

# ASSOCIATIVE MEMORY BY RECURRENT NEURAL NETWORKS WITH DELAY ELEMENTS

*Seiji Miyoshi<sup>†</sup>, Hiro-Fumi Yanai<sup>††</sup>, and Masato Okada<sup>†††,††††</sup>*

<sup>†</sup>Department of Electronic Engineering, Kobe City College of Technology

<sup>††</sup>Department of Media and Telecommunications, Faculty of Engineering, Ibaraki University

<sup>†††</sup>Exploratory Research for Advanced Technology, Japan Science and Technology

<sup>††††</sup>Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute

## ABSTRACT

The synapses of real neural systems seem to have delays. Therefore, it is worthwhile to analyze associative memory models with delayed synapses. Thus, a sequential associative memory model with delayed synapses is discussed, where a discrete synchronous updating rule and a correlation learning rule are employed. Its dynamic properties are analyzed by the statistical neurodynamics. In this paper, we first re-derive the Yanai-Kim theory, which involves macrodynamical equations for the dynamics of the network with serial delay elements. Since their theory needs a computational complexity of  $\mathcal{O}(L^4t)$  to obtain the macroscopic state at time step  $t$  where  $L$  is the length of delay, it is intractable to discuss the macroscopic properties for a large  $L$  limit. Thus, we derive steady state equations using the discrete Fourier transformation, where the computational complexity does not formally depend on  $L$ . We show that the storage capacity  $\alpha_C$  is in proportion to the delay length  $L$  with a large  $L$  limit, and the proportion constant is 0.195, i.e.,  $\alpha_C = 0.195L$ . These results are supported by computer simulations.

## 1. INTRODUCTION

Associative memory models of neural network can be mainly classified into two types [1, 2, 3]. The first is the auto-associative memory model where memory patterns are stored as equilibrium states of the network. The second type is the sequence processing model, which stores the sequence of memory patterns.

As a learning algorithm for storing memory patterns, the correlation learning algorithm based on Hebb's rule is well known. The storage capacity, that is, how many memory patterns can be stably stored with respect to the number of neurons, is the one of the most important properties of associative memory models. Hopfield [4] showed that the storage capacity of the auto-associative memory model using correlation learning is about 0.15 by computer simulation. On the other hand, many theoretical analyses have been done on the correlation learning type associative memory model[1]. As typical analytical methods,

there are the replica method [5, 6, 7] and the SCSNA[8] for the equilibrium state of the auto-associative memory models. There is also the statistical neurodynamics [9] for handling retrieval dynamics. By these theories, it became clear that the storage capacity of the auto-associative memory model is 0.138 and that of the sequence processing model is 0.269 [10, 11, 12]. Furthermore, it has become well-known that the analysis of the dynamics for the auto-associative model is more difficult than that for the sequence processing model. However, Okada [13] succeeded in explaining the dynamics of the retrieval process quantitatively by extending the statistical neurodynamics [9].

On the other hand, the synapses of real neural systems seem to have delays. Therefore, it is very important to analyze the associative memory model with delayed synapses. Computer simulation is a powerful method for investigating the properties of the neural network. However, there is a limit on the number of neurons. In particular, computer simulation for a network that has large delay steps is realistically impossible considering the required amount of calculation and memory. Therefore, the theoretical and analytical approach is indispensable to research on delayed networks.

A neural network in which each neuron has delay elements [2, 14, 15] was analyzed by Yanai and Kim [14]. They analyzed the delayed network using the statistical neurodynamics [9, 10] and derived macrodynamical equations for the dynamics. Their theory closely agrees with the results of computer simulation.

In this paper, after defining the model, the Yanai-Kim theory is re-derived by using the statistical neurodynamics [16]. We show that their macrodynamical equations make clear that the dynamics of the network and the phase transition points change with the initial conditions. The Yanai-Kim theory needs a computational complexity of  $\mathcal{O}(L^4t)$  to obtain the macrodynamics, where  $L$  and  $t$  are the length of delay and the time step, respectively. This means that it is indispensable to discuss the macroscopic properties for a large  $L$  limit. Thus, we derive the macroscopic steady state equations by using the discrete Fourier transformation. Using the derived steady state equations, the storage capacity can be quantitatively discussed even for a large  $L$  limit. Then, it becomes clear that the phase transition point calculated from the macroscopic steady state equations agrees with the phase transition point obtained by time-dependent

---

This research was partially supported by the Ministry of Education, Science, Sports and Culture of Japan, Grant-in-Aid for Scientific Research, 13780303, 2001.

calculation with a sufficient number of time steps from the optimum initial condition. Furthermore, it becomes clear that in the case of large delay length  $L$ , the storage capacity is in proportion to the length and the proportion constant is 0.195. These results are supported by computer simulation.

## 2. MODEL OF DELAYED NETWORK

The delayed network discussed in this paper has  $N$  neurons, and  $L-1$  serial delay elements are connected to each neuron. All neurons as well as all delay elements have synaptic connections with all neurons. In this neural network, all neurons and all delay elements change their states simultaneously. That is, this network employs a discrete synchronous updating rule. The output of each neuron is determined as

$$x_i^{t+1} = F(u_i^t), \quad (1)$$

$$u_i^t = \sum_{l=0}^{L-1} \sum_{j=1}^N J_{ij}^l x_j^{t-l}, \quad (2)$$

where  $x_i^t$  denotes the output of the  $i$ th neuron at time  $t$ , and  $J_{ij}^l$  denotes the connection weight from the  $l$ th delay elements of the  $j$ th neuron to the  $i$ th neuron.  $F(\cdot)$  is the sign function defined as

$$F(u) = \text{sgn}(u) = \begin{cases} +1, & u \geq 0. \\ -1, & u < 0. \end{cases} \quad (3)$$

Now, let's consider how to store the sequence of  $\alpha N$  memory patterns,  $\xi^1 \rightarrow \xi^2 \rightarrow \dots \rightarrow \xi^\mu \rightarrow \dots \rightarrow \xi^{\alpha N}$ . Here,  $\alpha$  and  $\alpha N$  are the loading rate and the length of the sequence, respectively. Each component  $\xi^\mu$  is assumed to be an independent random variable that takes a value of either  $+1$  or  $-1$  according to the following probabilities,

$$\text{Prob}[\xi_i^\mu = \pm 1] = \frac{1}{2}. \quad (4)$$

We adopt the following learning method using the correlation learning,

$$J_{ij}^l = \frac{c_l}{N} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^\mu, \quad (5)$$

where  $c_l$  is the strength of the  $l$ th delay step.

Correlation learning is an algorithm based on Hebb's rule. It is inferior to the error correcting learning in terms of storage capacity. However, as seen from eqn (5), when adding new patterns, it is not necessary to again learn all patterns that were stored in the past. Furthermore, correlation learning has been analyzed by many researchers due to its simplicity.

## 3. DYNAMICAL BEHAVIORS OF MACROSCOPIC ORDER PARAMETERS BY STATISTICAL NEURODYNAMICS AND DISCUSSION

As mentioned above, a neural network in which each neuron has delay elements [2, 14, 15] was analyzed by Yanai and Kim [14]. In this section, the Yanai-Kim theory is re-derived

by using the statistical neurodynamics [16]. Using the re-derived theory, we discuss the dynamical behaviors of a delayed network. Furthermore, the relationship between the phase transition point and the initial conditions of the delayed network is discussed.

In the case of a small loading rate  $\alpha$ , if a state close to one or some of the patterns stored as a sequence are given to the network, the stored sequence of memory patterns is retrieved. However, when loading rate  $\alpha$  increases, the memory is broken at a certain  $\alpha$ . That is, even if a state close to one or some of the patterns stored as a sequence is given to the network, the state of the network tends to leave the stored sequence of memory patterns. Moreover, even if one or some of the patterns themselves are given to the network, the state of the network tends to leave the stored sequence of memory patterns. This phenomenon, that is, the memory suddenly becoming unstable at a critical loading rate, can be considered a kind of phase transition.

We define the overlaps or direction cosine between a state  $\mathbf{x}^t = (x_i^t)$  appearing in a recall process at time  $t$  and an embedded pattern  $\xi^\mu = (\xi_i^\mu)$  as

$$m_\mu^t = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu x_i^t. \quad (6)$$

By using this definition, when the state of the network at time  $t$  and the  $\mu$ th pattern agree perfectly, the overlap  $m_\mu^t$  is equal to unity. When they have no correlation, the overlap  $m_\mu^t$  is equal to zero. Therefore, the overlap provides a way to measure recall quality.

Amari and Maginu[9] proposed the statistical neurodynamics. This analytical method handles the dynamical behavior of the recurrent neural network macroscopically, where cross-talk noise is regarded as a Gaussian random variable with a mean of zero and a time-dependent variance of  $\sigma_t^2$ . They then derived recursive relations for the variance and the overlap.

Using eqns (2), (5) and (6), we have derived the following macrodynamical equations.

$$\sigma_t^2 = \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l, t-l'}, \quad (7)$$

$$\begin{aligned} v_{t-l, t-l'} &= \alpha \delta_{l, l'} \\ &+ U_{t-l} U_{t-l'} \\ &\times \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1, t-l'-k'-1} \\ &+ \alpha (c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l}), \end{aligned} \quad (8)$$

$$U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{t-1}} \exp\left(-\frac{(s^{t-1})^2}{2\sigma_{t-1}^2}\right), \quad (9)$$

$$s^t = \sum_{l=0}^{L-1} c_l m_{t-l}, \quad (10)$$

$$m_{t+1} = \text{erf}\left(\frac{s^t}{\sqrt{2}\sigma_t}\right), \quad (11)$$

where  $m_t$  denotes  $m_t^t$ . If  $t < 0$ ,  $m_t = 0$  and  $U_t = 0$ . If  $k < 0$ ,  $c_k = 0$ . If either  $k < 0$  or  $k' < 0$ ,  $v_{k, k'} = 0$ . The

expression  $\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du$  denotes the error function.

Various cases can be considered as the initial condition of the network. For example, in the case of *the one step set initial condition*, only the states of neurons are set explicitly and those of delay elements are all zero. Then  $m_0 = m_{\text{init}}$  is the only initial condition that is set explicitly and  $v_{0,0} = \alpha$ . On the other hand, as one more extreme case, *the all steps set initial condition* can also be considered, where the states of all neurons and all delay elements are set to be close to the stored pattern sequences. In this case,  $m_l = m_{\text{init}}$ ,  $l = 0, \dots, L-1$  and  $v_{l,l} = \alpha$ ,  $l = 0, \dots, L-1$ . In the case where all neurons and all delay elements are set to  $m_{\text{init}} = 1$ , that is,  $m_l = 1$ ,  $l = 0, \dots, L-1$ , the maximum information for recalling the stored sequence of memory patterns is given. Therefore, this condition can be called *the optimum initial condition*. Furthermore, the storage capacity  $\alpha_C$  is defined as the critical loading rate where recalling becomes unstable under the optimum initial condition. We note that the derived macrodynamical equations and the Yanai-Kim theory[14] coincide with each other.

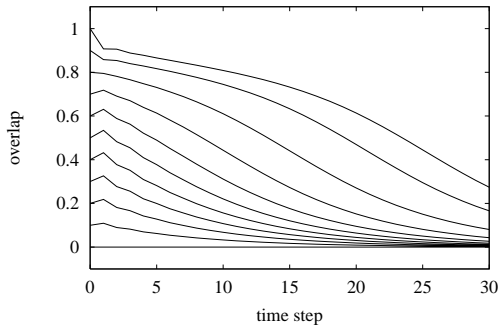


Figure 1: Dynamical behaviors of recall process ( $L = 2$ ,  $\alpha = 0.5$ : theory).

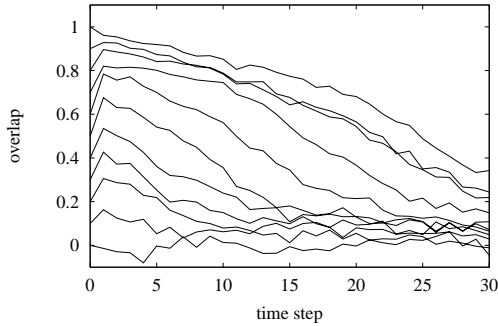


Figure 2: Dynamical behaviors of recall process ( $L = 2$ ,  $\alpha = 0.5$ : computer simulation with  $N = 2000$ ).

Some examples of the dynamical behaviors of recall processes by the above theory and computer simulations are shown in Figures 1-4. In these figures, the abscissa is time  $t$ , and the ordinate is the overlap  $m_t$ . These are the results when the all steps set initial conditions are given. Figures 1 and 3 are the results of 30-step time-dependent theoretical calculation by eqns (7)-(11) using various initial over-

laps. The details of the computer simulations shown in Figures 2 and 4 are as follows. First, a sequence of random patterns are generated. The length of the sequence is  $\alpha N = 1000$ . Each pattern is a vector in which the dimension is  $N = 2000$ . Therefore, the loading rate is  $\alpha = 0.5$ . Each element of the pattern vectors takes a value of either  $+1$  or  $-1$  with probability  $\frac{1}{2}$ . Next, the sequence is stored by correlation learning in two networks of which the number of neurons is  $N = 2000$  and the lengths of delay  $L$  are two and three. Then, various pattern sequences are generated where the initial overlaps with the 1st- $L$ th patterns stored are  $0.0 - 1.0$ . These are given as the initial state of the network. After that, 30-step calculation is carried out by using the updating rule of the network, that is, eqns (1)-(3).

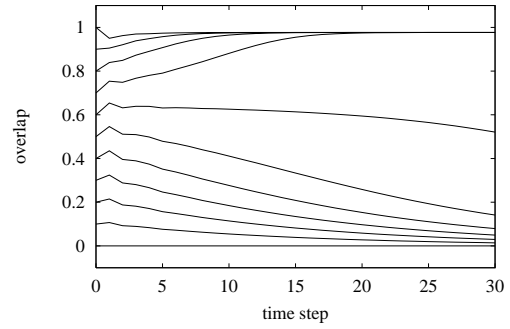


Figure 3: Dynamical behaviors of recall process ( $L = 3$ ,  $\alpha = 0.5$ : theory).

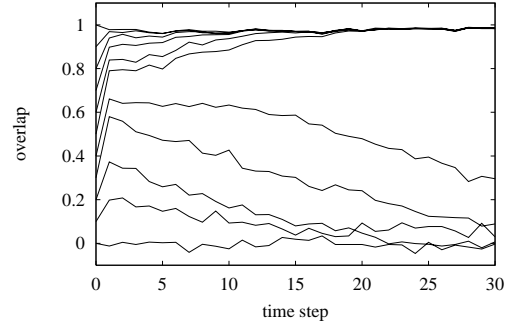


Figure 4: Dynamical behaviors of recall process ( $L = 3$ ,  $\alpha = 0.5$ : computer simulation with  $N = 2000$ ).

According to these figures, the dynamical behaviors of the overlaps obtained by the re-derived theory are in good agreement with those obtained by computer simulation. Figure 1 shows the dynamical behaviors of the overlaps in the case of  $L = 2$ . When the initial overlaps are  $0.1 - 0.7$ , the overlaps somewhat increase at the first time step. However, the recall processes eventually fail regardless of the initial overlaps. This fact indicates that the storage capacity  $\alpha_C$  in the case of  $L = 2$  is smaller than  $0.5$ . On the other hand, Figure 3 shows the dynamical behaviors of the overlaps in the case of  $L = 3$ . When the initial overlap is  $1.0$ , at the first time step the overlap somewhat decreases. However, then the overlap tends to a value near  $1.0$ . This fact indicates that the storage capacity  $\alpha_C$  in the case of  $L = 3$  is larger than  $0.5$ .

Figure 5 shows the relationship between the loading rates  $\alpha$  and the steady state overlaps  $m_\infty$  obtained by time-dependent theoretical calculations with a sufficient number of steps. Figure 6 shows the results of computer simulations under the same conditions as the theoretical calculations. In each figure, “(1)” and “(A)” denote the one step set initial condition and the all steps initial condition, respectively. In the computer simulations, the number of neurons is  $N = 500$ . The computer simulations have been carried out under five conditions :  $L = 1$ ,  $L = 3$  from the one step set initial condition,  $L = 3$  from the all steps initial condition,  $L = 10$  from the one step set initial condition, and  $L = 10$  from the all steps initial condition. Eleven simulations have been carried out at various loading rates  $\alpha$  under each condition. Here, the initial overlap  $m_{\text{init}} = 1$  in all cases. Therefore, the all steps set initial condition is equivalent to the optimum initial condition as described above. In Figure 6, data points  $\bullet$ ,  $\circ$ ,  $\blacksquare$ ,  $\square$ ,  $*$  indicate the medians of the 6th largest values in the eleven trials. Error bars indicate the third and the ninth largest values in the eleven trials.

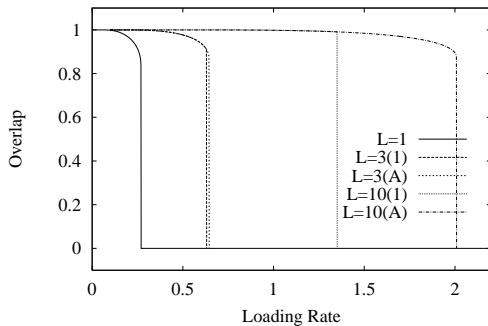


Figure 5: Relationship between loading rate  $\alpha$  and overlap  $m$ . “(1)” and “(A)” indicate one step set initial condition and all steps set initial condition, respectively (theory).

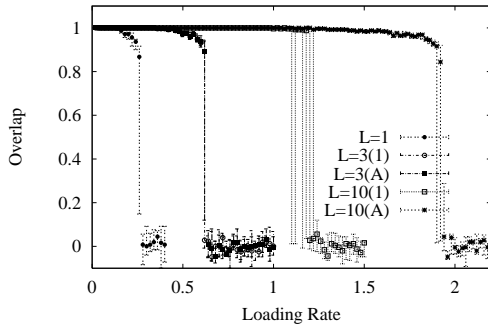


Figure 6: Relationship between loading rate  $\alpha$  and overlap  $m$ . “(1)” and “(A)” indicate one step set initial condition and all steps set initial condition, respectively (computer simulation with  $N = 500$ ).

These figures show that the steady states obtained by the re-derived theory are in good agreement with those obtained by computer simulation. In the case of  $L = 3$ , the difference between the phase transition point under the one step set initial condition and that under the optimum ini-

tial condition is small. However, in the case of  $L = 10$ , the difference is large. That is, the influence of the initial condition of the network increases as the length of delay  $L$  increases. We note that in the range between the phase transition point under the one step set initial condition and that under the optimum initial condition, there is the phenomenon that the attractor cannot be recalled from the former initial condition, although the sequence of memory patterns is certainly stored.

Figure 7 shows the relationship between the length of delay  $L$  and the critical loading rate where the phase transition occurs. This theory (Yanai-Kim theory) needs a computational complexity of  $\mathcal{O}(L^4 t)$  to obtain the macrodynamics, where  $L$  and  $t$  are the length of delay and the time step, respectively. Therefore, in this method, it is intractable to investigate the critical loading rate in the case of such a large delay  $L$ . Here, the results in the cases of  $L = 1, 2, \dots, 10$  are shown. This figure shows the following. When the initial conditions are optimum, the relationship between the length of delay  $L$  and the critical loading rate, that is, the storage capacity, seems to be almost linear. It is assumed that when the length of delay  $L$  further increases, this tendency would continue. These characteristics are analyzed in the next section. On the other hand, in the case of the one step set initial condition, the critical loading rate is not linear with the length of delay  $L$  but saturated. However, the reason for this phenomenon is the lack of initial information. We note that this saturation doesn't imply the essential saturation of the storage capacity of the delayed networks.

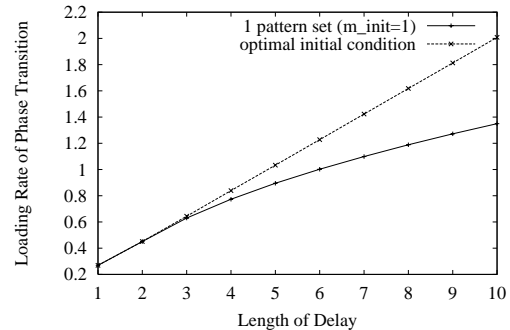


Figure 7: Length of delay  $L$  and loading rate of phase transitions. Solid line and dashed line show results for one step set initial condition and those for all steps set initial condition, respectively.

#### 4. MACROSCOPIC STEADY STATE ANALYSIS BY DISCRETE FOURIER TRANSFORMATION AND DISCUSSION

The Yanai-Kim theory re-derived in the previous section, that is, the macrodynamical equations obtained by the statistical neurodynamics, needs a computational complexity of  $\mathcal{O}(L^4 t)$  to obtain the macrodynamics shown in eqns (7) and (8), where  $L$  and  $t$  are the length of delay and the time step, respectively. Therefore, in this method, it is intractable to investigate the critical loading rate for a large  $L$  limit, that is, the asymptotic behavior of the storage capac-

$$\sigma^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\alpha [(1-U) \sin(\pi x) + U \sin \{(2L+1)\pi x\}] [1 - \cos(2L\pi x)]}{\sin(\pi x) [2 \sin^2(\pi x) - U^2 \{1 - \cos(2L\pi x)\}]} dx \quad (12)$$

ity in the large  $L$  limit. Thus, in this section, the Yanai-Kim theory in a steady state is considered. After that, we derive the macroscopic steady state equations of the delayed network by using the discrete Fourier transformation, where the computational complexity does not formally depend on  $L$ . Furthermore, the storage capacity is analytically discussed for a large  $L$  by solving the derived equations numerically.

In a steady state,  $\sigma_t$ ,  $U_t$  and  $m_t$  in eqns (7)-(11) can be expressed as  $\sigma$ ,  $U$  and  $m$ , respectively. In addition,  $v_{t-l, t-l'}$  can be expressed as  $v_{l-l'}$  because of the parallel symmetry in terms of  $v$ . Using the discrete Fourier transformation and the inverse discrete Fourier transformation, we can obtain the steady state equations in terms of the macroscopic variables of the network as eqns (12)-(15). Eqn(12) is shown at the top of this page.

$$U = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp\left(-\frac{s^2}{2\sigma^2}\right) \quad (13)$$

$$s = mL \quad (14)$$

$$m = \operatorname{erf}\left(\frac{s}{\sqrt{2}\sigma}\right) \quad (15)$$

Though the derived macroscopic steady state equations include a simple integral, their computational complexity does not formally depend on  $L$ . Therefore, we can easily perform numerical calculations for a large  $L$ .

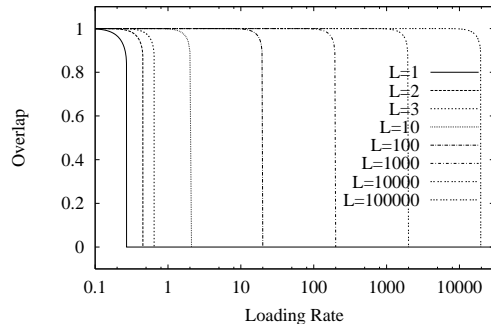


Figure 8: Relationship between loading rate  $\alpha$  and overlap  $m$ . These lines are obtained by solving steady state equations numerically. Storage capacity is  $0.195L$  with large  $L$  limit.

Figure 8 shows the relationship between the loading rate  $\alpha$  and the overlap  $m$ , which is obtained by solving these equations numerically.

Comparing Figure 5 and 8, especially considering the case of  $L = 1, 3, 10$ , we can see that the overlaps obtained by the macroscopic steady state equations in this section agree with those obtained by the dynamical calculations having sufficient time steps from the optimum initial conditions in the previous section. In other words, the phase transition points obtained by the macroscopic steady state equations

derived in this section agree with the storage capacity, that is, the phase transition points under the optimum initial condition. This fact shows that the solution to the dynamical equations from the optimum initial condition and the solution to the steady state equations support each other.

From Figure 8, we can see that the storage capacity increases in proportion to the length of delay  $L$  with a large  $L$  limit and the proportion constant is 0.195. That is, the storage capacity of the delayed network  $\alpha_C$  equals  $0.195L$  when the length of delay  $L$  is large. Though the result that the storage capacity of the delayed network is in proportion to the length of delay  $L$  is not nontrivial, the fact that this result has been proven analytically is significant. The proportion constant 0.195 is the mathematically significant number as the limit of the delayed network's storage capacity.

## 5. REFERENCES

- [1] M. Okada (1996), *Neural Networks*, 9, 1429-1458.
- [2] K. Fukushima (1973), *Kybernetik*, 12, 58-63.
- [3] S. Miyoshi and M. Okada (2000), *Trans. IEICE, J83-A* (11), 1330-1332. in Japanese.
- [4] J. J. Hopfield (1982), *Proc. National Academy of Sciences*, 79, 2554-2558.
- [5] D. Sherrington and S. Kirkpatrick (1975), *Physical Review Letters*, 35, 1792-1796.
- [6] D. J. Amit, H. Gutfreund, and H. Sompolinsky (1985), *Physical Review A*, 32, 1007-1018.
- [7] D. J. Amit, H. Gutfreund, and H. Sompolinsky (1985), *Physical Review Letters*, 55, 1530-1533.
- [8] M. Shiino and T. Fukai (1992), *Journal of Physics A: Mathematical and General*, 25, L375-L381.
- [9] S. Amari and K. Maginu (1988), *Neural Networks*, 1, 63-73.
- [10] S. Amari, S. Amari (1988), *Proceedings of IEEE International Conference on Neural Networks*, 1, 633-640.
- [11] A. Düring, A. C. C. Coolen, and D. Sherrington (1998), *J. Phys. A: Math. Gen.*, 31, 8607-8621.
- [12] M. Kawamura and M. Okada (2002), *J. Phys. A: Math. Gen.*, 35, 253-266.
- [13] M. Okada (1995), *Neural Networks*, 8 (6), 833-838.
- [14] H. Yanai and E. S. Kim (1993), *Technical report of IEICE, NC92-116*, 167-174. in Japanese.
- [15] S. Miyoshi and K. Nakayama (1995), *Proceedings of IEEE International Conference on Neural Networks*, 1955-1960.
- [16] S. Miyoshi and M. Okada (2002), *Trans. IEICE, J85-A* (1), 124-133. in Japanese.