Journal of the Physical Society of Japan
Vol. 75, No. 10, October, 2006, 104004

Distinction of Coexistent Attractors in an Attractor Neural Network Model Using a Relaxation Process of Fluctuations in Firing Rates —Analysis with Statistical Mechanics—

Sawako TANIMOTO, Masato OKADA[1,2,3], Tomoyuki KIMOTO[4] and Tatsuya UEZU*

Graduate School of Sciences and Humanities, Nara Women's University, Nara 630-8506
[1]*Division of Transdisciplinary Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561*
[2]*RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198*
[3]*JST PRESTO, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561*
[4]*Department of Electrical and Electronic Engineering, Oita National College of Technology, 1666 Oaza-Maki, Oita 870-0152*

(Received June 27, 2006; accepted August 21, 2006; published October 10, 2006)

Griniasty *et al.* introduced an attractor neural network of the temporal cortex based on the correlation-type associative memory model. In this model, there are parameter regions where the Hopfield attractor and the correlated attractor coexist. We study a method of distinguishing these two attractors. For this purpose, we examine the relaxations of neural firing rate fluctuations. In other words, we introduce sublattices and calculate the correlations of firing rate fluctuations in the sublattices using a statistical mechanical method and Monte Carlo simulations. As a result, we found that the relaxation time for the correlated attractor is longer than that for the Hopfield attractor. Therefore, two bistable attractors can be distinguished by observing relaxation times.

KEYWORDS: attractor neural network model, Hopfield attractor, correlated attractor, distinction of attractors, correlation function, relaxation time
DOI: 10.1143/JPSJ.75.104004

1. Introduction

In recent years, it has become possible to measure not only the average firing rates of neural spikes, but also higher order statistical quantities such as fluctuations around the average firing rate. It is expected that this experimental progress will have a great influence on the study of neural networks. However, there are few studies that have examined the structure and the dynamical states of neural networks in brains by observing higher order statistical quantities.[1,2] In this paper, using one of these higher order statistical quantities, i.e., correlation functions, we show that it is possible to distinguish coexistent attractors of a recurrent neural network.

In principle, it is practically impossible to uniquely determine the structure and dynamical states of a neural network using only the statistical quantities of neural spikes. However, frequently the candidates of a structure and the dynamical states are reduced to only a small number from anatomical, physiological and theoretical knowledge obtained previously. In such cases, there is a possibility that several models or states are distinguished by observing higher order statistical quantities.

In this paper, in order to look into this possibility, we study an attractor neural network model of the temporal cortex that was proposed by Griniasty *et al.* in Amit's group. In order to explain the characteristics of the neural firing rates in the temporal cortexes of monkeys that were observed while the monkeys were performing a delay sample matching task,[3,4] Griniasty *et al.* proposed the Amit model, an extension of the Hopfield model[5,6] that is the correlation-type associative memory model.[7,8]

In the Amit model, cross-correlation-type learning between two adjacent patterns is used together with ordinary auto-correlation-type learning. Thus, in this model, the notion of neighbors is introduced in the set of stored patterns. In this model, if the strength of the cross-correlation-type learning is weak, any stored pattern becomes an attractor as in the Hopfield model. This attractor is called a Hopfield attractor. On the other hand, if the strength of the cross-correlation-type learning exceeds a certain level, the firing pattern of the system becomes a mixture of the stored pattern and its adjacent patterns. Thus, diffusion takes place in the space of the stored patterns. This diffusion does not spread over an infinitely long range but stops at a finite range, and the distribution of overlaps between the stored pattern and adjacent patterns becomes a Gaussian like distribution. For each stored pattern, there corresponds an equilibrium state whose central pattern is the stored pattern. Thus, an equilibrium state of this type has correlations with other equilibrium states of the same type. The longer the distance between the central patterns of the equilibrium states, the smaller the correlation among them. Griniasty *et al.* called this type of the equilibrium state the correlated attractor.

Uezu *et al.* showed that the correlated attractor does not appear through the instability of the Hopfield attractor as the temperature changes, but exists when the Hopfield attractor is stable in some temperature region. Thus, in this model two attractors can coexist. Furthermore, Uezu *et al.* found that the system converges to a different attractor depending on initial condition in the coexistent region.[9,10]

*E-mail: uezu@cc.nara-wu.ac.jp

In this paper, we study the Amit model and investigate whether two attractors can be distinguished using relaxation times of the correlation functions of the firing rate fluctuations. The relaxation times of the correlation functions are related to the second partial derivatives of free energy. Since the attractor corresponds to the minimal value of the free energy, this means that attractors can be characterized by observing the fluctuations of the firing states after the system reaches equilibrium. We show that the relaxation time of the correlation functions for the correlated attractor is longer than that for the Hopfield attractor. Therefore, we can distinguish between two attractors by observing the relaxation times.

In the next section, we explain the present model. In §3 and §4, theoretical results and numerical results are presented, respectively. A summary and discussion are given in §5.

2. Model

In this section, we explain the Amit model[7] and summarize the previous results for the equilibrium states.[9] The instantaneous state of each neuron is expressed by s_i, which takes ± 1, where i labels the neuron ($i = 1, 2, \ldots, N$). The time evolution of s_i in the deterministic case is given by

$$s_i(t+1) = \mathrm{sgn}(h_i(t)), \tag{1}$$

where h_i is defined by

$$h_i(t) = \sum_{j(\neq i)} J_{ij} s_j(t), \tag{2}$$

and J_{ij} is the strength of the synaptic connection from the j-th neuron to the i-th neuron. We adopt asynchronous dynamics. In this paper, we consider the stochastic time evolution and introduce T, which is the noise level. We call T a "temperature". Then, the probability that $s_i(t+1)$ takes ± 1 is given by

$$\mathrm{Prob}[s_i(t+1) = \pm 1] = \frac{1 \pm \tanh(\beta h_i(t))}{2}, \tag{3}$$

where $\beta = 1/T$. Also, in this case we adopt asynchronous dynamics. In the Amit model, the synaptic weight J_{ij} is defined as

$$\begin{aligned} J_{ij} &= \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu (\xi_j^\mu + a(\xi_j^{\mu-1} + \xi_j^{\mu+1})) \\ &= \frac{1}{N} \sum_{\mu,\mu'} \xi_i^\mu D_{\mu\mu'} \xi_j^{\mu'} \quad (\text{for } i \neq j), \\ J_{ii} &= 0 \quad (\xi_i^0 = \xi_i^\mu,\ \xi_i^{p+1} = \xi_i^1), \end{aligned} \tag{4}$$

where ξ_i^μ represents the value of i-th neuron for the μ-th pattern $\boldsymbol{\xi}^\mu \equiv \{\xi_1^\mu, \xi_2^\mu, \ldots, \xi_N^\mu\}$ and it takes value $+1$ or -1 according to the following probability

$$\mathrm{Prob}[\xi_i^\mu = \pm 1] = \frac{1}{2}. \tag{5}$$

The parameter p refers to the total number of patterns. D is a $p \times p$ matrix defined as

$$D \equiv \{D_{\mu\nu}\} = \begin{pmatrix} 1 & a & 0 & \cdots & \cdots & 0 & a \\ a & 1 & a & 0 & \cdots & \cdots & 0 \\ 0 & a & 1 & a & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a & 1 & a & 0 \\ 0 & \cdots & \cdots & 0 & a & 1 & a \\ a & 0 & \cdots & \cdots & 0 & a & 1 \end{pmatrix}. \tag{6}$$

The equilibrium state is described by the canonical distribution with the following Hamiltonian.

$$H = -\frac{1}{2} \sum_{i,j} J_{ij} s_i s_j = -\frac{1}{2N} \sum_{i \neq j} \sum_{\mu,\mu'} \xi_i^\mu D_{\mu\mu'} \xi_j^{\mu'} s_i s_j. \tag{7}$$

The overlap m^μ between μ-th pattern $\boldsymbol{\xi}^\mu$ and a state of neurons $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$ is defined as

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu s_i(t) \quad (\mu = 1, 2, \ldots, p). \tag{8}$$

The free energy per neuron is

$$f = -\frac{1}{\beta N} \ln Z, \quad Z = \mathrm{Tr}_{\boldsymbol{s}} \exp[-\beta H],$$

where

$$\mathrm{Tr}_{\boldsymbol{s}} \equiv \sum_{\{s_i\}} \equiv \sum_{s_1=\pm 1} \sum_{s_2=\pm 1} \cdots \sum_{s_N=\pm 1}.$$

For $p \ll N$, f is obtained using the saddle point method and is expressed as

$$\begin{aligned} f = &\frac{\beta}{2} \sum_{\mu,\mu'} m^\mu D_{\mu\mu'} m^{\mu'} \\ &- \left\langle\!\!\left\langle \ln\left(2\cosh\left(\beta \sum_{\nu,\nu'} m^\nu D_{\nu\nu'} \xi^{\nu'}\right)\right)\right\rangle\!\!\right\rangle_{\boldsymbol{\xi}}. \end{aligned} \tag{9}$$

The saddle point equation becomes

$$m^\mu = \left\langle\!\!\left\langle \xi^\mu \tanh\left(\beta \sum_{\nu,\nu'} \xi^\nu D_{\nu\nu'} m^{\nu'}\right)\right\rangle\!\!\right\rangle_{\boldsymbol{\xi}}. \tag{10}$$

Here, $\langle\langle \cdots \rangle\rangle_{\boldsymbol{\xi}}$ is the average over $\boldsymbol{\xi} = \{\xi^1, \xi^2, \ldots, \xi^p\}$. The correlated attractor has the following symmetry $m^2 = m^{13}$, $m^3 = m^{12}, \ldots$, and $m^7 = m^8$ when m^1 is the largest value among m's. In the left panel of Fig. 1, we display the solutions of eq. (10). As seen from the figure, the Hopfield attractor and the correlated attractor coexist for $0 \leq T \leq 0.1$. In the right panel of Fig. 1, the free energy of each attractor is displayed.

3. Theory

In this section, we derive the time evolution equations of the correlation functions, and the equations for the correlation functions in the equilibrium. We assume that the number of neurons N is very large, and the number of patterns is much smaller than N, i.e., $N \gg 1$ and $p/N \ll 1$.

First, we study the retrieval dynamics. The input signal to the i-th neuron $h_i(t)$ is

$$h_i(t) = \sum_{j \neq i} J_{ij} s_j(t) \simeq \sum_{\mu,\mu'} \xi_i^\mu D_{\mu\mu'} m^{\mu'}(t). \tag{11}$$

We adopt asynchronous dynamics, and assume that the

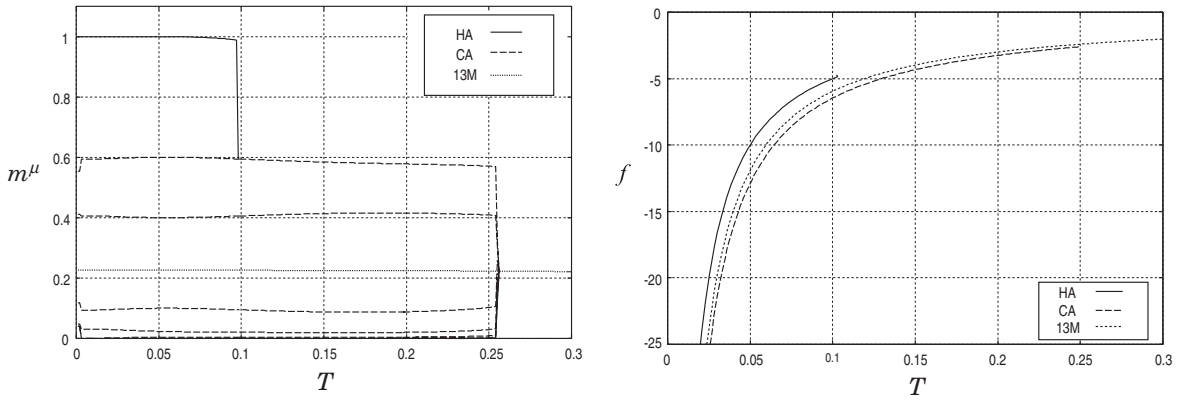

Fig. 1. Left panel: Equilibrium states [eq. (10)]. $a = 0.4$ and $p = 13$. The abscissa is T and the ordinate is m^μ ($\mu = 1, \ldots, p$). Solid curve: Hopfield attractor (HA) [$m^1 \simeq 1, m^\mu \simeq 0$ ($\mu = 2, \ldots, p$)], dashed curves: correlated attractor (CA), dotted curve: mixed state of thirteen patterns (13M). Right panel: Results of numerical calculation for the free energy [eq. (9)]. The abscissa is T and the ordinate is f. Line types are same as those in the left panel.

transition probability $w_k(\boldsymbol{s})$ from $\boldsymbol{s} = (s_1, \ldots, s_k, \ldots, s_N)$ to $F_k\boldsymbol{s} = (s_1, \ldots, -s_k, \ldots, s_N)$ takes the following form:

$$w_k(\boldsymbol{s}) = \frac{1 - s_k \tanh(\beta h_k(\boldsymbol{s}))}{2}, \tag{12}$$

where F_k is the flip operator of the k-th neuron. The probability distribution $P_t(\boldsymbol{s})$ of the state $\boldsymbol{s}$ at time t obeys the following master equation:[11,12]

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(\boldsymbol{s}) = \sum_{k=1}^{N} \{w_k(F_k\boldsymbol{s}) P_t(F_k\boldsymbol{s}) - w_k(\boldsymbol{s}) P_t(\boldsymbol{s})\}. \tag{13}$$

Then, the time evolution of overlap m^μ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle m^\mu \rangle = -\langle m^\mu \rangle + \left\langle\!\!\left\langle \xi^\mu \tanh\left(\beta \sum_{\tau,\tau'} \xi^\tau D_{\tau\tau'} m^{\tau'} \right) \right\rangle\!\!\right\rangle_{\boldsymbol{\xi}}, \tag{14}$$

where $\langle \cdots \rangle$ denotes the average over $P_t(\boldsymbol{s})$. From eq. (14), we obtain the equation for $\langle m^\mu \rangle$ at the steady state

$$\langle m^\mu \rangle = \left\langle\!\!\left\langle \xi^\mu \tanh\left(\beta \sum_{\nu,\nu'} \xi^\nu D_{\nu\nu'} \langle m^{\nu'} \rangle \right) \right\rangle\!\!\right\rangle_{\boldsymbol{\xi}}. \tag{15}$$

Since eq. (15) is the same as eq. (10), the steady state corresponds to the equilibrium state as it should.

Next, we calculate the correlation functions.[1)] Let $\delta s_i(t)$ be the fluctuation around the equilibrium value at time t, $\delta s_i(t) \equiv s(t) - \langle s_i \rangle$. The correlation function of fluctuations of the i-th neuron at time t and the j-th neuron at time $t + \tau$ is defined as

$$C_{ij}(t, t+\tau) \equiv \langle \delta s_i(t) \delta s_j(t+\tau) \rangle. \tag{16}$$

Then, we obtain the time evolution equations for the correlation functions as

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} C_{ij}(t,t) &= \frac{\mathrm{d}}{\mathrm{d}t} \langle \delta s_i(t) \delta s_j(t) \rangle \\ &= -2C_{ij}(t,t) + \langle \delta s_i(t) \delta \tanh(\beta h_j(t)) \rangle \\ &\quad + \langle \delta s_j(t) \delta \tanh(\beta h_i(t)) \rangle \quad (i \neq j) \end{aligned} \tag{17}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau} C_{ij}(t, t+\tau) &= \frac{\mathrm{d}}{\mathrm{d}\tau} \langle \delta s_i(t) \delta s_j(t+\tau) \rangle \\ &= -C_{ij}(t, t+\tau) + \langle \delta s_i(t) \delta \tanh(\beta h_j(t+\tau)) \\ &\qquad \text{(for any } i, j). \end{aligned} \tag{18}$$

The correlation function at the equilibrium is defined as

$$C_{ij}(\tau) \equiv \lim_{t \to \infty} C_{ij}(t, t+\tau). \tag{19}$$

We also study the correlation functions in the equilibrium state. From the definition, $C_{ii}(0)$ is obtained as

$$C_{ii}(0) = \langle (\delta s_i)^2 \rangle = 1 - \langle s_i \rangle^2. \tag{20}$$

Thus, this is the quantity of order N^0. The order of J_{ij} is N^{-1}. Now, we assume C_{ij} is the order of N^{-1} for $i \neq j$. Then, we obtain the equation of $C_{ij}(0)$ for $i \neq j$ by taking the limit $t \to \infty$ in eq. (17).

$$2C_{ij}(0) = \sum_k \tilde{J}_{jk} C_{ik}(0) + \sum_k \tilde{J}_{ik} C_{jk}(0) \quad (i \neq j), \tag{21}$$

where

$$\tilde{J}_{ik} \equiv \lim_{t \to \infty} \beta \cosh^{-2}(\beta \langle h_i(t) \rangle) J_{ik}.$$

Furthermore, the time evolution equation of $C_{ij}(\tau)$ for any i and j is obtained by taking the limit $t \to \infty$ in eq. (18)

$$\frac{\mathrm{d}}{\mathrm{d}\tau} C_{ij}(\tau) = -C_{ij}(\tau) + \sum_k \tilde{J}_{jk} C_{ik}(\tau). \tag{22}$$

Therefore, $C_{ii}(\tau)$ is expressed as

$$C_{ii}(\tau) = (1 - \langle s_i \rangle^2) \mathrm{e}^{-\tau}. \tag{23}$$

The number of the correlation functions $C_{ij}(\tau)$ is N^2. Thus, if N is very large, it is difficult to calculate all C_{ij} and their relaxation times. On the other hand, we can calculate correlation functions of the overlaps m^μ and their relaxation times. However, in order to calculate the correlation functions for the overlaps, we need information about stored patterns. In experiments, we do not have such information. Therefore, in order to avoid using information about stored patterns and to reduce the dimension of the relevant correlation matrix, we introduce sublattices and calculate the correlation functions of the firing rate fluctuations in the sublattices.

Now, let us define the sublattices. According to the value of $\xi_i^1, \xi_i^2, \ldots, \xi_i^p$, each neuron i is classified into 2^p sublattices. We denote these sublattices by $\Lambda_1, \ldots, \Lambda_{2^p}$. The l-th sublattice Λ_l is characterized by the p set of values $\eta_l^1, \ldots, \eta_l^p$, and the value ξ_i^μ at any neuron i in this sublattice takes the common value η_l^μ for all $\mu = 1, \ldots, p$. In the

equilibrium, the neurons that belong to the same sublattice Λ_l have the same input signal

$$h_i = \sum_{\mu,\nu} \xi_i^\mu D_{\mu\nu} m^\mu = \sum_{\mu,\nu} \eta_l^\mu D_{\mu\nu} m^\mu. \tag{24}$$

Therefore, the transition probability [eq. (12)] is the same for the neurons in the same sublattice. Next, we define the firing rate $m^{(l)}$ in the sublattice Λ_l and its fluctuation as follows:

$$m^{(l)}(t) = \frac{1}{|\Lambda_l|} \sum_{i\in\Lambda_l} s_i(t), \tag{25}$$

$$\delta m^{(l)}(t) = m^{(l)}(t) - \langle m^{(l)} \rangle, \tag{26}$$

where $|\Lambda_l|$ is the number of the elements in Λ_l. The correlation function of the firing rate fluctuations in sublattices Λ_{l_1} and Λ_{l_2} is defined as

$$\begin{aligned} L_{l_1 l_2}(\tau) &\equiv \lim_{t\to\infty} \langle \delta m^{(l_1)}(t) \delta m^{(l_2)}(t+\tau) \rangle \\ &= \frac{1}{|\Lambda_{l_1}||\Lambda_{l_2}|} \sum_{i\in\Lambda_{l_1}} \sum_{j\in\Lambda_{l_2}} C_{ij}(\tau). \end{aligned} \tag{27}$$

From eq. (21), we obtain the equation for $L_{l_1 l_2}(0)$ as

$$L_{l_1 l_2}(0) = \frac{1}{2} \sum_{l=1}^{2^p} (A_{l l_2} L_{l_1 l}(0) + A_{l l_1} L_{l_2 l}(0)) + \frac{2^p}{N} B_{l_1} \delta_{l_1,l_2}. \tag{28}$$

Furthermore, from eq. (22), we obtain the time evolution equation for $L_{l_1 l_2}(\tau)$ as

$$\frac{\mathrm{d}}{\mathrm{d}\tau} L_{l_1 l_2}(\tau) = -\sum_l (E_{l_2 l} - A_{l_2 l}) L_{l_1 l}(\tau). \tag{29}$$

Here, E is a $2^p \times 2^p$ unit matrix and A and B are defined as

$$A_{l_1 l_2} = \frac{\beta}{2^p} B_{l_1} \Xi_{l_1 l_2}, \quad B_l = \cosh^{-2}\left(\beta \sum_{\nu,\nu'} \eta_l^\nu D_{\nu\nu'} \langle m^{\nu'} \rangle \right), \tag{30}$$

where

$$\Xi_{l_1 l_2} \equiv \sum_{\mu\mu'} \eta_{l_1}^\mu D_{\mu\mu'} \eta_{l_2}^{\mu'}.$$

By defining $(\boldsymbol{L}_{l_1})_{l_2} = L_{l_1 l_2}$, eq. (29) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \boldsymbol{L}_{l_1}(\tau) = -(E - A) \boldsymbol{L}_{l_1}(\tau). \tag{31}$$

Let λ_i be the eigenvalue of A and $\boldsymbol{e}_i$ be the eigenvector belonging to λ_i. We expand $\boldsymbol{L}_{l_1}(0)$ by $\{\boldsymbol{e}_i\}$.

$$\boldsymbol{L}_{l_1}(0) = \sum_{i=1}^{2^p} a_i^{(l_1)} \boldsymbol{e}_i. \tag{32}$$

Then, the solution for eq. (31) is expressed as

$$\boldsymbol{L}_{l_1}(\tau) = \sum_{i=1}^{2^p} a_i^{(l_1)} \exp[-(1-\lambda_i)\tau] \boldsymbol{e}_i. \tag{33}$$

Since the matrix A depends on the overlap $\{m^\mu\}$, the matrix A differs from one attractor to another. From eq. (9), we calculate the curvature of the free energy and obtain

$$\frac{\partial^2 f}{\partial m^{(l_1)} \partial m^{(l_2)}} = \frac{\beta}{2^{2p}} \{\Xi(E - A)\}_{l_1 l_2}. \tag{34}$$

From eq. (33), we note that the eigenvalues of $E - A$ characterize the relaxation of $L_{l_1 l_2}(\tau)$. On the other hand, from eq. (34), we note that the second partial derivatives of the free energy are obtained from $E - A$ by operating Ξ. Thus, the relaxation times of the correlation functions are related to the curvatures of the free energy. Since the curvatures of the free energy depend on attractors, we expect that coexistent attractors can be distinguished by observing the relaxation times of the correlation functions.

4. Numerical Calculation

In this section, we compare the theoretical results with numerical results using Monte Carlo simulations (MCS) for asynchronous dynamics. First, we briefly describe parameters and the calculation method.

The parameters are fixed to the following values. The number of neurons is $N = 50000$ or 100000, the number of patterns is $p = 13$ and the strength of the synaptic connection between neighboring patterns is $a = 0.4$. As for the temperature, we adopted $T = 0.03$, 0.05, and 0.08.

An initial state of neurons is created according to the following probability

$$\mathrm{Prob}[s_i = \pm 1] = \frac{1 \pm m_0 \xi_i^1}{2}. \tag{35}$$

Thus, the initial condition of the μ-th overlap is

$$m^\mu(0) = m_0 \delta_{\mu,1}, \quad \mu = 1, \ldots, p. \tag{36}$$

As shown in Fig. 2, the system converges to different attractors depending on the values of m_0. If we put $m_0 = 0.5$, the system goes to the Hopfield attractor (the upper curve of Fig. 2), whereas if we put $m_0 = 0.1$, the system goes to the correlated attractor (the lower curve of Fig. 2). Using these values of m_0, we obtain two different attractors. We measure the relaxation time of the correlation function in each attractor. The correlation function is expressed by the sum of the exponential functions as in eq. (33). We denote the relaxation times of the correlation function at the Hopfield attractor and at the correlated attractor τ_h and τ_c, respectively. Assuming $C_{ij}(\tau) \propto \exp[-\tau/\tau_h]$ or $C_{ij}(\tau) \propto \exp[-\tau/\tau_c]$, we determine τ_h or τ_c from $\ln C_{ij}(\tau)$ using the least squares method. Next, we give the numerical results.

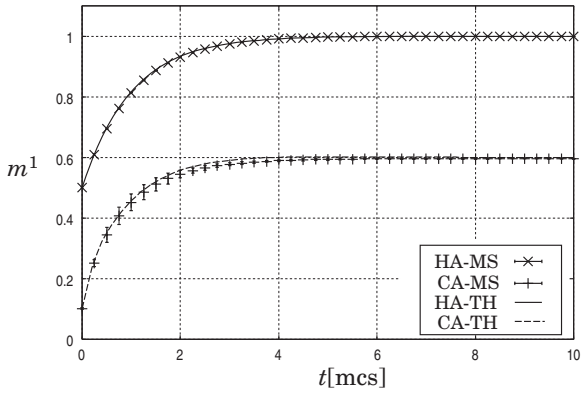

Fig. 2. The relaxation process to the equilibrium (the time evolution of m^1). The parameters are $a = 0.4$, $T = 0.05$ and $p = 13$. The abscissa is time t in the unit of the Monte Carlo step (mcs) and the ordinate is m^1. "-TH" denotes the result obtained by the Runge Kutta method (RK) of eq. (14) and "-MS" the result obtained by the Monte Carlo simulations ($N = 50000$). "HA" denotes the Hopfield attractor and "CA" the correlated attractor.

Table I. A comparison of the values of several correlation functions at $\tau = 0$ between the theory and the Monte Carlo simulations. $a = 0.4$ and $T = 0.05$. In the table, HA denotes the Hopfield attractor and CA the correlated attractor. MS denotes the Monte Carlo simulations and TH the theory, i.e., the solution in eq. (28).

HA		
	MS	TH
$L_{2822,2822}(0)$	0.159×10^{-3}	0.110×10^{-3}
$L_{2822,2566}(0)$	$\sim 10^{-9}$	$\sim 10^{-9}$
$L_{2822,2838}(0)$	$\sim 10^{-9}$	$\sim 10^{-9}$
$L_{2822,2886}(0)$	$\sim 10^{-9}$	$\sim 10^{-9}$
CA		
	MS	TH
$L_{2822,2822}(0)$	0.322×10^{-1}	0.579×10^{-1}
$L_{2822,2566}(0)$	0.494×10^{-2}	0.132×10^{-2}
$L_{2822,2838}(0)$	0.119×10^{-2}	0.161×10^{-2}
$L_{2822,2886}(0)$	0.732×10^{-2}	0.114×10^{-2}

4.1 $L_{l_1 l_2}(0)$

By solving eq. (28) numerically, we get values of $L_{l_1 l_2}(0)$. To check whether the results of the MCS are reliable, we compare the MCS results with the theoretical ones. We choose sublattices in which the value of the correlation function $L_{l_1 l_2}(0)$ is large and the transition probability is high, that is where $|\beta h_i|$ is small, so that we can easily observe the relaxation processes.

The results are shown in Table I. The cross-correlation functions for the same sublattice calculated theoretically and using the MCS are the order of 10^{-4} at the Hopfield attractor and the order of 10^{-2} at the correlated attractor, respectively. The cross-correlation functions for the different sublattices calculated theoretically and using the MCS at the correlated attractor are the order of 10^{-3}. The slight discrepancies between the theoretical results and the results using the MCS are due to the fact that eq. (28) is derived for the leading order of N, whereas the MCS is performed for finite N and includes all orders of N. Therefore, we consider the results achieved using the MCS are reliable. In these cases, the magnitude of correlation functions is large enough to observe the relaxation processes. On the other hand, the cross-correlation functions for the different sublattices calculated theoretically and using the MCS at the Hopfield attractor are the order of 10^{-9} and these values are the same order of numerical errors. In this case, it is difficult to observe the relaxation processes, and therefore, in order to obtain the relaxation times, we use only the theoretical results.

4.2 Relaxation time

We adopted $L_{2822,2822}$ as the cross-correlation function for the same sublattice and $L_{2822,2838}$ as the cross-correlation function for the different sublattices. Here, the sublattices $l = 2822$ and $l = 2838$ correspond to the following value sets of $(\eta^1, \eta^2, \ldots, \eta^{13})$, $(1, -1, 1, -1, -1, -1, -1, -1, 1, 1, -1, 1, -1)$ and $(1, -1, 1, -1, 1, -1, -1, -1, 1, 1, 1, -1, 1, -1)$, respectively. Only the values of η_5 are different in these sublattices.

First, we give the results of the cross-correlation function for the same sublattice $L_{2822,2822}$. In Fig. 3, we display the results for the Hopfield attractor (left panels) and for the correlated attractor (right panels) for $T = 0.05$. The upper panels are drawn for the correlation functions and the lower panels are for the natural logarithm of the correlation functions. In these panels, b_{MS} denotes the slope of $\ln L_{2822,2822}$ estimated from the MCS results using the least squares method for $\tau = 0$–4. b_{TH} denotes the slope of $\ln L_{2822,2822}$ calculated from eq. (33) using the least squares method. That is, bs are $1/\tau_a$ (a is h or c), which are estimated assuming that the correlation functions are proportional to $\mathrm{e}^{-\tau/\tau_a}$. In the Hopfield attractor, b_{TH} is obtained from data for $\tau = 0$–10, and in the correlated attractor, b_{TH1} is obtained from the data $\tau = 0$–4, b_{TH2} is for $\tau = 4$–8 and b_{TH3} is for $\tau = 8$–12.

First, we examine the results of the Hopfield attractor. Since the eigenvalues of the matrix A are $\lambda_i \simeq 0$, we expect the relaxation time to be $\tau_h \simeq 1$. Using the theoretical results for $\tau = 0$–10, we found $\tau_h = 1/b_{\mathrm{TH}} = 1.0001 \pm 2 \times 10^{-8} \sim 1$. On the other hand, for the MCS, we found $\tau_h = 1/b_{\mathrm{MS}} = 0.85 \pm 0.01$ and this value is less than 1. We also obtained similar results for other sublattices.

Next, we examine the correlated attractor. Since the correlation function is theoretically expressed as $L_{l_1 l_2}(\tau) = \sum_{i=1}^{2^p} a_i^{(l_1)} \exp[-(1 - \lambda_i)\tau](\boldsymbol{e}_i)_{l_2}$ [eq. (33)], its relaxation time seems to be characterized by the largest value of the eigenvalues λ_i of the matrix A. For $l = 2822$, the largest eigenvalue is $\lambda_i \simeq 0.48$. If this mode had large contribution to the relaxation process, the relaxation time would be $\tau_c^{(\max)} = 1/(1 - 0.48) \simeq 1.92$. However, we obtained faster relaxation than we expected, that is, $\tau_c = 1/b_{\mathrm{MS}} = 1.036 \pm 0.002$ using the MCS and $\tau_c = 1/b_{\mathrm{TH1}} = 1.055 \pm 0.001$ using the theoretical calculation. In order to clarify the reason, we calculated the relaxation modes. Among the 2^p modes of the relaxation in the correlated attractor, 13 modes had relaxation times τ_i longer than 1 and had slower relaxation than the Hopfield attractor. See Fig. 4. Furthermore, five modes have relaxation times that are longer than 1.3. Other modes had $\tau_i \sim 1$. Moreover, the coefficient $a_i^{(l_1)}$ of the slowest mode was the order of 10^{-4}, while the largest coefficient of fast modes was the order of 10^{-3}. Thus, when the time lag of the correlation function is small ($\tau = 0$–4), since the contributions of fast modes are large, the relaxation time becomes shorter than we expected. Even in this situation, the relaxation time of the correlated attractor is longer than that of the Hopfield attractor because the correlated attractor contains modes with longer relaxation times than 1. On the other hand, the theoretically estimated relaxation time tends to $\tau_c^{(\max)}$ as we estimate it for a later time period. ($\tau_{\mathrm{TH2}} = 1/b_{\mathrm{TH2}} = 1.222 \pm 0.004$, $\tau_{\mathrm{TH3}} = 1/b_{\mathrm{TH3}} = 1.539 \pm 0.004$).

Furthermore, we calculated the relaxation at different temperatures. In the correlated attractor, the result of the MCS at $T = 0.08$ is $\tau_c = 1.11 \pm 0.02$ and that at $T = 0.03$ is $\tau_c = 1.01 \pm 0.01$. On the other hand, in the Hopfield attractor, the results of the MCS at $T = 0.08$ and $T = 0.03$ are $\tau_h \simeq 1$. This indicates that when the temperature is increased, the difference in the relaxation times between two attractors becomes large.

Next, we give the results of the cross-correlation function for the different sublattices $L_{2822,2838}$ for $T = 0.05$. As in Fig. 3, we display the results for the Hopfield attractor (left

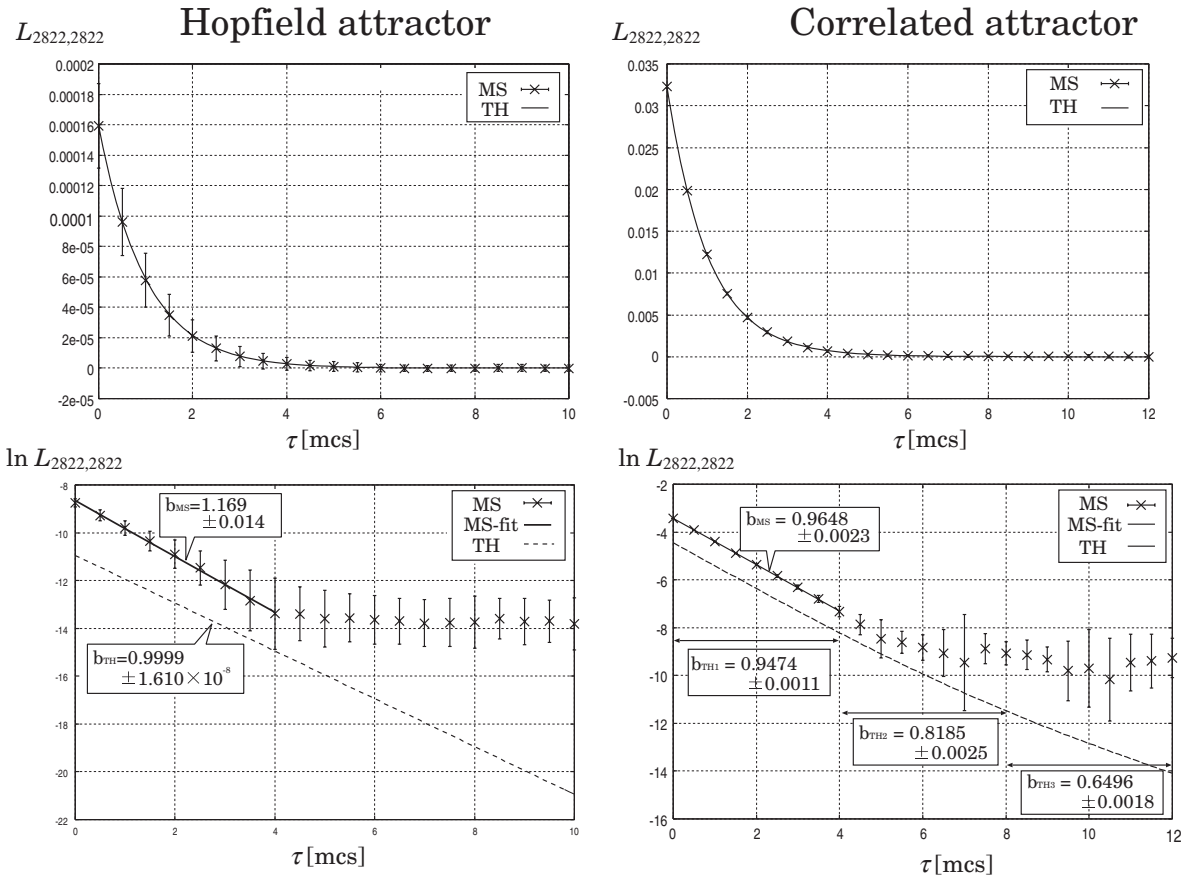

Fig. 3. The cross-correlation function for the same sublattice of each attractor. $a = 0.4$, $T = 0.05$. Upper panels: the correlation function of each attractor. The abscissa is τ and the ordinate is $L_{2822,2822}$. The crosses with error bars (MS) represent the MCS results and the solid curve (TH) the theoretical result in eq. (33). Lower panels: the τ dependence of $\ln L_{2822,2822}$. The crosses with error bars (MS) represent the MCS results, and the solid line (MS-fit) is obtained from the MCS results using the least squares method. The dashed curve (TH) is the theoretical result, which is shifted below in order to be easily compared with the MCS results. b_{TH} and b_{MC} are the theoretical and numerical estimates of the slope of $\ln L_{2822,2822}$, respectively.

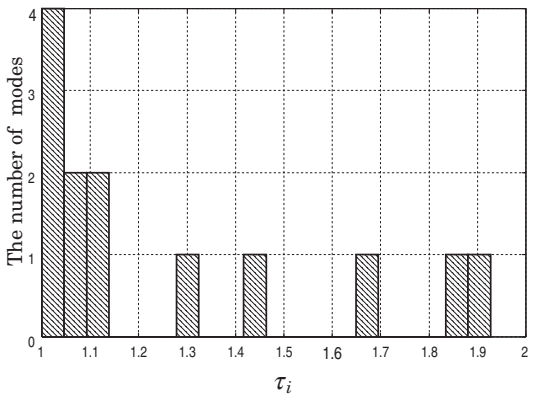

Fig. 4. The distribution of the modes of relaxation times that are longer than 1 for $a = 0.4$, $p = 13$ and $T = 0.05$ in the correlated attractor. The abscissa is $\tau_i = 1/(1 - \lambda_i)$ and the ordinate is the number of modes. Here, λ_i are eigenvalues of A.

panels) and for the correlated attractor (right panels) for $T = 0.05$ in Fig. 5. As for the Hopfield attractor, the relaxation time was theoretically estimated to be $\tau_h = 1/b_{\mathrm{TH}} = 0.97 \pm 5.0 \times 10^{-5}$, but it could not be estimated using the MCS because the magnitude of $L_{2822,2838}(0)$ is the order of 10^{-9} and is the same order of numerical errors. In term of the correlated attractor, the relaxation time was estimated to be $\tau_c = 1/b_{\mathrm{TH1}} = 1.94 \pm 0.01$ theoretically and $\tau_c = 1/b_{\mathrm{MS}} = 2.2 \pm 0.1$ using the MCS. Thus, the relaxation time of the cross-correlation function of the correlated attractor is nearly equal to $\tau_c^{(\max)}$. From these results, we conclude that $\tau_h \sim 1$ and $\tau_c \sim 2$ and the difference between τ_h and τ_c is very large in the cross-correlation function for the different sublattices.

5. Summary and Discussion

We have studied a method of distinguishing coexistent attractors by observing the correlation functions in a recurrent-type neural network model called the Amit model.

The Amit model has a parameter region where the Hopfield attractor and the correlated attractor coexist. Since the difference in the curvatures of the free energy between two attractors is considered to reflect the relaxation of the correlation functions, we expected that different types of attractors could be distinguished by observing the relaxation times of the correlation functions.

In order to calculate the correlation functions for the overlaps, we need information about stored patterns. During experiments, we do not have such information. On the other hand, theoretical calculations of the correlation functions and relaxation times for neurons are impossible when the number of neurons N becomes of the order of 10^4. Therefore, in order to avoid using stored pattern information and to reduce the dimensions of the relevant correlation

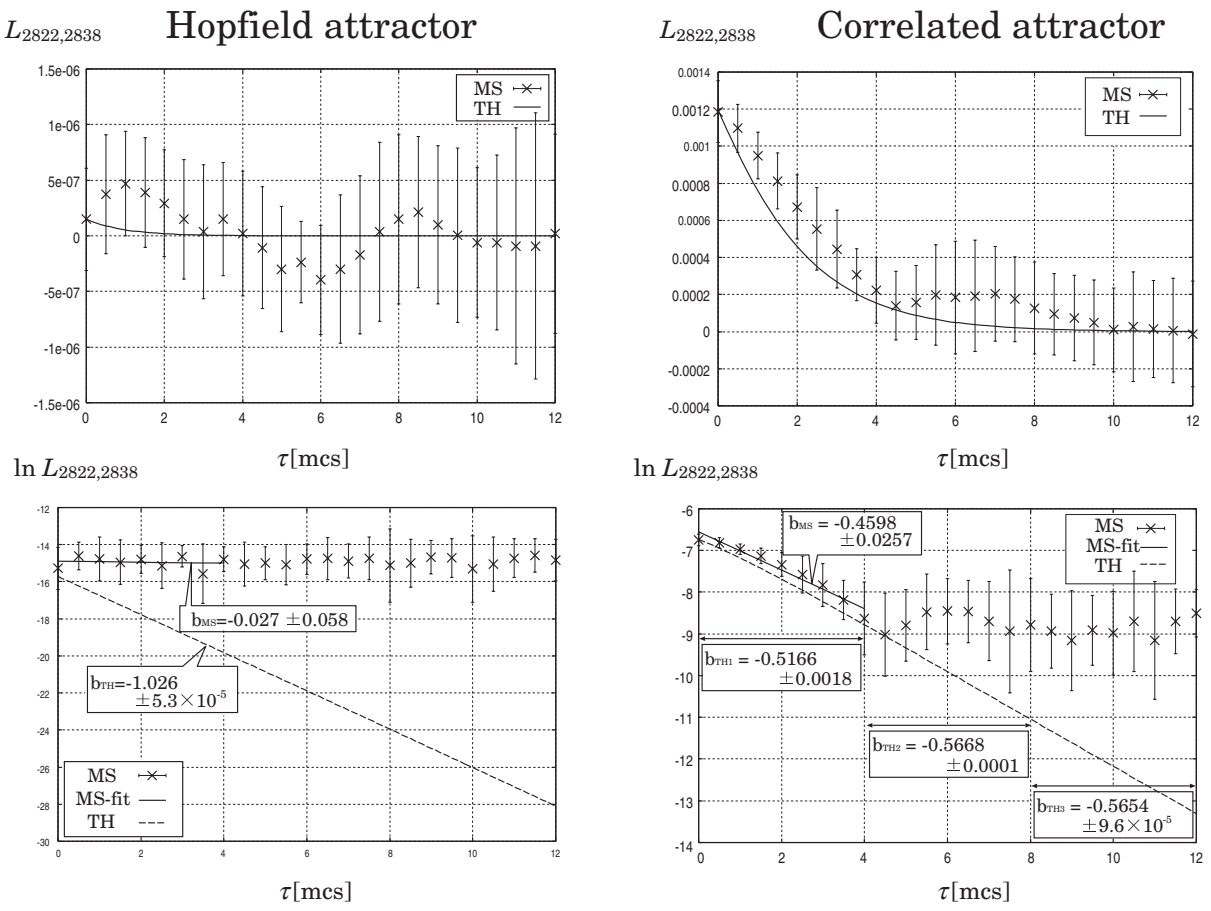

Fig. 5. The cross-correlation function for the different sublattices of each attractor. $a = 0.4$, $T = 0.05$. Upper panels: the correlation function of each attractor. The abscissa is τ and the ordinate is $L_{2822,2838}$. The crosses with error bars (MS) represent the MCS results and the solid curve (TH) the theoretical result, eq. (33). Lower panels: the τ dependence of $\ln L_{2822,2838}$. The crosses with error bars (MS) represent the MCS results, and the solid line (MS-fit) is obtained from the MCS results using the least squares method. The dashed curve (TH) is the theoretical result, which is shifted below in order to be easily compared with the MCS results. b_{TH} and b_{MC} are the theoretical and numerical estimates of the slope of $\ln L_{2822,2838}$, respectively.

matrix, we introduced a 2^p set of sublattices and used the correlation functions of the fluctuations of firing rates in the sublattices. We investigated the correlation functions of bistable attractors and calculated the relaxation times of these correlation functions theoretically and using the MCS.

Theoretically, the relaxation time of correlation functions is characterized by the largest relaxation time $\tau^{(\max)}$ among $\tau_i = 1/(1-\lambda_i)$, which is the relaxation time of the i-th mode. The distribution of relaxation times is nearly equal to the delta function $\delta(\tau - 1)$ in the Hopfield attractor, whereas in the correlated attractor, there are several τ_i larger than 1. Thus, we expected that the relaxation of the correlated attractor would be longer than that of the Hopfield attractor. In order to investigate whether this was true or not, we calculated the cross-correlation function for the same sublattice $L_{2822,2822}$ at $T = 0.03$, 0.05, and 0.08, and the cross-correlation function for the different sublattices $L_{2822,2838}$ at $T = 0.05$, by fixing $a = 0.4$ and $p = 13$. The sublattices $l = 2822$ and $l = 2838$ were chosen because the magnitude of the correlation functions was large enough to observe the relaxation process.

First, we summarize the results for $L_{2822,2822}$ at $T = 0.05$.

In the Hopfield attractor, we obtained $\tau_h \sim 1$ theoretically and $\tau_h = 0.85 \pm 0.01$ using the MCS, whereas in the correlated attractor we obtained $\tau_c = 1.055 \pm 0.001$ theoretically and $\tau_c = 1.036 \pm 0.002$ with MCS. The relaxation time in the correlated attractor was shorter than $\tau_c^{(\max)}$. The reason for this is considered to be that the number of the modes whose relaxation times are larger than 1 was about 10 in 2^p and their coefficients are one order smaller than those of the faster modes. However, since there are slower relaxation modes that do not exist in the Hopfield attractor, a longer relaxation time was observed in the correlated attractor than in the Hopfield attractor. Thus, the two attractors can be distinguished by the relaxation times.

The above analysis was performed at $T = 0.05$. Since the relative difference in the relaxation time between the Hopfield attractor and the correlated attractor is $(\tau_c - \tau_h)/\tau_c \sim 0.05$, which is not large, there is a possibility that two attractors cannot be distinguished under some external noise. However, the relative difference becomes larger as T is increased. In fact, we obtained the relaxation times in the correlated attractor using the MCS as $\tau_c = 1.11 \pm 0.02$ at $T = 0.08$ and $\tau_c = 1.01 \pm 0.01$ at $T = 0.03$. In this model, since the temperature range for two attractors to coexist is small, i.e., $0 \leq T \leq 0.1$ for $a = 0.4$ and $p = 13$, the relative difference in the relaxation times of two attractors is at most 10%. We think that there are parameter regions where the relative difference is bigger and where two attractors are distinguished easily. We think the reason that the relaxation

time increases as temperature does in the correlated attractor is that the curvatures of the free energy become smaller for the larger temperatures.

Next, let us summarize the results for the cross-correlation function $L_{2822,2838}$ at $T = 0.05$.

The relaxation time of the Hopfield attractor was estimated as $\tau_h \sim 1$ theoretically, but it was difficult to estimate it with the MCS because the magnitude of the correlation function is the same order of the numerical errors. On the other hand, the relaxation time of the correlated attractor was estimated as $\tau_c = 1.94 \pm 0.01$ theoretically and as $\tau_c = 2.2 \pm 0.1$ using the MCS. It turned out that the difference between τ_h and τ_c was very large. In this case, the relaxation time is characterized by the largest τ_i, $\tau_c^{(\max)}$ in Fig. 4. We consider the reason the relaxation in the cross-correlation function for the different sublattices of the correlated attractor is characterized by $\tau_c^{(\max)}$ is that the auto-correlation function $C_{ii}(\tau)$, whose relaxation time is 1 and whose order is 1, is not included in the cross-correlation function for the different sublattices.

In conclusion, in the Amit model, two coexistent attractors can be distinguished using the relaxation of the correlation functions of fluctuations around the averaged firing rates in the sublattices. In particular, it turned out that there existed a striking difference of relaxation times of the cross-correlation function for the different sublattices between two attractors. Since the cross-correlation function $C_{ij}(\tau)$ between two neurons in the different sublattices is expected to behave similarly to the cross-correlation function between the corresponding sublattices, we believe that it is possible to distinguish two attractors by observing $C_{ij}(\tau)$ experimentally.

1) I. Ginzburg and H. Sompolinsky: Phys. Rev. E **50** (1994) 3171.
2) M. Tatsuno and M. Okada: Neural Comput. **16** (2004) 737.
3) Y. Miyashita: Nature **335** (1988) 817.
4) Y. Miyashita and H. S. Chang: Nature **331** (1988) 68.
5) J. J. Hopfield: Proc. Natl. Acad. Sci. U.S.A. **79** (1982) 2554.
6) M. Okada: Neural Networks **9** (1996) 1429.
7) M. Griniasty, M. V. Tsodyks and D. J. Amit: Neural Comput. **5** (1993) 1.
8) D. J. Amit, N. Brunel and M. V. Tsodyks: J. Neurosci. **14** (1994) 6435.
9) T. Uezu, A. Hirano and M. Okada: J. Phys. Soc. Jpn. **73** (2004) 867.
10) T. Fukai, T. Kimoto, M. Doi and M. Okada: J. Phys. A **32** (1999) 5551.
11) R. J. Glauber: J. Math. Phys. **4** (1963) 294.
12) M. Suzuki and R. Kubo: J. Phys. Soc. Jpn. **24** (1968) 51.