Phase Diagram and Storage Capacity of Sequence Storing Neural Networks

A. Düring
Dept. of Physics
Oxford University
Oxford OX1 3NP
United Kingdom
a.during1@physics.oxford.ac.uk

A. C. C. Coolen
Dept. of Mathematics
King's College
London WC2R 2LS
United Kingdom
tcoolen@mth.kcl.ac.uk

D. Sherrington
Dept. of Physics
Oxford University
Oxford OX1 3NP
United Kingdom
d.sherrington1@physics.oxford.ac.uk

Abstract

We solve the dynamics of Hopfield–type neural networks which store sequences of patterns, close to saturation. The asymmetry of the interaction matrix in such models leads to violation of detailed balance, ruling out an equilibrium statistical mechanical analysis. Using generating functional methods we derive exact closed equations for dynamical order parameters, viz. the sequence overlap and correlation and response functions, in the limit of an infinite system size. We calculate the time translation invariant solutions of these equations, describing stationary limit–cycles, which leads to a phase diagram. The effective retarded self-interaction usually appearing in symmetric models is here found to vanish, which causes a significantly enlarged storage capacity of $\alpha_c \approx 0.269$, compared to $\alpha_c \approx 0.139$ for Hopfield networks storing static patterns. Our results are tested against extensive computer simulations and excellent agreement is found.

1 INTRODUCTION AND DEFINITIONS

We consider a system of N neurons $\boldsymbol{\sigma}(t) = \{\sigma_i(t) = \pm 1\}$, which can change their states collectively at discrete times (parallel dynamics). Each neuron changes its state with a probability $p_i(t) = \frac{1}{2}[1-\tanh\beta\sigma_i(t)[\sum_j J_{ij}\sigma_j(t)+\theta_i(t)]]$, so that the transition matrix is

$$W[\boldsymbol{\sigma}(s+1)|\boldsymbol{\sigma}(s)] = \prod_{i=1}^{N} \mathrm{e}^{\beta\sigma_i(s+1)\left[\sum_{j=1}^{N} J_{ij}\sigma_j(s)+\theta_i(s)\right]-\ln 2\cosh\left(\beta\left[\sum_{j=1}^{N} J_{ij}\sigma_j(s)+\theta_i(s)\right]\right)} \tag{1}$$

with the (non-symmetric) interaction strengths J_{ij} chosen as

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{\mu+1}\xi_j^{\mu}, \tag{2}$$

The ξ_i^{μ} represent components of an ordered sequence of patterns to be stored[1]. The gain parameter β can be interpreted as an inverse temperature governing the noise level in the dynamics (1) and the number of patterns is assumed to scale as N, i. e. $p = \alpha N$. If the interaction matrix would have been chosen symmetrically, the model would be accessible to methods originally developed for the equilibrium statistical mechanical analysis of physical spin systems and related models [1, 2], in particular the replica method. For the nonsymmetric interaction matrix proposed here this is ruled out, and no exact solution exists to our knowledge, although both models have been first mentioned at the same time and an approximate solution compatible with the numerical evidence at the time has been provided by Amari [3]. The difficulty for the analysis is that a system with the interactions (2) never reaches equilibrium in the thermodynamic sense, so that equilibrium methods are not applicable. One therefore has to apply dynamical methods and give a dynamical meaning to the notion of the recall state. Consequently, we will for this paper employ the dynamical method of path integrals, pioneered for spin glasses by de Dominicis [4] and applied to the Hopfield model by Rieger *et al.* [5].

We point out that our choice of parallel dynamics for the problem of sequence recall is deliberate in that simple sequential dynamics will not lead to stable recall of a sequence. This is due to the fact that the number of updates of a single neuron per time unit is not a constant for sequential dynamics. Schemes for using delayed asymmetric interactions combined with sequential updates have been proposed (see e. g. [6] for a review), but are outside the scope of this paper.

Our analysis starts with the introduction of a generating functional $Z[\boldsymbol{\psi}]$ of the form

$$Z[\boldsymbol{\psi}] = \sum_{\boldsymbol{\sigma}(0)\ldots\boldsymbol{\sigma}(t)} p[\boldsymbol{\sigma}(0),\ldots,\boldsymbol{\sigma}(t)]\,\mathrm{e}^{-\mathrm{i}\sum_{s<t}\boldsymbol{\sigma}(s)\cdot\boldsymbol{\psi}(s)}, \tag{3}$$

which depends on real fields $\{\psi_i(t)\}$. These fields play a formal role only, allowing for the identification of interesting order parameters, such as

$$m_i(s) = \langle\sigma_i(s)\rangle = \mathrm{i}\lim_{\boldsymbol{\psi}\to 0}\frac{\partial Z[\boldsymbol{\psi}]}{\partial\psi_i(s)}$$

$$G_{ij}(s,s') = \frac{\partial}{\partial\theta_j(s')}\langle\sigma_i(s)\rangle = \mathrm{i}\lim_{\boldsymbol{\psi}\to 0}\frac{\partial^2 Z[\boldsymbol{\psi}]}{\partial\psi_i(s)\partial\theta_j(s')}$$

$$C_{ij}(s,s') = \langle\sigma_i(s)\sigma_j(s')\rangle = -\lim_{\boldsymbol{\psi}\to 0}\frac{\partial^2 Z[\boldsymbol{\psi}]}{\partial\psi_i(s)\partial\psi_j(s')}.$$

[1] Upper (pattern) indices are understood to be taken modulo p unless otherwise stated.

for the average activation, response and correlation functions, respectively. Since this functional involves the probability $p[\boldsymbol{\sigma}(0),\ldots,\boldsymbol{\sigma}(t)]$ of finding a 'path' of neuron activations $\{\boldsymbol{\sigma}(0),\ldots,\boldsymbol{\sigma}(t)\}$, the task of the analysis is to express this probability in terms of the macroscopic order parameters itself to arrive at a set of closed macroscopic equations.

The first step in rewriting the path probability is to realise that (1) describes a one–step Markov process and the path probability is therefore just the product of the single–time transition probabilities, weighted by the probability of the initial state: $p[\boldsymbol{\sigma}(0),\ldots,\boldsymbol{\sigma}(t)] = p(\boldsymbol{\sigma}(0))\prod_{s=0}^{t-1} W[\boldsymbol{\sigma}(s+1)|\boldsymbol{\sigma}(s)]$. Furthermore, we will in the course of the analysis frequently isolate interesting variables by introducing appropriate δ-functions, such as

$$1 = \int \mathrm{d}\mathbf{h}(s) \prod_{i=1}^{N} \delta\left[h_i(s) - \left(\sum_{j=1}^{N} J_{ij}\sigma_j(s) + \theta_i(s)\right)\right]$$
$$= \int \frac{\mathrm{d}\mathbf{h}(s)\,\mathrm{d}\hat{\mathbf{h}}(s)}{(2\pi)^N} \prod_{i=1}^{N} \mathrm{e}^{\mathrm{i}\hat{h}_i(s)\left(h_i(s)-\sum_{j=1}^{N} J_{ij}\sigma_j(s)-\theta_i(s)\right)}$$

The variable $h_i(t)$ can be interpreted as the local field (or presynaptic potential) at site i and time t and their introduction transforms $Z[\psi]$ into

$$Z[\psi] = \sum_{\boldsymbol{\sigma}(0)\ldots\boldsymbol{\sigma}(t)} p(\boldsymbol{\sigma}(0)) \int \frac{\mathrm{d}\mathbf{h}\,\mathrm{d}\hat{\mathbf{h}}}{(2\pi)^{Nt}} \prod_{s=0}^{t-1} \Big[\mathrm{e}^{\beta\boldsymbol{\sigma}(s+1)\cdot\mathbf{h}(s) - \sum_i \ln 2\cosh(\beta h_i(s))}$$
$$\mathrm{e}^{\mathrm{i}\left(\hat{\mathbf{h}}(s)\cdot\mathbf{h}(s) - N^{-1}\sum_{i,j}\hat{h}_i(s)\sum_\mu \xi_i^{\mu+1}\xi_j^\mu\sigma_i(s) - \hat{\mathbf{h}}(s)\cdot\boldsymbol{\theta}(s) - \boldsymbol{\psi}(s)\cdot\boldsymbol{\sigma}(s)\right)}\Big]. \qquad (4)$$

This expression is the last general form of $Z[\psi]$ we consider. To proceed with the analysis, we have to make a specific ansatz for the system behaviour.

2 DYNAMIC MEAN FIELD THEORY

As sequence recall is the mode of operation we are most interested in, we make the ansatz that, for large systems, we have an overlap of order $\mathcal{O}\left(N^0\right)$ between the pattern $\boldsymbol{\xi}^s$ at time s, and that all other patterns are overlapping with order $\mathcal{O}\left(N^{-1/2}\right)$ at most. Accordingly, we introduce the macroscopic order parameters for the condensed pattern $m(s) = N^{-1}\sum_i \xi_i^s\sigma_i(s)$ and for the quantity $k(s) = N^{-1}\sum_i \xi_i^s\hat{h}_i(s)$, and their noncondensed equivalents $y^\mu(s) = N^{-1/2}\sum_i \xi_i^\mu\sigma_i(s)$ and $x(s) = N^{-1/2}\sum_i \xi_i^\mu\hat{h}_i(s)$ $(\mu \neq s)$, where the scaling ansatz is reflected in the normalisation constants. Introducing these objects using δ functions, as with the local fields $h_i(s)$, removes the product of two patterns in the last line of eq. (4), so that the exponent will be linear in the pattern bits.

Because macroscopic observables will in general not depend on the microscopic realisation of the patterns, the values of these observables do not change if we average $Z[\psi]$ over the realisations of the patterns. Performing this average is complicated by the occurrence of some patterns in both the condensed and the noncondensed overlaps, depending on the current time index, which is an effect not occurring in the standard Hopfield model. Using some simple scaling arguments, this difficulty can be removed and we can perform the average over the noncondensed patterns. The disorder averaged $Z[\psi]$ acquires the form

$$Z[\psi] = \int \mathrm{d}\mathbf{m}\,\mathrm{d}\hat{\mathbf{m}}\,\mathrm{d}\mathbf{k}\,\mathrm{d}\hat{\mathbf{k}}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\hat{\mathbf{q}}\,\mathrm{d}\mathbf{Q}\,\mathrm{d}\hat{\mathbf{Q}}\,\mathrm{d}\mathbf{K}\,\mathrm{d}\hat{\mathbf{K}}\,\mathrm{e}^{N\left(\Psi[\ldots]+\Phi[\ldots]+\Omega[\ldots]\right)+\mathcal{O}\left(N^{1/2}\right)} \qquad (5)$$

where we have introduced the new observables $q(s,s') = 1/N \sum_i \sigma_i(s)\sigma_i(s')$, $Q(s,s') = 1/N \sum_i \hat{h}_i(s)\hat{h}_i(s')$, and $K(s,s') = 1/N \sum_i \sigma_i(s)\hat{h}_i(s')$, and their corresponding conjugate variables. The functions in the exponent turn out to be

$$\Psi[\mathbf{m},\hat{\mathbf{m}},\mathbf{k},\hat{\mathbf{k}},\mathbf{q},\hat{\mathbf{q}},\mathbf{Q},\hat{\mathbf{Q}},\mathbf{K},\hat{\mathbf{K}}] = \mathrm{i}\sum_{s<t}\Big[\hat{m}(s)m(s) + \hat{k}(s)k(s) - m(s)k(s)\Big] + \mathrm{i}\sum_{s,s'<t}\Big[\hat{q}(s,s')q(s,s') + \hat{Q}(s,s')Q(s,s') + \hat{K}(s,s')K(s,s')\Big], \quad (6)$$

$$\Phi[\mathbf{m},\mathbf{k},\hat{\mathbf{q}},\hat{\mathbf{Q}},\hat{\mathbf{K}}] = \frac{1}{N}\sum_i \ln\Bigg[\sum_{\sigma(0)\ldots\sigma(t)} p_i(\sigma(0)) \int \prod_{s<t}\left[\frac{\mathrm{d}h(s)\,\mathrm{d}\hat{h}(s)}{2\pi}\right] e^{\sum_{s<t}\left[\beta\sigma(s+1)h(s) - \ln 2\cosh(\beta h(s))\right]} \times e^{-\mathrm{i}\sum_{s,s'<t}\left[\hat{q}(s,s')\sigma(s)\sigma(s') + \hat{Q}(s,s')\hat{h}(s)\hat{h}(s') + \hat{K}(s,s')\sigma(s)\hat{h}(s')\right]} \times e^{\mathrm{i}\sum_{s<t}\hat{h}(s)\left[h(s) - \theta_i(s) - \hat{k}(s)\xi_i^{s+1}\right] - i\sum_{s<t}\sigma(s)\left[\hat{m}(s)\xi_i^s + \psi_i(s)\right]}\Bigg], \quad (7)$$

and

$$\Omega[\mathbf{q},\mathbf{Q},\hat{\mathbf{Q}}] = \frac{1}{N}\ln\int\prod_{s<t}\left[\frac{\mathrm{d}u(s)\,\mathrm{d}v(s)}{(2\pi)^{(p-t)}}\right] e^{\mathrm{i}\sum_{\mu>t}\sum_{s<t}u_{\mu+1}(s)v_\mu(s)} \times e^{-\frac{1}{2}\sum_{\mu>t}\sum_{s,s'<t}\left[u_\mu(s)Q(s,s')u_\mu(s') + u_\mu(s)K(s',s)v_\mu(s') + v_\mu(s)K(s,s')u_\mu(s') + v_\mu(s)q(s,s')v_\mu(s')\right]}. \quad (8)$$

The first of these expressions is just a result of the introduction of δ functions, while the second will turn out to represent a probability measure given by the evolution of a *single* neuron under *prescribed* fields and the third reflects the disorder contribution to the local fields in that single neuron measure[2]. We have thus reduced the original problem involving N neurons in a one-step Markov process to one involving just a single neuron, but at the cost of introducing two–time observables.

3 DERIVATION OF SADDLE POINT EQUATIONS

The integral in (5) will be dominated by saddle points, in our case by a unique saddle point when causality is taken into account. Extremising the exponent with respect to all occurring variables gives a number of equations, the most important of which give the physical meanings of three observables: $q(s,s') = C(s,s')$, $K(s,s') = iG(s,s')$,

$$m(s) = \lim_{N\to\infty}\frac{1}{N}\sum_i \overline{\langle\sigma_i(s)\rangle}\xi_i^s \quad (9)$$

with

$$C(s,s') = \lim_{N\to\infty}\frac{1}{N}\sum_i \overline{\langle\sigma_i(s)\sigma_i(s')\rangle} \qquad G(s,s') = \lim_{N\to\infty}\frac{1}{N}\sum_i \frac{\partial\overline{\langle\sigma_i(s)\rangle}}{\partial\theta_i(s')}. \quad (10)$$

[2]We have assumed $p(\boldsymbol{\sigma}(0)) = \prod_i p_i(\sigma_i(0))$.

which are the single-site correlation and response functions, respectively. The overline $\overline{\ldots}$ is taken to represent disorder averaged values. Using also additional equations arising from the normalisation $Z[0] = 1$, we can rewrite the single neuron measure Φ as

$$\langle f[\{\sigma\}]\rangle_* = \sum_{\sigma_0 \ldots \sigma(t)} \int \prod_{s<t} \left[\frac{\mathrm{d}h(s)\,\mathrm{d}\hat{h}(s)}{2\pi}\right] p(\sigma(0)) f[\{\sigma\}] \mathrm{e}^{\sum_{s<t}\left[\beta\sigma(s+1)h(s) - \ln 2\cosh(\beta h(s))\right]}$$
$$\times\, \mathrm{e}^{\mathrm{i}\sum_{s<t}\hat{h}(s)\left[h(s)-\theta(s)-m(s)\right] - \frac{1}{2}\alpha\sum_{s,s'<t} R(s,s')\hat{h}(s)\hat{h}(s')} \qquad (11)$$

with the short-hand $\mathbf{R} = \sum_{l=0}^{\infty} \mathbf{G}^{\dagger l}\mathbf{C}\mathbf{G}^{l}$. To simplify notation, we have here assumed that the initial probabilities $p_i(\sigma_i(0))$ are uniform and that the external fields $\theta_i(s)$ are so-called staggered ones, i. e. $\theta_i(s) = \theta\xi_i^{s+1}$, which makes the single neuron measure site–independent. This single neuron measure (11) represents the essential result of our calculations and is already properly normalised (i.e. $\langle 1\rangle_* = 1$).

When one compares the present form of the single neuron measure with that obtained for the symmetric Hopfield network, one finds in the latter model an additional term which corresponds to a retarded self–interaction. The absence of such a term here suggests that the present model will have a higher storage capacity. It can be explained by the constant change of state of a large number of neurons as the network goes through the sequence, which prevents the build-up of microscopic memory of past activations.

However, as is the case for the standard Hopfield model, the measure (11) is still too complicated to find explicit equations for the observables we are interested in. Although it is possible to evaluate the necessary integrals numerically, we instead concentrate on the interesting behaviour when transients have died out and time–translation invariance is present.

4 STATIONARY STATE

We will now concentrate on the behaviour of the network at the stage when transients have subsided and the system is on a macroscopic limit cycle. Then the relations

$$m(s) = m \qquad C(s,s') = C(s-s') \qquad G(s,s') = C(s-s'). \qquad (12)$$

hold and also $R(s,s') = R(s-s')$. We can then for simplicity shift the time origin $t_0 = -\infty$ and the upper temporal bound to $t = \infty$. Note, however, that this state is not to be confused with microscopic equilibrium in the thermodynamic sense. The stationary versions of the measure (11) for the interesting observables are then given by the following expressions (note that $C(0) = 1$):

$$m = \int \prod_s \frac{\mathrm{d}v(s)\,\mathrm{d}w(s)}{2\pi} \mathrm{e}^{\mathrm{i}\mathbf{v}\cdot\mathbf{w} - \frac{1}{2}\mathbf{w}\cdot\mathbf{R}\mathbf{w}} \tanh\beta\left[m + \theta + \alpha^{\frac{1}{2}} v(0)\right]$$

$$C(\tau \neq 0) = \int \prod_s \frac{\mathrm{d}v(s)\,\mathrm{d}w(s)}{2\pi} \mathrm{e}^{\mathrm{i}\mathbf{v}\cdot\mathbf{w} - \frac{1}{2}\mathbf{w}\cdot\mathbf{R}\mathbf{w}} \times$$
$$\tanh\beta\left[m + \theta + \alpha^{\frac{1}{2}} v(\tau)\right] \tanh\beta\left[m + \theta + \alpha^{\frac{1}{2}} v(0)\right]$$

$$G(\tau) = \beta\delta_{\tau,1}\left[1 - \int \prod_s \frac{\mathrm{d}v(s)\,\mathrm{d}w(s)}{2\pi} \mathrm{e}^{\mathrm{i}\mathbf{v}\cdot\mathbf{w} - \frac{1}{2}\mathbf{w}\cdot\mathbf{R}\mathbf{w}} \tanh^2\beta\left[m + \theta + \alpha^{\frac{1}{2}} v(0)\right]\right] \qquad (13)$$

and we notice that the response function is now limited to a single time step, which again reflects the influence of the uncorrelated flips induced by the sequence recall. These equations can be solved by separating the persistent and fluctuating parts of $C(\tau)$ and $R(\tau)$,

$$C(\tau) = q + \tilde{C}(\tau), \qquad R(\tau) = r + \tilde{R}(\tau), \qquad \lim_{\tau=\pm\infty} \tilde{C}(\tau) = \lim_{\tau=\pm\infty} \tilde{R}(\tau) = 0.$$

Doing so eventually leads us to the coupled equations

$$\rho = \left[1 - \beta^2(1-\tilde{q})^2\right]^{-1} \tag{14}$$

$$m = \int \mathrm{D}z \, \tanh\beta\left[m + \theta + z\sqrt{\alpha\rho}\right] \tag{15}$$

$$\tilde{q} = \int \mathrm{D}z \, \tanh^2\beta\left[m + \theta + z\sqrt{\alpha\rho}\right] \tag{16}$$

$$q = \int \mathrm{D}z \left[\int \mathrm{D}x \, \tanh\beta\left[m + \theta + z\sqrt{\alpha q\rho} + x\sqrt{\alpha(1-q)\rho}\right]\right]^2. \tag{17}$$

Note that the three equations (14—16) form a closed set, from which the persistent correlation q simply follows.

5 PHASE DIAGRAM AND STORAGE CAPACITY

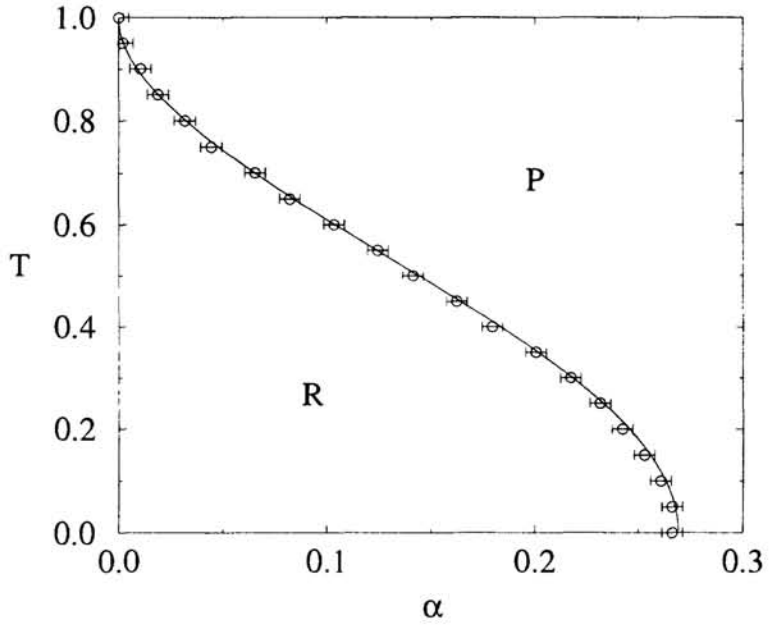

Figure 1: Phase diagram of the sequence storage network, in which one finds two phases: a recall phase (R), characterized by $\{m \neq 0,\ q > 0,\ \tilde{q} > 0\}$, and a paramagnetic phase (P), characterized by $\{m = 0,\ q = 0,\ \tilde{q} > 0\}$. The solid line separating the two phases is the theoretical prediction for the (discontinuous) phase transition. The markers represent simulation results, for systems of $N = 10,000$ neurons measured after $2,500$ iteration steps, and obtained by bisection in α. The precision in terms of α is at least $\Delta\alpha = 0.005$ (indicated by error bars); the values for T are exact.

The coupled equations (14—17) can be solved numerically for $\theta = 0$ to find the area in the α—T plane where solutions $m \neq 0$ — corresponding to sequence recall — exist. The boundary of this area describes the storage capacity of the system. This theoretical curve can then be compared with computer simulations directly performing the neural dynamics

given by (1) and (2). We show the result of doing both in the same accompanying diagram. We find that there are only two types of solutions, namely a recall phase R where $m \neq 0$ and $q \neq 0$, and a paramagnetic phase where $m = q = 0$. Unlike the standard Hopfield model, the present model does not have a spin glass phase with $m = 0$ and $q \neq 0$. The agreement between simulations (done here for $N = 10,000$ neurons) and theoretical results is excellent and separate simulations of systems with up to $N = 50,000$ neurons to assess finite size effects confirm that the numerical data are reliable.

6 DISCUSSION

In this paper, we have used path integral methods to solve in the infinite system size limit the dynamics of a non-symmetric neural network model, designed to store and recall a sequence of patterns, close to saturation. This model has been known for over a decade from numerical simulations to possess a storage capacity roughly twice that of the symmetric Hopfield model, but no rigorous analytic results were available. We find here that in contrast to equilibrium statistical mechanical methods, which do not apply due to the absence of detailed balance, the powerful path integral formalism provides us with a solution and a transparent explanation of the increased storage capacity. It turns out that this higher capacity is due to the absence of a retarded self-interaction, viz. the absence of microscopic memory of activations.

The theoretically obtained phase diagram can be compared to the results of numerical simulations and we find excellent agreement. Our confidence in this agreement is supported by additional simulations to study the effect of finite size scaling. Full details of the calculations will be presented elsewhere [7].

References

[1] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1972

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

[3] Amari S and Maginu K 1988 *Neural Networks* **1** 63

[4] de Dominicis G 1978 *Phys. Rev. B* **18** 4913

[5] Rieger H, Schreckenberg M, and Zittartz J 1988 *J. Phys. A: Math. Gen.* **21** L263

[6] Kühn R and van Hemmen J L 1991 *Temporal Association* ed E Domany, J L van Hemmen, and K Schulten (Berlin, Heidelberg: Springer) p 213

[7] Düring A, Coolen A C C, and Sherrington D 1998 *J. Phys. A: Math. Gen.* **31** 8607