Statistical Neurodynamics for sequence processing neural networks with finite dilution⋆

Pan Zhang ⋆⋆ and Yong Chen⋆ ⋆ ⋆

Institute of Theoretical Physics, Lanzhou University, 730000 Lanzhou, China

Abstract. We extend the statistical neurodynamics to study transient dynamics of sequence processing neural networks with finite dilution, and the theoretical results are supported by extensive numerical simulations. It is found that the order parameter equations are completely equivalent to those of the Generating Functional Method, which means that crosstalk noise follows normal distribution even in the case of failure in retrieval process. In order to verify the gaussian assumption of crosstalk noise, we numerically obtain the cumulants of crosstalk noise, and third- and fourth-order cumulants are found to be indeed zero even in non-retrieval case.

1 Introduction

Models of attractor neural networks for processing sequences of patterns, as a realization of a temporal association, have been of great interest over some time [1–5]. On contrary to Hopfield model [6, 7], the asymmetry of the interaction matrix in this model leads to violation of detailed balance, ruling out an equilibrium statistical mechanics analysis. Usually, Generating Functional Method and Statistical Neurodynamics are two ways to study this model. Generating Functional Method [2, 13–15] is also known as e.g. path integral method, and it allows the exact solution of the dynamics and to search all relevant physical order parameters at any time step via the derivatives of a generating functional. Statistical Neurodynamics (so called Signal-To-Noise Analysis) [3, 8, 9, 12], starts from splitting of the local field into a signal part originating from the pattern to be retrieved and a noise part arising from the other patterns, and uses simple assumption that crosstalk noise is normally distributed with zero mean and evaluated variance. In Hopfield model, this treatment is proved to be just an approximation and the Gaussian form of crosstalk noise only holds when pattern recall occurs [10, 11, 17]. But Kawamuran et al. suggest that Statistical Neurodynamics is exact in fully connected sequence processing model [3]. In this paper we show that treatment of Statistical Neurodynamics is also exact correct in the case of diluted connection.

⋆⋆ Email: july.lzu@gmail.com
⋆ ⋆ ⋆ Corresponding author. Email: ychen@lzu.edu.cn

In the nature of brain, fully-connected structure of network is biological un-realistic. Therefore, as a relaxation of conventional, unrealistic condition that every neuron is connected to every other, diluted networks have received lots of attention [18–22]. There are several kinds of dilution, with different aver-age connections cN, where c is connection probability. If $cN = O(\ln n)$, it is called extreme dilution, and in this case, network has a local Caley-Tree struc-ture and almost all pairs of neurons have entirely different sets of ancestors. Thus, the correlations in noise may be neglected. The reported first that the extremely diluted asymmetric network can be calculated exactly is Derrida et al [19]. Extremely diluted symmetric Hopfield network was also studied using replica-symmetric calculation. When $cN = O(1)$, network only have finite con-nection regardless of its scale, and the equilibrium properties has been calculated using replica-symmetric approximation [23]. In this work, we focus on the case that $cN = O(N)$, and we will show that our theoretical result also suitable for extremely diluted network $cN \ll N$. Recently, Theumann [4] studied this type of dilution using Generating Functional Method, and we will discuss the relationship between our result and those obtained by Theumann.

This paper was organized as follows. In Section 2 we recall the sequence processing neural networks with finite dilution. In Section 3 we use statistical neurodynamics to obtain the order parameter equations which describe dynam-ics of network, and compare our theoretical result with numerical simulations. Section 4 discuss the relationship between our results and those obtained by Generating Functional Method. Section 5 contains conclusion and discussion.

2   Model definition

Let us consider a sequence processing model that consists of N spins of neurons, when $N \to \infty$. The state of spins takes $S_i(t) = \pm 1$ and updates the state synchronously with following probability [24]:

$$\mathbf{Prob}[S_i(t+1)\,|h_i(t)] = \frac{1}{2}[1 + S_i(t+1)\tanh \beta h_i(t)], \tag{1}$$

where $\beta = 1/T$ is inverse temperature and local field is given by

$$h_i(t) = \sum_{j=1}^{N} J_{ij} S_j(t). \tag{2}$$

When we use $F(\cdot)$ to express transfer function, the parallel dynamics is expressed by

$$S_i(t+1) = F(h_i(t)). \tag{3}$$

We store $P = \alpha N$ random patterns $\xi^\mu = (\xi_1^\mu, ..., \xi_N^\mu)$ in network, where α is the loading ratio. Interaction matrix J_{ij} is chosen to retrieve the patterns as $\xi^1 \to \xi^2 \to ...\xi^P \to \xi^1$ sequentially. For instance, it is given by [5]

$$J_{ij} = \frac{c_{ij}}{cN} \sum_{\mu=1}^{P} \xi_i^{\mu+1} \xi_j^\mu, \tag{4}$$

with $\xi^{(P+1)} = \xi^1$. c_{ij} takes its value as 0 or 1 probabilities

$$\mathbf{Prob}\,(c_{ij} = 1) = 1 - \mathbf{Prob}\,(c_{ij} = 0) = c \qquad (5)$$

with $(0 \leq c \leq 1)$ and $c_{ij} = c_{ji}$ for symmetric dilution.

Let us consider the case to retrieve ξ^q. We define $m^q\,(t)$ as the overlap parameter between the network state $\mathbf{S}\,(t)$ and the condensed pattern ξ^q

$$m^q\,(t) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^q S_i\,(t)\,. \qquad (6)$$

3 Statistical Neurodynamic for diluted networks

We start from Somplinsky's idea that in the limit $N \to \infty$ the synaptic matrix in (4) can be written as a fully connected model with synaptic noise [22]

$$J_{ij}^{\text{eff}} = \frac{1}{N} \sum_{\mu=0}^{\alpha N} \xi_i^{\mu+1} \xi_j^\mu + \eta_{ij}, \qquad (7)$$

where η_{ij} is a complex random variable following Gaussian distribution with mean 0 and variance $\alpha(1 - c)/c$. Theumann also proved this relationship by taking expansion of averaging $e^{-i\hat{\mathbf{h}}(t)\cdot\mathbf{J}\mathbf{S}(t)}$ over disorder, where $\hat{\mathbf{h}}$ is auxiliary local field (see section 4).

Then local field is decribed by

$$h_i\,(t) = \xi_i^{q+1} m^q\,(t) + Z_i\,(t)\,. \qquad (8)$$

Here $Z_i\,(t)$ denotes the crosstalk noise term from uncondensed patterns,

$$Z_i\,(t) = \frac{1}{N} \sum_{\mu \neq q}^{\alpha N} \sum_{j \neq i}^{N} \xi_i^{\mu+1} \xi_j^\mu S_j\,(t) + \sum_{j=1}^{N} S_j\,(t)\,\eta_{ij}. \qquad (9)$$

We assume that the noise term $Z_i\,(t)$ is normally distributed with mean 0 and variance $\sigma_i^2\,(Z_i\,(t))$. For Hopfield model, the assumption is shown to be valid within statistical errors by Monte Carlo simulations as long as the memory retrieval is successful [10]. And for sequence processing neural networks, the assumptions is shown to be hold even in the non-retrieval case [3]. We also keep this assumption in our model with presence of c_{ij}, which is verified by numerical simulations of cumulants of $Z_i\,(t)$.

To obtain the variance $\sigma^2\,(Z_i\,(t+1))$ in a recursive form, we express the noise term of the transfer function $F\,(\cdot)$,

$$Z_i\,(t+1) = \frac{1}{N} \sum_{\mu \neq q}^{\alpha N} \sum_{j \neq i}^{N} \xi_i^{\mu+1} \xi_j^\mu F\,(h_j\,(t)) + \sum_{j=1}^{N} F\,(h_j\,(t))\,\eta_{ij}$$

$$= Y_i\,(t+1) + \sum_{j=1}^{N} F\,(h_j\,(t))\,\eta_{ij}, \qquad (10)$$

where mean of $Y_i(t+1)$ is zero, and

$$h_j(t) = \xi_j^q m^{q-1}(t) + \frac{1}{N}\sum_{k\neq j}^{N}\sum_{\nu\neq q-1}^{\alpha N}\xi_j^{\nu+1}\xi_k^{\nu}S_k(t-1) + \sum_j^N S_k(t-1)\eta_{jk}. \quad (11)$$

Since $h_j(t)$ depends strongly on ξ_j^μ, and the term $\frac{1}{N}\sum_{k\neq j}^{N}\xi_j^\mu \xi_k^{\mu-1}S_k(t)$ can be rationally considered as stochastic order $O(1/\sqrt{N})$, one can expand the transfer function $F(\cdot)$ up to the first order, then obtain

$$Y_i(t+1) = \frac{1}{N}\sum_{j\neq i}^{N}\sum_{\mu\neq q}^{\alpha N}\xi_i^{\mu+1}\xi_j^\mu F\left(\hat{h}_j(t)\right) + U(t+1)Y_i(t), \quad (12)$$

where

$$\hat{h}_j(t) = \xi_j^q m^{q-1}(t) + \frac{1}{N}\sum_{k\neq j}^{N}\sum_{\nu\neq q-1,\mu}^{\alpha N}\xi_j^{\nu+1}\xi_k^{\nu}S_k(t-1) + \sum_j^N S_k(t-1)\eta_{jk} \quad (13)$$

and

$$U(t+1) = \frac{1}{N}\sum_{j\neq i}^{N}F'\left(\hat{h}_j(t)\right). \quad (14)$$

Since crosstalk noise is assumed to be zero mean, one can calculate variance of crosstalk noise directly by

$$\sigma^2(Z_i(t+1)) = \sigma^2(Y_i(t+1)) + \alpha(1-c)/c. \quad (15)$$

Firstly, we have to determine $\sigma^2(Y_i(t+1))$,

$$\sigma^2(Y_i(t+1)) = \frac{1}{N^2}E\left[\sum_j\sum_\mu\left(\xi_i^{\mu+1}\right)^2\left(\xi_j^\mu\right)^2\left(F\left(\hat{h}_j(t)\right)\right)^2\right] + U^2(t+1)\sigma^2(Y_i(t))$$

$$+ E\left[2\frac{U(t+1)}{N}\sum_{j,k}\sum_{\mu,\nu}\xi_i^{\mu+1}\xi_i^{\nu+1}\xi_j^\mu \hat{S}_j(t+1)\xi_k^{\nu-1}S_k(t)\right], \quad (16)$$

where we use $\hat{S}_j(t+1) = F\left(\hat{h}_j(t)\right)$. According to the literature [3, 12], when expand $S_k(t)$ up to $t = 0$, the last term in Eq. (16) becomes

$$\frac{2}{N}E\left[\sum_{n=1}^{t+1}\left(\prod_{\tau=1}^{n}U(t+1-\tau)\sum_j\hat{S}_j(t+1)\hat{S}_j(t+1-n)\right)\sum_\mu\left(\xi_i^{\mu+1}\right)^2\xi_j^\mu\xi_j^{\mu-n}\right]. \quad (17)$$

Here all terms in Eq. (17) are independent of each other except for $n = \alpha N, 2\alpha N, 3\alpha N....$ So when $N \to \infty$, the dependence vanishes, and the last term in Eq. (16) is ignored. The variance of $Y_i(t+1)$ term becomes

$$\sigma_i^2(Y_i(t+1)) = \alpha + U^2(t+1)\sigma_i^2(Y_i(t)). \quad (18)$$

When variance of crosstalk noise is determined using Eqs. (15, 18), all order parameters can be expressed by the following closed equations

$$m\left(t+1\right) = \int Dz \left\langle \xi F\left(\xi m\left(t\right) + \sigma\left(Z_i\left(t\right)\right)z\right)\right\rangle_\xi, \tag{19}$$

$$U\left(t+1\right) = \frac{1}{\sigma\left(Z_i\left(t\right)\right)}\int Dz z \left\langle F\left(\xi m\left(t\right) + \sigma\left(Z_i\left(t\right)\right)z\right)\right\rangle_\xi, \tag{20}$$

$$\sigma^2\left(Z_i\left(t\right)\right) = \frac{\alpha}{c} + U^2\left(t\right)\left(\sigma^2\left(Z_i\left(t-1\right)\right) - \alpha\left(1-c\right)/c\right), \tag{21}$$

where $\left\langle\cdot\right\rangle_\xi$ stands for the average over the retrieval pattern ξ, and $Dz = \frac{1}{\sqrt{2\pi}}exp\left(-z^2/2\right)dz$.

In the case that the temperature is absolute zero, $F\left(\cdot\right) = sgn\left(\cdot\right)$, we get following expressions of order parameters

$$m\left(t+1\right) = \mathtt{erf}\left(\frac{m\left(t\right)}{\sigma\left(Z_i\left(t\right)\right)}\right), \tag{22}$$

$$U\left(t+1\right) = \frac{1}{\sigma\left(Z_i\left(t\right)\right)}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{m^2}{2\sigma^2\left(Z_i\left(t\right)\right)}\right), \tag{23}$$

where

$$\mathtt{erf}\left(u\right) = \sqrt{2/\pi}\int_0^u \exp\left(-x^2/2\right)dx. \tag{24}$$

This finishes statistical neurodynamics treatment of Sequence Processing Neural Networks with finite synaptic dilution. The above equations form a recursive scheme to calculate the dynamical properties of the systems with an arbitrary time step. The time evolution of overlaps obtained both in theory and numerical simulation are plotted in Fig. 1. Initial overlaps range from 0.1 to 1.0. when initial overlap is smaller than 0.5, network will fail to retrieval, and overlap will finally vanishes. Fig. 2 shows the basin of attraction both in theory and in numerical simulations.

4   Generating Functional Method and Statistical Neurodynamics

The idea of generating functional method is to concentrate on the moment generating function $Z\left[\psi\right]$, which fully captures the statistics of paths,

$$Z\left[\psi\right] = \sum_{\sigma(0),...,\sigma(t)} P\left[\sigma\left(0\right),...,\sigma\left(t\right)\right] e^{-i\sum_{s<t}\mathbf{S}(s)\cdot\psi(s)}. \tag{25}$$

The generating function $Z\left[\psi\right]$ involves the overlap parameter $m\left(t\right)$. The response functions $G\left(t,t'\right)$ and the correlation functions $C\left(t,t'\right)$ are

$$m\left(t\right) = i\lim_{\psi\to 0}\frac{1}{N}\sum_{i=1}^N \xi_i^t \frac{\partial Z\left[\psi\right]}{\partial\psi_i\left(t\right)} = \frac{1}{N}\sum_{i=1}^N \xi_i^t \left\langle S_i\left(t\right)\right\rangle, \tag{26}$$
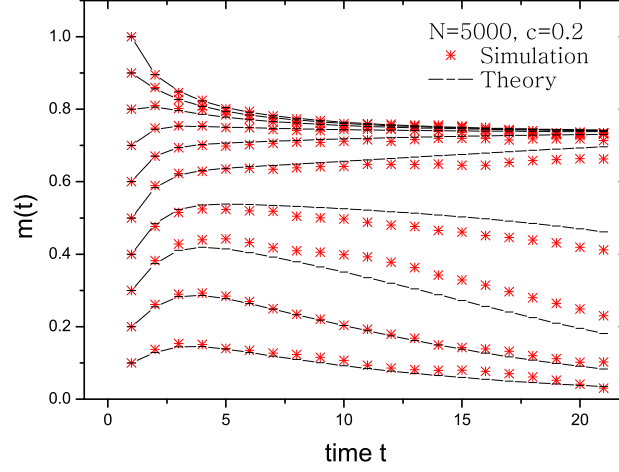
Fig. 1. Temporal evolution of overlap $m(t)$, initial overlap ranging from 1.0 to 0.1 (up to down). The parameters of networks are $N = 5000$, $c = 0.2$ and $\alpha/c = 0.38$.
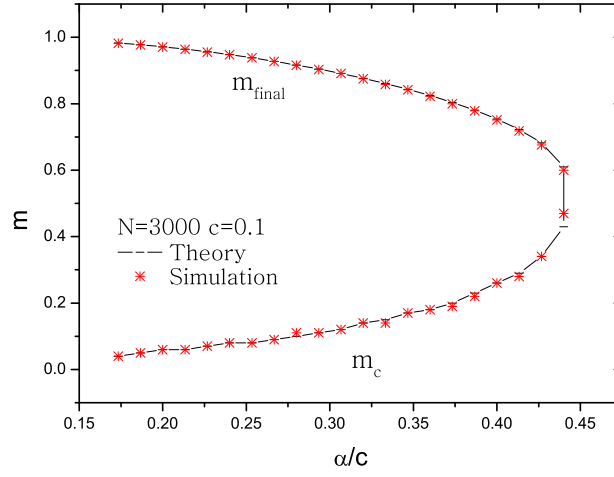

Fig. 2. Basin of attraction obtained by theory and numerical simulations. The parameters of networks are $N = 5000$ and $c = 0.1$.

$$G\left(t,t'\right) = i \lim_{\psi \to 0} \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2 Z\left[\psi\right]}{\partial \psi_i\left(t\right) \partial \theta_i\left(t'\right)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \langle S_i\left(t\right)\rangle}{\partial \theta_i\left(t'\right)}, \qquad (27)$$

$$C\left(t,t'\right) = -\lim_{\psi \to 0} \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2 Z\left[\psi\right]}{\partial \psi_i\left(t\right) \partial \psi_i\left(t'\right)} = \frac{1}{N} \sum_{i=1}^{N} \langle S_i\left(t\right) S_i\left(t'\right)\rangle. \qquad (28)$$

Düring et al. first discussed the sequence processing model using Generating Functional Method and obtained dynamical equations in the form of a multiple Gaussian integral [2], which is too complex to calculate. Then Kawamura et al. simplified those equations and obtained a tractable description of dynamical equations with single Gaussian integral [3]. Theumann discussed the case of finite dilution using Generating Functional Method [4], with idea that

$$\left\langle e^{-i\hat{\mathbf{h}}(s) \cdot \mathbf{J}^{\mathbf{eff}} \mathbf{S}(t)} \right\rangle_{c_{ij}} = e^{-i\hat{\mathbf{h}}(t) \cdot \mathbf{J} \mathbf{S}(t)} e^{-\Delta^2 \hat{\mathbf{h}}^2(t)/2}, \qquad (29)$$

where $\Delta^2 = \alpha\left(1 - c\right)/c$ presents variance of independent Gaussian random variables, which is exactly the idea of Sompolinsky that $J_{ij}^{eff} = J_{ij} + \eta_{ij}$ [22]. With this formula, Theumann obtained the temporal evolution of order parameters using the same scheme to that in [2, 3],

$$m\left(t\right) = \int Dz \left\langle tanh\beta \left[m\left(t-1\right) + \theta\left(t-1\right) + z\sqrt{\alpha D\left(t-1, t-1\right)}\right]\right\rangle_{\xi}, \qquad (30)$$

$$G\left(t, t-1\right) = \beta \left\{1 - \int Dz \left\langle tanh^2\beta \left[m\left(t-1\right) + \theta\left(t-1\right) + z\sqrt{\alpha D\left(t-1, t-1\right)}\right]\right\rangle_{\xi}\right\}, \qquad (31)$$

and $C\left(t,t\right) = 1$, where $Dz = 1/\sqrt{2\pi} \exp\left(-z^2/2\right)$. Covariance matrix of crosstalk noise is given by

$$D\left(t,t\right) = R\left(t,t\right) + \left(1 - c\right)/c, \qquad (32)$$

$$R\left(t,t\right) = 1 + G^2\left(t, t-1\right) R\left(t-1, t-1\right). \qquad (33)$$

From Eq. (19) to Eq. (30), Note that $\sigma^2\left(Z_i\left(t\right)\right)$ corresponds to $\alpha D\left(t,t'\right)$, and $U\left(t\right)$ corresponds to $G\left(t, t-1\right)$. It is easy to find that Statistical Neurodynamics and Generating Functional Method present the same order parameter equations of temporal evolution. It means that the Gaussian form of crosstalk noise holds and Statistical Neurodynamics can give the exact solution, comparing with Hopfield model that the crosstalk noise is normally distributed only in retrieval case [10]. To verify the distribution of crosstalk noise, the first, second, third, and fourth cumulants $c_1\left(t\right), c_2\left(t\right), c_3\left(t\right), c_4\left(t\right)$ are evaluated numerically, and the third and fourth cumulants are found to be zero even when network fails in retrieval (see Fig. 3).

5   Conclusion and discussion

In this paper, Statistical Neurodynamics is extended to study the retrieval dynamics of sequence processing neural networks with random synaptic dilution.
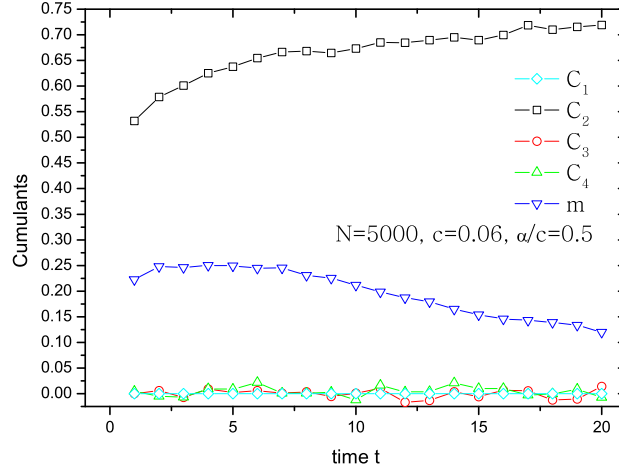
Fig. 3. Temporal evolution of cumulants $C_1(t)$, $C_2(t)$, $C_3(t)$, $C_4(t)$, and overlap $m(t)$. The initial overlap is $m(0) = 0.2$. The parameters of networks are $N = 5000$, $c = 0.06$, and $\alpha/c = 0.5$.

Our theoretical results are exactly consistent with numerical simulations. The order parameter equations are obtained which is complete equivalent to those obtained by Generating Functional Method. We also present the first, second, third and forth-order cumulants of crosstalk noise to verify the Gaussian distribution of noise.

Finally, note that, in fully connected network or extremely diluted network, one can also obtain the order parameter equations from our Eqs. (19-21). For $c = 1$, $\sigma^2(Z_i(t)) = \sigma^2(Y_i(t))$ and equation (21) corresponds to the order parameter equations in fully connected networks [3]. For the limit $cN \ll N$, $\alpha \ll \alpha/c$, then the variance of crosstalk noise is always α/c, that is exactly the result obtained in extremely diluted network where the local Cayley-Tree structure is held and all correlations in noise are neglected [19].

Acknowledgment

References

1. Sompolinsky H., Kanter I.: Temporal Association in Aymmetric Neural Networks. Phys. Rev. Lett. **57** (1986) 2861-2864.
2. Düring A., Coolen A. C. C., Sherrington D.: Phase Diagram And Storage Capacity of Sequence Processing Neural Networks. J. Phys. A: Math. Gen. **31** (1998) 8607-8621.
3. Kawamura M., Okada M.: Transient Dynamics for Sequence Processing Neural Networks . J. Phys. A: Math. Gen. **35** (2002) 253-266.
4. Theumann W. K.: Mean-field Dynamics of Sequence Processing Neural Networks with Finite Connectivity. Physica A. **328** (2003) 1-12.
5. Yong C., Hai W. Y., Qing, Y. K.: The Attractors in Sequence Processing Neural Networks, Int. J. Modern Phys. C **11** (2000) 33-39.
6. Hopfield J. J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. Proc. Nat. Acad. Sci. **79** (1982) 2554-2558.
7. Amit D. J., Gutfreund H., Sompolinsky H.: Spin-glass Models of Neural Networks. Phys. Rev. A. **32** (1985) 1007-1018.
8. Amari S.: Statistical Neurodynamics of Associative Memory. Proc. IEEE Conference on Neural Networks. **1** (1988) 633-640.
9. Okada M.: A Hierarchy of Macrodynamical Equations for Associative Memory. Neural Networks. **8** (1995) 833-838.
10. Nishimori H., Ozeki T.: Retrieval Dynamics of Associative Memory of the Hopfield Type. J. Phys. A: Math. Gen. **26** (1993) 859-871.
11. Ozeki T., Nishimori, H.: Noise Distributions in Retrieval Dynamics of the Hopfield Model. J. Phys. A: Math. Gen. **27** (1994) 7061-7068.
12. Kitano K., Aoyagi T.: Retrieval Dynamics of Neural Networks for Sparsely Coded Sequential Patterns. J. Phys. A:Math. Gen. **31** (1998) L613-L620.
13. Gardner E., Derrida B., Mottishaw, P.: Zero Temperature Parallel Dynamics for Infinite Range Spin Glasses and Neural Networks. J. Physique **48** (1987) 741-755.
14. Sommers H. J.: Path-integral Approach to Ising Spin-glass Dynamics. Phys. Rev. Lett. **58** (1987) 1268-1271.
15. Gomi S., Yonezawa F.: A New Perturbation Theory for the Dynamics of the Little-Hopfield Model. J. Phys. A: Math. Gen. **28** (1995) 4761-4775.
16. Koyama H., Fujie N., Seyama H.: Results From the Gardner-Derrida-Mottishaw Theory of Associative Memory. Neural Networks. **12** (1999) 247-257.
17. Coolen A.C.C.: Statistical Mechanics of Recurrent Neural Networks II. Dynamics. cond-mat/0006011.
18. Watkin T. L. H., Sherrington D.: The Parallel Dynamics of a Dilute Symmetric Hebb-rule Network. J. Phys A: Math. Gen. **24** (1991) 5427-5433.
19. Derrida B., Gardner E., Zippelius A.: An Exactly Solvable Asymmetric Neural Network Model. Europhys. Lett. **4** (1987) 167-173.
20. Patrick A. E., Zagrebnov V. A.: Parallel Dynamics for an Extremely Diluted Neural Network. J. Phys. A: Math. Gen. **23** (1990) L1323-L1329.
21. Castillo I. P., Skantzos N. S.: The Little-Hopfield Model on a Random Graph. cond-mat/0307499.
22. Sompolinsky H.: Neural networks with Nonlinear Synapses and Static Noise. Phys. Rev. A. **34** (1986) 2571-2574.
23. Wemmenhove B., Coolen A. C. C.: Finite Connectivity Attractor Neural Networks. J. Phys A: Math. Gen. **36** (2003) 9617-9633.
24. Chen Y., Wang Y. H., Yang K. Q.: Macroscopic Dynamics in Separable Neural Networks. Phys. Rev. E **63** (2001) 041901-4.