An exactly solvable model for competitive networks

1989 J. Phys. A: Math. Gen. 22 2047

(http://iopscience.iop.org/0305-4470/22/12/010)

J. Phys. A: Math. Gen. **22** (1989) 2047-2056. Printed in the UK

An exactly solvable model for competitive networks

Adam Bennett†

Department of Experimental Psychology, University of Oxford, 1 South Parks Road, Oxford, UK

Received 25 January 1989

Abstract. An energy function is proposed whose long-time dynamic behaviour is believed to resemble that of a realistic large competitive neural network. A simple model which gives rise to this energy function is described. The model is solved exactly for a finite number of patterns in the thermodynamic limit using mean-field theory. Simulations of this model are presented. The behaviour of large competitive networks is discussed.

1. Introduction

In this paper a Hamiltonian is introduced whose behaviour resembles that of competitive networks. A model is described which gives rise to this Hamiltonian. This model is a highly idealised version of real neural networks and consequently possesses many unphysiological features; however, it is believed that it captures many of the important long-term properties of large competitive networks. In this respect it plays a similar role for competitive networks as the Hopfield model does for auto-associative memories. In fact this model shares many other features in common with the Hopfield model although it describes a very different neural system.

Competitive networks were first introduced to explain the firing patterns found in the visual cortex; see, for example, [1-6]. Figure 1 shows a schematic representation of this type of network. In its simplest form a competitive network consists of a set of N_a afferent axons (here shown coming in from the left) which synapse onto a set of N_d (vertical) dendrites. The synaptic coupling strength between axon i and dendrite p we denote by c_i^p. A set of N_p input patterns are presented to the network. The presynaptic potential produced by pattern μ on axon i we denote by ξ_i^μ. The presynaptic potentials create postsynaptic potentials which are accumulated by the dendrites causing the cell bodies to fire. However, the efferent axons (those leaving the cell bodies) synapse with interneurons which feedback onto all the neurons in the vicinity, altering the final firing rate (denoted by $S^{p,\mu}$).

In our competitive network model we make the simplifying assumption that the synaptic weights and the input patterns are Ising variables (± 1). The dynamics of the weights is assumed to give rise to a Hamiltonian $\mathcal{H}[c]$ consisting of three pieces

$$\mathcal{H}[c] = -\frac{1}{2N_a}\sum_{\mu=1}^{N_p}\sum_{i,j=1}^{N_a}\sum_{p=1}^{N_d}\xi_i^\mu\xi_j^\mu c_i^p c_j^p + \frac{1}{2N_a}\sum_{\mu=1}^{N_p}\sum_{i,j=1}^{N_a}\sum_{p,q=1}^{N_d}\lambda^{p,q}\xi_i^\mu\xi_j^\mu c_i^p c_j^q - J\sum_{\mu=1}^{N_p}\sum_{i=1}^{N_a}\sum_{p=1}^{N_d}\xi_i^\mu c_i^p. \tag{1}$$

† Work started in the Department of Physics, University of Edinburgh, UK.

0305-4470/89/122047+10$02.50

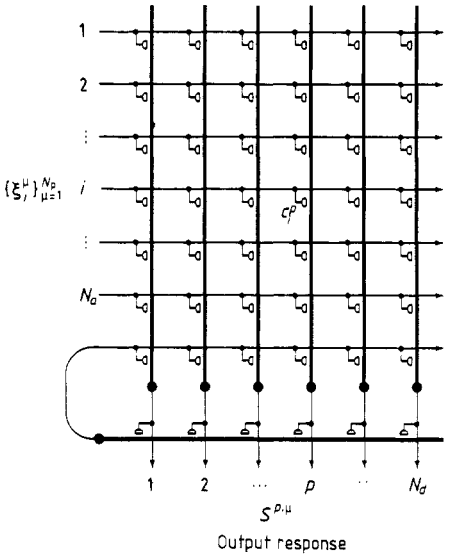

Figure 1. Schematic representation of a competitive network. Input stimuli ξ_i^μ are presented along the horizontal axons; these synapse with the vertical dendrites which sum the postsynaptic potentials causing the cell body to fire. The firing of the cells is detected by the horizontal inhibitory inter-neuron which moderates the cell firing rates.

The effect of the three parts of this equation can be roughly interpreted as follows: the first term is minimised when the synaptic weights down each dendrite are identical to the input patterns (a similar energy term has been discussed in [7, 8]); the second 'competition' term causes each dendrite to learn a different pattern; finally the last term prevents the inverse of each pattern being learned. The inhibition $\lambda^{p,q}$ is thought to be a function of the distance between dendrites p and q. In the later parts of this paper we will make the simplifying assumption that the inhibition is constant (i.e. $\lambda^{p,q} = \lambda$). I will assume that the dendrites inhibit themselves—an unnecessary but convenient assumption. Provided $\lambda \ll 1$ this assumption is of little consequence. The equilibrium behaviour of this model can be calculated exactly when the number of axons becomes very large. Before solving this model we will describe an actual model whose dynamics gives rise to exactly this Hamiltonian.

2. The model

We consider a network with the architecture described above which is presented with a series of input patterns. A single input pattern will cause a postsynaptic potential on a dendrite p which is assumed to be the sum over all axons of the presynaptic potentials ξ_i^μ multiplied by the synaptic weights c_i^p

$$V^{p,\mu} = \frac{1}{N_a} \sum_{i=1}^{N_a} \xi_i^\mu c_i^p. \tag{2}$$

In a more realistic model a masking term could be introduced to take account of missing synapses—for simplicity we will assume the network to be fully connected.

The cell bodies are connected to each other through a set of interneurons which moderate the cell firing. The stable cell firing rate is assumed to depend on three pieces: a linear function of the postsynaptic potential at the cell body; an inhibition term proportional to the postsynaptic potential of cell bodies in its vicinity; and a threshold (or spontaneous firing rate) term assumed to be the same for each body

$$S^{p,\mu} = V^{p,\mu} - \sum_{q=0}^{N_d} \lambda^{p,q} V^{q,\mu} + 2J^p. \tag{3}$$

The synapse between axon i and dendrite p experiences a (linear) 'Hebbian field' $h_i^{p,\mu}$ proportional to the presynaptic potential and the firing rate of the postsynaptic cell

$$h_i^{p,\mu} = \xi_i^\mu S^{p,\mu}. \tag{4}$$

Each synapse is assumed to integrate the Hebbian fields of all the input patterns: thus after receiving all the patterns each synapse holds in short term memory the 'total Hebbian field'

$$h_i^p = \sum_{\mu=1}^{N_p} h_i^{p,\mu}. \tag{5}$$

Only after receiving all the input a synapse will update itself from $c_i^p \to \bar{c}_i^p$ with probability

$$p(\bar{c}_i^p = \pm 1) = \frac{1}{1+\exp(-\beta h_i^p[c]\bar{c}_i^p)}. \tag{6}$$

Where the total Hebbian field $h_i^p[c]$ depends on the old synaptic weights.

The set of patterns is repeatedly presented, and a synapse is updated each time. After sufficiently many presentations the distribution of synaptic weight values will relax into the Gibb's distribution

$$P_{eq} \sim \exp\left(-\frac{\beta}{2}\sum_{i=1}^{N_s}\sum_{p=1}^{N_d} h_i^p[c]c_i^p\right) = \exp(-\beta\mathcal{H}[c]) \tag{7}$$

where $\mathcal{H}[c]$ is the Hamiltonian given in equation (1).

This model suffers from many deficiencies as a realistic or useful competitive network, which arise mainly from the requirement that the Hamiltonian have a quadratic form. Most importantly the synapses must integrate the Hebbian fields for each pattern before deciding whether to update. In a more biologically plausible model the synaptic weights would be continuous valued and update by a small amount after each pattern is presented. Since we have used discrete-valued synapses this is not possible in our model. If we naively use continuous synapses we no longer have a sufficient amount of non-linearity in the system to learn interesting mappings. (In real systems this non-linearity would enter through the sigmoid firing responses of the neurons).

It is hoped that our model approximates the most important feature of a real (or useful) competitive network. Some encouragement that at least some of the differences between these models do not effect the long term dynamics comes from the apparent closeness between the energy function of equation (1) and the Hopfield function. (The Hopfield model describes the dynamics of the firing rates of the neurons; the synapses are treated as the quenched variables. Nevertheless our Hamiltonian is formally identical to the Hopfield Hamiltonian when there is no mutual inhibition between the neurons.) The long-time behaviour of the Hopfield model is known to be very similar to that of the Little model which differs from the Hopfield model in that it undergoes

parallel updating as opposed to serial updating. It is hoped that the long-term behaviour in our model is similarly close to a model which updates in parallel. Indeed, simulations of our model with parallel updating appears to show long-time behaviour similar to serially updated simulations, although their detailed behaviour, of course, differs.

Another question is: what is the effect of allowing the presynaptic potentials only Ising values? There is some evidence from the Hopfield model that at least some of the important features are common to models with continuous and discrete variables (see, for example, the effects of clipping described in [9]). However it should also be noted that the energy surface is much more stable to overloading when the quenched variables are chosen from some continuous distributions (see [10]).

3. Mean-field theory

The equilibrium states for the Hamiltonian of equation (1) can be found in the thermodynamic limit ($N_a \to \infty$) using mean-field theory. We consider the simpler case of a finite number of input patterns. This calculation closely resembles that of [10] for the Hopfield model although the meanings of the two models are very different. In our model the input patterns are treated as quenched variables while the synaptic weights are the annealed variables.

The partition function is given by

$$Z(\beta) = \operatorname*{Tr}_{\{c\}} \exp(-\beta \mathcal{H}[c]) \tag{8}$$

where $\mathcal{H}$ is the energy function of equation (1). To perform the trace we first decouple the synaptic weights by introducing a magnetisation $m^{p,\mu}$ for each pattern and each dendrite. The partition function becomes

$$Z(\beta) = C\left(\prod_{\mu=1}^{N_p}\prod_{p=1}^{N_d}\int_{-\infty}^{\infty} dm^{p,\mu}\right)\exp\left(-\frac{\beta N_a}{2}\sum_{\mu}\sum_{p,q} m^{p,\mu}T^{p,q}m^{q,\mu}\right)$$
$$\times \exp\left\{\sum_{i}\sum_{p}\ln\left[2\cosh\left(\beta\sum_{\mu}\sum_{q}(m^{q,\mu}T^{pq}+J^{p}\delta^{pq})\xi_i^{\mu}\right)\right]\right\} \tag{9}$$

where C is a constant term which does not contribute to the free energy in the thermodynamic limit and

$$T^{p,q} = \delta^{p,q} - \lambda^{p,q}. \tag{10}$$

The free energy per synapse averaged over all possible sets of input patterns is

$$f(\beta) = \frac{-1}{\beta N_a N_d}\langle\langle \ln Z(\beta)\rangle\rangle \tag{11}$$

where $\langle\langle \ldots \rangle\rangle$ signifies the ensemble average over the quenched variables ξ_i^{μ}. In the limit of an infinite number of patterns the fluctuations in the log-cosh term can be neglected (since they are of order $1/\sqrt{N_a}$) and thus this term can be replaced by its average value, i.e. the free energy and magnetisation self-average. The free energy is therefore found from the saddle point evaluation of equation (9)

$$f(\beta) = \frac{1}{2N_d}\sum_{\mu=1}^{N_p}\sum_{p=1}^{N_d}(m^{p,\mu})^2 - \frac{1}{2N_d}\sum_{\mu=1}^{N_p}\sum_{p=1}^{N_d}\sum_{q=1}^{N_d}\lambda^{p,q}m^{p,\mu}m^{q,\mu}$$
$$-\frac{1}{\beta N_d}\sum_{p=1}^{N_d}\left\langle\left\langle \ln\left[2\cosh\left(\beta\sum_{\mu}(m^{p,\mu}-\sum_{q}\lambda^{p,q}m^{q,\mu}+J)\xi^{\mu}\right)\right]\right\rangle\right\rangle \tag{12}$$

where ξ^μ is the presynaptic potential of a typical axon and the magnetisations satisfy the stationary point condition

$$m^{p,\mu} = \left\langle\!\!\left\langle \xi^\mu \tanh\left(\beta \sum_{\nu=1}^{N_p} (m^{p,\nu} - \lambda M^\nu + J)\xi^\nu \right) \right\rangle\!\!\right\rangle. \tag{13}$$

It is straightforward to show that the magnetisations are equal to

$$m^{p,\mu} = \left\langle\!\!\left\langle \frac{1}{N_a} \sum_{i=1}^{N_a} \xi_i^\mu \langle c_i^p \rangle \right\rangle\!\!\right\rangle \tag{14}$$

where $\langle \ldots \rangle$ denotes the ensemble average over the annealed variables c_i^p. The magnetisation $m^{p,\mu}$ thus measures how closely the synaptic weights on a dendrite p correspond to a particular input pattern $\boldsymbol{\xi}^\mu$. In the rest of this paper we will for simplicity consider isotropic systems where $\lambda^{p,q} = \lambda$.

4. Zero temperature

At zero temperature ($\beta \to \infty$) the free energy becomes

$$E = \frac{-1}{2N_d} \sum_\mu \sum_p (m^{p,\mu})^2 + \frac{\lambda}{2N_d} \sum_\mu (M^\mu)^2 - \frac{J}{N_d} \sum_\mu M^\mu \tag{15}$$

where M^μ is the total magnetisation of the dendrites for a given input pattern:

$$M^\mu = \sum_{p=1}^{N_d} m^{p,\mu} \tag{16}$$

and the magnetisation $m^{p,\mu}$ satisfies

$$m^{p,\mu} = \left\langle\!\!\left\langle \operatorname{sgn}\left(\sum_{\nu=1}^{N_p} (m^{p,\nu} - \lambda M^\nu + J)\xi^\nu \xi^\mu \right) \right\rangle\!\!\right\rangle. \tag{17}$$

The equilibrium solutions of this model correspond to the minima of the energy surface of equation (15).

To illustrate the meaning of these two field equations we examine systems with a very small number of input patterns and dendrites. In these systems all the solutions can be found. The simplest non-trivial case we can examine is when we have two input patterns and two dendrites. There are only eight possible values for the magnetisations for each dendrite ($\boldsymbol{m}^p = (m^{p,1}, m^{p,2})$) which are $(\pm 1, 0)$, $(0, \pm 1)$, and $(\pm\frac{1}{2}, \pm\frac{1}{2})$, although some combinations of these magnetisations will only be solutions for specific values of the inhibition λ and the threshold J. The two possible 'grandmother cell' solutions ($\boldsymbol{m}^1 = (1,0)$, $\boldsymbol{m}^2 = (0, 1)$ and vice versa) will be solutions to these equations provided $1 - 2\lambda + 2J > 0$; this inequality comes from ensuring equation (17) is consistent. The energy for this solution is found, from a straightforward substitution of the magnetisation into equation (15), to be

$$E = -\tfrac{1}{2} + \tfrac{1}{2}\lambda - J. \tag{18}$$

By calculating the energies of all the other possible solutions one can determine which are the ground states of the system. The grandmother cell solutions will be the ground states of this system when $\lambda > 0$ and $J > \lambda/2$. If the inhibition λ is negative then the ground state consists of the two dendrites learning the same input. If $J < \lambda/2$

then one dendrite learns one of the patterns, $\boldsymbol{\xi}^1$ say, while the other dendrite learns its inverse $-\boldsymbol{\xi}^1$ (learning the inverse pattern is a peculiarity of subtractive inhibition). The mixture states $\boldsymbol{m}^p = (\pm\frac{1}{2}, \pm\frac{1}{2})$ are always local maxima.

In the case when there are three patterns and three dendrites the mixture states $\boldsymbol{m}^p = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ will be local minima of the system. Indeed this mixture state will be the ground state when $3J - 5\lambda > 0$ and $\lambda > 0$. The grandmother cell solutions ($\boldsymbol{m}^1 = (1, 0, 0)$, $\boldsymbol{m}^2 = (0, 1, 0)$, $\boldsymbol{m}^3 = (0, 0, 1)$ and the five other permutations) will be local minima provided $1 - 3\lambda + 5J > 0$. They will be the *ground-state* solutions when $4J - 5\lambda < 0$ and $5J - 2\lambda < 0$. When this second constraint is not satisfied the ground state will consist of one dendrite learning one pattern and a second the inverse pattern while the third dendrite can learn any of the three patterns.

We must also consider the case when the number of patterns differs from the number of dendrites. The easier case occurs when we underload the net. For example when we have two patterns and four dendrites then the state equivalent to the grandmother cell solutions are $\boldsymbol{m}^1 = (1, 0)$, $\boldsymbol{m}^2 = (1, 0)$, $\boldsymbol{m}^3 = (0, 1)$ and $\boldsymbol{m}^4 = (0, 1)$ and its various possible permutations. These solutions will be the ground state of the system when $\lambda > 0$ and $2J - 3\lambda > 0$. When $2J - 3\lambda < 0$ one of the dendrites will learn the inverse of a pattern until $2J - \lambda < 0$ when the system will learn two patterns and two inverse patterns.

Of more interest is when the network is overloaded and particularly when some of the patterns are correlated. To examine this we consider a network with two dendrites being presented with four patterns; the patterns form two classes where the correlation within each class is c and between members of different classes is zero (we assume patterns 1 and 2 are correlated and likewise patterns 3 and 4). In this case there are two types of solutions of interest: the 'feature detector' solution, for example $\boldsymbol{m}^1 = (1, c, 0, 0)$ and $\boldsymbol{m}^2 = (0, 0, 1, c)$, where each dendrite learns one of the four possible patterns; and the 'categoriser' solution, for example $\boldsymbol{m}^1 = \frac{1}{4}(1 + c)(2, 2, 1, 1)$ and $\boldsymbol{m}^2 = \frac{1}{4}(1 + c)(1, 1, 2, 2)$ where each dendrite learns only to differentiate between patterns of different classes. The conditions under which these solutions exist and are ground states are more complicated than before, especially as the inequalities now contain three variables. In essence, the feature detector solution will be the ground state only for small correlations and when J/λ is small, although for very small J-to-λ ratio the inverse pattern will be learned. The categoriser solution will be the lowest energy state when J/λ is large although for very large J-to-λ ratio the fully symmetric solution (each dendrite firing equal to each input pattern) will be the global minimum.

Although we have determined the ground states, what is more interesting is the size of the basin of attraction—it is this that gives the probability of learning a particular state (assuming that the initial synaptic weights are random and the patterns are applied equally often). We believe that the size of the basins of attraction are related to the depths of the energy wells. This is believed to be the case for the Hopfield model and agrees qualitatively with simulations on this model.

5. Simulations

We have carried out simulations for various different sets of input patterns. The simulations agree with the behaviour of the model described above. Finite-size effects enter most transparently by causing a small correlation between patterns and a slight bias of each pattern—these can be avoided if the number of patterns is small by using

specially constructed input patterns which have no bias and are exactly orthogonal. However, provided the inhibition and threshold constants are sufficiently large and the number of patterns sufficiently small these finite-size effects do not significantly alter the behaviour of the system from that predicted by the theory.

The graphs in figure 2 illustrate the typical behaviour of this model when random patterns are presented. In this example four random patterns are applied to a network consisting of four dendrites and 2048 axons with $\lambda = 0.05$, $J = 0.05$ and $\beta = 4.00$. The magnetisation in this case is just

$$m^{p,\mu} = \frac{1}{N_a} \sum_{i=1}^{N_a} c_i^p \xi_i^\mu . \tag{19}$$

One cycle corresponds to all the synapses being updated. Notice that each dendrite 'learns' a different pattern (i.e. grandmother cell encodes), in this case the system has found one of the global minima. The speed of learning depends on the initial state of the system although typically these simulations are observed to settle down within 5 to 30 cycles.

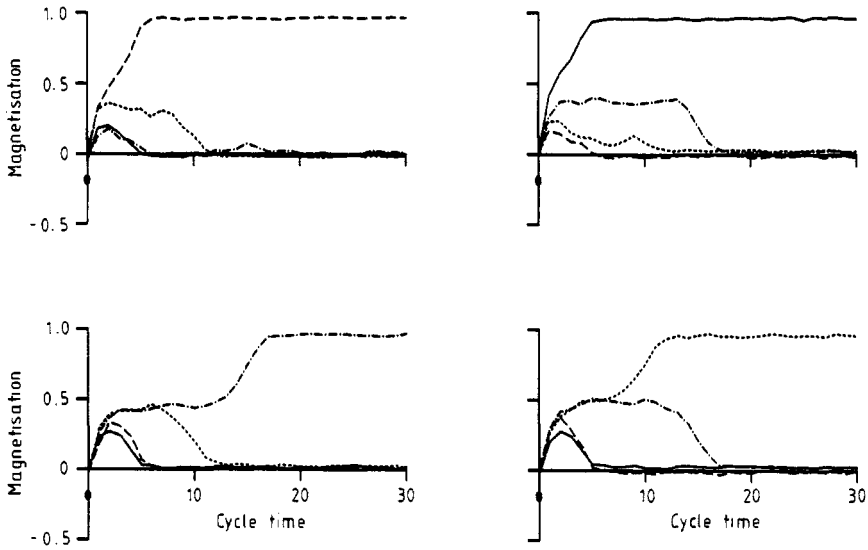

Figure 2. Typical simulation of a competitive network with four dendrites and four random patterns. Each dendrite is represented by a different graph.

Figure 3 shows a graph of the magnetisations for one of sixteen dendrites receiving sixteen patterns again with 2048 axons and $\lambda = 0.05$, $J = 0.05$ and $\beta = 4.0$. In this case the system learnt thirteen patterns only (three patterns being learnt by two dendrites). This may be a local minimum which the system is trapped in or may possibly be the global minimum for this set of patterns—this would then be a finite-size effect caused by the small biases and slight correlations between the patterns which is sufficient to alter the energy surface so that the grandmother cell solutions are not the global minima.

We can also simulate the behaviour of this model with correlated input patterns. Figure 4 shows the magnetisation for two systems both with 2048 axons and three dendrites and both experiencing nine patterns which form three classes of three correlated patterns each. For small correlations the system can be made to work as a

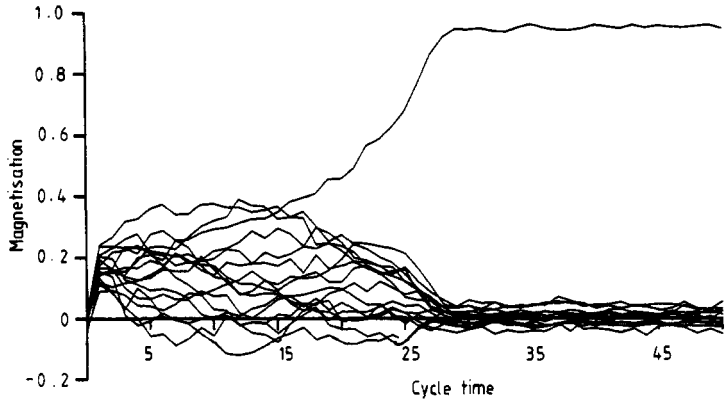

Figure 3. Simulation of a competitive network with sixteen input patterns and sixteen dendrites. The graph shows the magnetisations on just one dendrite.

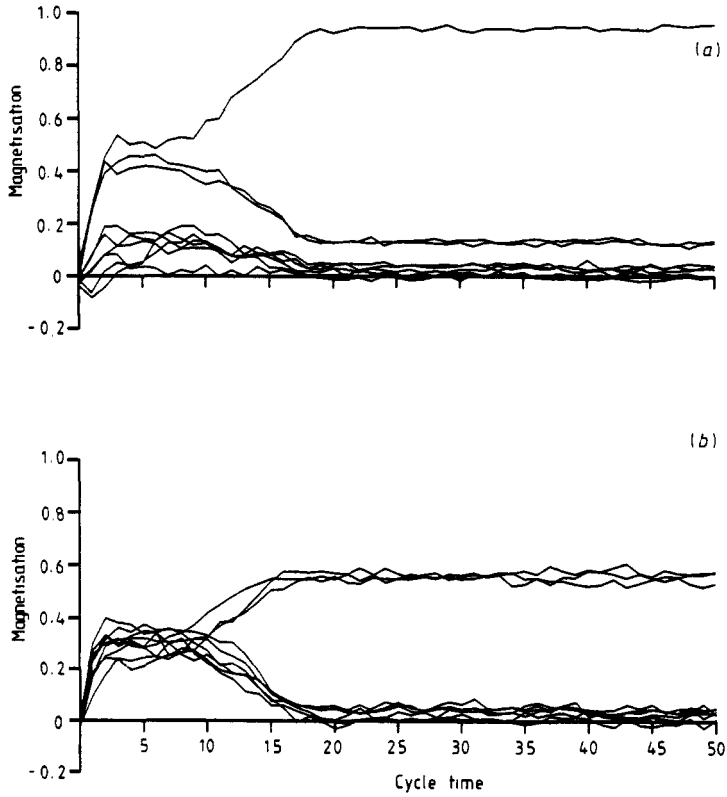

Figure 4. Two simulations of networks with three dendrites and three classes of three correlated patterns: (*a*) the interclass correlations are all around $\frac{1}{10}$, $\lambda = 0.02$, $J = 0.01$ and $\beta = 4.0$; (*b*) the interclass correlations are all around $\frac{1}{5}$, $\lambda = 0.05$, $J = 0.02$ and $\beta = 4.0$.

'feature detector'—each dendrite learns just one pattern but will of course have an overlap with the patterns in the same class. This case is shown in the top graph where the average correlation between members of the same class is 0.1 and $\lambda = 0.02$ and $J = 0.01$. If the interclass correlation is larger then the 'feature detector' will only be the ground-state solution for very small values of λ and J, but in the size of system simulated the biases that arise as finite-size effects are sufficiently large to upset this solution. A much easier solution to obtain is the 'categoriser' solution in which each dendrite learns a mixture state of all three pattern in a class. The graph in figure 4(*b*) shows an example of this solution; in this case the average correlation within a class is 0.2, $\lambda = 0.05$, $J = 0.02$ and again $\beta = 4.0$. Both the solutions shown compete with the completely mixed state (all dendrites responding equally to every input pattern) which will be the ground state for large J-to-λ ratios. For small J-to-λ ratios dendrites will tend to learn the inverse of patterns as well as the patterns themselves.

Parallel updating gives similar graphs to those shown. Increasing the temperature has the effect of removing some of the spurious states but also slows down learning. Above a certain temperature (β less than about 2.5) a 'pseudo-paramagnetic' phase is found; this is a finite-size effect, since no real paramagnetic phase exists for $J \neq 0$.

6. Discussion

The Hamiltonian of equation (1) is reminiscent of the Hopfield Hamiltonian: it is perhaps appropriate to stress some of the differences. Competitive networks can function as feature detectors (e.g. when they have grandmother cells) or as categorisers. They respond immediately to an input stimulus. The dynamics of interest for these networks are those of the synapses. The Hopfield energy describes the dynamics of the neuron firing rates for an auto-associative memory. The synaptic weights are the quenched rather than the annealed variables. The Hopfield network undergoes several iterations before it reaches a useful state. Learning mechanisms are not treated in the classical Hopfield model.

Although there are many more realistic models of competitive networks (see the list of references in the introduction) little is known about their large-scale behaviour. If, as in our model, the number of equilibrium states for the synapse grows exponentially with the size of the system then different behaviour might arise in large systems. It is hoped that this model will provide a new means for tackling these problems. When the number of patterns scales as the number of axons we can no longer ignore the correlations between the patterns. In this case the replica symmetry trick (see [11]) should be employed to solve this model (work in progress). From the close analogy with the Hopfield model we can speculate on some of the properties of large systems. (Note that when there is no inhibition the energy function is identical in form to the Hopfield energy function). In particular we would expect that there will be an upper capacity for these networks when learning random patterns (analogous to $\alpha_c N$ for the Hopfield model) above which the magnetisations are no longer close to one (i.e. grandmother-like solutions will not be found). It may also be possible to use the natural tendencies of the spin-glass phase to have an ultrametric structure of solutions to increase the capacity and allow for an effective coding of correlated input patterns (see [12]).

In understanding networks within the brain many questions must be answered. The importance of sparse and highly correlated inputs as well as low synaptic contact

probabilities must be examined. The influence of noise must also be studied. Another area that has been studied in simulations by other authors is the effect of having local inhibition (or 'clusters'). In most of this paper we have been looking at grandmother cell solutions but in the brain ensemble encoding is used (in which case we may not want the grandmother cell solutions at all), and this poses a further problem in our understanding of how competitive networks operate. Finally we must determine if this model does indeed resemble the long-term behaviour of real competitive networks. In real competitive networks the competition is achieved by a combination of the non-linear firing of neurons followed by a moderation of the firing across all the neurons caused by the inhibitory interneurons. The behaviour of this sort of net will be somewhat different; for example the inverse patterns will not be learnt. How many of these differences are crucial to the long-term behaviour has still to be determined.

We hope that this model will provide a new stimulus into acquiring a rigorous understanding of this important class of neural networks.

Acknowledgments

The work in this paper was started with Elizabeth Gardner. Sadly she died before its completion. Those people who know her work will appreciate how crucial her early collaboration was and how much this work has suffered through her loss. We both owe a debt of gratitude to Dr E T Rolls for stimulating discussions on the brain. I would also like to personally thank Professor M Virasoro for a fruitful discussion and useful comments. This work is supported by the Medical Research Council.

References

[1] von der Malsberg C 1973 *Kybernetik* **14** 85
[2] Perez R, Glass L and Shlaer R 1975 *J. Math. Biol.* **1** 275
[3] Fukushima K 1975 *Biol. Cybern.* **20** 121
[4] Grossberg S 1976 *Biol. Cybern.* **23** 121
[5] Kohonen T 1982 *Proc. 6th Int. Conf. on Pattern Recognition* ed M Lang (Silver Spring, MD: IEEE Computer Society Press) pp 114-25
[6] Rumelhart D E and Zipser D 1985 *Parallel Distributed Processing* ed J L McClelland and D E Rumelhart (Cambridge, MA: MIT Press) pp 151-93
[7] Linsker R 1986 *Proc. Natl Acad. Sci. USA.* **83** 7508, 8390, 8779
[8] Linsker R 1988 *Computer IEEE* 105
[9] Sompolinsky H 1987 *Heidelberg Colloquium on Glassy Dynamics* ed J L van Hemmen and I. Morgenstern (Berlin: Springer) pp 485-527
[10] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev.* A **32** 1007
[11] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
[12] Parga N and Virasoro M A 1986 *J. Physique* **47** 1857