our expansions since terms of the form $\frac{w}{t^3}$ are in fact of order t^3. Putting this expression back into equations 3.21 and 3.22 gives us,

$$\begin{aligned} m_c &\simeq 2^{\frac{2}{3}}\sqrt{\left(\frac{3}{\pi}\right)}w^{\frac{1}{6}} \\ r_c &\simeq 1 + \sqrt[3]{\frac{w}{4}} \end{aligned} \tag{3.25}$$

and,

$$\begin{aligned} \alpha_c &\simeq \frac{2}{\pi}\left(1 - \frac{3}{2^{\frac{2}{3}}}w^{\frac{1}{3}}\right) \\ C_c &\simeq 1 - (2w)^{\frac{1}{3}} \end{aligned} \tag{3.26}$$

for networks where the connectivity is random and w is small. Thus the maximum possible value of α_c is $\frac{2}{\pi}$ and this occurs in the limit $w \longrightarrow 0$. The one sixth power in the expression for m_c explains why the value of m_c holds up as w becomes small before rapidly dropping to zero as w approaches zero (see figure 3.4). Another interesting result here is that as $w \longrightarrow 0$, $r_c \longrightarrow 1$ which is the same value as q. It turns out that for the randomly connected model in this limit $r \longrightarrow q$ at all temperatures for the memory phase but we will leave further discussion of this until section 3.7.

3.3 Maximum Information Capacity per Connection in Partially Connected Networks at Zero Temperature

We saw in the last section how for different connection architectures we obtained different values for the storage capacity per connection α and accuracy of storage m. We now wish to find some way of directly comparing the information storage capacity of systems of different connection architectures which takes into account both the number of states stored as well as their accuracy of storage. The application of information theory techniques will allow us to calculate an expression which takes into account both these factors. We will then maximize this expression for different hypercubic connection architectures and hence be in a position to directly compare the performance of different connection architectures. We will thus be able to directly compare systems storing a number of states very accurately with those storing many more states but less accurately. As we have already seen though, the more partially connected networks will always perform best. In what follows we will use the same techniques as Amit *et al* [7].

If we consider an N bit vector then the amount of information contained in that vector is defined to be the log of the total number of permutations possible with an N bit vector. This gives the information content of an N bit vector as $\ln 2^N$. We can understand the form of this quite easily from basic intuitive ideas about information. Firstly we expect the longer a vector is the more information it must contain hence the 2^N factor. Secondly we expect information to be additive property. In terms of entropy the information of an N bit vector is just the entropy associated with the ensemble of all possible states of the vector. Now suppose we have an N bit vector which has a certain number of bits W which are wrong. What is the information content of this vector ? We proceed in a similar way to the information content of the N bit vector and define the information lost by having W bits wrong as the log of the total number of possible ways of choosing W bits from N bits. Therefore the information content of this vector is given by,

$$information \quad = N\ln 2 - \ln\left(\frac{N!}{W!(N-W)!}\right) \tag{3.27}$$

The number of bits wrong is related to the overlap m by,

$$W = N\frac{(1-m)}{2} \tag{3.28}$$

In the thermodynamic limit Stirling's approximation becomes exact and can be used to calculate the factorials in equation 3.27 and we get, for the information stored in a neural network per connection,

$$I(\alpha) = \frac{\alpha}{2\ln 2}[(1+m)\ln(1+m) + (1-m)\ln(1-m)] \tag{3.29}$$

The factor $\frac{\alpha}{2\ln 2}$ is a normalization factor so that the information $I(\alpha)$ equals α if all the states are stored exactly ($m = 1$). The maximum values of $I(\alpha)$ are obtained by relaxing α below α_c. In the case of the randomly connected model in the limit $w \rightarrow 0$, $m_c \rightarrow 0$ so there is no information stored in the network at α_c. In the case of the fully connected network the maximum value of $I(\alpha)$ is obtained at $\alpha_{info} = 0.134$ below $\alpha_c = 0.138$.

Figure 3.7 shows the values of α_{info} giving maximum storage capacity per connection, for different connection architectures, and the corresponding values of m_{info}. We can see by comparing these curves with figure 3.4 how relaxing α below α_c always leads to an increase in m_{info} and maximizes $I(\alpha)$. The higher the dimensionality and the lower w, the more α has to be relaxed below α_c to maximize the information storage capacity per connection. Figure 3.8 shows the maximum values of $I(\alpha)$ for the different connection architectures. Again we see that the the curve for the randomly connected model and the curve for the fully connected model envelope all the other connection architecture curves. The maximum information capacity per connection is achieved with a randomly connected network in the limit $w \rightarrow 0$ which stores about 70% more information per connection than a fully connected network.

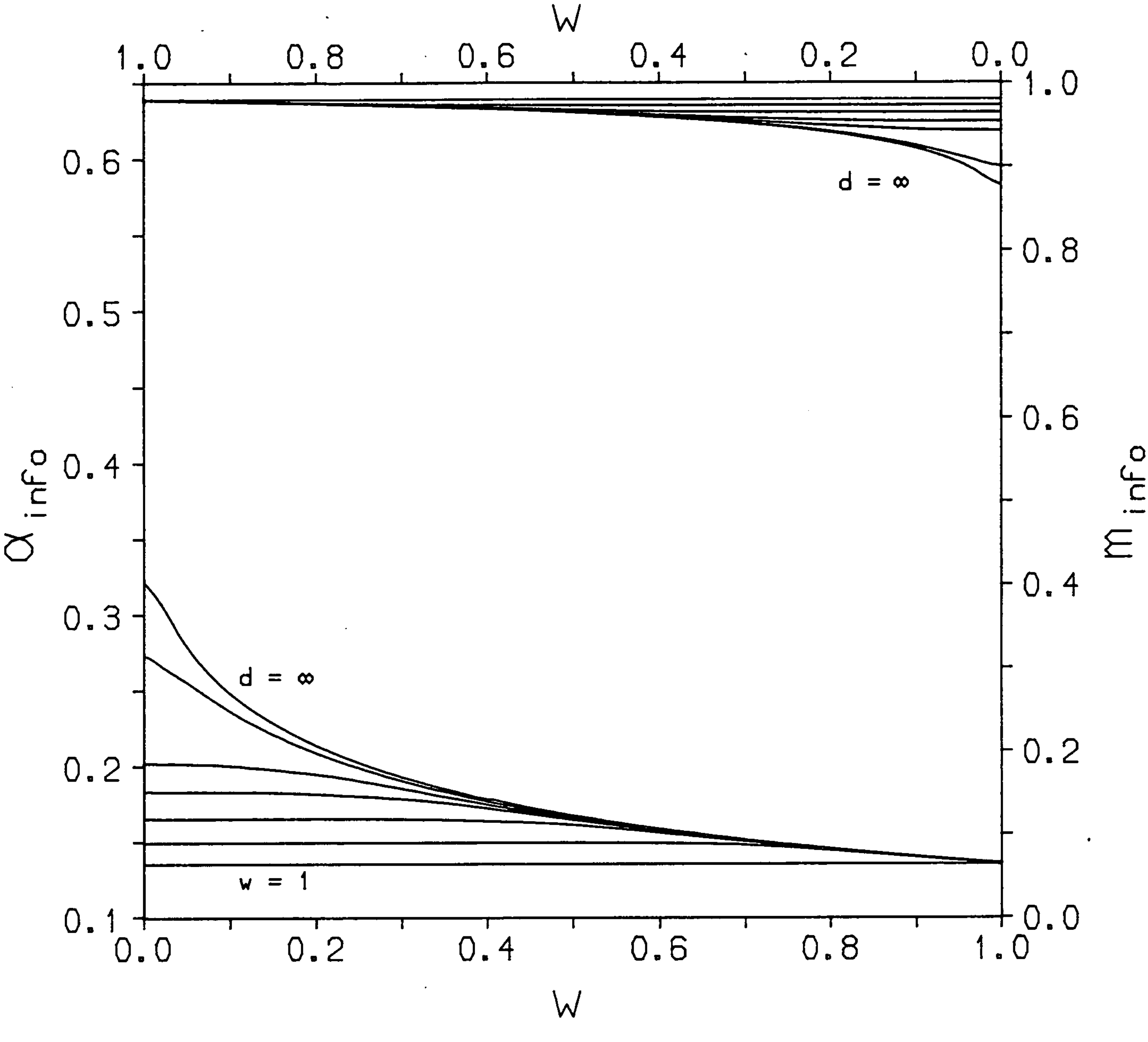

Figure 3.7: The values of α and m giving maximum information storage. On the lower axis are plotted the values for α_{info} for the connection architectures $w = 1, d = 1, 2, 3, 4, 8$ to $d = \infty$. On the upper axis are plotted the values of m_{info} from $w = 1$ at the top through to $d = \infty$.

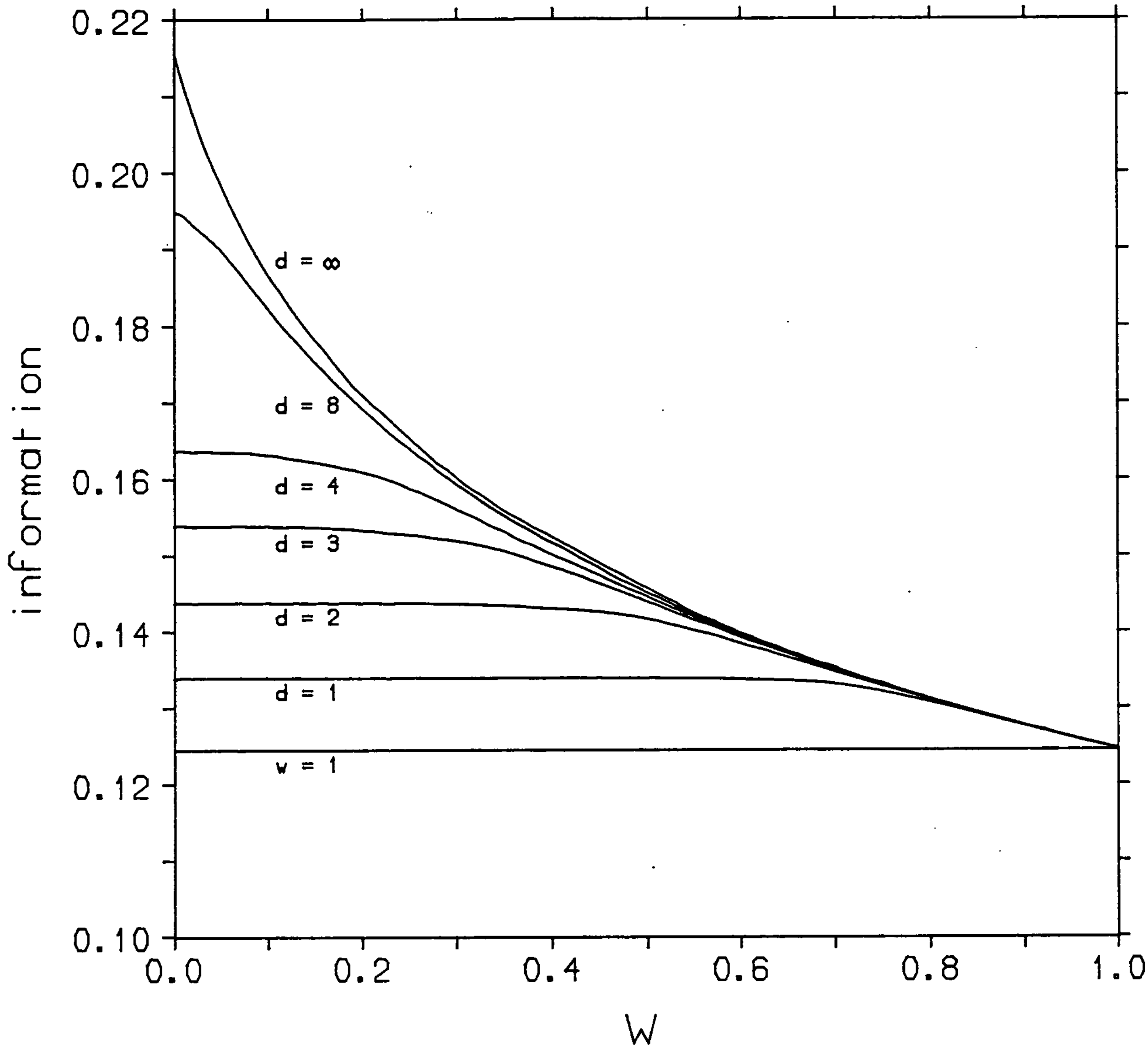

Figure 3.8: The values of maximum information storage capacity per connection as defined by equation 3.29 for different connection architectures.

3.4 Phase Boundaries for Randomly Connected Networks in Replica-symmetric Theory

As we have seen with the zero temperatures studies of different networks, the randomly connected network and the fully connected network represent two limits between which the results for all other connection architectures lie. We are therefore going to look at the randomly connected model at finite temperature and in particular derive the phase diagram for the $w = 0$ model. After that we will qualitatively discuss the phase diagrams for other connection architectures, although it is possible to numerically calculate the phase diagram for any model by using the random walk technique to calculate r and then solving the order parameter equations numerically.

When we have a second order phase transition from the paramagnetic phase ($m = q = r = 0$), to an ordered phase, some or all of the order parameters will change continuously across the phase boundary. The phase transition lines can then be determined analytically by expanding all the order parameter equations in small m, q and r and solving them to first order. This technique was illustrated in section 1.5 where we used it to calculate the ferromagnetic phase boundary for the infinite range Ising model.

Firstly we will start by looking for a second order phase boundary between the paramagnetic phase and the spin glass phase. Since $m = 0$ across this phase boundary we only have two order parameter equations in q and r (see equations 3.1 and 3.12) to expand and solve. Expanding the order parameter equations in q and r gives,

$$\begin{aligned} r &= q\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right] - q^2\frac{2w\beta}{(1-\beta)^3} + O(q^3) \\ q &= r\beta^2\alpha - 2r^2\beta^4\alpha^2 + O(r^3) \end{aligned} \tag{3.30}$$

eliminating r from these two equations gives to second order in q,

$$q = q\beta^2\alpha\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right] - q^2\left(\frac{2w\alpha\beta^3}{(1-\beta)^3} + 2\beta^4\alpha^2\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right]^2\right) \tag{3.31}$$

$q = 0$ is always a solution of this equation but below a certain temperature T_g there are also finite q solutions. Above T_g, $q = 0$ is a stable solution but below T_g this solution becomes unstable. Solutions of this equation to first order will give us candidates for the spin glass phase boundary and the second order solution will give us the value of q close to the phase boundary. To first order we have,

$$f(\beta) = 1 - \beta^2\alpha\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right] = 0 \qquad (3.32)$$

Solutions of this equation will yield candidates for the spin glass phase boundary. For the fully connected model $w = 1$, we get two solutions $T = 1 \pm \sqrt{\alpha}$ with the highest curve $T_g = 1 + \sqrt{\alpha}$ being the phase boundary. Providing $T \neq 1$ we can rewrite the above equation as a quartic in T giving,

$$T^4 - 2T^3 + (1-\alpha)T^2 + 2\alpha(1-w)T + \alpha(w-1) = 0, \quad T \neq 1 \qquad (3.33)$$

Depending on the value of w this quartic is of irreducible form for $\alpha \leq \alpha_i$ where $\alpha_i \leq 1$, and $\alpha_i \to 0$ as $w \to 1$ and $\alpha_i \to 1$ as $w \to 0$ (see reference [49] for more information on irreducible polynomials). α_i depends only on the value of w. We can therefore not explicitly write down the roots of this quartic in terms of w and α for what turns out to be the most important area of the phase diagram. The sum of the four roots of the quartic is two and the product of the four roots is $\alpha(w-1)$ which is always negative for $w \neq 1$. The complex roots of polynomials with real coefficients only occur in conjugate pairs so for the quartic equation 3.33 there must always be at least two real roots to give a valid phase boundary. Therefore, since the product of a conjugate pair is always positive if there are only two real roots one of them must always be negative therefore not a possible candidate for a phase boundary. In this instance the real positive root will give the phase boundary. In the range $0 < \alpha < \alpha_i$ the quartic always has three positive real roots and one negative real root. Two of these positive real roots always lie between zero and one and merge at α_i becoming complex for $\alpha > \alpha_i$. These roots correspond to a minima that always lies between the two asymptotes $T = 0$ and $T = 1$ of equation 3.32. At $\alpha < \alpha_i$ this minima always lies below the $f(\beta) = 0$ axis but as α increases this minima rises up passing through the $f(\beta) = 0$ axis when $\alpha = \alpha_i$. Thus neither of these two roots give continuous values of T_g for all α and therefore cannot correspond to the spin glass phase boundary. The largest root, which is always greater than one, must always be the spin glass phase boundary T_g. This means that at temperatures above the value of this root $q = 0$ must be the only solution and at temperatures below the root, positive real solutions for q must exist. We will now look at some solutions of equation 3.33 for different values of w.

In the case of $w = \frac{1}{2}$ with some intuitive skill the quartic can be broken up into two quadratic terms giving,

$$\left(T^2 - T - \frac{\alpha}{2} + \frac{\sqrt{(\alpha^2 + 2\alpha)}}{2}\right)\left(T^2 - T - \frac{\alpha}{2} - \frac{\sqrt{(\alpha^2 + 2\alpha)}}{2}\right) \tag{3.34}$$

The first quadratic term gives the two real roots between zero and one for $0 \leq \alpha < \frac{1}{4}$ which become complex when $\alpha \geq \frac{1}{4}$. The second quadratic factor gives two real roots for all $\alpha \geq 0$, one of the roots always being negative and the other positive and larger than one. For $w = \frac{1}{2}$ the original quartic is termed of irreducible form in the region $\alpha \leq \alpha_i$ where $\alpha_i = \frac{1}{4}$ and all the roots are real. The positive root of the second quadratic gives the spin glass phase boundary T_g where,

$$T_g = \frac{1 + \sqrt{\left[1 + 2(\alpha + \sqrt{\alpha^2 + 2\alpha})\right]}}{2} \tag{3.35}$$

This root and the other two roots are plotted in figure 3.9 along with numerical solutions of the quartic equation for $w = 0.001$. The two solutions always lie below the fully connected solution $T_g = 1 + \sqrt{\alpha}$ except in the limits $\alpha \rightarrow \infty$ and $\alpha \rightarrow 0$ where they give the same result. In fact if we take the limit $\alpha \rightarrow \infty$ in the quartic equation 3.33 we only get two solutions $T = \pm\sqrt{\alpha}$ for all values of w with $T_g = +\sqrt{\alpha}$ being the physically meaningful solution. Similarly if we take the limit $\alpha \rightarrow 0$ in equation 3.33 then $T \rightarrow 1$ or 0 for all values of w with $T = 1$ being the physically meaningful phase boundary. So the spin glass phase boundary for all values of w must start at $T = 1$ and tend to $\sqrt{\alpha}$ for large α.

We will now look at the case $w \rightarrow 0$, where it turns out that $T = 1$ is a solution so we must work from the original expression equation 3.32. The $T = 1$ solution comes from the interplay of limits $w \rightarrow 0$ and $\beta \rightarrow 1$ on the term $\frac{w}{(1-\beta)^2}$ which keeps it finite. We will therefore look for solutions of the form $T = 1 + x$ where x is small when w is small. Solving for x in terms of w the limit $w \rightarrow 0$ will then be well defined on T. Putting $T = 1 + x$ into equation 3.32 and assuming w is of order x^2 gives,

$$x^2 = x^2\alpha + w\alpha + O(x^3) \tag{3.36}$$

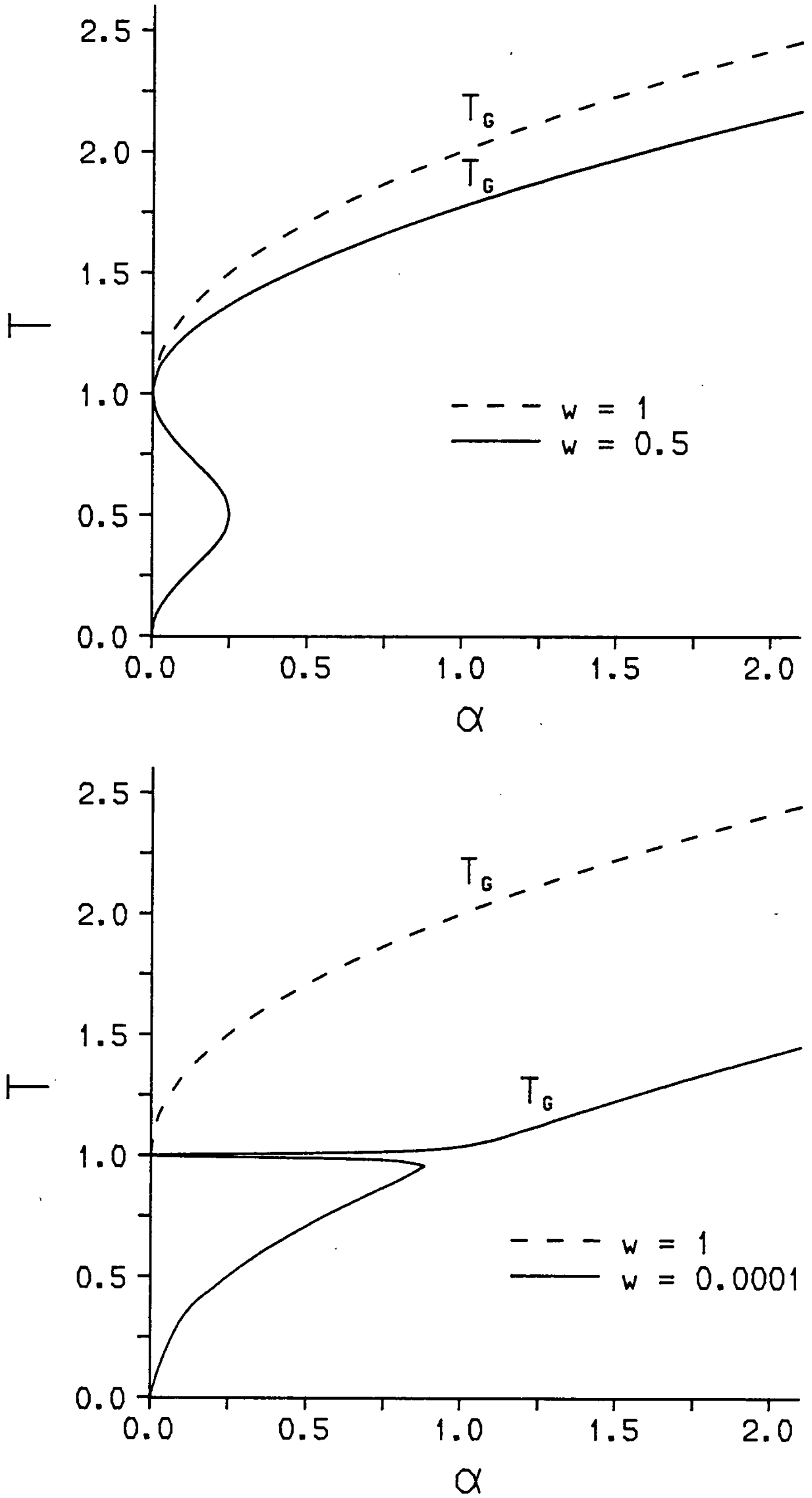

Figure 3.9: Positive real roots of the quartic equation 3.33 for w = 0.5 (top) and w = 0.001 (bottom). The spin glass phase boundary T_g is given by the largest root. The spin glass phase boundary for the $w = 1$ model is also presented for comparison.

Ignoring terms of order x^3 gives,

$$x = \pm\sqrt{\frac{\alpha w}{1-\alpha}} \tag{3.37}$$

so our assumption that w was of order x^2 was self-consistent. We therefore have two solutions of which the largest one gives the phase boundary T_g where,

$$T_g = 1 + \sqrt{\frac{\alpha w}{1-\alpha}}, \qquad \alpha < 1 \tag{3.38}$$

for small w. We can now take the limit $w \rightarrow 0$ where we find that the two solutions of equation 3.37 merge into one, giving,

$$T_g = 1, \quad for \;\; w = 0, \quad \alpha < 1 \tag{3.39}$$

This only gives us part of the phase boundary corresponding to $\alpha < 1$. For larger values of α we already know that $T_g \rightarrow \sqrt{\alpha}$. Also since the solutions for $w = 1$ and $w = \frac{1}{2}$ approach $\sqrt{\alpha}$ from above as α becomes large we will look for solutions of the form $T_g = \sqrt{\alpha}(1+x)$ for w small where we expect x to be small and positive. Solving equation 3.32 to first order in x and w gives,

$$T_g = \sqrt{\alpha}\left(1 + \frac{w}{2}\left[\frac{\alpha}{(\sqrt{\alpha}-1)^2} - 1\right]\right) \quad \alpha \neq 1 \tag{3.40}$$

We can now take the limit $w \rightarrow 0$ and we get,

$$T_g = \sqrt{\alpha}, \;\; for \;\; w = 0, \; \alpha \neq 1 \tag{3.41}$$

The other solution obtained from this expansion is $-\sqrt{\alpha}$ which is physically not meaningful. Therefore for the $w = 0$ model we only have one solution for $\alpha > 1$ but we have two possible solutions $\sqrt{\alpha}$ or 1 for $\alpha < 1$ which both give continuous phase boundaries.. The $\sqrt{\alpha}$ is not physically meaningful for $\alpha < 1$ since it implies that as $\alpha \rightarrow 0$ then $q \rightarrow 0$ on the phase boundary. At $T = 0$ this is not possible since $q = 1$ and therefore the phase boundary would have to be first order which is not self-consistent. We therefore have for the spin glass phase boundary for the $w = 0$ randomly connected model,

$$w = 0, \;\; T_g = \begin{cases} \sqrt{\alpha} & \alpha > 1 \\ 1 & \alpha \leq 1 \end{cases} \tag{3.42}$$

It is very interesting to note how the smaller positive roots play a secondary role in the build up of the discontinuity in the curvature of the phase boundary as

$w \rightarrow 0$ (see the $w = 0.001$ roots in figure 3.9). As $w \rightarrow 0$ the smaller of the two positive roots $\rightarrow \sqrt{\alpha}$ and the other root $\rightarrow 1$ from below. As $\alpha \rightarrow 1$ the $\sqrt{\alpha}$ root merges with the root just below one and these two roots become complex (see figure 3.9). The largest root, which for $0 < \alpha < 1$ has stayed close to one, then suddenly takes over the behaviour of the $\sqrt{\alpha}$ root for $\alpha > 1$. Therefore the curvature of the phase boundary is always continuous except in the limit $w \rightarrow 0$. Equation 3.40 actually corresponds to two different roots of the quartic depending on whether $\alpha > 1$ or $\alpha < 1$ which explains the singularity at $\alpha = 1$.

The second order solution of equation 3.31 will give us the form of q close to the phase boundary. This gives for q,

$$q \simeq \frac{\beta^2\alpha\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right] - 1}{\frac{2w\alpha\beta^3}{(1-\beta)^3} + 2\beta^4\alpha^2\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right]^2} \tag{3.43}$$

so when this term is negative, $q = 0$ is the only physically meaningful solution of the order parameter equations and when it is positive q becomes finite corresponding to the spin glass phase. We can check the validity of the spin glass phase boundary we obtained from the first order solutions of equation 3.31. At temperatures above T_g equation 3.43 should only give negative unphysical values for q corresponding to $q = 0$, being the only physically meaningful solution of equation 3.31 and at temperatures below T_g it should give finite positive values of q. This means that the numerator in equation 3.43 should control the change in sign of q, negative above T_g and positive below T_g, and the denominator should always be positive across the phase boundary. Since the phase boundary derived from the first order equation is always at $\beta < 1$ the denominator in equation 3.43 is always positive across the phase boundary and the numerator changes sign in the correct direction. In the case of the $w = 0$ model the value of q, close to the phase boundary, is given by,

$$w = 0, \quad q \simeq \begin{cases} \frac{\beta^2\alpha - 1}{2\beta^4\alpha^2} & \alpha > 1 \\ 1 - \frac{1}{\beta} & \alpha \leq 1 \end{cases} \tag{3.44}$$

For any infinite range connection architecture, the spin glass phase boundary will be given by the root of the polynomial in β,

$$\beta^2\alpha\sum_{k=0}^{n}(k+1)\beta^k a_k(w) - 1 = 0 \tag{3.45}$$

which is less than one and positive. The series is always monotonic decreasing for the required root and therefore the number of terms n, needed to evaluate the root accurately will depend on the values of the $a_k(w)$'s. In general more terms will be needed if the $a_k(w)$'s are small since in this case the required solution for $\alpha < 1$ gives β closer to one.

It should be noted that in the calculation of the spin glass phase boundary by expanding in the order parameters we have assumed that this is a second order phase boundary which is entered from the paramagnetic phase as the temperature is lowered. This appears to be the case for all connection architectures although, as we shall see in the next section, for the randomly connected $w = 0$ model the memory phase boundary coincides with the spin glass phase boundary for $\alpha < 1$.

As we have seen in the zero temperature studies the phase transition point for the memory phase becomes second order as $w \to 0$ for the randomly connected model. We will therefore use the same method as we used for the spin glass phase boundary to look for a second order memory phase boundary at finite temperature. This method will only work if the memory phase boundary is coincident with the spin glass phase boundary, otherwise q will be finite across the phase boundary. The whole of the memory phase always overlaps the spin glass phase for all connection architectures. Expanding all the order parameter equations 3.1 and 3.12 to third order in m and noting that q and r are of order m^2 we have,

$$\begin{aligned} m &= \beta m - \beta^3 \alpha r m - \frac{\beta^3 m^3}{3} + O(m^5) \\ q &= \beta^2 \alpha r + \beta^2 m^2 + O(m^4) \\ r &= q\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right] + O(m^4) \end{aligned} \tag{3.46}$$

Solving these equations for m gives,

$$m = m\beta - m^3\left(\frac{\beta^5\alpha\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right]}{1 - \beta^2\alpha\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right]} + \frac{\beta^3}{3}\right) + O(m^5) \tag{3.47}$$

The first order solutions of this equation will give us potential candidates for the memory phase boundary, while solving it to cubic order will tell us the value of m close to the phase boundary. We can see from the above equation that the only candidate for a second order phase boundary is $T = 1$ for all values of w. This is below the spin glass phase boundary for all values of w except in the limit $w \rightarrow 0$. Therefore it can only represent a valid second order phase boundary for the $w = 0$ model in the region $\alpha < 1$. If we solve the equation to cubic order and then take the limit $w \rightarrow 0$, we obtain an expression for m which is valid close to the phase boundary,

$$m \simeq \pm\sqrt{\frac{3(\beta-1)(1-\beta^2\alpha)}{\beta^3(1+2\beta^2\alpha)}} \quad \alpha < 1 \tag{3.48}$$

The positive root corresponds to spins being aligned with the nominated memory state and the negative root corresponds to alignment with the nominated states image. This expression is only valid for $\alpha < 1$ due to the $(1 - \beta^2\alpha)$ term, which since β is close to one, will be negative if $\alpha > 1$ giving $m = 0$ as the only solution. This is in agreement with its co-existence with the spin glass boundary in the region $\alpha < 1$. The $m = 0$ solution of equation 3.47 exists at all temperatures but gives a maximum of the free energy at temperatures below T_m corresponding to an unstable state. So far for the $w = 0$ model we have only calculated a section of the memory phase boundary,

$$w = 0, \quad T_m = 1, \quad \alpha < 1 \tag{3.49}$$

The other section of the phase boundary at finite temperature which separates the spin glass phase from the co-existence phase must be calculated numerically as the spin glass order parameter q remains finite across this boundary. At zero temperature we were able to analytically calculate the phase point (see section 3.2) for the $w = 0$ randomly connected model and it is given by $(T_m = 0, \alpha = \frac{2}{\pi})$. This part of the phase boundary is also second order and the numerical results for it, along with the other phase boundaries are shown in figure 3.10 in the next section. For all the other connection architectures the memory phase boundary lies below the spin glass phase boundary at all values of α except zero therefore, q is finite across the phase boundary. Numerical techniques are therefore required to evaluate T_m. These numerical solutions will not actually be carried out but we will discuss what we expect the solutions to be at the end of the next section.

3.5 Replica Symmetry Broken Phases in Partially Connected Networks

All the calculations for the phase boundaries and information capacity we have done so far have been in replica-symmetric theory. In section 2.4 we calculated an eigenvalue which determines in what areas of the phase diagram the replica-symmetric solutions are unstable. These areas are usually referred to as replica broken phases and are determined by the inequality,

$$\frac{\beta^2 \alpha r}{q} \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sech}^4 \beta(\sqrt{\alpha r} z + m) < 1 \tag{3.50}$$

In spin glasses and the fully connected Hopfield model the spin glass phase is always unstable to replica symmetry breaking. For partially connected networks the spin glass phase is also unstable to replica symmetry breaking but in general this has to be proved numerically. This broken symmetry is to be expected since, as discussed in section 1.6, the very nature of the spin glass phase can only be described within a replica broken theory. We can however, explicitly examine the stability of the spin glass phase close to T_g by expanding the inequality equation 3.50 in the order parameters. Expanding sech^4 gives,

$$\operatorname{sech}^4(\beta\sqrt{\alpha r} z) = 1 - 2\beta^2 \alpha r z^2 + \frac{7}{3}(\beta^2 \alpha r)^2 z^4 + \cdots \tag{3.51}$$

Carrying out Gaussian integrals term by term gives for the stability condition,

$$q > \beta^2 \alpha r - 2(\beta^2 \alpha r)^2 + 7(\beta^2 \alpha r)^3 + \cdots \tag{3.52}$$

Expanding the order parameter equation in q (see equation 3.1) to order r^3 gives,

$$q = \beta^2 \alpha r - 2(\beta^2 \alpha r)^2 + \frac{17}{3}(\beta^2 \alpha r)^3 + \cdots \tag{3.53}$$

Therefore the stability condition becomes,

$$0 > \frac{4}{3}(\beta^2 \alpha r)^3 + \cdots \tag{3.54}$$

Since the order parameter r is positive this inequality is violated by terms of order r^3 and so the spin glass phase close to T_g is unstable to replica symmetry breaking. The stability of the spin glass phase can also be studied close to $T = 0$ by making the change of variable,

$$x = \beta\sqrt{\alpha r} z \tag{3.55}$$

in the integral in equation 3.50. We can then expand in $\frac{1}{\beta}$ giving for the integral,

$$\frac{1}{\beta\sqrt{2\pi\alpha r}}\int_{-\infty}^{+\infty} dx \exp\left(-\frac{x^2}{2\alpha r\beta^2}\right)\ \mathrm{sech}^4(x) = \frac{1}{\beta\sqrt{2\pi\alpha r}}\left[2 + O\left(\frac{1}{\beta^2}\right)\right] \qquad (3.56)$$

The stability condition now becomes,

$$\frac{\beta\sqrt{\alpha r}}{q\sqrt{2\pi}}\left[2 + O\left(\frac{1}{\beta^2}\right)\right] < 1 \qquad (3.57)$$

This is violated at low temperatures where the spin glass phase is replica broken. The instability of the spin glass phase in other regions of the phase diagram can only be shown numerically.

Unlike the spin glass phase the memory phase is split into two regions by replica symmetry breaking. The region at higher temperatures is stable to replica symmetry breaking and the region at lower temperatures has replica symmetry broken. The energy surface associated with the Hamiltonian has a basket of minima associated with each stored state which, at higher temperatures, merge into a single minimum in the free energy (see section 2.4). The line which separates these two phases can only be found by numerically solving the order parameter equations and plugging the values into the stability condition. The replica-symmetric phase diagram for the $w = 0$ randomly connected model with replica broken phases is shown in figure 3.10 along with the phase diagram for the $w = 1$ model [6] for comparison. These two phase diagrams represent the limits between which the phase diagrams of all other infinite range connection architectures lie. In the fully connected model only a very small section of the spin glass co-existence phase boundary lies in a replica broken area. In the $w = 0$ model the whole of the phase boundary between the spin glass phase and the co-existence phase is unstable to replica symmetry breaking since the whole boundary lies in a replica broken phase. This means that the position of the phase boundary predicted by replica-symmetric theory is incorrect, although the point $(T_m = 1, \alpha = 1)$ on the phase boundary is correct since it is coincident with the replica symmetry breaking line. We expect the exact results to always diverge continuously from replica-symmetric results when we cross into a replica broken region. In the range $\frac{2}{\pi} < \alpha < 1$ the system has a memory phase at high temperatures but there is only a spin glass phase at low temperatures. This behaviour is strange since we expect the more ordered memory state to be stable

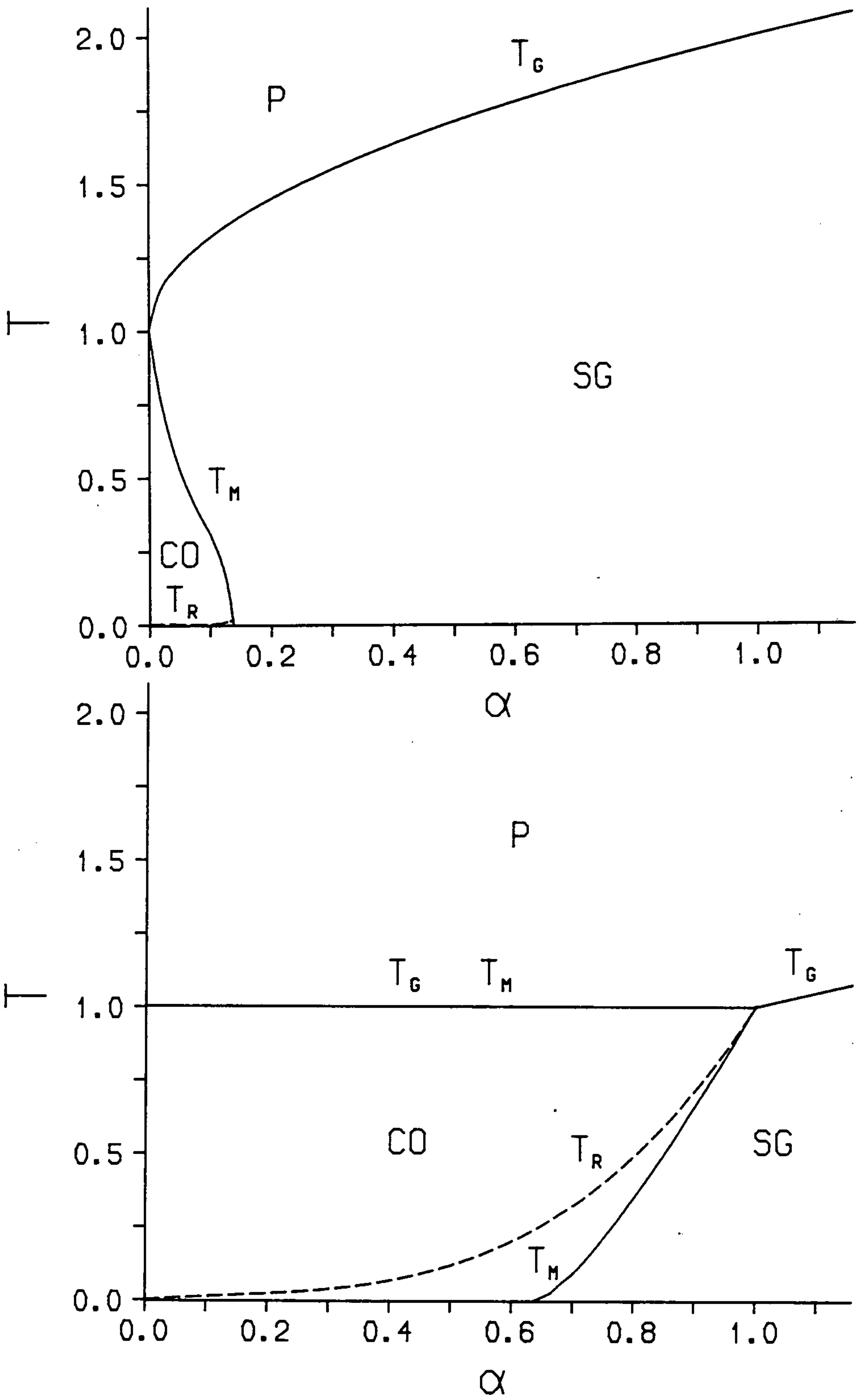

Figure 3.10: Replica-symmetric phase diagrams for the $w = 1$ model (top) and the $w = 0$ randomly connected model (bottom). P =paramagnetic phase, SG = spin glass phase, CO = co-existence phase (which contains both memory and spin glass phases). T_M and T_G are the memory and spin glass phase boundaries and T_R is the replica symmetry breaking line in the memory phase.

at lower temperatures while at higher temperatures, where entropy plays a role, we expect the spin glass states to be stable. As we shall see in the next section, by analogy with the SK model, we expect the true phase boundary to be vertical from $(T = 0, \alpha = 1)$ to $(T = 1, \alpha = 1)$. Therefore, the true phase boundary does not have this re-entrant memory phase and $\alpha_c = 1$ not $\frac{2}{\pi}$. The replica symmetry breaking effect is therefore quite large for the $w = 0$ model, unlike the fully connected model where replica symmetry breaking only increases α_c from 0.138 to 0.145 [11] (see also Chapter 4 section 3 of this thesis).

For other connection architectures we expect the replica symmetry breaking line to lie between the two limits of the fully connected model and the $w = 0$ randomly connected model. The lower the dimensionality and the higher the connectivity the smaller the effect of replica symmetry breaking will be. Replica symmetry breaking also increases the value of m for a given α in the replica broken phase. The results we obtained for the information storage capacity and the critical storage ratio α_c in sections 3.2 and 3.3 were therefore below the true values with the error being larger, the more partially connected and random the network is.

Although we have not numerically calculated the phase boundaries for other than randomly connected networks, the zero temperature results for different hypercubic connection architectures suggest the kind of results we would expect. The importance of the shorter correlation loops in determining the properties of the system also give us a good guide to the behaviour of different hypercubic connection architectures (see section 3.1). The spin glass phase boundary T_g, is always fixed at both ends $(\alpha = 0, T_g = 1)$ and $(\alpha \to \infty, T_g \to \sqrt{\alpha})$, but moves downwards in its central section as w decreases. The higher w and the lower the dimensionality of connectivity the more we expect the phase boundary to be similar to the fully connected model. Similarly the more random the connectivity and the lower w the more the phase boundary will be like the randomly connected $w = 0$ model. All possible infinite range connection architectures will produce a family of spin glass phase boundaries that lie between the two extremes of fully connected and the $w = 0$ randomly connected model (see figure 3.10).

In the co-existence part of the phase diagram of the network, the memory states give the system content addressable storage but the spin glass states contain no information. The number of spin glass states grows exponentially with the system size but the number of memory states only grows linearly. As w decreases, the co-existence phase increases in size as does the storage capacity per connection. In the limit of random connectivity and $w = 0$ the storage area of the phase diagram increases to about ten times the size of the equivalent area in the fully connected model. This seemingly large increase in storage is offset by the fact that the accuracy of storage m decreases at the phase boundary as $w \rightarrow 0$. The phase boundary T_m is first order for all connection architectures except for the randomly connected model in the limit $w \rightarrow 0$ where it continuously approaches a second order phase boundary. The point $(\alpha = 0, T_m = 1)$ is second order for all connection architectures and the spin glass phase boundary and the memory phase boundary always meet at this point. We also expect the curvature of the memory phase boundary to only become discontinuous in the limit $w \rightarrow 0$ like the spin glass phase boundary. For other connection architectures the phase boundary will also be most similar to the randomly connected $w = 0$ model the higher the dimensionality of connectivity and the lower w is (see figure 3.10).

In this section we have derived some phase boundaries for different connection architectures and the full replica-symmetric phase diagram for the $w = 0$ randomly connected model. In deriving the spin glass phase boundary we have always assumed that even though the spin glass phase has broken replica symmetry, the phase boundary predicted by it is always in the corrected place. We have argued this only by analogy with spin glasses. In the case of the spin glass, Parisi [41,42,43,44] calculated a full replica solution for the spin glass phase which shows a continuous divergence from the replica symmetric theory as the replica broken spin glass phase is entered. This calculation therefore explicitly showed the equivalence of the spin glass phase boundaries predicted by the two theories. In the next section we will derive the spin glass phase boundary from replica theory to show the equivalence of the replica and replica-symmetric theories at the phase boundary. This can be done by looking at the stability of the paramagnetic phase which becomes unstable at the spin glass phase boundary.

3.6 The Stability of the Paramagnetic Phase in Replica Theory

To study the stability of replica solutions would normally be extremely complicated but in the case of the paramagnetic phase all the replica order parameters are zero ($m^\rho = q^{\rho\sigma} = r^{\rho\sigma} = 0$). Therefore, all the off-diagonal spin averages in the second derivatives of the free energy are also zero and only the diagonal terms have to be evaluated. The expressions for all the non-zero second derivatives (see equation 2.70) where we are only considering states with a single macroscopic overlap, are therefore,

$$\begin{aligned}
\frac{n}{w}\frac{\partial^2 f}{\partial(m^\rho)^2} &= 1-\beta \\
\frac{n}{w}\frac{\partial^2 f}{\partial(q^{\rho\sigma})^2} &= -\frac{\beta^3\alpha}{w}\mathop{\mathrm{Tr}}_{ij}(\mathbf{r}^{\rho\rho}\mathbf{r}^{\sigma\sigma}) \quad \rho<\sigma \\
\frac{n}{w}\frac{\partial^2 f}{\partial(r^{\rho\sigma})^2} &= -\beta^3\alpha^2 \quad \rho<\sigma \\
\frac{n}{w}\frac{\partial^2 f}{\partial q^{\rho\sigma}\partial r^{\rho\sigma}} &= \alpha\beta \quad \rho<\sigma
\end{aligned} \tag{3.58}$$

The evaluation of the trace in the second of these derivatives is similar to the replica theory case (see Appendix B), except that now, $q^{\rho\sigma} = 0, \;\; \rho \neq \sigma$ instead of $q^{\rho\sigma} = q$. So setting $\mathbf{r} = 0$ and $q = 0$ in equation B.12 gives,

$$\mathop{\mathrm{Tr}}_{ij}(\mathbf{r}^{\rho\rho}\mathbf{r}^{\sigma\sigma}) = \left(\frac{\mathbf{D}}{\beta N}\right)^2 \left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right)^{-2} \tag{3.59}$$

Now, assuming $\beta < 1$, we can expand this expression in β and we find that, evaluating the full second order fluctuation in the free energy, the line which signifies the instability of the paramagnetic phase at low temperatures is the same as that predicted by replica-symmetric theory. We also find that the eigenvalue which becomes negative first is the one that controls the fluctuations in $q^{\rho\sigma}$ so the instability signifies the onset of a spin glass phase with $q^{\rho\sigma}$ finite. Since this phase boundary always occurs at $\beta < 1$ the expansion in β is always valid. For simplicity we will show the equivalence of the two theories for the randomly connected network where the trace term in the derivatives can be evaluated explicitly giving,

$$\frac{n}{w}\frac{\partial^2 f}{\partial(q^{\rho\sigma})^2} = -\alpha\beta\left[1 + w\left(\frac{1}{(1-\beta)^2} - 1\right)\right] \tag{3.60}$$

This gives for the second order fluctuations in the free energy for the paramagnetic phase,

$$\begin{aligned}\frac{n}{w}\delta^2 f &= (1-\beta)\sum_\rho(\delta m^\rho)^2 - \beta^3\alpha^2\sum_{\rho<\sigma}(\delta r^{\rho\sigma})^2 \qquad (3.61)\\ &\quad -\alpha\beta\left[1+w\left(\frac{1}{(1-\beta)^2}-1\right)\right]\sum_{\rho<\sigma}(\delta q^{\rho\sigma})^2 + 2\alpha\beta\sum_{\rho<\sigma}\delta q^{\rho\sigma}\delta r^{\rho\sigma}\end{aligned}$$

The fluctuations in $r^{\rho\sigma}$ run along a contour in the direction of the imaginary axis and we can shift this contour in order to make it run through the saddle point. The shift necessary is,

$$\delta r^{\rho\sigma} = \frac{1}{\alpha\beta^2}(\delta q^{\rho\sigma} + i\delta s^{\rho\sigma}) \qquad (3.62)$$

This gives for the second order fluctuations in the free energy,

$$\begin{aligned}\frac{n}{w}\delta^2 f &= (1-\beta)\sum_\rho(\delta m^\rho)^2 + \frac{1}{\beta}\sum_{\rho<\sigma}(\delta s^{\rho\sigma})^2\\ &\quad + \left(\frac{1}{\beta} - \alpha\beta\left[1+w\left(\frac{1}{(1-\beta)^2}-1\right)\right]\right)\sum_{\rho<\sigma}(\delta q^{\rho\sigma})^2 \qquad (3.63)\end{aligned}$$

When any of the three eigenvalues which factor the fluctuations become negative this signifies instability of the paramagnetic phase and the system is entering a new ordered phase. The point at which the first eigenvalue becomes zero gives a phase boundary. For any finite value of w as T is decreased the eigenvalue in front of the $q^{\rho\sigma}$ fluctuations term becomes negative first which signifies the onset of the spin glass phase. The spin glass phase boundary is therefore given by solutions of,

$$\frac{1}{\beta} - \alpha\beta\left[1+w\left(\frac{1}{(1-\beta)^2}-1\right)\right] = 0 \qquad (3.64)$$

This expression is exact and signifies the onset of the spin glass phase for the system of n replicas. Unlike the replica-symmetric equations we have made no assumptions about the form of the solutions within the phase. This expression is exactly the same as the one we derived for the spin glass phase boundary in replica-symmetric theory by solving the order parameter equations to first order in q (see equation 3.32). This calculation therefore shows the equivalence of the two theories in determining the spin glass phase boundary. From equation 3.63 we can gain no information about what happens below the spin glass phase boundary as this depends on what other states become available to the system.

For example, as we have seen from the replica theory, at lower values of temperature we enter a co-existence phase with memory states as well as spin glass states. We can also gain no information from equation 3.63 about the form of the order parameters across the phase boundary. In the limit $w \rightarrow 0$ as we have seen in section 3.5 the solution of equation 3.64 gives $T_g = 1$. This means that the eigenvalue in front of the m^ρ fluctuations becomes negative at the same value of temperature as the $q^{\rho\sigma}$ fluctuations' eigenvalue. We therefore, as we would expect, see replica theory predicting the coincident memory and spin glass phase boundary for the $w = 0$ randomly connected model in the region $\alpha < 1$.

3.7 A Comparison of the SK Spin Glass and the Randomly Connected $w = 0$ Model

Now that we have derived and solved the order parameter equations for different neural network architectures we are in a position to see the similarities between neural networks, particularly the $w = 0$ model, and the SK spin glass [27]. The SK spin glass was described with references in section 1.4. The interactions J_{ij} for the SK model are chosen from a Gaussian distribution with first and second moments given by $\frac{J_o}{N}$ and $\frac{J^2}{N}$. The replica-symmetric theory for the spin glass only has two order parameters: the magnitization m and the EA spin glass order parameter q. The two replica-symmetric order parameter equations for the SK model are given by,

$$\begin{aligned} m &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(Jq^{\frac{1}{2}}z + J_o m) \\ q &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2 \beta(Jq^{\frac{1}{2}}z + J_o m) \end{aligned} \qquad (3.65)$$

The expression in brackets $(Jq^{\frac{1}{2}}z + J_o m)$, is called the local field term. The second part $J_o m$, is called the ferromagnetic term since it is responsible for the ferromagnetic behaviour of the spin glass. With $J = 0$ the first order parameter equation reduces to the order parameter equation for the infinite range Ising ferromagnet (see equation 1.34). The term $Jq^{\frac{1}{2}}z$ is responsible for the spin glass behaviour of the system as would be expected since J measures the standard deviation of the interactions. It is the fluctuations in the values of the J_{ij}'s, as we have seen in section 1.6, that causes the spin glass behaviour. With $J_o = 0$ the system exhibits no ferromagnetic behaviour.

The overlap m in neural networks plays a very similar role to the magnitization in spin glasses. The main difference between neural networks and spin glasses is the existence of the extra order parameter r which plays the same role as q in the order parameter equations for m and q (see equation 3.1). Thus the local field for a neural network consists of two parts; a memory part m resulting from the single condensed overlap, and a spin glass part $\sqrt{r\alpha}z$, generated by the random overlaps with the rest of the patterns. The interactions for a neural network have,

$$[T_{ij}^2]_{av} - [T_{ij}]_{av}^2 = \frac{\alpha}{N} \tag{3.66}$$

so $\sqrt{\alpha}$ is the normalized standard deviation for neural network interactions and plays the same role as J does for spin glasses. The SK model with $J_o = 1$ is closest to neural network models and its phase diagram (see figure 3.11) is presented in a similar form to the neural network phase diagrams in figure 3.10.

The reason why a neural network behaves similarly to a spin glass with $J_o = 1$ is quite easy to understand. We will work with the fully connected model for simplicity. If we nominate one state for condensation $\{\xi_i^s\}$ then the connection strengths T_{ij} can be broken up into two terms giving,

$$T_{ij} = \frac{\xi_i^s \xi_j^s}{N} + \frac{1}{N} \sum_{\mu \neq s}^{p} \xi_i^\mu \xi_j^\mu \quad i \neq j \tag{3.67}$$

If we now consider a single site k, then all the connections into that site have the same first order term of size $\frac{1}{N}$ which can align the state of the system at that site with the nominated pattern. Unlike the spin glass, the sign of the aligning term is local to the site so for a given site k we have,

$$[T_{ik}]_{av} = \pm\frac{1}{N} \tag{3.68}$$

where the average $[\ ]_{av}$ is only over sites i. If the average had been over all sites then the mean value would just have been zero.

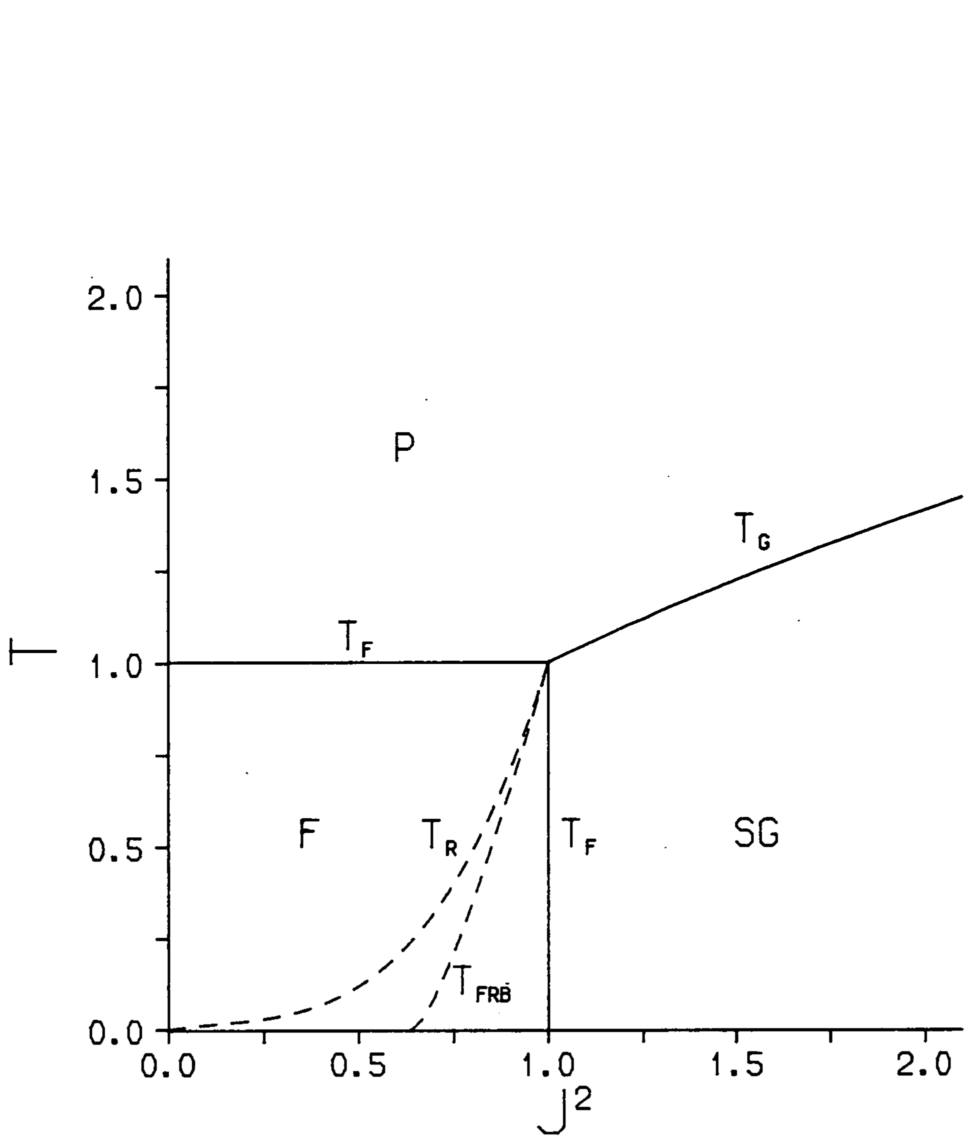

Figure 3.11: Phase diagram for the SK spin glass. P,F and SG stand for paramagnetic, ferromagnetic and spin glass phase. T_R is the replica symmetry breaking line and T_{FRB} is the ferromagnetic phase boundary predicted by replica-symmetric theory. T_F is the true ferromagnetic phase boundary predicted by Parisi from replica broken calculations.

The phase diagram for the SK spin glass (see figure 3.11), is very similar to the phase diagram for the $w = 0$ randomly connected model. The only difference is that the co-existence phase in the neural network model is a pure ferromagnetic phase in the SK spin glass. This difference can be understood by studying the order parameter equations for the randomly connected neural network which are,

$$\begin{aligned} m &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(\sqrt{\alpha r} z + m) \\ q &= \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2 \beta(\sqrt{\alpha r} z + m) \\ r &= q\left[1 + w\left(\frac{1}{(1-C)^2} - 1\right)\right] \end{aligned} \tag{3.69}$$

In the case of the spin glass phase to the left of the co-existence phase boundary, the limit $w \to 0$ is not well defined on the term $\frac{w}{(1-C)^2}$ away from the spin glass phase boundary. In section 3.4 equation 3.38, we found that the limit $w \to 0$ on the term gives,

$$\lim_{w \to 0} \frac{w}{(1-C)^2} = \frac{1}{\alpha} - 1 \tag{3.70}$$

at the spin glass phase boundary. Away from the phase boundary the limit must be derived numerically. In all cases though, the limit does produce a finite value which means $r \neq q$ and the model does not behave the same as a spin glass. This also means that $C = 1$ in the co-existence part of the spin glass phase for the $w = 0$ randomly connected model. Therefore the value of q in replica theory is given by,

$$q = 1 - T \tag{3.71}$$

in the co-existence part of the phase diagram. This therefore accounts for the larger spin glass phase for a neural network which extends down to $\alpha = 0$.

In the case of the memory phase the limit $w \to 0$ is well defined on the term $\frac{w}{(1-C)^2}$, and always gives zero. We have already seen this analytically at zero temperature in section 3.2 of this chapter where $1 - C$ is of the order $\sqrt[3]{w}$ on the phase boundary (see equation 3.26) and so the limit $w \to 0$ on the term $\frac{w}{(1-C)^2}$, gives zero. This gives $r = q$ and hence we only have two order parameter equations in m and q to solve which are identical to the SK spin glass order parameter equations (see equation 3.65). The replica-symmetric memory phase of the $w = 0$ model is therefore the same shape as the replica-symmetric ferromagnetic phase of the SK model. On the co-existence phase boundary the memory solutions turn continuously into spin glass solutions therefore, the parameters of the two different types of solution must be the same at this line. This explains why $C = 1$ on this line for both types of solution in the limit $w \to 0$, although they approach this limit in different ways.

The replica-symmetric stability condition equation 3.50, can also be calculated in this limit and it reduces to the same form as the stability condition for the SK model [37], and so the replica symmetry breaking line is in the same place for both models. The replica symmetric free energy for the memory phase can also be calculated from equation 2.65 where only the first term in the series is non-zero. This gives the free energy per site for the state associated with a single condensed pattern as,

$$\begin{aligned} \frac{f}{w} &= \frac{1}{2}m^2 - \frac{\alpha\beta(1-q)^2}{4} \\ &\quad -\frac{1}{\beta}\frac{dz}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\ln[2\cosh\beta(\sqrt{\alpha q}z + m)] \end{aligned} \qquad (3.72)$$

which is exactly the same expression as the replica-symmetric free energy per site for the SK spin glass [27]. Thus the behaviour of the memory phase for a single macroscopic overlap appears to be identical to the ferromagnetic phase of the SK model. We must remember though, that the neural network at low α also has memory phases having overlaps with more than one of the patterns nominated for storage. It is only the particular case of the single overlap memory phase where the model behaves the same as the ferromagnetic phase of the SK model. Since the replica symmetry breaking line for both models is in the same place we may expect the true behaviour of the neural network in the replica broken part of the memory phase to be the same as the SK spin glass. Parisi's [41,42,43,44] replica broken solution for the ferromagnetic phase boundary of the SK spin glass is believed to be correct and predicts a vertical line from $(T_m = 1, J = 1)$ to $(T_m = 0, J = 1)$ (see figure 3.11). We can therefore, by analogy, draw in the phase boundary for the $w = 0$ model in the same place giving the phase diagram shown in figure 3.12.

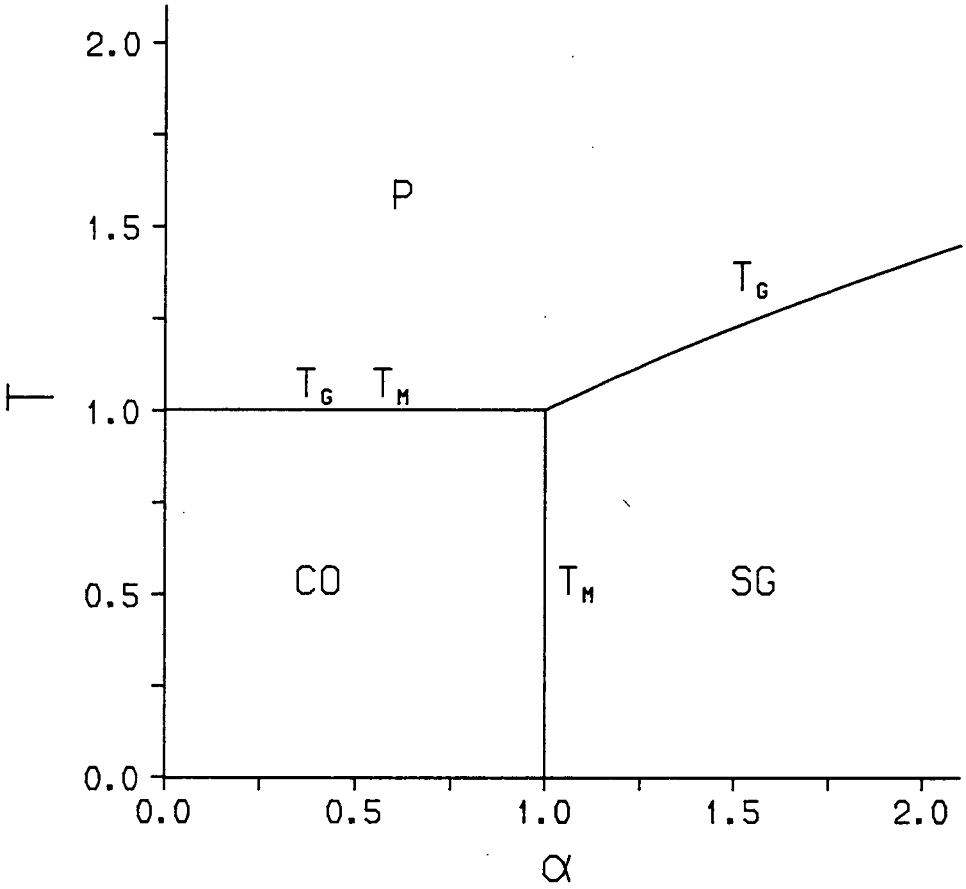

Figure 3.12: Expected true phase diagram for the randomly connected $w = 0$ Hopfield model by analogy with the SK spin glass. P,SG and CO are the paramagnetic, spin glass and co-existence phases. T_M and T_G are the memory and spin glass phase boundaries.

Chapter 4

Numerical Results

4.1 Introduction

The most important result from the previous chapter is the existence of a critical storage ratio α_c above which there are no stable memory states. The phase transition at α_c is always of first order except in the limit of the randomly connected $w = 0$ model and the maximum value of α_c occurs at zero temperature where replica symmetry is broken. Therefore the theoretical results for the maximum value of α_c are expected to be in error with the size of the error increasing the further the critical point is from the replica symmetry breaking line. In this chapter we are going to study this phase transition for for a fully connected and one dimensionally connected system by numerical simulations on the DAP computer which is well suited to the study of boolean systems. The DAP's architecture, programming languages and some of the programming techniques used for the simulations in this chapter are discussed in Appendix D.

4.2 Finite Size Scaling of First Order Phase Transitions

The theoretical concept of a first order phase transition requires the thermodynamic limit $N \to \infty$ to be taken so, how do we expect the critical parameters to behave in a finite system ? If we consider some numerically measurable parameter X, of a finite system, its value will change continuously as we allow the external parameters of the system to move through their critical values. If we then simulate bigger and bigger systems then we may expect the value of X to change more rapidly as we move through the critical values of the external parameters eventually reaching a discontinuity only in the limit of an infinite system. Therefore by studying different system sizes we may hope to extrapolate to the case of an infinite system and so determine the critical values of the external parameters of the system. A typical function which captures the expected main features of a first order transition in a finite size system is,

$$X = A \exp B(\alpha_c - \alpha)N \tag{4.1}$$

where α is the value of the external parameter and α_c is its critical value. N is the size of the system and A and B are constants whose values will depend on the type of system under study. The fitting of numerical data to functions of the type given in equation 4.1 to determine critical values of parameters is termed finite size scaling.

4.3 Numerical Studies at α close to α_c

The numerical studies in this section are of a very similar nature to those carried out by Amit *et al* [6] and Bruce *et al* [12]. The majority of the simulations of a fully connected system were carried out prior to the publication of Amit *et al*'s work on the same subject and our results are mainly in agreement with his.

Numerical simulations were performed on systems of size 1024,2048,3072 and 4096 for a fully connected and a one dimensionally connected network. The simulations were carried out at values of α close to the zero temperature critical values α_c calculated from replica-symmetric theory (see section 3.2). For each set of values; N and α, 10 different sets of simulations were carried out for the 4096 network increasing to 20 for the smallest size network. In each set of simulations for the larger size systems 128 of the patterns nominated for storage were iterated to stability but for smaller systems, where the number of patterns was less than 128, all of them were iterated. From now on, for simplicity, we will refer to the patterns nominated for storage as the learnt patterns although this does not necessarily mean that they or patterns closely associated with them are stored in the network. We will refer the stable states, which are closely associated with the learnt patterns, as memory states.

Starting from the learnt pattern the network was updated by serial single site update, using the update algorithm of equation 1.2, until a stable state was reached. The distribution of the m values for the overlaps between the initial learnt states and the final stable states was typically found to be of the form shown in figure 4.1. The results of other system sizes being very similar to Amit *et al's* results [6]. The distribution has two peaks: one close to $m = 1$ corresponding to patterns closely associated with the learnt patterns being stable and another peak at about $m = 0.35$. As α increases the weight of the first peak is transferred to the second peak. We will assume that the iterated learnt states which form the second peak at $m = 0.35$ mean that there are no stable states closely associated with these learnt states. Since the values of m for the memory states are close to one we think it is unlikely that a vector starting at a learnt state will not be trapped in the associated memory state's basin of attraction if one exists. A more detailed discussion of the possible discrepancies between the theoretical results and a simulation of this type are given in Bruce *et al* [12]. Gardner [16] showed that there are other stable states clustered around the memory states which, though higher in energy than the memory state, are still stable to single spin flip dynamics. There also exists an exponentially large number of spin glass states so there will always be spin glass states which have a finite overlap with any of the learnt patterns. We expect the peak at $m = 0.35$ to be caused by the iterative scheme terminating at either of these types of states.

In figure 4.1 the values of $-\ln P$, where P is the weight under the high m peak

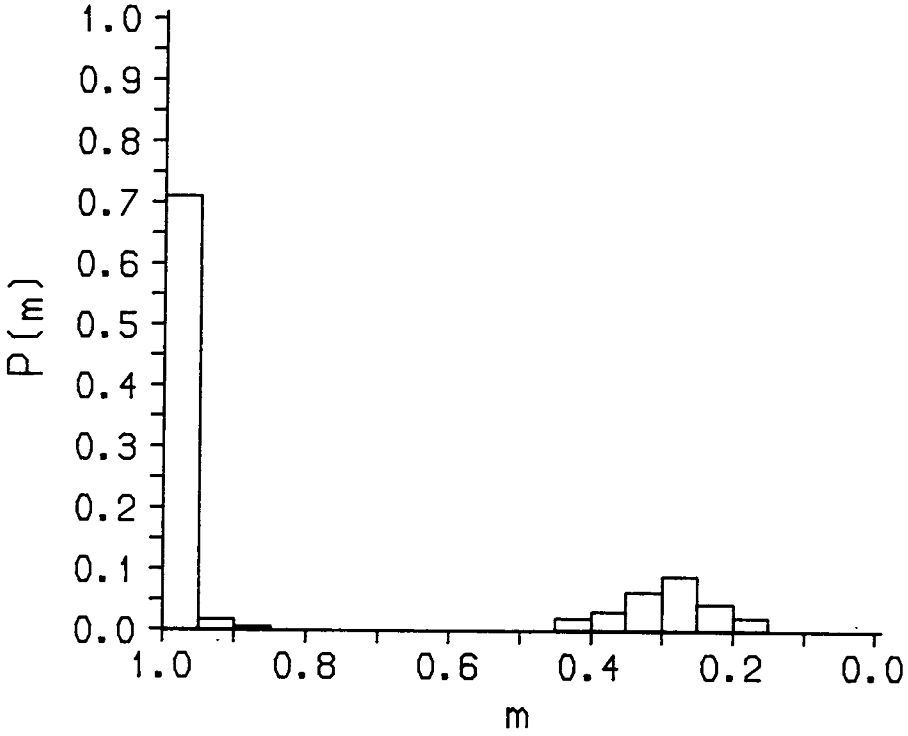

Figure 4.1: Histogram of the overlap of the retrieval state with the initial state for $N = 4096, \alpha = 0.1465$.

are plotted against N for different values of α. For an infinite size system replica theory predicts that P should change discontinuously at α_c so we hope to use this parameter for finite size scaling. In all our simulations the two peaks in the m distribution (see figure 4.1), were well separated so there was no ambiguity in selecting which iterated states belonged to which peaks. For $\alpha = 0.1367$ the value of P increased as N increased, within the error bars, while for $\alpha = 0.1465$ it decreased implying that $0.1367 < \alpha_c < 0.1465$. The results for α larger than α_c showed a much sharper change in the value of P as N increased than the results for α lower than α_c. This is partly because, in the system sizes we studied, the change over from positive to negative gradient occurred at values of P close to one. Therefore possible increases in the value of P as N increased were restricted to a very small range of values. This also meant that even small errors in P could mask the scaling properties of the system for values of α below α_c. The scaling of P with the system size is therefore much more distinguishable in the two higher values of α (see figure 4.2). The two sets of points for these two α values were found to fit extremely well to the exponential form in equation 4.1.

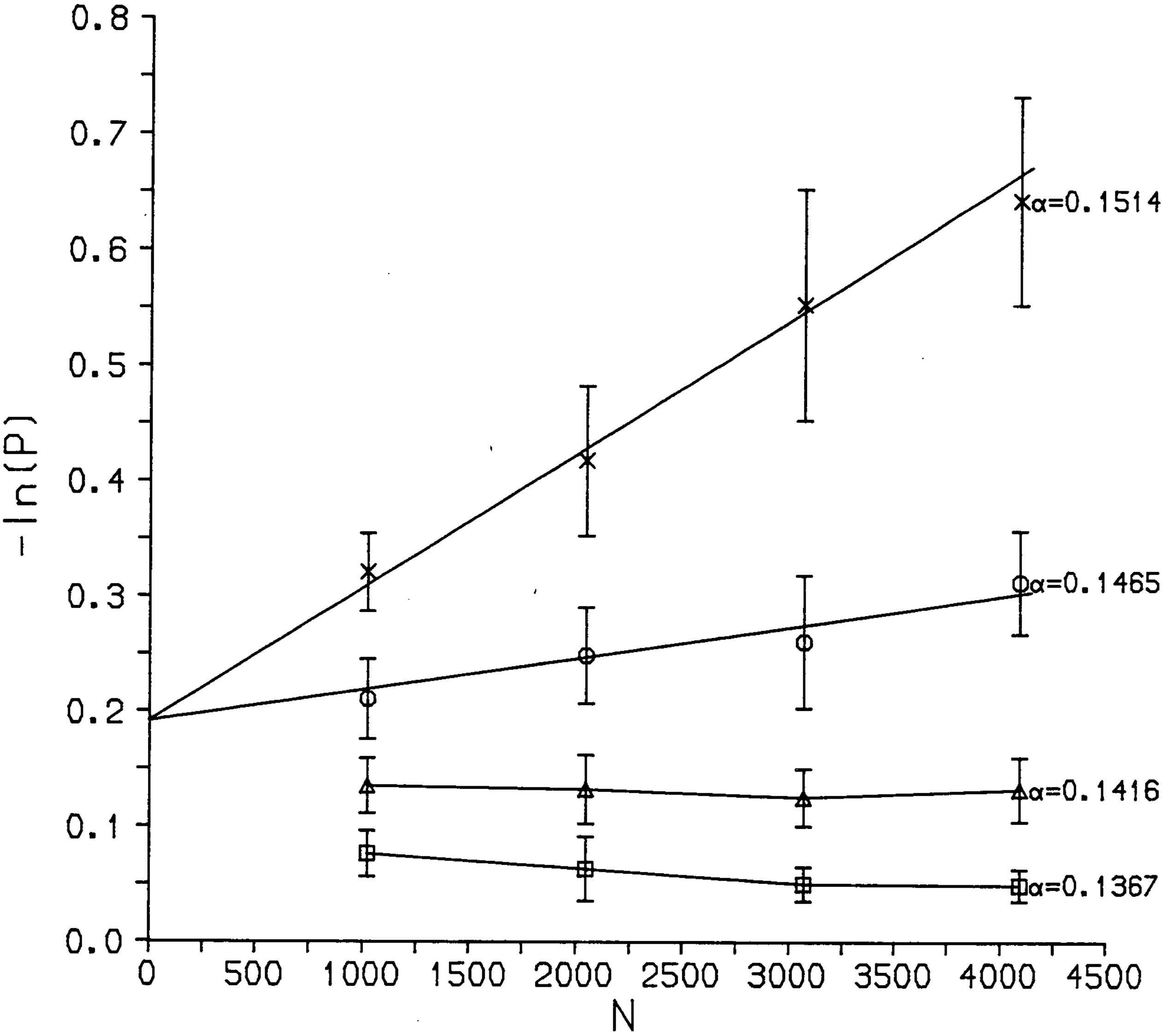

Figure 4.2: The logarithm of the weight of the peak close to $m = 1$, for various α values, plotted against N, for a fully connected network. The two upper sets of points are fitted to equation 4.1 using weighted linear regression.

Taking logs of equation 4.1 we obtain,

$$-\ln P = -\ln A + B(\alpha - \alpha_c)N \tag{4.2}$$

Putting $p = \alpha N$ we obtain,

$$-\ln P = -\ln A + Bp - B\alpha_c N \tag{4.3}$$

which reduces the expression for P to linear form in p and N. We can now perform a weighted, multiple linear regression, which corresponds to fitting our data points at $\alpha = 0.1514$ and 0.1465 to a plane. The values of A, B and α_c obtained from this fit were,

$$\alpha_c = 0.1450 \pm 0.0003, \quad A = 0.836 \pm 0.009, \quad B = 0.0181 \pm 0.0007 \tag{4.4}$$

where α_c is calculated from the coefficients of p and N. The errors in these coefficients are not independent and the relative error in α_c was found to be much smaller than either of the relative errors in the two coefficients.

Even within the error bars the value of α_c is not in agreement with the theoretically predicted value of $\alpha_c = 0.138$. A possible explanation of this discrepancy is the effect of replica symmetry breaking. Crisanti *et al* [11] determined α_c from a one step replica symmetry breaking calculation and found it to be 0.145 although how much breaking symmetry once approximates the true solution is very difficult to estimate. A full replica solution would require all the replica order parameters to take continuous values rather than just two possible values. This calculation does however suggest that the effect of replica symmetry breaking is to increase α_c. Amit *et al* [6] obtained, by numerical simulations, $\alpha_c = 0.145 \pm 0.01$ in close agreement with our result. There calculations were performed on six systems of size 500 to 3000 although only five sets of 100 patterns were iterated for each value of α compared to our simulations of typically 15 sets of 128 patterns.

The value of m at $\alpha = 0.138$ was found to be, for $N = 4096$ (averaged over 256 patterns), $m = 0.978 \pm 0.008$ and within the error bars remained unchanged for lower values of N. Again the value is higher than the theoretical value of $m_c = 0.968$ suggesting that replica symmetry breaking also has the effect of increasing m. Crisanti *et al* [11] also found with their replica broken calculation that the value of m increased.

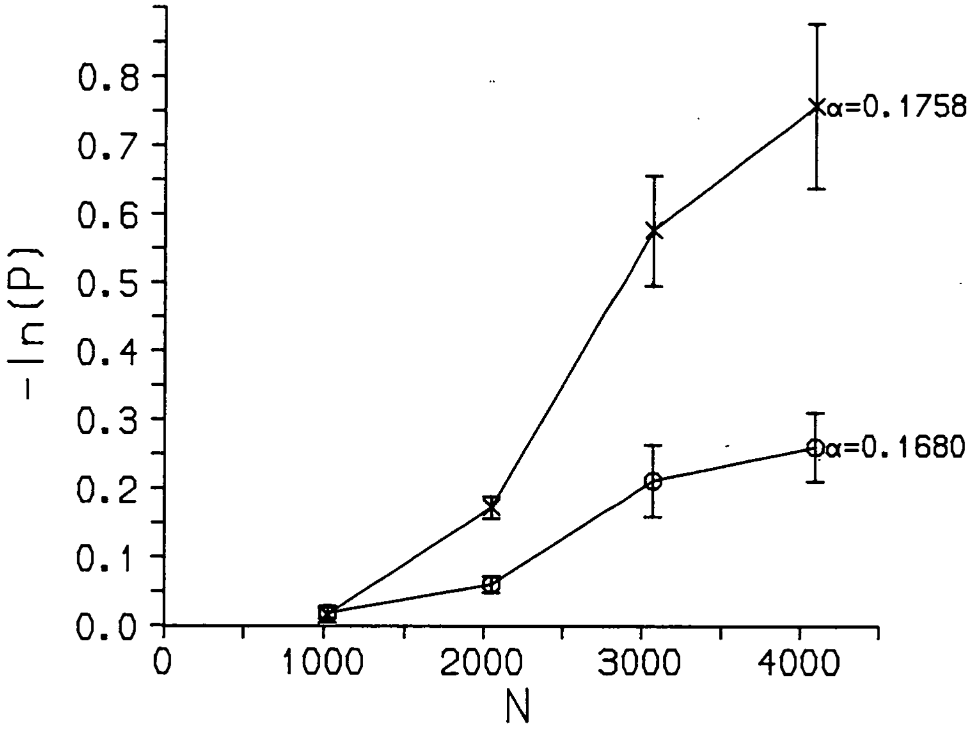

Figure 4.3: The logarithm of the weight of the peak close to $m = 1$, for various α values, plotted against N, for a one dimensionally connected network with $w = 0.494$.

A similar set of simulations to those carried out for the fully connected model were performed on a system with a one dimensional connectivity architecture. The simulations were carried out with $w = 0.494$ at two values of α (0.1680,0.1758), both above the critical value $\alpha_c = 0.154$ predicted by replica-symmetric theory (see section 2.2). The results of these simulations are shown in figure 4.3. The most significant result from these numerical studies is the increase in the storage capacity per connection over the fully connected system. For example, in the one dimensional system with $\alpha = 0.168, N = 4096$, we find $P = 0.771$ while for the fully connected system at $\alpha = 0.1514, N = 4096$, we find $P = 0.526$. This backs up our theoretical calculations in Chapter 3 which also showed an increase in the storage capacity per connection for partially connected systems. The results in figure 4.3 do not fit the scaling form of equation 4.1. We find a similar scaling to that found by Bruce *et al* [12] for the V model network where the values of P increase much more rapidly as N decreases than the scaling form of equation 4.1 predicts. A possible explanation of this is a finite size effect that gives increased storage capacity in systems with fewer connections and becomes particularly significant for systems with fewer than 1000 connections. Therefore small systems and partially connected systems would have improved storage as N decreases. The simulations in the one dimensionally connected system have about half the number of connections as the fully connected model which could explain why the finite size effect is much more prevalent in these simulations. Simulations by Bruce *et al* were only carried out on one system size greater than 1024 which could also explain why their results did not fit well to the form of equation 4.1. To really understand the finite size size effects present in the system it would be necessary to perform detailed simulations on a large range of system sizes which would be very demanding on computer time. To obtain the numerical results in this chapter about 200 hours of computer time was used. The simulations on system sizes of 3072 and 4096 (see figure 4.3) show signs of fitting the form of equation 4.1. A least squares fit on these four points gave $\alpha_c = 0.165 \pm 0.012$ which again is higher than the theoretical value of $\alpha_c = 0.1514$. The value of m at $\alpha = 0.138$ was found to be 0.988 ± 0.007 which is higher than the value obtained for the fully connected system. Again this gives some support to our theoretical predictions that a partially connected system has a higher value of m for a given α than a fully connected system.

Conclusions and Discussion

The main result of the work in this thesis is that partially connected versions of Hopfield networks can store more patterns per connection than a fully connected network. However, our calculations were restricted to models of infinite size, with an infinite number of connections and having the same connection architecture at each site. The increased storage is reflected in the increased size of the memory phase (which co-exists with the spin glass phase) and hence α_c as the network's connectivity is diluted. The increase in α_c is partly offset by the decrease in overlap m_c at the phase boundary but, for a given value of α, m is always higher for a partially connected network.

The specific position of the phase boundaries for both the co-existence phase and the spin glass phase are controlled by the probabilities associated with different sized sets of neurons being connected in closed loops. The shorter loops being the most important in determining the thermodynamic properties of the system. The two limiting cases of the probability values associated with the loops are the fully connected model, which has all the probabilities equal to one, and the randomly connected $w = 0$ model, which has all the probabilities equal to zero except the two site loop. The probability associated with this loop is always one due to the symmetric choice of the interactions. All other infinite range connection architectures were found to lie between these two cases. We therefore studied the randomly connected model in detail in section 3.4 where we found that the memory phase boundary becomes second order in the limit $w \rightarrow 0$. In Chapter 3 we also outlined a numerical method using random walks by which the phase diagram for any infinite range connection architecture could be determined.

All the phase diagram calculations in this work were carried out within the framework of replica-symmetric theory. However, in section 3.5 we determined in which areas of the phase diagram replica symmetry breaking is present corresponding to the breakdown of replica-symmetric theory. Replica symmetry breaking breaking was found to play an increasingly important role the higher the dimensionality of the connectivity and the lower the value of w. In the case of the randomly connected $w = 0$ model the whole of the memory phase boundary predicted by replica-symmetric theory lay in a replica broken area and was therefore incorrect. By analogy to the SK spin-glass we then supposed that the effect of replica symmetry breaking is to change the phase diagram to the form in figure 3.12.

In Chapter 4 we performed some numerical simulations on finite systems that in general agreed with our theoretical predictions for infinite systems. These results also suggested that the true values of α_c and m_c in the one dimensionally connected and fully connected model were slightly above our replica-symmetric theoretical values. We expect these discrepancies to be due to replica symmetry breaking which our replica symmetry breaking studies in section 3.5 suggested, had a smaller effect on systems with a lower dimensionality of connectivity and higher w value.

If we consider the resources requirements for Hopfield neural networks we can see that a partially connected system requires many more neural units, for the same number of connections, to have significantly more storage than a fully connected system. It is only when we consider restrictions of space and communication times that the major advantages of a partially connected system can be seen. In the "neural chips" which have been built so far at Bell laboratories [52] and also in the brain the neural units occupy negligible space compared to the connections. Therefore a partially connected network, particularly with some form of local connectivity, would be the most efficient use of space, reduce communication times and increase storage capacity per connection as well.

There are many other possible areas of research in partially connected networks which have as yet not been studied. Firstly the differences in size of basins of attraction for different architectures could be studied by a similar method to that followed by Forrest [19] for a fully connected model. We may expect the basins of attraction for the stored states to be larger due to the less crowded nature of the phase space. Fewer states are stored in the same size of phase space for a partially connected system than a fully connected system with the same number of nodes. The extent to which these results extend to Hopfield type networks with other learning algorithms (see [17,47,19]) which improve on the basic Hebb rule used in this paper could also be studied. Compared with complete connectivity random dilution Gardner [18] has found improved storage per connection for the perceptron learning algorithms of Gardner [17], Krauth [46] and Forrest [19]. The ability of partially connected networks to store information with short range correlations would also be worth investigating particularly if the connection range is chosen to be of a similar range to the correlations. Detailed studies of other types of neural network models could determine whether the results presented in this paper are valid beyond Hopfield networks. Do all partially connected systems have improved properties if more neural units are used with the same numbers of connections ?

Appendix A

The Kronecker Product of Two Matrices.

Let $\mathbf{A}$ be a square matrix of dimension c and $\mathbf{B}$ be a square matrix of dimension k. The Kronecker product of theses matrices is a matrix of dimension ck and can be expressed in block form as,

$$\mathbf{A} \times \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & a_{13}\mathbf{B} & \cdots & a_{1c}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & & \vdots \\ a_{31}\mathbf{B} & \vdots & \ddots & & \\ \vdots & & & & \vdots \\ a_{c1}\mathbf{B} & \cdots & & \cdots & a_{cc}\mathbf{B} \end{pmatrix} \tag{A.1}$$

Some important properties of the Kronecker product which are used in Chapter 2 are,

$$\begin{aligned} \underset{ab,mn}{\mathrm{Tr}}\ (\mathbf{A} \times \mathbf{B}) &= \underset{ab}{\mathrm{Tr}}(\mathbf{A})\ \underset{mn}{\mathrm{Tr}}(\mathbf{B}) \\ \underset{mn}{\mathrm{Tr}}(\mathbf{A} \times \mathbf{B}) &= \mathbf{A}\,\underset{mn}{\mathrm{Tr}}(\mathbf{B}) \end{aligned} \tag{A.2}$$

where a and b are indices of matrix $\mathbf{A}$ and m and n are indices of matrix $\mathbf{B}$. We can also define the determinant or cofactor, with respect to the first two indices, of a blocked matrix formed from Kronecker products of matrices of the same form as $\mathbf{A}$ and $\mathbf{B}$. This will be a matrix of the same dimension as the second matrices in the products (see Appendix B). In the case of the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ we obtain,

$$|\mathbf{A} \times \mathbf{B}|_{ab} = |\mathbf{A}|\mathbf{B} \tag{A.3}$$

which is clearly a matrix of dimension k.

Appendix B

Calculation of $\mathbf{r}$, $\mathbf{r}_d$ and $(\mathbf{r}-\mathbf{r}_d)^2$.

We wish to calculate $\mathbf{r}^{\rho\sigma}$ in replica symmetric theory where we have from equation 2.67,

$$\mathbf{r}^{\rho\sigma} = \frac{1}{\beta}\left(\frac{\frac{1}{N}\mathbf{D}\mathbf{Q}^{\rho\sigma}}{\left|\mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q}\times\mathbf{D}\right|_{\rho\sigma}}\right) \quad \forall\rho,\sigma \tag{B.1}$$

The first step of the replica theory is to set $q^{\rho\sigma} = q$, this gives us for $\mathbf{Q}$, remembering that $q^{\rho\rho} = 1$ still holds,

$$\mathbf{Q} = \mathbf{I}_{nN} - \frac{\beta}{wN}\mathbf{q}\times\mathbf{D} = \left(\begin{array}{ccccc} \mathbf{Y} & \mathbf{X} & \mathbf{X} & \cdots & \mathbf{X} \\ \mathbf{X} & \mathbf{Y} & \cdots & & \vdots \\ \mathbf{X} & \vdots & \ddots & & \\ \vdots & & & & \vdots \\ \mathbf{X} & \cdots & & \cdots & \mathbf{Y} \end{array}\right) \tag{B.2}$$

where,

$$\begin{aligned} \mathbf{Y} &= \mathbf{I}_N - \frac{\beta}{wN}\mathbf{D} \\ \mathbf{X} &= -\frac{\beta q}{wN}\mathbf{D} \end{aligned} \tag{B.3}$$

are matrices of dimension N. We have represented $\mathbf{Q}$ in a blocked structure of $n\times n$ matrices all of dimension N. We now wish to calculate the determinant of $\mathbf{Q}$ with respect to the replica indices ρ and σ. This can be done by treating $\mathbf{X}$ and $\mathbf{Y}$ like elements in an n dimensional matrix. The first step in factorising the determinant of $\mathbf{Q}$ is to add each row to the top row so that every element in the top row is $\mathbf{Y}+(n-1)\mathbf{X}$. If we now subtract column one from every other column and take out the factor $\mathbf{Y}+(n-1)\mathbf{X}$ from the top row we get,

$$|\mathbf{Q}|_{\rho\sigma} = (\mathbf{Y}+(n-1)\mathbf{X}) \begin{vmatrix} \mathbf{I}_N & \mathbf{0}_N & \mathbf{0}_N & \cdots & \mathbf{0}_N \\ \mathbf{X} & \mathbf{Y}-\mathbf{X} & \mathbf{0}_N & \cdots & \vdots \\ \mathbf{X} & \mathbf{0}_N & \mathbf{Y}-\mathbf{X} & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X} & \cdots & & \cdots & \mathbf{Y}-\mathbf{X} \end{vmatrix}_{\rho\sigma} \tag{B.4}$$

where $\mathbf{0}_N$ is the N dimensional zero matrix. We can now easily expand the determinant of $\mathbf{Q}$ about row one giving,

$$|\mathbf{Q}|_{\rho\sigma} = (\mathbf{Y}+(n-1)\mathbf{X})(\mathbf{Y}-\mathbf{X})^{n-1} \tag{B.5}$$

If we now carry out the second stage of the replica assumption $n \to 0$ the determinant becomes,

$$|\mathbf{Q}|_{\rho\sigma} = (\mathbf{Y}-\mathbf{X})(\mathbf{Y}-\mathbf{X})^{-1} = \mathbf{I}_N \tag{B.6}$$

So the required determinant is simply the identity matrix. To calculate $\mathbf{r}$ and $\mathbf{r}_d$ we also need to evaluate the cofactors of $\mathbf{Q}$ with respect to the replica indices. The first stage of the replica assumption $q^{\rho\sigma} = q$ reduces the number of distinct cofactors to two depending on whether it is a diagonal or off-diagonal cofactor. The diagonal cofactor will be required to evaluate $\mathbf{r}_d$ and the off-diagonal will give us $\mathbf{r}$. We will look at the off-diagonal cofactor first and for simplicity we will consider $\mathbf{Q}^{12}$ where,

$$\mathbf{Q}^{12} = - \begin{vmatrix} \mathbf{X} & \mathbf{X} & \mathbf{X} & \cdots & \mathbf{X} \\ \mathbf{X} & \mathbf{Y} & \cdots & \cdots & \vdots \\ \mathbf{X} & \mathbf{X} & \mathbf{Y} & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X} & \cdots & & \cdots & \mathbf{Y} \end{vmatrix}_{\rho\sigma} \tag{B.7}$$

If we now subtract column one from all the other columns and then expand in row one this gives,

$$\mathbf{Q}^{\rho\sigma} = -\mathbf{X}(\mathbf{Y}-\mathbf{X})^{n-2} \quad \rho \neq \sigma \tag{B.8}$$

So in the limit $n \to 0$ this gives for $\mathbf{r}$,(see equation B.1) using the result of equation B.6,

$$\mathbf{r} = -\frac{1}{\beta N}\mathbf{D}\mathbf{X}(\mathbf{Y}-\mathbf{X})^{-2} = \frac{q}{w}\left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q)\right)^{-2}\left(\frac{\mathbf{D}}{N}\right)^{2} \tag{B.9}$$

The cofactor of the diagonal terms is the determinant of a matrix which has exactly the same form as $\mathbf{Q}$ but of blocked dimension $n-1$ rather than n. The cofactor is therefore, from equation B.5,

$$\mathbf{Q}^{\rho\rho} = (\mathbf{Y} - (n-2)\mathbf{X})(\mathbf{Y} - \mathbf{X})^{n-2} \tag{B.10}$$

Taking the limit $n \rightarrow 0$ this gives,

$$\mathbf{r}_d = \frac{\mathbf{D}}{\beta} N \left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-2q) \right) \left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q) \right)^{-2} \tag{B.11}$$

for the diagonal term. In order to look at the stability conditions for replica theory it is necessary to evaluate the difference of the two terms squared. This is found to be,

$$(\mathbf{r} - \mathbf{r}_d)^2 = \left(\frac{\mathbf{D}}{\beta N} \right)^2 \left(\mathbf{I}_N - \frac{\beta}{wN}\mathbf{D}(1-q) \right)^{-2} \tag{B.12}$$

This is very closely related to $\mathbf{r}$ (equation B.9), giving,

$$(\mathbf{r} - \mathbf{r}_d)^2 = \frac{w\mathbf{r}}{q\beta^2} \tag{B.13}$$

Appendix C

Analytical Calculation of $a_1(w)$ for Hypercubic Connectivity Architectures

From equation 3.2 $a_1(w)$ is given by,

$$a_1(w) = \frac{1}{N^2 w^2} \sum_{jk} D_{ij} D_{jk} D_{ki} \tag{C.1}$$

where we are choosing site i to be fixed. Since the connection architecture is the same at every site there is no loss of generality in making this choice. $a_1(w)$ only has contributions from sites j and k which are connected to site i, and to each other. The probability that site j is in the connection space of site i is w therefore, the probability that site j and site k are in the connection space of i is w^2. This means that $a_1(w)$ is the probability that site j is connected to site k given that both sites are connected to site i. For simplicity we will first of all look at the one dimensional case. Fig C.1 is a symbolic representation of a one dimensional connectivity architecture.

Given that j and k are within the connection space of i there are two distinct situations possible for the positions of j and k each with probabilty $\frac{1}{2}$.

1) Both sites are on the same side of i.

In this case j is always connected to k and therefore every term contributes to $a_1(w)$ giving a $\frac{1}{2}$ contribution in total.

2) Each site is on opposite sides of i.

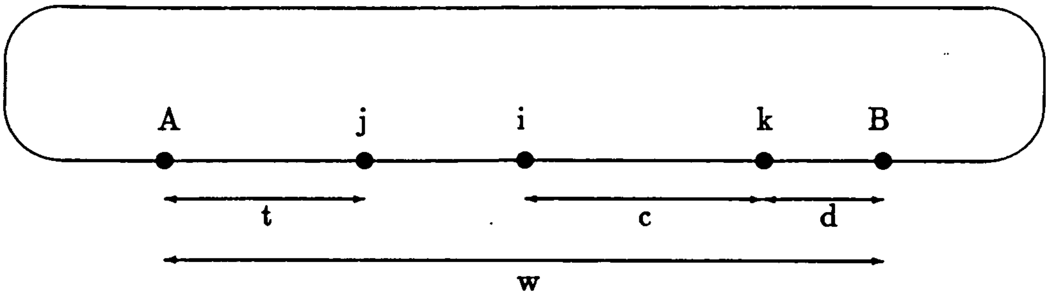

Figure C.1: Representation of a one dimensionally connected network with connectivity ratio w by a loop of length one unit. This diagram shows the relative positions of sites i, j and k. Points A and B are the limits of connectivity of site i.

In this case j is not necessarily connected to k.

Let t be the fractional distance of j from the left end of the connection space of i (see figure C.2), then,

$$t = \frac{\vec{Aj}}{\vec{Bi}} \tag{C.2}$$

There are now two different possible ways in which j and k could be connected.

a) The two sites are connected through i.

If the two sites are connected in this way then k must lie a fractional distance $c \leq t$ to the right of site i, since the connection space to one side of any site is of length $\frac{w}{2}$.

b) The two sites are connected through A and B.

If the two sites are connected in this way then k must lie within a fractional distance d from B where,

$$t\frac{w}{2} + (1 - w) + d\frac{w}{2} \leq \frac{w}{2} \tag{C.3}$$

therefore,

$$d \leq (3 - \frac{2}{w}) - t \tag{C.4}$$

Since d and t are both positive this requires $w > \frac{2}{3}$ for there to be any possibility of the two sites being connected in this way. To obtain the contributions to $a_1(w)$ from situations (a) and (b) we must integrate over all the possible positions of site k relative to site j. Thus integrating over the allowed values of c and d gives for the contributions from (a) and (b) to $a_1(w)$, remembering that the probability of situation 2 is $\frac{1}{2}$,

$$\frac{1}{2}\int_0^1 tdt + \frac{1}{2}\int_0^{3-\frac{2}{w}} \left(3 - \frac{2}{w} - t\right) \theta \left(w - \frac{2}{3}\right) dt \tag{C.5}$$

where θ is the Heaviside function. After integration the above expression becomes,

$$\frac{1}{4}\left[1+\left(3-\frac{2}{w}\right)^2\right]\theta\left(w-\frac{2}{3}\right)+\frac{1}{4}\theta\left(\frac{2}{3}-w\right) \tag{C.6}$$

Adding the contributions from situations 1 and 2 together gives, for the value of $a_1(w)$ in the case of one dimensional connectivity,

$$a_1(w)=\begin{cases}\frac{3}{4} & w<\frac{2}{3}\\ \frac{3}{4}+\frac{1}{4}\left(3-\frac{2}{w}\right)^2 & w>\frac{2}{3}\end{cases} \tag{C.7}$$

For higher dimensional hypercubic connectivity architectures the required calculation is equivalent to carrying out this calculation for each of the dimensions with a connection space length of $\sqrt[n]{w}$. The required probability is then given by the product of the probabilities associated with each of the dimensions. In general this gives,

$$a_k^n(w)=\left(a_k^1(\sqrt[n]{w})\right)^n \tag{C.8}$$

where $a_k^n(w)$ is the value of $a_k(w)$ for an n dimensional connection space. We therefore have from equation C.7, for a network with an n dimensional hypercubic connection architecture,

$$a_1(w)=\begin{cases}\left(\frac{3}{4}\right)^n & w<\left(\frac{2}{3}\right)^n\\ \left(\frac{3}{4}+\frac{1}{4}\left(3-\frac{2}{\sqrt[n]{w}}\right)^2\right)^n & w>\left(\frac{2}{3}\right)^n\end{cases} \tag{C.9}$$

Appendix D

The DAP computer

The DAP computer is a single instruction multiple data stream computer (SIMD) machine having an array of 64 × 64 single bit processors. Each processing element has associated with it a memory area of 4096 bits which it has direct access to [1]. The processors have nearest neighbour connections and, in addition, a slower data bus system connects processors by rows and columns. Using these channels the processors can pass data to each other.

The DAP has two programming languages at present; Fortran-plus which is a parallel language based on Fortran, and APAL which is a parallel assembly code language. Fortran-plus incorporates array and vector constructs so that if $\mathbf{A},\mathbf{B}$ and $\mathbf{C}$ are matrices then $\mathbf{A} = \mathbf{B} * \mathbf{C}$ will produce $A_{ij} = B_{ij} * C_{ij}$ on all the processing elements in parallel. These operations can be performed under a logical masking matrix so that we only obtain results on the required processors. The array of processing elements have a Q,C and A plane associated with them. The A plane is an activity control plane which is used for the masking operations while the Q and C planes correspond to an accumulator and carry plane. APAL instructions involve bit manipulation using these planes and are typically constructs of the form **CQPCQS M1**. This instruction, reading from left to right, adds the **S**,**Q** and **C** plane together putting the carry in **C** and the least significant bit in **Q**. The **S** plane can be any of the 4096 store planes and

[1] The DAP is now manufactured by AMT Reading and has 32 × 32 processors with 32k bits of memory each. Some of the simulations in this work were carried out on this new machine. AMT also plans in the future to build more 64 × 64 machines.

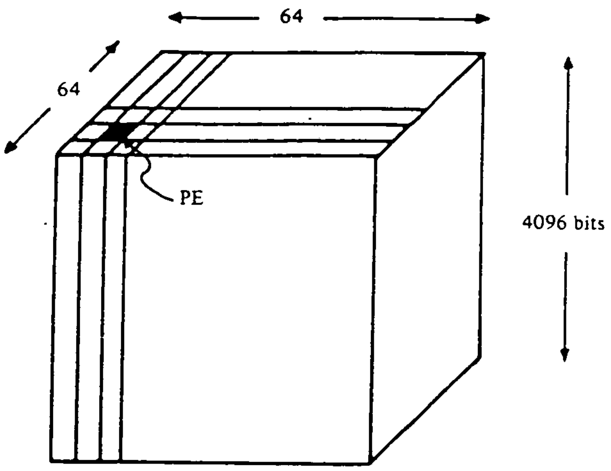

Figure D.1: Schematic representation of the DAP.

its exact address is specified by the number stored in the master control unit register **M1**. All Fortran-plus commands call macros of APAL code.

Since the DAP is constructed from bit processors it can cope with integers and real numbers of different bit lengths. Therefore Fortran-plus supports code for integer bit lengths from 8 bit to 64 bit in 8 bit intervals and real bit lengths from 32 to 64 bits. In general code for shorter length numbers, which the DAP is very well suited to, will run faster than longer length numbers. The DAP is fastest at bit manipulation where it out performs most other super-computers with a typical APAL instruction taking 200ns (100ns on the new 32 × 32 DAP).

In our simulations of neural networks in Chapter 4 we, in the case of the 4096 network, mapped each neuron onto each processor to exploit the parallelism of the DAP. For smaller system sizes we interleaved more than one simulation (eg. 4 simulations for the 1024 size system) and processed them simultaneously. We, where possible, always worked with logical variables representing the states of the nodes and short integers for the connection strengths. Where short integer operations were not well supported in Fortran-plus sections of code were written in APAL. This was particularly true in the case of the **SUM** instruction which sums the values of an array and typically takes much longer than other Fortran-plus matrix instructions. This instruction was used to sum the inputs to each neuron so that its new state could be determined. The part of the code which contained this sum was always in the innermost loop of the program and therefore represented the bottle neck of the program. A new version of the **SUM** function was written in APAL that summed integer lengths of 11 bits. In all our simulations we never stored more than about 630 states so 11 bit numbers were sufficient for the connection strengths. In some cases the new section of code produced a speed up factor of as much as 2.5 on the original Fortran-plus code enabling about 8000 nodes to be updated per second in the case of the 4096 system [2].

[2] Since this work a faster version of the standard 16 bit **SUM** function has been implemented on the 32 × 32 DAP which is about 1.7 times faster than the old version.

Bibliography

[1] McCulloch, W.S and Pitts, W.A., (1943) *Bull. Math Biophys.* **5**, 115.

[2] Hebb, D.O., (1949) The Organisation of Behaviour (Wiley, New York).

[3] Hopfield, J.J., (1982) *Proc. Natl. Acad. Sci. U.S.A* **79**, 2544.

[4] Amit, D.J., Gutfreund, H. and Sompolinsky, H., (1985) *Phys. Rev. Lett.* **55**, 1530.

[5] Amit, D.J., Gutfreund, H. and Sompolinsky, H., (1985) *Phys. Rev.* **A32**, 1007.

[6] Amit, D.J., Gutfreund, H. and Sompolinsky, H., (1987) *Annl. Phys.* **173**, 30.

[7] Amit, D.J., Gutfreund, H. and Sompolinsky, H., (1987) *Phys. Rev.* **A32**, 2293.

[8] Sompolinsky, H., (1986) *Phys. Rev.* **A34**, 2571.

[9] Amit, D.J., (1986) Heidelberg Symposium on Glassy Dynamics. eds: I. Morgenstern and J.L. van Hemmen, (Berlin:Springer).

[10] Sompolinsky, H., (1986) The Theory of Neural Networks: The Hebb Rule and Beyond. Heidelberg Colloquium on Glassy Dynamics and Optimisation., (June 1986), Springer-Verlag.

[11] Crisanti, A., Amit, D.J. and Gutfreund, H., (1986) *Europhys. Lett.* **2**, 337.

[12] Bruce, A.D., Gardner, E. and Wallace, D.J., (1987) *J. Phys.* **A20**, 2909.

[13] Bruce, A.D., Canning, A., Forrest, B., Gardner, E. and Wallace, D.J., (1986) In Proceedings of the Conference on Neural Networks for Computing (Snowbird Utah).

[14] Hinton, G.E. and Anderson, J.A., (1981) eds. *Parallel Models of Associative Memory*, Lawrence Erlbaum, New Jersey.
Rumelhart, D.E., McClelland, J.L. and the PDP research group (1986), *Parallel Distributed Processing: Explorations in the Micro-Structure of Cognition,* Vols. 1 and 2, Bradford Books, Cambridge, MA.
Grossberg. S., (1987), ed., *The Adaptive Brain,* Vols. 1 and 2, North Holland.

[15] Gardner-Medwin, A.R., (1976) *Proc. R. Soc. Lond.* B. **194**, 375.

[16] Gardner, E., (1986) *J. Phys.* **A19**, L1047.

[17] Gardner, E., (1987) *Europhys. Lett.* **4**, 481.

[18] Gardner, E., (1988) Edinburgh Preprint

[19] Forrest, B., (1988) *J. Phys.* **A21**, 245.

[20] Ising, E., (1925) *Z. Phys.* **31**, 253.

[21] Onsager, L., (1944) *Phys. Rev.* **65**, 117.

[22] Wallace, D.J. and Zia, R.K.P., (1978) The Renormalisation Group Approach to Scaling in Physics, *Rep. Prog. Phys.* **41**, 1.

[23] Kac,M., (1968) *Statistical Physics, Phase Transitions and Superfluidity,* Vol 1, (eds. M. Chretien, E.P. Gross and S. Desser), Gordon and Breach, New York, p241.

[24] Kac, M., (1968) Trondheim Theoretical Physics Seminar, Nordita, Publ. No. 286.

[25] Berlin, T.H. and Kac, M., (1952) *Phys. Rev.* **86**, 821.

[26] Edwards, S.F. and Anderson, P.W, (1975) *J. Phys.* **F5**, 965.

[27] Ruderman, M.A. and Kittel, C., (1954) *Phys. Rev.* **96**, 99.
Kasuya, T., (1956) *Prog. Theoret. Phys.* **16**, 45.
Yoshida, K., (1957) *Phys. Rev.* **106**, 893.

[28] Sherrington, D. and Kirkpatrick, S., (1975) *Phys. Rev. Lett.* **32**, 1792.
(1978) *Phys. Rev.* **B17**, 4384.

[29] Binder, K. and Young, A.P., (1986) *Rev. Mod. Phys.* **Vol 58, No. 4**, 801.

[30] Toulouse, G., (1977) *Comm. Phys.* **2**, 115.

[31] Mattis, D.C., (1976) *Phys. Lett.* **A36**, 421.

[32] Mezard, M., Parisi, G., Sourlas, N., Toulouse, G. and Virasoro, M., (1984) *J. Physique.* **45**, 843.

[33] Brout, R., (1959) *Phys. Rev.* **115**, 824.

[34] Palmer, R.G. and Pond, C.M., (1979) *J. Phys.* **F9**, 1451.

[35] van Hemmen, J.L. and Palmer. R.G., (1979) *J. Phys.* **A12**, 563.

[36] van Hemmen, J.L. and Palmer. R.G., (1982) *J. Phys.* **A15**, 3881.

[37] Lautrup, B., (1988) The Theory of the Hopfield Model, Neils Bohr Institute Preprint NBI-HE-88-06.

[38] De Almeida, J.R.L., and Thouless, D.J., (1978) *Phys. Rev.* A11, 983.

[39] Blandin, A., (1978) *J. Phys.* (Paris) Colloq. **C6-39**, 1499.

[40] Blandin, A., Gabay, M. and Berker, N., (1980) *J. Phys.* **C13**, 403.

[41] Bray, A.J. and Moore, M.A., (1978) *Phys. Rev. Lett.* **41**, 1068.

[42] Parisi, G., (1980) *J. Phys.* **A13**, L115.

[43] Parisi, G., (1980) *J. Phys.* **A13**, 1101.

[44] Parisi, G., (1980) *J. Phys.* **A13**, 1887.

[45] Parisi, G., (1983) *Phys. Rev. Lett.* **50**, 1946.

[46] Derrida, B., Gardner, E. and Zippelius, A., (1987) *Europhys. Lett.* **4** (2), 167.

[47] Krauth, W. and Mezard, M., (1987) *J. Phys.* **A20**, L745.

[48] Bellman, R., (1960) Introduction to Matrix Analysis, (The Maple Press Company, York, PA.), McGraw-Hill Series in Matrix Theory.

[49] Handbook of Mathimatical Functions (1965), p17, eds. M. Abramowitz and I.A. Stegun. (Dover Publications, Inc., New York).

[50] Khanin, K.M. and Sinai, Y.G., (1979) *J. Stat. Phys.* **20**, 573.

[51] Mezard, M. and Virasoro, M.A., (1985) *J. Phys.* **46**, 1293.

[52] Jackel, L.D., Denker, H.P., Graf, H.P., Howard, R.E., Hubbard, W., Schwartz, D., Straughn, B.L. and Tennant, D.M., (1986) *The Physics and Fabrication of Microstructures*, March 25, p453.