Partially connected models of neural networks

1988 J. Phys. A: Math. Gen. 21 3275

(http://iopscience.iop.org/0305-4470/21/15/016)

J. Phys. A: Math. Gen. 21 (1988) 3275-3284. Printed in the UK

Partially connected models of neural networks

A Canning and E Gardner

Department of Physics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK

Received 11 March 1988

Abstract. A partially connected Hopfield neural network model is studied under the restriction that w, the ratio of connections per site to the size of the system, remains finite as the size $N \to \infty$ with the connection structure at each site being the same. The replica symmetric mean field theory equations for the order parameters are derived. The zero-temperature forms of these equations are then solved numerically for a few different 'local' connectivity architectures showing phase transitions at different critical storage ratios, α_c, where the states which we are trying to store in the network become discontinuously unstable. We show that the information capacity per connection improves for partially connected systems.

1. Introduction

In the physics community over the past few years there has been a lot of research carried out on neural networks (Wallace 1985, Amit *et al* 1985a, b, 1987a, b, Crisanti *et al* 1986, Sompolinsky 1986, 1987, Bruce *et al* 1987). The main stimulus for this work was when Hopfield (1982) formulated the Hamiltonian for a type of network that had been studied in one form or another as far back as McCulloch and Pitts (1943) and Hebb (1949). Physicists then realised it was a generalisation of the infinite-range spin glass studied by Sherrington and Kirkpatrick (1975) and could be studied with the same theoretical techniques. So far the Hopfield model has only been studied in its fully connected and randomly diluted form (Sompolinsky 1986, Derrida *et al* 1987, Kanter 1987, Kanter and Sompolinsky 1987).

There are several motivations for studying more structured systems. Fully or randomly connected systems require long-range connections and are therefore difficult to build and have slower communication times between neurons than a system with more compact local connectivity. Correlations in real problems are likely to be local also. Building networks with structured neighbourhoods would be much simpler and take up less space than a fully connected network. They would be more useful than fully connected networks provided the loss of storage capacity due to loss of connections does not outweigh these factors. In biological systems, neurons have some topology of neighbourhood connectivity embedded in three-dimensional space which is more local than random or fully connected systems. The correlation between the brain and recurrent networks of which Hopfield's is only one example is discussed with references by Gardner-Medwin (1976).

The N neurons in this model can take values 1 or -1 as in the model studied by Amit *et al* (1985a, b) and the connection strengths are defined by

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

0305-4470/88/153275+10$02.50

where p is the number of nominated configurations $\{\xi_i^\mu = \pm 1\}$, $i = 1, \ldots, N$, to be stored and N is the size of the system. $\boldsymbol{D}$ is a symmetric matrix which defines the connection architecture, having a 1 or 0 at position i, j depending on whether site i is connected to site j or not. In this paper we will only consider models where the connection architecture at each site is the same; thus we can define the connectivity ratio of the network w to be:

$$w = \left(\sum_{j=1}^{N} D_{ij} \right) N^{-1} \qquad \forall i. \tag{2}$$

In order to perform the mean field theory calculations w is kept finite in the limit as $N \to \infty$.

2. Mean field theory

We calculate the free energy using the replica symmetric method employed by Sherrington and Kirkpatrick (1986) for infinite-range spin glasses. The average free energy per spin is then given by:

$$f = \lim_{n \to 0} \lim_{N \to \infty} \frac{-1}{\beta n N} (\langle\langle Z^n \rangle\rangle - 1) \tag{3}$$

where $\langle\langle\ \rangle\rangle$ represents averaging over the quenched distribution of the ξ and $\langle\ \rangle$ will be used to represent the thermal average.

The partition function of n (labelled by $\rho = 1, 2, \ldots, n$) replicas is then

$$\langle\langle Z^n \rangle\rangle = \left\langle\left\langle \operatorname*{Tr}_{S^\rho} \exp\left(\frac{\beta}{2N} \sum_{ij\mu\rho} (\xi_i^\mu S_i^\rho) D_{ij} (\xi_j^\mu S_j^\rho) - \tfrac{1}{2}\beta p n \right) \right\rangle\right\rangle. \tag{4}$$

The $\frac{1}{2}\beta p n$ term comes from the $i = j$ term and we therefore set $D_{ii} = 1$. The sites are decoupled by introducing variables $m_{\rho i}^\mu$ for each nominated configuration μ, each replica ρ and each site i. The same procedure as Amit *et al* (1987a) is then followed by splitting the sum over μ configurations into two parts corresponding to a finite set of s patterns having macroscopic overlap with the nominated configurations and an infinite set with microscopic overlap. The variables $q_i^{\rho\sigma}$ and $r_i^{\rho\sigma}$ are then introduced to allow us to do the integrals over the infinite set of configurations with microscopic overlap. We then have for the partition function:

$$\begin{aligned} \langle\langle Z^n \rangle\rangle = \int \prod_\nu \mathrm{d}m_{\rho i}^\nu \int \prod_{i,\, \rho > \sigma} \mathrm{d}q_i^{\rho\sigma}\, \mathrm{d}r_i^{\rho\sigma} \\ \times \exp\left\{ N \left[-\tfrac{1}{2}\beta n \alpha - \tfrac{1}{2}\beta \sum_{\nu\rho ij} m_{\rho i}^\nu D_{ij}^{-1} m_{\rho j}^\nu - \tfrac{1}{2}\alpha \operatorname*{Tr}_{\rho\sigma ij} \ln\left(\delta_{\rho\sigma}\delta_{ij} - \frac{\beta D_{ij}}{N} Q_i^{\rho\sigma} \right) \right.\right. \\ - \tfrac{1}{2}\alpha\beta^2 \frac{1}{N} \sum_{i\rho\sigma} r_i^{\rho\sigma} q_i^{\rho\sigma} \\ \left.\left. + \left\langle\left\langle \ln \operatorname*{Tr}_{S_i^\rho} \exp\left(\tfrac{1}{2}\alpha\beta^2 \frac{1}{N} \sum_{i\rho\sigma} S_i^\rho S_i^\sigma r_i^{\sigma\rho} + \frac{\beta}{N} \sum_{\nu\rho i} m_{\rho i}^\nu \xi_i^\nu S_i^\rho \right) \right\rangle\right\rangle \right] \right\} \end{aligned} \tag{5}$$

where ρ and σ are replica indices and range from $1, \ldots, n$, Σ_ν is over the finite set of s patterns, $\alpha = p/N$ is the storage ratio and Q is a matrix defined, $Q_i^{\rho\sigma} = q_i^{\rho\sigma} \rho \neq \sigma$, $Q_i^{\rho\rho} = 1$.

As $N \to \infty$ the integral is dominated by its saddle point. The following equations for the physical meaning of the order parameters are determined from the saddle point:

$$\begin{aligned} m_{\rho i}^{\nu} &= \frac{1}{N}\left\langle\!\!\left\langle \sum_j D_{ij}\xi_j^{\nu}\langle S_j^{\rho}\rangle \right\rangle\!\!\right\rangle \\ q_i^{\rho\sigma} &= \langle\!\langle \langle S_i^{\rho}\rangle\langle S_i^{\sigma}\rangle \rangle\!\rangle \\ r_i^{\rho\sigma} &= \frac{1}{\alpha}\sum_{\mu=s+1}^{\alpha N} \langle\!\langle m_{\rho i}^{\mu} m_{\sigma i}^{\mu}\rangle\!\rangle . \end{aligned} \tag{6}$$

We will now look at the possible choices of D_{ij} in more detail. If we consider the neurons to be sitting at sites on a hypercubic lattice then in the limit $N \to \infty$ we can think of the lattice as being continuous with a neuron sitting at each point. For any neuron the other neurons connected to it will define a shape or shapes on the lattice. We will henceforth refer to this as the connection space of that neuron. If this connection space was chosen to be the same at every site, it is reasonable to expect that site-independent solutions for the order parameters will exist. In this paper only site-independent solutions will be considered. For example if we had chosen $\boldsymbol{D}$ to be structured with one neuron having many more connections to it than any other we would expect $m_{\rho i}^{\nu}$ to always depend on i. Therefore for site-independent solutions we can set:

$$\begin{aligned} m_{\rho i}^{\nu} &= m_{\rho}^{\nu} \\ r_i^{\rho\sigma} &= r^{\rho\sigma} \\ Q^{\rho\sigma} &= \frac{1}{N}\sum_i Q_i^{\rho\sigma} \qquad q^{\rho\sigma} = \frac{1}{N}\sum_i q_i^{\rho\sigma}. \end{aligned} \tag{7}$$

The finite sum of s condensed patterns self-averages (Amit *et al* 1985a) and taking the limit as $n \to 0$ we obtain the replica symmetric equation for the free energy per spin:

$$\begin{aligned} f = {} & \tfrac{1}{2}\alpha + \frac{1}{2w}\sum_{\nu}(m^{\nu})^2 \sum_{ij} D_{ij}^{-1} + \frac{\alpha\beta r}{2}(1-q) \\ & \times \frac{\alpha}{2\beta}\left\{ \operatorname*{Tr}_{ij} \ln\left(\delta_{ij} - \frac{\beta D_{ij}}{N}(1-q)\right) - \operatorname*{Tr}_{ij}\left[\frac{\beta D_{ij} q}{N}\left(1 - \frac{\beta D_{ij}}{N}(1-q)\right)^{-1}\right]\right\} \\ & - \frac{1}{\beta}\int \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)\left\langle\!\!\left\langle \ln 2\cosh\beta\left(\sqrt{\alpha r}z + \sum_{\nu} m^{\nu}\xi^{\nu}\right)\right\rangle\!\!\right\rangle . \end{aligned} \tag{8}$$

If we rescale some of the parameters, i.e. $m^{\nu} \to w m^{\nu}$, $r \to wr$ and $\alpha \to w\alpha$, then the order parameters regain the same form as for a fully connected network. We now have from the saddle point the following equations for the order parameters:

$$\begin{aligned} m^{\nu} &= \left\langle\!\!\left\langle \int \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \xi^{\nu} \tanh \beta w(\sqrt{\alpha r}z + \boldsymbol{m}\cdot\boldsymbol{\xi})\right\rangle\!\!\right\rangle \\ q &= \left\langle\!\!\left\langle \int \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tanh^2 \beta w(\sqrt{\alpha r}z + \boldsymbol{m}\cdot\boldsymbol{\xi})\right\rangle\!\!\right\rangle \\ r &= \operatorname*{Tr}_{ij}\left[\frac{q}{w}\left(1 - \frac{\beta D_{ij}(1-q)}{N}\right)^{-2}\left(\frac{D_{ij}}{N}\right)^2 q\right]. \end{aligned} \tag{9}$$

The physical meaning of these order parameters is then from (6):

$$m^\nu = \frac{1}{N}\sum_{i=1}^{N} \langle\langle \xi_i^\nu \langle S_i \rangle \rangle\rangle$$

$$q = \frac{1}{N}\sum_{i=1}^{N} \langle\langle \langle S_i \rangle^2 \rangle\rangle \tag{10}$$

$$r = \frac{1}{\alpha}\sum_{\mu=s+1}^{\alpha N} \langle\langle (m^\mu)^2 \rangle\rangle .$$

Thus m^ν measures the correlations of the states of the system with the nominated configurations and q is the Edwards-Anderson (1975) order parameter. α is now a measure of the storage per connection and is defined as

$$\alpha = \frac{p}{wN}. \tag{11}$$

3. Zero-temperature mean field equations

We now look at the zero-temperature limit corresponding to $\beta \to \infty$. We assume that r can be expanded in $\beta \boldsymbol{D}(1-q)/N$. This expansion was found to be valid when the mean field equations were solved numerically except for random connectivity in the limit $w \to 0$ when the sum can be done analytically (see § 4 and the appendix). The mean field equations for the zero-temperature model for states with a single non-vanishing overlap are then:

$$m = 2\,\mathrm{erf}\frac{m}{\sqrt{\alpha r}}$$

$$q = 1$$

$$r = 1 + \sum_{k=1}^{\infty} C^k (k+1) a_k(w) \tag{12}$$

$$C = \beta w (1-q)$$

$$a_k(w) = w\,\mathrm{Tr}\left(\frac{\boldsymbol{D}}{wN}\right)^{k+2}.$$

Thus the mean field equations for m, q and C have the same form as for a fully connected network and only a_k explicitly contains information about the connection architecture of the system. The solutions of these equations yield the storage capacity and accuracy of storage of a network with architecture specified by $\boldsymbol{D}$. For a given choice of $\boldsymbol{D}$ with the required restrictions we find varying values of α_c above which no storage takes place. With $a_k(w) = 1$ for all k we recover the order parameter equation $r = (1-C)^{-2}$ for the fully connected model (Amit *et al* 1987a). With $a_k(w) = w$ for all k corresponding to random connectivity we obtain $r = 1 + w[(1-C^{-2}-1]$ which is the same result Sompolinsky (1986) obtained by a different route for a randomly diluted system. The above equations were solved for hypercubic lattices of neurons with the connection space of each neuron being a hypercube of neurons centred on that neuron. This, for example in two dimensions, would give us a square lattice of neurons with

each neuron connected to a square of neurons about it. It was found that for dimensionality higher than 16 the behaviour was very close to that of a randomly connected system. The results presented here will be for dimensions $d = 1, 2, 3, 4, 8$ and randomly connected which will be referred to as the $d = \infty$ model. The results for a fully connected model ($w = 1$) will also be presented for comparison. We will now look at the form of $a_k(w)$ in more detail which, from (13), is

$$a_k(w) = N^{-(k+2)} w^{-(k+1)} \underbrace{\sum_{i_1=1} \sum_{i_2=1} \cdots \sum_{i_{k+2}=1}^{N} D_{i_1 i_2} D_{i_2 i_3} \cdots D_{i_{k+2} i_1}}_{S} . \tag{13}$$

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_1 is connected in a loop of $k+2$ connections back to itself through neurons i_2 to i_{k+2}. Thus S/N^{k+2} is the probability that $k+2$ neurons chosen at random are connected together in a loop where S contains all the possible ways of choosing the $k+2$ neurons from the total set of N neurons. The less likely a loop is complete the lower the value of $a_k(w)$ and the correspondingly higher the value of α_c. Hence random connectivity is the best choice of connection architecture for maximising α_c since, for a given w this minimises $a_k(w)$. This is not the whole story, however, as m decreases as α_c increases but this will be discussed in more detail later. For a discussion of the numerical results for $a_k(w)$ see the appendix.

3.1. Results

Numerical solutions of (12) with single non-vanishing overlap yields a family of curves of α against m. A few of these curves for different architecture and connectivity ratio are plotted in figure 1. The maximum values give α_c above which $m = 0$ is the only solution. An important property of these curves is that the higher the dimensionality of the connectivity and the smaller w the higher the value of α is for a given m and

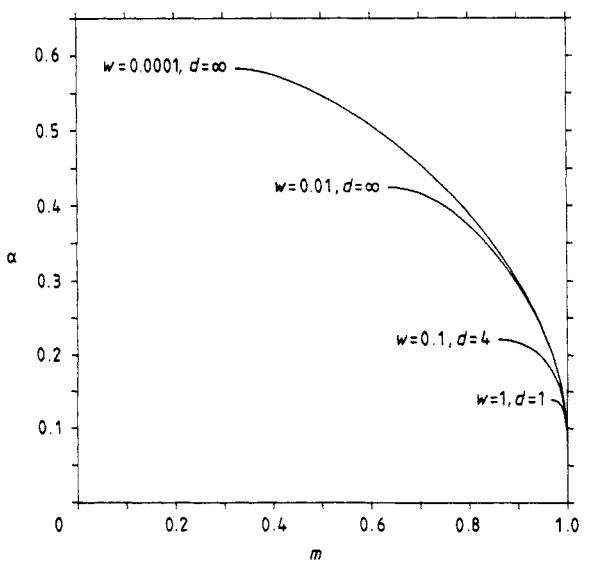

Figure 1. α against m for some different connection architectures given by the solutions of (12).

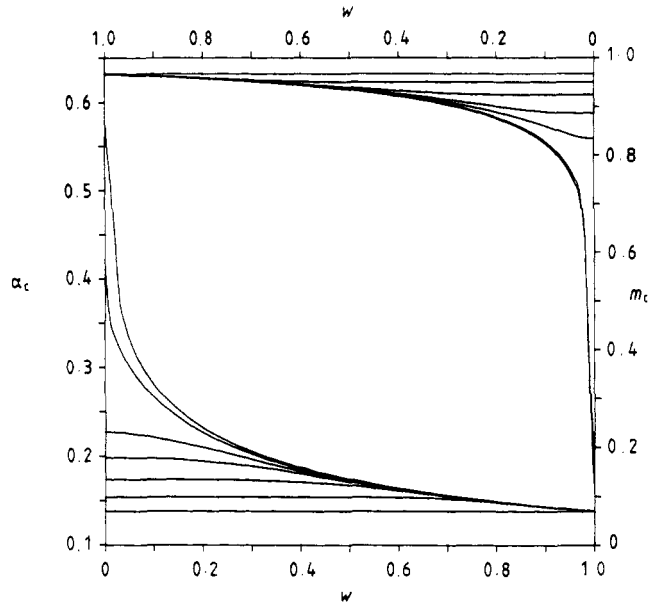

Figure 2. Critical values of the order parameters α and m are plotted against w, the connectivity ratio. On the left and bottom axes are plotted α_c against w for different hypercubic connection spaces. The curves are, from bottom to top, $w = 1$, $d = 1, 2, 3, 4, 8, \infty$. Here $w = 1$ represents a fully connected network and $d = \infty$ a randomly connected network. On the right-hand and upper axes m_c is plotted against w where the curves are, from top to bottom, $w = 1$, $d = 1, 2, 3, 4, 8, \infty$.

hence the higher the storage capacity. Figure 2 shows the critical values α_c and m_c associated with the phase transitions for different hypercubic connection spaces. From these graphs we can see how α_c is always constrained to lie between 0.138 for $w = 1$ and $2/\pi$ ($= 0.637$) for $d = \infty$ as $w \to 0$. For higher dimensionality of connectivity where the highest values of α_c are obtained m_c holds up very well as w decreases but falls off sharply as w approaches zero. We can expand the order parameter equations for small w giving:

$$\alpha_c \approx \frac{2}{\pi}\left(1 - \frac{3}{2^{2/3}} w^{1/3}\right)$$
$$m_c \approx \left(\frac{432w}{\pi^3}\right)^{1/6}. \tag{14}$$

Numerical simulations and theoretical calculations on a fully connected network with one broken replica symmetry by Crisanti *et al* (1986) suggested that the effect of replica symmetry breaking is to increase α_c and also to increase the accuracy of storage. This increase is quite small with α_c increasing from 0.138 to 0.145 and it is expected that a similar effect will be found in partially connected systems if m is close to one.

4. Information content

If we wish to compare the information stored in networks with different architectures we must take into account not only the number of patterns stored but also the accuracy

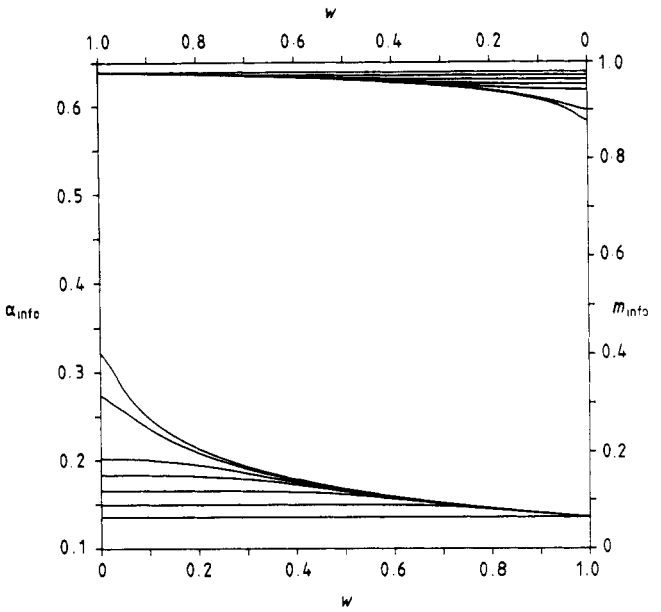

Figure 3. This plot has the same layout as figure 2 but shows the values of α and m giving maximum information storage as defined by (15).

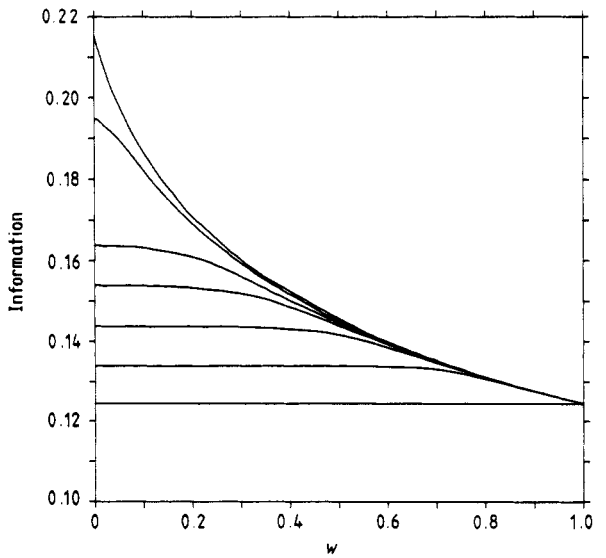

Figure 4. The values of maximum information storage as defined by (15) against w, the connectivity ratio for different hypercubic connection architectures. The curves are, from bottom to top, $w = 1$, $d = 1, 2, 3, 4, 8, \infty$.

with which they are stored. For example even though with a randomly connected network we can achieve an α_c of $2/\pi$ as $w \to 0$ we also find $m \to 0$ so there is no information stored in the network. Amit *et al* (1987b) derived equations for the information stored in a network. The basic idea of this is that the information content of an N-bit vector equals the entropy associated with all the possible ways of choosing an N-bit vector. Thus the information content of an N-bit vector is $N \ln 2$. The information lost when the vector is not fully known is then the ln of the number of ways of choosing this N-bit vector with a certain fraction of bits wrong and this must be subtracted from $N \ln 2$. This then gives for a measure of information per connection:

$$I(\alpha) = \frac{\alpha}{2 \ln 2} [(1+m) \ln(1+m) + (1-m) \ln(1-m)] \tag{15}$$

which is normalised so that when $m = 1$, $I(\alpha) = \alpha$. Figure 3 shows the values of α and m giving the maximum values of $I(\alpha)$. It can be seen clearly from this how it pays to relax α below α_c, particularly for small w, to increase information storage. Figure 4 shows the maximum values of $I(\alpha)$ for the different architectures studied showing that, as $w \to 0$, the best results are obtained for information storage with random connectivity having the highest storage capacity.

5. Discussion

This paper has shown that partially connected networks have better storage per connection than fully connected networks, the storage increasing the more sparse and random the connections are. If we consider the resource requirements for these networks we can see that a partially connected system requires many more neural units to have any significant storage over a fully connected system. It is when we consider restrictions of space and communication times that the major advantages of a partially connected system are seen. In the 'neural chips' which have been built so far (Jackel *et al* 1986), 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 (1988) for the fully connected model. We would expect the basins of attraction for the stored states to be larger due to the less crowded nature of the phase space. The extent to which these results extend to Hopfield-type networks with other learning algorithms (Wallace 1985, Bruce *et al* 1986, Gardner *et al* 1987) which improve on the basic Hebb rule used in this paper could also be studied. Compared with complete connectivity, random dilution (Gardner 1988b) improves storage per connection for the perceptron learning algorithms of Gardner (1988a), Krauth and Mézard (1987) and Forrest (1987). The ability of partially connected networks to store information with short-range correlation 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?

Acknowledgments

We would like to thank David J Wallace for suggesting looking at partially connected networks and Douglas McKnight for useful discussion on integrated circuit design. Financial support for Elizabeth Gardner was from a Dewar Fellowship from Edinburgh University and from SERC. Financial support for Andrew Canning was from SERC and Active Memory Technology Ltd., Reading, UK.

Appendix

We wish to calculate $a_k(w)$ (equation (13)) in the limit as $N \to \infty$. In order to illustrate the method of calculating it we will work, for simplicity, with a two-dimensional system but the method extends easily to any dimension. Let a square of side $\sqrt{1/w}$ with cyclic boundary conditions represent the lattice of neurons. Then each neuron which is represented by a point has a connection space of a square of side 1 around it. The first step is then to choose a point i_1 which can be any point on the lattice (all points are equivalent due to cyclic boundary conditions) and then randomly choose a point i_2 in the connection space around i_1. We continue this process for $k+1$ steps until we reach the point i_{k+2} then $a_k(w)$ is the probability that the final point is in the connection space of the initial point i_1. At each step because we only choose a point in the connection space of the previous point rather than a random point, we introduce a factor $1/w$ into the probability that sites are connected. The $w^{-(k+1)}$ factor in $a_k(w)$ comes from the $k+1$ steps. Therefore the calculation of $a_k(w)$ is reduced to the probability that a bounded random walk of $k+1$ steps ends in the connection space of the starting point. Thus, we are calculating the traces of powers of the connectivity matrix $\boldsymbol{D}$ for a system with size of the order of the inverse of the precision of the computer used. The calculations of the $a_k(w)$ were carried out on the ICL Distributed Array Processor machine (DAP) which is a SIMD machine with 4096 bit processors†. On this computer about a quarter of a million random steps plus the calculation of $a_k(w)$ all in double precision could be carried out per second. There were three main possible sources of error in this calculation but these were overcome in the following way.

Firstly, by calculating $a_k(w)$ from about two million random walks the standard deviation was reduced to a negligible size. Secondly, the matrix we are working with is of finite size but for certain values of $a_1(w)$ we can analytically calculate it with the restriction that we have hypercubic connection spaces in n dimensions. Specifically when

$$w < (\tfrac{2}{3})^n \qquad a_1(w) = (\tfrac{3}{4})^n. \tag{A1}$$

The numerical results for $a_1(w)$ were found to agree to within about 0.01% with the theoretical results. Finally the number of terms required in the series was very dependent upon the value of C. If C was large, which occurred with high dimensionality

† The DAP is now manufactured by AMT Ltd and has 1024 processors but a clock cycle approximately twice as fast as the 4096 DAP.

and low w, more terms were needed. As the number of steps in the random walk tends to infinity the final position is unrelated to the starting position. Thus $a_k(w) \to w$ as $k \to \infty$. Therefore when C is large we calculate enough terms in the sequence until $a_k(w) \approx w$ and then add in a correction term corresponding to summing the remaining terms in the sequence with $a_k(w)$ set to w. It was found in all the models studied that no more than twenty terms were needed to evaluate the sequence very accurately.

References

Amit D J, Gutfreund H and Sompolinsky H 1985a *Phys. Rev. Lett.* **55** 1530

—— 1985b *Phys. Rev. Lett.* **55** 1530; *Phys. Rev.* A **32** 1007

—— 1987a *Ann. Phys., NY* **173** 30

—— 1987b *Phys. Rev.* A **32** 2293

Bruce A D, Canning A, Forrest B, Gardner E and Wallace D J 1986 *Proc. Conf. on Neural Networks for Computing, Snowbird, UT, 1986 (AIP Conf. Proc.* **151**) pp 65-70

Bruce A D, Gardner E and Wallace D J 1987 *J. Phys. A: Math. Gen.* **20** 2909

Crisanti A, Amit D J and Gutfreund H 1986 *Europhys. Lett.* **2** 337

Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167

Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965

Forrest B 1988 *J. Phys. A: Math. Gen.* **21** 245

Gardner E 1988a *J. Phys. A: Math. Gen.* **21** 257

—— 1988b *Edinburgh preprint* 88/440

Gardner E, Stroud N and Wallace D J 1987 *Edinburgh preprint* 87/410

Gardner-Medwin A R 1976 *Proc. R. Soc.* B **194** 375

Hebb D O 1949 *The Organisation of Behaviour* (New York: Wiley)

Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2544

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* p 453

Kanter I 1987 *Preprint* Bar-Ilan University

Kanter I and Sompolinsky H 1987 *Phys. Rev. Lett.* **58** 164

Krauth W and Mezard M 1987 *J. Phys. A: Math. Gen.* **20** 745

McCulloch W S and Pitts W A 1943 *Bull. Math. Biophys.* **5** 115

Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **32** 1792

Sompolinsky H 1986 *Phys. Rev.* A **34** 2571

—— 1987 *The Theory of Neural Networks: The Hebb Rule and Beyond. Heidelberg Colloq. on Glassy Dynamics and Optimisation, June 1986* (Berlin: Springer)

Wallace D J 1985 *Advances in Lattice Gauge Theory* ed D W Duke and J F Owens (Singapore: World Scientific) p 326