Storage Capacity Diverges with Synaptic Efficiency in an Associative Memory Model with Synaptic Delay and Pruning

Seiji Miyoshi*, *Member, IEEE,* and Masato Okada †

Abstract

It is known that storage capacity per synapse increases by synaptic pruning in the case of a correlation-type associative memory model. However, the storage capacity of the entire network then decreases. To overcome this difficulty, we propose decreasing the connecting rate while keeping the total number of synapses constant by introducing delayed synapses. In this paper, a discrete synchronous-type model with both delayed synapses and their prunings is discussed as a concrete example of the proposal. First, we explain the Yanai-Kim theory by employing the statistical neurodynamics. This theory involves macrodynamical equations for the dynamics of a network with serial delay elements. Next, considering the translational symmetry of the explained equations, we re-derive macroscopic steady state equations of the model by using the discrete Fourier transformation. The storage capacities are analyzed quantitatively. Furthermore, two types of synaptic prunings are treated analytically: random pruning and systematic pruning. As a result, it becomes clear that in both prunings, the storage capacity increases as the length of delay increases and the connecting rate of the synapses decreases when the total number of synapses is constant. Moreover, an interesting fact becomes clear: the storage capacity asymptotically approaches $2/\pi$ due to random pruning. In contrast, the storage capacity diverges in proportion to the logarithm of the length of delay by systematic pruning and the proportion constant is $4/\pi$. These results theoretically support the significance of pruning following an overgrowth of synapses in the brain and strongly suggest that the brain prefers to store dynamic attractors such as sequences and limit cycles rather than equilibrium states.

keywords

associative memory, neural network, delay, synaptic pruning, statistical neurodynamics

1 Introduction

Robustness against noise and damage is often given as a positive feature of neural networks. Therefore, it is important to analyze neural networks with respect to synaptic pruning. Particularly in the case of correlation-type associative memory [1], models with randomly pruned synapses have been discussed in detail [2, 3, 4, 5]D As a result, it became quantitatively clear that synapse efficiency, which is defined by storage capacity per synapse, increases by synaptic pruning, although storage capacity of the entire network decreases.

On the other hand, it has often been observed that synapses are pruned following an overgrowth in real neural systems [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. Though the functional significance of this phenomenon is not known, Chechik et al. recently proposed the following hypothesis [5]. They considered cutting synapses that are lightly weighted after learning

with an excess of synapses, expecting synapse efficiency to increase by such systematic pruning. Therefore, they hypothesized that increasing synapse efficiency in this way adds functional significance to synaptic pruning following an overgrowth. They used a correlation-type auto-associative memory model to verify this hypothesis. After correlation learning, they left heavily weighted synapses in the model, in which all neurons are fully connected. Through computer simulations, they showed that the synapse efficiency increased by obtaining storage capacity. In this paper, synaptic pruning as described above is called systematic pruning. Although the hypothesis of Chechik et al. is interesting as neuroscience, there are some unclear or imperfect points from the theoretical viewpoint: for example, what degree of systematic pruning is more efficient than random pruning. Accordingly, Mimura et al. [4] analyzed this system by using the self-consistent signal to noise analysis (SCSNA) [21], which is a method of statistical mechanics. They showed that systematic pruning increased synapse efficiency by the order of $-\ln(1-R)$ over random pruning at the limit when R approached unity, where R $(0 \geq R \geq 1)$ was the rate of synaptic pruning. The important point in this case is that the storage capacity of the entire network decreased, though synapse efficiency increased by random pruning or systematic pruning.

To overcome this difficulty, we propose decreasing the connecting rate while keeping the total number of synapses constant by introducing delayed synapses with respect to a discrete synchronous-type model. In this model, the storage capacity is expected to grow with increases in synapse efficiency because synapse efficiency increases by synaptic pruning, while the total number of synapses remains constant. The discrete synchronous-type model with delayed synapses [16, 24, 28] was proposed by Fukushima [16]. Yanai and Kim [24] theoretically analyzed this model with the statistical neurodynamics [22]. Their theory closely agrees with the results of our computer simulation.

In this paper, after defining the model, we explain the Yanai-Kim theory [24, 25, 26] using the statistical neurodynamics [22], which involves macrodynamical calculations for a network with delayed synapses. The Yanai-Kim theory needs a computational complexity of $O(L^4 t)$ to obtain the macrodynamics, where L and t are the length of delay and the time step, respectively. Therefore, this theory is intractable for discussing macroscopic properties at the limit where L is extremely large [25, 27]. Thus, considering the translational symmetry of time steps, which holds in the steady state of the Yanai-Kim theory, we re-derive the macroscopic steady state equations by employing the discrete Fourier transformation, where the computational complexity does not formally depend on L [25, 26]. Using the re-derived steady state equations, storage capacities can be quantitatively discussed even for a large L limit.

Next, synaptic pruning in the delayed network is investigated theoretically, and storage capacities are evaluated quantitatively. We deal with two types of pruning: random pruning and systematic pruning. As a result, it becomes clear that in both types of pruning, storage capacity increases as the length of delay increases, while the connecting rate of synapses decreases where the total number of synapses is constant. Moreover, an interesting fact becomes clear: the storage capacity asymptotically approaches $2/\pi$ by random pruning. In contrast, the storage capacity diverges in proportion to the logarithm of the length of delay L, that is, $(4/\pi)\ln L$, by systematic pruning.

2 Delayed Network

2.1 Model

The structure of the delayed network discussed in this paper is shown in Figure 1. This figure corresponds to the case of fully synaptic connections, meaning no synaptic pruning. The network 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, i.e., this network employs

a discrete synchronous updating rule. The output of each neuron is determined by

$$x_i^{t+1} = F\left(u_i^t\right), \tag{1}$$
$$F\left(\cdot\right) = \text{sgn}\left(\cdot\right), \tag{2}$$
$$u_i^t = \sum_{l=0}^{L-1}\sum_{j=1}^{N} J_{ij}^l x_j^{t-l}, \tag{3}$$

where x_i^t denotes the output of the ith neuron at time t, and J_{ij}^l denotes the connection weight from the lth delay elements of the jth neuron to the ith neuron. Here, sgn is the sign function defined as

$$\text{sgn}\left(u\right) = \begin{cases} +1, & u \geq 0, \\ -1, & u < 0. \end{cases} \tag{4}$$

In this paper, the limit $N \to \infty$ is used unless stated otherwise.

Let us consider the storing sequence of αN memory patterns, $\boldsymbol{\xi}^1 \to \boldsymbol{\xi}^2 \to \cdots \to \boldsymbol{\xi}^\mu \to \cdots \to \boldsymbol{\xi}^{\alpha N}$. Here, α and αN are the loading rate and the length of the sequence, respectively. Each component of $\boldsymbol{\xi}^\mu$ is assumed to be an independent random variable that takes a value of either $+1$ or -1 according to the following probabilities.

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

The synaptic weight J_{ij}^l is determined by correlation learning:

$$J_{ij}^l = \frac{c_l}{N}\sum_\mu \xi_i^{\mu+1+l}\xi_j^\mu, \tag{6}$$

where c_l is the strength of the lth delay step.

Correlation learning is an algorithm based on the Hebb rule, and it is inferior to the error correcting learning in terms of storage capacity. However, as seen in (6), it is not necessary to re-learn all patterns that were stored in the past when adding new patterns. Furthermore, correlation learning has been analyzed by many researchers due to its simplicity.

2.2 Dynamical Behaviors of Macroscopic Order Parameters by Statistical Neurodynamics

In the case of a small loading rate α, if a state close to one or a set of the patterns stored as a sequence is given to the network, the stored sequence of memory patterns is retrieved. However, when the loading rate α increases, the memory fails at a certain α. That is, even if a state close to one or a set of the patterns stored as a sequence is given to the network, the state of the network tends to omit the stored sequence of memory patterns. Moreover, even if one or a set of the patterns itself is given to the network, the state of the network tends to leave the stored sequence of memory patterns. This phenomenon of the memory suddenly becoming unstable at a critical loading rate can be considered a kind of phase transition. Here, the storage capacity α_C is defined as the critical loading rate where recall becomes unstable.

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

$$m_\mu^t = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu x_i^t. \tag{7}$$

Using this definition, when the state of the network at time t and the μth pattern agree perfectly, the overlap m_t^μ is equal to unity. When they have no correlation, the overlap m_t^μ is equal to zero. Therefore, the overlap provides a means of measuring recall quality.

3

Amari and Maginu proposed the statistical neurodynamics for the associative memory model [22, 23, 31]. This analytical method handles the dynamical behavior of the associative memory model macroscopically, where cross-talk noise is regarded as a Gaussian random variable with a mean of zero and a time-dependent variance of σ_t^2. They then derived recursive relations for the variance and the overlap.

Yanai and Kim applied this method to the present model, and succeeded in obtaining macroscopic state transition equations[24]. We will briefly explain their derivation as follows.

The total input of the ith neuron at time t is given as

$$
\begin{aligned}
u_i^t &= \sum_{l=0}^{L-1} \sum_{j=1}^{N} J_{ij}^l x_j^{t-l} \\
&= s^t \xi_i^{t+1} + z_i^t, &(8)\\
s^t &= \sum_{l=0}^{L-1} c_l m_{t-l}^{t-l}, &(9)\\
z_i^t &= \sum_{l=0}^{L-1} c_l \sum_{\nu \neq t} \xi_i^{\nu+1} m_{\nu-l}^{t-l}. &(10)
\end{aligned}
$$

The first term in (8) is the signal useful for recall, while the second term is cross-talk noise that prevents ξ_i^{t+1} from being recalled. This procedure is called a signal-to-noise analysis.

We can then use the first-order Taylor expansion regarding $F(\cdot)$ to obtain

$$
\begin{aligned}
m_\mu^t &= \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu x_i^t \\
&= \bar{m}_\mu^t + U_t \sum_{l'=0}^{L-1} c_{l'} m_{\mu-l'-1}^{t-l'-1}, &(11)\\
\bar{m}_\mu^t &= \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu F \left(\sum_{l=0}^{L-1} \sum_{j=1}^{N} \frac{c_l}{N} \right. \\
&\qquad\qquad \left. \times \sum_{\nu \neq \mu-l-1} \xi_i^{\nu+1+l} \xi_j^\nu x_j^{t-l-1} \right), &(12)\\
U_t &= \frac{1}{N} \sum_{i=1}^{N} F' \left(\sum_{l=0}^{L-1} \sum_{j=1}^{N} \frac{c_l}{N} \right. \\
&\qquad\qquad \left. \times \sum_{\nu \neq \mu-l-1} \xi_i^{\nu+1+l} \xi_j^\nu x_j^{t-l-1} \right), &(13)
\end{aligned}
$$

where $F'(\cdot)$ is the differentiation of $F(\cdot)$.

Taking the correlation in the cross-talk noise z_i^t into account, we can derive the following macrodynamical equations using (1)-(13) (see Appendix A).

$$
\begin{aligned}
\sigma_t^2 &= \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l,t-l'}, &(14)\\
v_{t-l,t-l'} &= \alpha \delta_{l,l'} \\
&\quad + U_{t-l} U_{t-l'} \\
&\quad \times \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1,t-l'-k'-1}
\end{aligned}
$$

4

$$+ \quad \alpha \left(c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l} \right), \tag{15}$$

$$U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{t-1}} \exp\left(-\frac{\left(s^{t-1} \right)^2}{2\sigma_{t-1}^2} \right), \tag{16}$$

$$s^t = \sum_{l=0}^{L-1} c_l m_{t-l}, \tag{17}$$

$$m_{t+1} = \operatorname{erf}\left(\frac{s^t}{\sqrt{2}\sigma_t} \right), \tag{18}$$

where m_t denotes m_t^t. $v_{t-l,t-l'} = \sum_{\mu \neq t} m_{\mu-l}^{t-l} m_{\mu-l'}^{t-l'}$. σ_t^2 is the variance of the cross-talk noise. U_t is a kind of susceptibility, which measures the sensitivity of neuron output with respect to the external input. 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 $\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-u^2\right) du$ denotes the error function. In this paper, the initial condition is that the states of all neurons and all delay elements are set to be the stored pattern sequences. In this case, $m_l = 1 \ (l = 0, \cdots, L-1)$ and $v_{l,l} = \alpha \ (l = 0, \cdots, L-1)$.

2.3 Macroscopic Steady State Analysis by Discrete Fourier Transformation and Discussion

The Yanai-Kim theory explained in the previous section, which involves the macrodynamical equations obtained by the statistical neurodynamics, needs a computational complexity of $O(L^4 t)$ to obtain the macrodynamics shown in (14) and (15), where L and t are the length of delay and the time step, respectively [24, 25, 26]. Therefore, in this method, it is difficult to investigate the critical loading rate for a large L limit, i.e., the asymptotic behavior of the storage capacity in a large L limit. Thus, Miyoshi, Yanai and Okada considered the Yanai-Kim theory in a steady state and derived the macroscopic steady state equations of the delayed network. Furthermore, the storage capacity was analyzed for a large L by solving the derived equations numerically [25, 26].

We will briefly explain the derivation of the macroscopic steady state equations to make the present paper self-contained.

For simplicity, let us assume that $c_l = 1$, $l = 0, \cdots, L-1$. In a steady state, $v_{t-l,t-l'}$ can be expressed as $v_{l-l'}$ because of the translational symmetry in terms of time step. Therefore, by modifying (14) and (15), we obtain

$$\sigma^2 = \sum_{n=1-L}^{L-1} (L - |n|) v(n), \tag{19}$$

$$v(n) = \alpha \delta_{n,0}$$
$$+ \quad U^2 \sum_{i=1-L}^{L-1} (L - |i|) v(n-i) + U \alpha d(n), \tag{20}$$

$$d(n) = \begin{cases} 1, & |n| = 1, 2, \cdots, L \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where $n = l - l'$, $i = k - k'$, $v(n)$ denotes v_n and δ is Kronecker's delta.

Using the discrete Fourier transformation, we can obtain the steady state equations in terms of the network's macroscopic variables as (22)-(25) (see Appendix B).

$$\sigma^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\alpha \left[(1-U) \sin(\pi x) + U \sin\{(2L+1)\pi x\} \right] \left[1 - \cos(2L\pi x) \right]}{\sin(\pi x) \left[2\sin^2(\pi x) - U^2 \{1 - \cos(2L\pi x)\} \right]} dx, \tag{22}$$

5

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

$$s = mL, \tag{24}$$

$$m = \operatorname{erf}\left(\frac{s}{\sqrt{2}\sigma}\right). \tag{25}$$

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 2 shows the results of theoretical calculations in cases where $L = 1, 3$ and 10, which are obtained by solving these equations numerically. Figure 3 shows the results of computer simulations. In these Figures, the abscissa is the loading rate α. In the computer simulations, the number of neurons is $N = 500$. The initial condition is that the states of all neurons and all delay elements are set to be the stored pattern sequences. The steady state overlaps m_∞ are obtained by calculations with a sufficient number of steps. Eleven simulations were carried out for each combination of loading rate α and lengths of delay L. Data points ● , ○ , ■ indicate the medians of the sixth largest values for $L =$1,3 and 10, respectively, in the eleven trials. Error bars indicate the third and ninth largest values in the eleven trials. In each trial, the loading rate is increased by adding new patterns.

These figures show that the steady states obtained by the derived theory agree closely with those obtained by computer simulation. Therefore, in the case of a large L, only the theoretical calculations are executed. Figure 4 shows the results, while Figure 5 shows the relationship between the length of delay L and the storage capacity α_C.

From these figures, we can see that the storage capacity increases in proportion to the length of delay L with a large L limit and a proportion constant of 0.195. In other words, the storage capacity of the delayed network α_C equals $0.195L$ when the length of delay L is large [25, 26]. Although the result indicating that the delayed network's storage capacity is in proportion to the length of delay L may be trivial, the fact that this result has been proven analytically is significant. Moreover, the proportion constant 0.195 is a mathematically significant number because it represents the limit of the delayed network's storage capacity.

3   Synaptic Pruning

3.1   Necessity of Analyzing Synaptic Pruning

During brain development, the phenomenon of synaptic pruning following overgrowth [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] can be observed. Since this pruning following overgrowth seems to be a universal phenomenon occurring in almost all areas – visual cortex, motor area, association area, and so on – it is important to analyze synaptic pruning and to discuss its properties quantitatively.

In real neural systems, some synaptic delay is inevitable. This property can be analyzed with a model that involves both delay elements and synaptic pruning. For example, Figure 6 shows that a delay of three time steps can be represented by pruning the first, second, fourth and fifth synapses, and a five-time-step delay can be represented by pruning the first, second, third and fourth synapses with a model whose length of delay is five. From this perspective, analyzing a model with both delay elements and synaptic pruning is significant.

Moreover, in the case of a delayed network with no pruning, it is obvious that storage capacity increases as the length of delay L increases. On the contrary, it is interesting to analyze the storage capacity of a delayed network that has a constant number of synapses by introducing synaptic pruning.

It has been reported that the synapse efficiency, which is defined as storage capacity per synapse, increases due to synaptic pruning in networks with no delay elements [2, 5]. Two types of pruning can be considered, namely random pruning and systematic pruning, which are typical

methods [2, 3, 5]. Mimura et al. [4] showed that synapse efficiency converged to $\frac{2}{\pi}$ by random pruning and diverged as $\frac{2}{\pi}(-2\ln c)$ by systematic pruning at the limit where the connecting rate c is extremely small. Here, the relation between connecting rate c and pruning rate R was given by $c = 1 - R$. The important point here is that the storage capacity of the entire network decreases, since the number of synapses decreases.

In the following discussion, a delayed network with synaptic pruning is analyzed on the basis of the macrodynamical equations and macroscopic steady state equations re-derived in the former section. We consider two types of pruning – random pruning and systematic pruning – for synaptic pruning.

3.2 Random Pruning

In this section, synapses of a delayed network are randomly pruned. Random pruning of synapses can be realized without any complicated control mechanism, so it is important to investigate its effect on the dynamical behavior of pattern recall and storage capacity.

In the random synaptic pruning model, synaptic connections are constituted as

$$J_{ij}^l = \frac{c_l c_{ij}^l}{Nc} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu}, \tag{26}$$

$$\text{Prob}[c_{ij}^l = 1] = 1 - \text{Prob}[c_{ij}^l = 0] = c, \tag{27}$$

where c is the connecting rate.

Modifying (26), we obtain

$$J_{ij}^l = \frac{c_l}{N} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu}$$

$$+ \frac{c_l \left(c_{ij}^l - c \right)}{Nc} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu}. \tag{28}$$

Using (28), we obtain the total input of the ith neuron at time t as

$$u_i^t = \sum_{l=0}^{L-1} \sum_{j=1}^{N} J_{ij}^l x_j^{t-l}$$

$$= s^t \xi_i^{t+1} + z_i^t$$

$$+ \sum_{l=0}^{L-1} \sum_{j \neq i} \frac{c_l \left(c_{ij}^l - c \right)}{Nc} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu} x_j^{t-l}, \tag{29}$$

$$s^t = \sum_{l=0}^{L-1} c_l m_{t-l}, \tag{30}$$

$$z_i^t = \sum_{l=0}^{L-1} c_l \sum_{\nu \neq t} \xi_i^{\nu+1} m_{\nu-l}^{t-l}. \tag{31}$$

As in the case of a fully connected network, the first, second and third terms of (29) are useful signals for recall, cross-talk noise and new noise generated by synaptic pruning, respectively.

In the third term of (29),

$$x_j^{t-l} = F\left(u_j^{t-l-1} \right)$$

$$= \sum_{l'=0}^{L-1} \sum_{k \neq j} J_{jk}^{l'} x_k^{t-l-1-l'}$$

$$= F\left(\sum_{l'=0}^{L-1}\sum_{k\neq j}\frac{c_{l'}c_{jk}^{l'}}{Nc}\sum_{\nu}\xi_j^{\nu+1+l'}\xi_k^{\nu}x_k^{t-l-1-l'}\right)$$

$$= x_j^{t-l,(\mu)} + \left(\sum_{l'=0}^{L-1}\sum_{k\neq j}\frac{c_{l'}c_{jk}^{l'}}{Nc}\xi_j^{\mu+1+l'}\xi_k^{\mu}x_k^{t-l-1-l'}\right)x_j'^{t-l,(\mu)}, \tag{32}$$

where

$$x_j^{t-l,(\mu)} = F\left(\sum_{l'=0}^{L-1}\sum_{k\neq j}\frac{c_{l'}c_{jk}^{l'}}{Nc}\sum_{\nu\neq\mu}\xi_j^{\nu+1+l'}\xi_k^{\nu}x_k^{t-l-1-l'}\right), \tag{33}$$

$$x_j'^{t-l,(\mu)} = F'\left(\sum_{l'=0}^{L-1}\sum_{k\neq j}\frac{c_{l'}c_{jk}^{l'}}{Nc}\sum_{\nu\neq\mu}\xi_j^{\nu+1+l'}\xi_k^{\nu}x_k^{t-l-1-l'}\right). \tag{34}$$

Using (32), the third term of (29) becomes

$$\sum_{l=0}^{L-1}\sum_{j\neq i}\frac{c_l\left(c_{ij}^l-c\right)}{Nc}\sum_{\mu}\xi_i^{\mu+1+l}\xi_j^{\mu}x_j^{t-l}$$

$$= \sum_{l=0}^{L-1}\sum_{j\neq i}\frac{c_l\left(c_{ij}^l-c\right)}{Nc}\sum_{\mu}\xi_i^{\mu+1+l}\xi_j^{\mu}x_j^{t-l,(\mu)}$$

$$+ \sum_{l=0}^{L-1}\sum_{j\neq i}\frac{c_l\left(c_{ij}^l-c\right)}{Nc}\sum_{\mu}\xi_i^{\mu+1+l}\xi_j^{\mu}$$

$$\sum_{l'=0}^{L-1}\sum_{k\neq j}\frac{c_{l'}c_{jk}^{l'}}{Nc}\xi_j^{\nu+1+l'}\xi_k^{\mu}x_k^{t-l-1-l'}x_j'^{t-l,(\mu)}. \tag{35}$$

The second term of (35) becomes

$$\frac{1}{c^2}\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}c_lc_{l'}\sum_{\mu}\xi_i^{\mu+1+l}E\left[\xi_k^{\mu}x_k^{t-l-1-l'}\right]E\left[\left(c_{ij}^l-c\right)c_{jk}^{l'}\xi_j^{\mu}x_j'^{t-l,(\mu)}\right], \tag{36}$$

where $E\left[\xi_k^{\mu}x_k^{t-l-1-l'}\right]$ and $E\left[\left(c_{ij}^l-c\right)c_{jk}^{l'}\xi_j^{\mu}x_j'^{t-l,(\mu)}\right]$ obey $N\left(0,1/N\right)$ and $N\left(0,O\left(1/N\right)\right)$, respectively. Therefore, the second term of (35) becomes 0 for a large N limit. Here, $E[\cdot]$ and $N\left(a,\sigma^2\right)$ stand for an average and a Gaussian distribution with average a and variance σ^2, respectively.

Using this result and (35), the third term of (29) becomes [3]

$$\sum_{l=0}^{L-1}\sum_{j\neq i}\frac{c_l\left(c_{ij}^l-c\right)}{Nc}\sum_{\mu}\xi_i^{\mu+1+l}\xi_j^{\mu}x_j^{t-l,(\mu)} \sim N\left(0,\frac{\alpha(1-c)}{c}\sum_{l=0}^{L-1}c_l^2\right). \tag{37}$$

As a result, we can obtain the macrodynamical equations for random pruning as follows.

$$\tilde{\sigma}_t^2 = \sigma_t^2 + \frac{\alpha\left(1-c\right)}{c}\sum_{l=0}^{L-1}c_l^2, \tag{38}$$

$$\sigma_t^2 = \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}c_lc_{l'}v_{t-l,t-l'}, \tag{39}$$

$$
\begin{aligned}
v_{t-l,t-l'} &= \alpha \delta_{l,l'} \\
&+ U_{t-l} U_{t-l'} \\
&\times \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1,t-l'-k'-1} \\
&+ \alpha \left(c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l} \right),
\end{aligned} \tag{40}
$$

$$
U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_{t-1}} \exp\left(-\frac{\left(s^{t-1}\right)^2}{2\tilde{\sigma}_{t-1}^2} \right), \tag{41}
$$

$$
s^t = \sum_{l=0}^{L-1} c_l m_{t-l}, \tag{42}
$$

$$
m_{t+1} = \mathrm{erf}\left(\frac{s^t}{\sqrt{2}\tilde{\sigma}_t} \right), \tag{43}
$$

where the initial conditions are the same as in the case of a fully connected network. δ is Kronecker's delta. Equation (38) means that the variance $\tilde{\sigma}_t^2$ after pruning is the sum of the variance of cross-talk noise among patterns and the variance of new noise generated by pruning. Using the discrete Fourier transformation, as for a fully connected network, the macroscopic steady state equations in the case of random pruning become

$$
\tilde{\sigma}^2 = \sigma^2 + \frac{\alpha(1-c)}{c} \sum_{l=0}^{L-1} c_l^2, \tag{44}
$$

$$
U = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}} \exp\left(-\frac{s^2}{2\tilde{\sigma}^2} \right), \tag{45}
$$

$$
s = mL, \tag{46}
$$

$$
m = \mathrm{erf}\left(\frac{s}{\sqrt{2}\tilde{\sigma}} \right), \tag{47}
$$

where σ^2 is given by (22).

It is obvious that storage capacity increases with the length of delay L if the connecting rate c is constant. Therefore, the storage capacity α_C is investigated under the condition that $c \times L$ is constant. This means that (44)-(47) are solved numerically and that the steady state overlaps m_∞ are investigated by using $c = 1/L$, where $c_l = 1$. Figure 7 shows the results of theoretical calculations and computer simulations when $L = 1, 2, 3, 5$ and 10. In this figure, the abscissa is loading rate α. In the computer simulations, the number of neurons is $N = 500$. The initial condition is that the states of all neurons and all delay elements are set to be the stored pattern sequences. The steady state overlaps m_∞ are obtained by calculations with a sufficient number of steps. Eleven simulations were carried out for each combination of loading rates α and lengths of delay L. Data points ● , ○ , ■ , □ , * indicate the medians of the sixth largest values for $L =$1,2,3,5 and 10, respectively, in the eleven trials. Error bars indicate the third and the ninth largest values in the eleven trials. In each subsequent trial, the loading rate is increased by adding new patterns.

Figure 7 displays the following results. In the case of $L = 1 \left(c = \frac{1}{L} = 1.0 \right)$, which is fully connected with no delay elements, the recurrent neural network's storage capacity α_C for sequential association is 0.269. This agrees with the results of the previous works [29, 30, 31]. As the length of delay L increases, storage capacity α_C increases even though the total number of synapses is constant. This phenomenon is due to the time lag of synaptic inputs by delays, which reduces the statistical correlation among synaptic inputs. As a result, variance of the noise component decreases. This figure shows that theoretical results closely agree with the simulation results. Therefore, only a theoretical calculation is executed when the length of delay L is large, and the results of this calculation are shown in Figure 8.

9

The properties of $L = \infty$ in Figure 8 are obtained as follows. The first term of the r.h.s. of (44) can be written as $\sigma^2 = \alpha r$ from (22). Therefore, we first numerically investigated the dependence of r on L. Figure 9 shows the results. The straight line in this figure shows a first-order approximation, which is obtained by using a least squares method, of the relation between $\log L$ and $\log r$ at the phase transition point. We can see that r is $O(L^{1.45})$ by reading the slope of the line. On the other hand, considering $c = 1/L$ and $c_l = 1$, when L is extremely large, the second term of the r.h.s. of (44) becomes αL^2, that is $O(L^2)$. Therefore, only the second term is effective in the r.h.s. of (44) when L is extremely large, and (44) becomes

$$\tilde{\sigma}^2 \rightarrow \alpha L^2. \tag{48}$$

Based on these considerations, the properties of $L = \infty$ in Figure 8 are obtained by ignoring the first term in the r.h.s. of (44). Figure 8 shows that the steady state overlap asymptotically approaches that of $L = \infty$ obtained above as L becomes large.

Now, in the case of $L = \infty$, the storage capacity can be obtained analytically as follows. Substituting (46) into (47), we obtain

$$m = \text{erf}\left(\frac{mL}{\sqrt{2}\tilde{\sigma}}\right). \tag{49}$$

Equation (49) has nontrivial solutions $m \neq 0$ within the range where the slope of the r.h.s. at $m = 0$ is greater than 1. Here, the slope of the r.h.s. of (49) regarding m can be written as

$$\frac{d}{dm}\text{erf}\left(\frac{mL}{\sqrt{2}\tilde{\sigma}}\right) = \frac{L}{\tilde{\sigma}}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{m^2 L^2}{2\tilde{\sigma}^2}\right). \tag{50}$$

Therefore, we can obtain the critical value of the noise $\tilde{\sigma}_c^2$ as

$$\tilde{\sigma}_c^2 = \frac{2}{\pi}L^2. \tag{51}$$

From (48) and (51), the storage capacity α_C of random pruning at the limit when L approaches ∞ is obtained as

$$\alpha_C = \frac{2}{\pi} \simeq 0.637. \tag{52}$$

Figure 8 shows that the storage capacity approaches this value asymptotically as L increases.

3.3 Systematic Pruning

Chechik et al.[5] discussed the functional significance of synaptic pruning following overgrowth on the basis of a correlation-type associative memory model. They pointed out that synapse efficiency, which is storage capacity per synapse, increases by cutting synapses that are lightly weighted after correlation learning.

This type of systematic pruning can be expressed by nonlinear function $f(\cdot)$ shown in Figure 10. Synapses in the range of $-z_{th} < z < +z_{th}$ are pruned by $f(\cdot)$. In this case, synaptic connections are constituted by

$$J_{ij}^l = \frac{c_l\sqrt{\alpha N}}{N}f\left(T_{ij}^l\right), \tag{53}$$

$$T_{ij}^l = \frac{1}{\sqrt{\alpha N}}\sum_\mu \xi_i^{\mu+1+l}\xi_j^\mu. \tag{54}$$

Equation (54) is a stochastic variable that obeys normal distribution $N(0, 1)$. Therefore, the relationship between the connection rate c and z_{th} is given by

$$c = \int_{\{z|f(z)\neq 0\}} Dz = 1 - \text{erf}\left(\frac{z_{th}}{\sqrt{2}}\right), \tag{55}$$

10

where Dz stands for $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz$, and the integral is from $-\infty$ to $+\infty$.

Modifying the connection weight J_{ij}^l[3], we obtain

$$
\begin{aligned}
J_{ij}^l &= \frac{c_l\sqrt{\alpha N}}{N}f\left(T_{ij}^l\right) \\
&= \frac{c_l\sqrt{\alpha N}}{N}\left(JT_{ij}^l + \left(f\left(T_{ij}^l\right) - JT_{ij}^l\right)\right),
\end{aligned}
\tag{56}
$$

where

$$
\begin{aligned}
J &= \int Dx\, f'(x) \\
&= \int_{-\infty}^{\infty}\frac{dx}{\sqrt{2\pi}}\exp-\frac{x^2}{2}f'(x) \\
&= \int_{-\infty}^{\infty}\frac{dx}{\sqrt{2\pi}}\exp-\frac{x^2}{2}xf(x) \\
&= \int Dx\, x f(x).
\end{aligned}
\tag{57}
$$

Using these modification, we obtain the total input of the ith neuron at time t is given as

$$
\begin{aligned}
u_i^t &= \sum_{l=0}^{L-1}\sum_{j=1}^{N}J_{ij}^l x_j^{t-l} \\
&= \sum_{l=0}^{L-1}\sum_{j=1}^{N}\frac{c_l\sqrt{\alpha N}}{N}\left(JT_{ij}^l + \left(f\left(T_{ij}^l\right) - JT_{ij}^l\right)\right)x_j^{t-l} \tag{58} \\
&= \sum_{l=0}^{L-1}\sum_{j=1}^{N}\frac{c_l J}{N}\sum_\mu \xi_i^{\mu+1+l}\xi_j^\mu x_j^{t-l} \\
&\quad + \sum_{l=0}^{L-1}\sum_{j=1}^{N}\frac{c_l\sqrt{\alpha N}}{N}\left(f\left(T_{ij}^l\right) - JT_{ij}^l\right)x_j^{t-l} \tag{59} \\
&= \left(J\sum_{l=0}^{L-1}c_l m_{t-l}^{t-l}\right)\xi_i^{t+1} \\
&\quad + J\sum_{l=0}^{L-1}c_l\sum_{\nu\neq t}\xi_i^{\nu+1}m_{\nu-l}^{t-l} \\
&\quad + \sum_{l=0}^{L-1}\sum_{j=1}^{N}\frac{c_l\sqrt{\alpha N}}{N}\left(f\left(T_{ij}^l\right) - JT_{ij}^l\right)x_j^{t-l}. \tag{60}
\end{aligned}
$$

As in the case of a fully connected network, the first, second and third terms of (60) are useful signals for recall, cross-talk noise and new noise generated by nonlinear transformation, respectively. Here, the average of the third term equals 0, and the variance equals [3]

$$
\begin{aligned}
&E\left[\left(\sum_{l=0}^{L-1}\sum_{j=1}^{N}\frac{c_l\sqrt{\alpha N}}{N}\left(f\left(T_{ij}^l\right) - JT_{ij}^l\right)x_j^{t-l}\right)^2\right] \\
&= E\left[\alpha\sum_{l=0}^{L-1}c_l^2\frac{1}{N}\sum_j\left(f\left(T_{ij}^l\right) - JT_{ij}^l\right)^2\right]
\end{aligned}
$$

11

$$= \alpha \sum_{l=0}^{L-1} c_l^2 \int Dx \, (f(x) - J\tilde{x})^2$$

$$= \alpha \sum_{l=0}^{L-1} c_l^2 \left(\int Dx f(x)^2 - 2J \int Dxx f(x) + J^2 \int Dxx^2 \right)$$

$$= \alpha \sum_{l=0}^{L-1} c_l^2 \left(\int Dx f(x)^2 - 2J^2 + J^2 \right)$$

$$= \alpha \left(\tilde{J}^2 - J^2 \right) \sum_{l=0}^{L-1} c_l^2, \tag{61}$$

where $\tilde{J}^2 = \int Dx \, (f(x))^2$. As a result, we can obtain the macrodynamical equations for systematic pruning as follows.

$$\tilde{\sigma}_t^2 = \sigma_t^2 + \alpha \left(\tilde{J}^2 - J^2 \right) \sum_{l=0}^{L-1} c_l^2, \tag{62}$$

$$\sigma_t^2 = J^2 \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l,t-l'}, \tag{63}$$

$$\tilde{J}^2 = \int Dx \, (f(x))^2, \tag{64}$$

$$J = \int Dxx f(x), \tag{65}$$

$$
\begin{aligned}
v_{t-l,t-l'} =\ & \alpha \delta_{l,l'} \\
+\ & U_{t-l} U_{t-l'} \\
& \times \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1,t-l'-k'-1} \\
+\ & \alpha \left(c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l} \right),
\end{aligned} \tag{66}
$$

$$U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_{t-1}} \exp \left(-\frac{(s^{t-1})^2}{2\tilde{\sigma}_{t-1}^2} \right), \tag{67}$$

$$s^t = J \sum_{l=0}^{L-1} c_l m_{t-l}, \tag{68}$$

$$m_{t+1} = \mathrm{erf} \left(\frac{s^t}{\sqrt{2}\tilde{\sigma}_t} \right), \tag{69}$$

where the initial conditions are the same as in the case of a fully connected network. δ is Kronecker's delta. Equation (62) means that the variance $\tilde{\sigma}_t^2$ after pruning is the sum of the variance σ_t^2 of cross-talk noise among patterns and the variance $\alpha \left(\tilde{J}^2 - J^2 \right) \sum_{l=0}^{L-1} c_l^2$ of new noise generated by pruning. Using the discrete Fourier transformation, as in the case of full connections, the macroscopic steady state equations in the case of systematic pruning become

$$\tilde{\sigma}^2 = \sigma^2 + \alpha \left(\frac{\tilde{J}^2}{J^2} - 1 \right) \sum_{l=0}^{L-1} c_l^2, \tag{70}$$

$$U = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}} \exp \left(-\frac{s^2}{2\tilde{\sigma}^2} \right), \tag{71}$$

$$s = mL, \tag{72}$$

$$m = \mathrm{erf} \left(\frac{s}{\sqrt{2}\tilde{\sigma}} \right), \tag{73}$$

12

where σ^2 is given by (22).

As for random pruning, the storage capacity α_C is investigated under the condition that $c \times L$ is constant. This means that (70)-(73) are solved numerically, and the steady state overlaps m_∞ are investigated by using $c = 1/L$, where $c_l = 1$. Figure 11 shows the results of theoretical calculations and computer simulations when $L = 1, 2, 3, 5$ and 10. In this figure, the abscissa is the loading rate α. In the computer simulations, the number of neurons is $N = 500$, and the steady state overlaps m_∞ are obtained by calculations with a sufficient number of steps. Eleven simulations were carried out for each combination of loading rate α and length of delays L. Data points $\bullet$, $\circ$, $\blacksquare$, $\square$, $*$ indicate the medians of the sixth largest values for L =1,2,3,5 and 10, respectively, in the eleven trials. Error bars indicate the third and the ninth largest values in the eleven trials. In each trial, the loading rate is increased by adding new patterns.

Figure 11 shows that as the length of delay L increases, storage capacity α_C increases, though the total number of synapses is constant. This figure also shows that theoretical results closely agree with the simulation results. Therefore, only a theoretical calculation is executed when the length of delay L is large. Figure 12 shows the results. Figure 13 shows the relationship between the length of delay L and the storage capacities.

Here, we investigate the dependence of the first and second terms of the r.h.s. of (70) on L in the same manner as random pruning to obtain the asymptotic storage capacity analytically when L is extremely large.

The first term of the r.h.s. of (70) can be written as $\sigma^2 = \alpha r$ from (22). Therefore, we first numerically investigate the dependence of r on L. Figure 14 shows the results. The straight line in this figure shows a first-order approximation, which is obtained by using a least squares method, of the relation between $\log L$ and $\log r$ at the phase transition point. We can see that r is $O(L^{1.33})$ by reading the slope of the line.

On the other hand, the dependence of the second term on L can be obtained as follows. When L is extremely large, that is, when $c = 1/L$ is extremely small and z_{th} is extremely large, the connection rate c of (55) is as follows.

$$
\begin{aligned}
c &= \int_{\{z|f(z)\neq 0\}} Dz \\
&= 1 - \mathrm{erf}\left(\frac{z_{th}}{\sqrt{2}}\right) \\
&\to \sqrt{\frac{2}{\pi}} z_{th}^{-1} \exp\left(-\frac{z_{th}^2}{2}\right), \quad z_{th} \to \infty.
\end{aligned}
\tag{74}
$$

J and $\tilde{J}^2$ of (65) and (64) become

$$
\begin{aligned}
J &= 2\int_t^\infty Dz z^2 \\
&= \sqrt{\frac{2}{\pi}} z_{th} \exp\left(-\frac{z_{th}^2}{2}\right) + 1 - \mathrm{erf}\left(\frac{z_{th}}{\sqrt{2}}\right) \\
&\to \sqrt{\frac{2}{\pi}} z_{th} \exp\left(-\frac{z_{th}^2}{2}\right), \quad z_{th} \to \infty,
\end{aligned}
\tag{75}
$$

$$
\begin{aligned}
\tilde{J}^2 &= 2\int_t^\infty Dz z^2 \\
&= J.
\end{aligned}
\tag{76}
$$

Considering $c = 1/L$ and $c_l = 1$, from (74)C (75) and (76), the second term of (70) can be transformed as

$$
\alpha\left(\frac{\tilde{J}^2}{J^2} - 1\right)\sum_{l=0}^{L-1} c_l^2 = \alpha\left(\frac{1}{J} - 1\right) L
$$

13

$$\rightarrow \quad \alpha \left(\frac{1}{\sqrt{\frac{2}{\pi}} z_{th} \exp\left(-\frac{z_{th}^2}{2}\right)} - 1 \right) L$$

$$\rightarrow \quad \alpha \left(\frac{1}{z_{th}^2 c} - 1 \right) L$$

$$\rightarrow \quad \alpha \frac{1}{z_{th}^2 c} L$$

$$\rightarrow \quad \alpha \frac{1}{-2c \ln c} L$$

$$= \quad \alpha \frac{L^2}{2 \ln L}. \tag{77}$$

Since (77) is $O(L^2)$, only the second term is effective in the r.h.s. of (70) when L is extremely large. Based on these considerations, the storage capacity of systematic pruning can be obtained as follows. Substituting (72) into (73), we obtain

$$m = \mathrm{erf}\left(\frac{mL}{\sqrt{2}\tilde{\sigma}}\right). \tag{78}$$

Equation (78) has nontrivial solutions $m \neq 0$ within the range where the slope of the r.h.s. at $m = 0$ is greater than 1. Here, the slope of the r.h.s. of (78) regarding m can be written as

$$\frac{d}{dm}\mathrm{erf}\left(\frac{mL}{\sqrt{2}\tilde{\sigma}}\right) = \frac{L}{\tilde{\sigma}}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{m^2 L^2}{2\tilde{\sigma}^2}\right). \tag{79}$$

Therefore, we can obtain the critical value of the noise $\tilde{\sigma}_c^2$ as

$$\tilde{\sigma}_c^2 = \frac{2}{\pi}L^2. \tag{80}$$

From (77) and (80), the storage capacity α_C of systematic pruning at the limit when L approaches ∞ is obtained as

$$\alpha_C = \frac{4}{\pi}\ln L. \tag{81}$$

Figures 13 shows that the storage capacity of large L is parallel with the line of $\frac{4}{\pi}\ln L$ when L is large. This means that the storage capacity approaches $\frac{4}{\pi}\ln L$ relatively, and this result supports the derived theory.

As the length of delay L increases, storage capacity α_C increases, even though the total number of synapses remains constant; the tendency of increase is different from that of random pruning. Storage capacity is in proportion to the logarithm of the length of delay L, and the proportion constant is $\frac{4}{\pi}$. In other words, for systematic pruning, storage capacity diverges with the increase in the length of delay L. It is amazing that the storage capacity diverges regardless of whether the total number of synapses is constant.

4   Conclusions

We analyzed a discrete synchronous-type model that adopts correlation learning by using the statistical neurodynamics and discussed sequential associative memory by recurrent neural networks with synaptic delay and pruning. First, we explained the Yanai-Kim theory [24], which involves macrodynamical equations for the dynamics of a network with serial delay elements. Next, considering the translational symmetry of the explained equations, we explained the macroscopic steady state equations of the model by using the discrete Fourier transformation [25, 26]. The storage capacity was analyzed quantitatively. As a result, we showed that the storage capacity

is in proportion to the length of delay L when the L limit is large and the proportion constant is 0.195. Furthermore, two types of synaptic prunings were analyzed: random pruning and systematic pruning. As a result, it became clear that under both pruning conditions, the storage capacity grows with an increase in delay and a decrease in the connecting rate when the total number of synapses is constant. Moreover, an interesting fact became clear: the storage capacity approaches $2/\pi$ asymptotically by random pruning. In contrast, the storage capacity diverges in proportion to the logarithm of the length of delay by systematic pruning, and the proportion constant is $4/\pi$. These results theoretically support the significance of pruning following overgrowth of synapses in the brain [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] and strongly suggest that the brain prefers to store dynamic attractors such as sequences or limit cycles rather than equilibrium states.

Appendix A: Derivations of the Macrodynamical Equations of the Delayed Network

Using (10) and (11), we obtain

$$z_i^t = z_A + z_B, \tag{82}$$

$$z_A = \sum_{l=0}^{L-1} c_l \sum_{\mu \neq t} \xi_i^{\mu+1} \frac{1}{N} \sum_j \xi_j^{\mu-l} x_j^{t-l,(\mu-l)}, \tag{83}$$

$$z_B = \sum_{l=0}^{L-1} c_l \sum_{\mu \neq t} \xi_i^{\mu+1} U_{t-l} \sum_{l'=0}^{L-1} c_{l'} m_{\mu-l-l'-1}^{t-l-l'-1}, \tag{84}$$

where $x_j^{t-l,(\mu-l)}$ is the variable obtained by removing the influence of $\xi_j^{\mu-l}$ from x_j^{t-l}. Using (82),(83) and (84), we obtain

$$E\left[z_i^t\right] = 0, \tag{85}$$

$$\therefore \quad \sigma_t^2 = E\left[\left(z_i^t\right)^2\right] \tag{86}$$

$$= E\left[z_A^2 + z_B^2 + 2z_A z_B\right]. \tag{87}$$

Transforming z_A^2, z_B^2 and $z_A z_B$ with consideration given to their correlation, we obtain

$$E\left[z_A^2\right] = \alpha \sum_{l=0}^{L-1} c_l^2, \tag{88}$$

$$E\left[z_B^2\right] = \sum_{\mu \neq t} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_l c_{l'} c_k c_{k'}$$
$$\times U_{t-l} U_{t-l'} m_{\mu-l-k-1}^{t-l-k-1} m_{\mu-l'-k'-1}^{t-l'-k'-1}, \tag{89}$$

$$E\left[2z_A z_B\right] = \alpha \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'}$$
$$\times \left(c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l}\right), \tag{90}$$

where

$$v_{t-l,t-l'} = \sum_{\mu \neq t} m_{\mu-l}^{t-l} m_{\mu-l'}^{t-l'}. \tag{91}$$

Using (87)-(90), we obtain

$$\sigma_t^2 = \alpha \sum_{l=0}^{L-1} c_l^2$$

$$
\begin{aligned}
+ \ & \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}\sum_{k=0}^{L-1}\sum_{k'=0}^{L-1} c_l c_{l'} c_k c_{k'} \\
& \times U_{t-l} U_{t-l'} v_{t-l-k-1,t-l'-k'-1} \\
+ \ & \alpha \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} c_l c_{l'} \left(c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l} \right).
\end{aligned}
\tag{92}
$$

Using (10) and (86), we obtain

$$
\sigma_t^2 = \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l,t-l'}.
\tag{93}
$$

Comparing (92) and (93) as identical equations regarding $c_l c_{l'}$, we obtain

$$
\begin{aligned}
v_{t-l,t-l'} \ = \ & \alpha \delta_{l,l'} \\
+ \ & U_{t-l} U_{t-l'} \\
& \times \sum_{k=0}^{L-1}\sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1,t-l'-k'-1} \\
+ \ & \alpha \left(c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l} \right),
\end{aligned}
\tag{94}
$$

where δ is Kronecker's delta. Using (13), we obtain

$$
\begin{aligned}
U_t \ = \ & \frac{1}{N}\sum_{i=1}^{N} F' \left(\sum_{l=0}^{L-1}\sum_{j=1}^{N} \frac{c_l}{N} \right. \\
& \left. \times \sum_{\nu\neq\mu-l-1} \xi_i^{\nu+1+l}\xi_j^{\nu} x_j^{t-l-1} \right) \\
= \ & E\left[F'\left(u^{t,(\mu)} \right) \right] \\
= \ & E\left[F'\left(u^t \right) \right] \\
= \ & \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \ll F'\left(u^t \right) \gg \\
= \ & \frac{1}{\sigma}\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} z \ll F\left(u^t \right) \gg \\
= \ & \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_{t-1}}\exp\left(-\frac{\left(s^{t-1} \right)^2}{2\sigma_{t-1}^2} \right),
\end{aligned}
\tag{95}
$$

where $u^{t,(\mu)}$ is the variable obtained by removing the influence of $\boldsymbol{\xi}^\mu$ from u^t. Here, $\ll \cdot \gg$ stands for the average over pattern $\boldsymbol{\xi}$.

As a result, we can obtain the macrodynamical equations for overlap m, that is, (14)-(18).

Appendix B: Derivations of the Macroscopic Steady State Equations by Discrete Fourier Transformation

Using the discrete Fourier transformation, we re-derive the general term of $v(n)$, which is expressed by the recurrence formula in (20) [25, 26]. Applying the discrete Fourier transformation to (20) and (21), we obtain

$$
V(r) \ = \ \alpha + U^2 \sum_{i=1-L}^{L-1} (L-|i|)V(r)e^{-j2\pi\frac{ri}{2T+1}} + \alpha U D(r),
\tag{96}
$$

$$D(r) = \sum_{n=-T}^{T} d(n) e^{-j2\pi \frac{rn}{2T+1}}$$

$$= \sum_{n=1}^{L} \left(e^{-j2\pi \frac{rn}{2T+1}} + e^{j2\pi \frac{rn}{2T+1}} \right), \tag{97}$$

where $V(r)$ and $D(r)$ are the discrete Fourier transformations of $v(n)$ and $d(n)$, respectively.

Solving (96) and (97) in terms of $V(r)$, we obtain

$$V(r) = \frac{\alpha \left(1 + U \sum_{n=1}^{L} \left(e^{-j2\pi \frac{rn}{2T+1}} + e^{j2\pi \frac{rn}{2T+1}} \right) \right)}{1 - U^2 \sum_{i=1-L}^{L-1} (L - |i|) e^{-j2\pi \frac{ri}{2T+1}}}. \tag{98}$$

Summations in (98) are calculated as follows.

$$\sum_{n=1}^{L} \left(e^{-j2\pi \frac{rn}{2T+1}} + e^{j2\pi \frac{rn}{2T+1}} \right)$$

$$= \begin{cases} \dfrac{\sin\left((2L+1)\frac{\pi r}{2T+1}\right)}{\sin\left(\frac{\pi r}{2T+1}\right)} - 1, & r \neq 0, \\ 2L, & r = 0, \end{cases} \tag{99}$$

$$\sum_{i=1-L}^{L-1} e^{-j2\pi \frac{ri}{2T+1}} = \begin{cases} \dfrac{\sin\left((2L-1)\frac{\pi r}{2T+1}\right)}{\sin\left(\frac{\pi r}{2T+1}\right)}, & r \neq 0, \\ 2L - 1, & r = 0. \end{cases} \tag{100}$$

When $r \neq 0$,

$$\sum_{i=1-L}^{L-1} |i| e^{-j2\pi \frac{ri}{2T+1}}$$

$$= \sum_{i=1}^{L-1} i e^{-j2\pi \frac{ri}{2T+1}} + \sum_{i=1-L}^{-1} (-i) e^{-j2\pi \frac{ri}{2T+1}}$$

$$= j\frac{2T+1}{2\pi} \frac{\partial}{\partial r} \sum_{i=1}^{L-1} \left(e^{-j2\pi \frac{ri}{2T+1}} - e^{j2\pi \frac{ri}{2T+1}} \right)$$

$$= j\frac{2T+1}{2\pi} \frac{\partial}{\partial r} \frac{\sin\left((L-1)\frac{\pi r}{2T+1}\right)}{\sin\left(\frac{\pi r}{2T+1}\right)} \left(e^{-j\pi \frac{rL}{2T+1}} - e^{j\pi \frac{rL}{2T+1}} \right)$$

$$= \frac{L\cos\left((L-1)\frac{2\pi r}{2T+1}\right) - (L-1)\cos\left(\frac{2L\pi r}{2T+1}\right) - 1}{2\sin^2\left(\frac{\pi r}{2T+1}\right)}. \tag{101}$$

When $r = 0$,

$$\sum_{i=1-L}^{L-1} |i| e^{-j2\pi \frac{ri}{2T+1}} = L(L-1). \tag{102}$$

Substituting (99)-(102) into (98), we obtain

$$V(r)$$

$$= \begin{cases} \dfrac{2\alpha \sin\left(\frac{\pi r}{2T+1}\right)\left((1-U)\sin\left(\frac{\pi r}{2T+1}\right) + U\sin\left((2L+1)\frac{\pi r}{2T+1}\right)\right)}{2\sin^2\left(\frac{\pi r}{2T+1}\right) - U^2\left(1 - \cos\left(\frac{2L\pi r}{2T+1}\right)\right)}, & r \neq 0, \\ \dfrac{\alpha(1+2UL)}{1 - U^2 L^2}, & r = 0. \end{cases} \tag{103}$$

17

Since the inverse discrete Fourier transformation of (103) equals $v(n)$, we obtain

$$v(n) = \lim_{T \to \infty} \frac{1}{2T+1} \sum_{r=-T}^{T} V(r) e^{j2\pi \frac{rn}{2T+1}}. \tag{104}$$

Substituting (104) into (19), we obtain

$$\sigma^2 = \lim_{T \to \infty} \frac{1}{2T+1} \sum_{r=-T}^{T} V(r) \sum_{n=1-L}^{L-1} (L - |n|) e^{j2\pi \frac{rn}{2T+1}}. \tag{105}$$

Using (100)-(105) and rewriting $\frac{r}{2T+1} \to x$, $\frac{1}{2T+1} \to dx$, we can express σ^2 as a form using a simple integral like (22). As a result, we can obtain the steady state equations in terms of the macroscopic variables of the network as (22)-(25).

Acknowledgment

References

[1] M. Okada, "Notions of associative memory and sparse coding," *Neural Networks*, vol. 9, pp. 1429–1458, 1996.

[2] H. Sompolinsky, "Neural networks with nonlinear synapses and a static noise," *Physical Review A*, vol. 34, pp. 2571–2574, 1986.

[3] M. Okada, T. Fukai and M. Shiino, "Random and systematic dilutions of synaptic connections in a neural network with a nonmonotonic response function," *Physical Review E*, vol. 57, pp. 2095–2103, 1998.

[4] K. Mimura, T. Kimoto and M. Okada, "Synapse efficiency diverges due to synaptic pruning following over-growth," (cond-mat/0207545)

[5] G. Chechik, I. Meilijson and E. Ruppin, "Synaptic pruning in development: A computational account," *Neural Computation*, vol. 10, pp. 1759–1777, 1998.

[6] P. R. Huttenlocker, "Synaptic density in human frontal cortex. Development changes and effects of age," *Brain Res.*, vol. 163, pp. 195–205, 1979.

[7] P. R. Huttenlocker, C. De Courten, L. J. Garey and H. Van der Loos, "Synaptogenesis in human visual cortex —evidence for synapse elimination during normal development," *Neuroscience Letters*, vol. 33, pp. 247–252, 1982.

[8] J. P. Bourgeois and P. Rakic, "Changing of synaptic density in the primary visual cortex of the rhesus monkey from fetal to adult age," *J. Neurosci.*, vol. 13, pp. 2801–2820, 1993.

[9] J. Takacs and J. Hamori, "Development dynamics of Purkinje cells and dendritic spines in rat cerebellar cortex," *J. of Neuroscience Research*, vol. 38, pp. 515–530, 1994.

[10] G. M. Innocenti, "Exuberant development of connections and its possible permissive role in cortical evolution," *Trends Neurosci.*, vol. 18, pp. 397–402, 1995.

[11] M. F. Eckenhoff and P. Rakic, "A quantative analysis of synaptogenesis in the molecular layer of the dentate gyrus in the resus monkey," *Developmental Brain Research*, vol. 64, pp. 129–135, 1991.

[12] P. Rakic, J. P. Bourgeois and P. S. Goldman-Rakic, "P. S. Synaptic development of the cerebral cortex: Implementation for learning, memory and mental illness," *Progress in Brain Research*, vol. 102, pp. 227–243, 1994.

[13] M. P. Stryker, "Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex," *J. of Neuroscience*, vol. 6, pp. 2117–2133, 1986.

[14] A. W. Roe, S. L. Pallas, J. O. Hahm and M. Sur, "A map of visual space induced in primary auditory cortex," *Science*, vol. 250, pp. 818–820, 1990.

[15] J. R. Wolff, R. Laskawi, W. B. Spatz and M. Missler, "Structural dynamics of synapses and synaptic components," *Behavioural Brain Research*, vol. 66, pp. 13–20, 1995.

[16] K. Fukushima, "A model of associative memory in the brain," *Kybernetik*, vol. 12, pp. 58–63, 1973.

[17] S. Miyoshi and M. Okada, "A theory of syn-fire chain model," *Transactions of the Institute of Electronics, Information and Communication Engineers*, vol. J83-A, pp. 1330-1332, 2000, in Japanese.

[18] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. National Academy of Sciences*, vol. 79, pp. 2554–2558, 1982.

[19] D. J. Amit, H. Gutfreund and H. Sompolinsky, "Spin-glass model of neural networks," *Physical Review A*, vol. 32, pp. 1007–1018, 1985.

[20] D. J. Amit, H. Gutfreund and H. Sompolinsky, "Storing infinite numbers of patterns in a spin-glass model of neural networks," *Physical Review Letters*, vol. 55, pp. 1530–1533, 1985.

[21] M. Shiino and T. Fukai, "Self-consistent signal-to-noise analysis and its application to analogue neural networks with asymmetric connections," *Journal of Physics A: Mathematical and General*, vol. 25, L375–L381, 1992.

[22] S. Amari and K. Maginu, "Statistical neurodynamics of associative memory," *Neural Networks*, vol. 1, pp. 63–73, 1988.

[23] M. Okada, "A hierarchy of macrodynamical equations for associative memory," *Neural Networks*, vol. 8, no. 6, pp. 833–838, 1995.

[24] H. F. Yanai and E. S. Kim, "Dynamics of neural nets with delay-synapses," in *Technical report of the Institute of Electronics, Information and Communication Engineers*, NC92-116, pp. 167–174, 1993, in Japanese.

[25] S. Miyoshi, H. F. Yanai and M. Okada, "Associative memory by recurrent neural networks with delay elements," *Proc. 9th Int. Conf. on Neural Information Processing*, vol. 1, pp. 70–74, 2002.

[26] S. Miyoshi, H. F. Yanai and M. Okada, "Associative memory by recurrent neural networks with delay elements," *Neural Networks* (submitted).

[27] S. Miyoshi and M. Okada, "Associative memory by neural networks with delays and pruning," *Transactions of the Institute of Electronics, Information and Communication Engineers*, vol. J85-A, pp. 124-133, 2002, in Japanese.

[28] S. Miyoshi and K. Nakayama, "A recurrent neural network with serial delay elements for memorizing limit cycles," in *Proc. of ICNN'95*, pp. 1955–1960, 1995.

[29] M. Kawamura and M. Okada, "Transient dynamics for sequence processing neural network," *Journal of Physics A: Mathematical and General*, vol. 35, pp. 253–266, 2002.

[30] A. Düring, A. C. C. Coolen and D. Sherrington, "Phase diagram and storage capacity of sequence processing neural networks," *Journal of Physics A: Mathematical and General*, vol. 31, pp. 8607–8621, 1998.

[31] S. Amari, "Statistical neurodynamics of various versions of correlation associative memory," in *Proc. ICNN'88* pp. 633–640, 1998.

Figure Captions
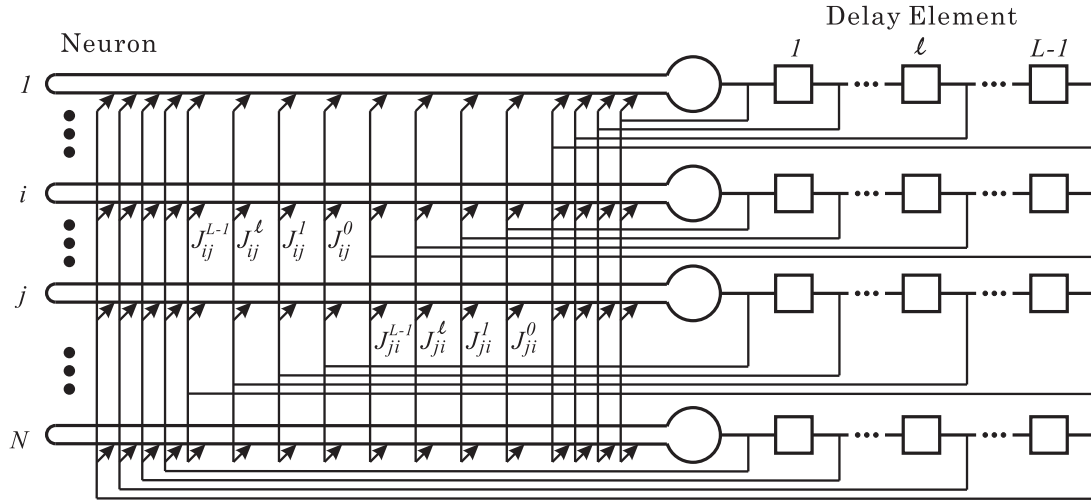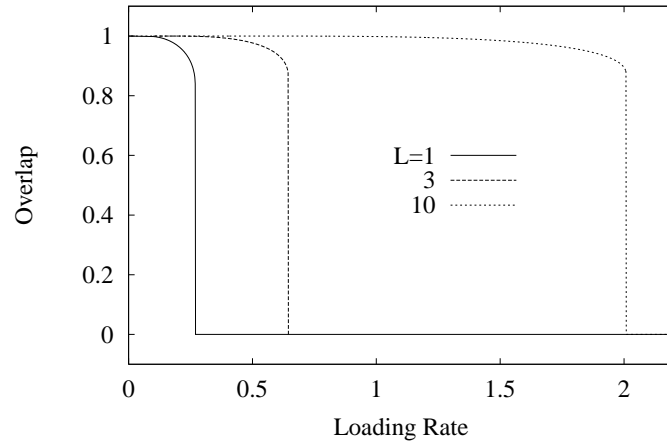
1. Structure of delayed network.

2. Relationship between loading rate α and overlap m (theory).
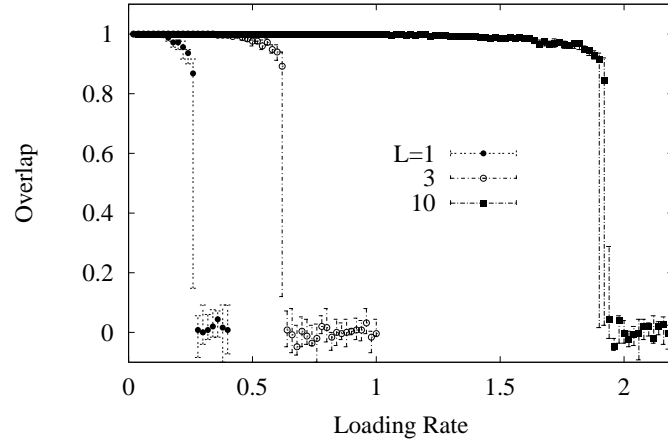
3. Relationship between loading rate α and overlap m (computer simulation).

4. Relationship between loading rate α and overlap m. These lines are obtained by solving steady state equations numerically.

5. Relationship between length of delay L and storage capacity α_C. This line is obtained by solving the steady state equations numerically. Storage capacity is $0.195L$, with a large L limit.

6. Representation of delayed synapses by pruning: (a) length of delay is three, and (b) length of delay is five.

7. Relationship between loading rate α and overlap m when synapses are randomly pruned (theory(t) and computer simulation(s)).

8. Relationship between loading rate α and overlap m when synapses are randomly pruned (theory).

9. Relationship between $\log L$ and $\log r$ when synapses are randomly pruned.

10. Nonlinear function for systematic pruning.

11. Relationship between loading rate α and overlap m when synapses are systematically pruned (theory(t) and computer simulation(s)).

12. Relationship between loading rate α and overlap m when synapses are systematically pruned (theory).

13. Relationship between length of delay L and storage capacity α_C when synapses are systematically pruned.

14. Relationship between $\log L$ and $\log \sigma^2$ when synapses are systematically pruned.

Figures

Figure 1: Structure of delayed network.

Figure 2: Relationship between loading rate α and overlap m (theory).

Figure 3: Relationship between loading rate α and overlap m (computer simulation).
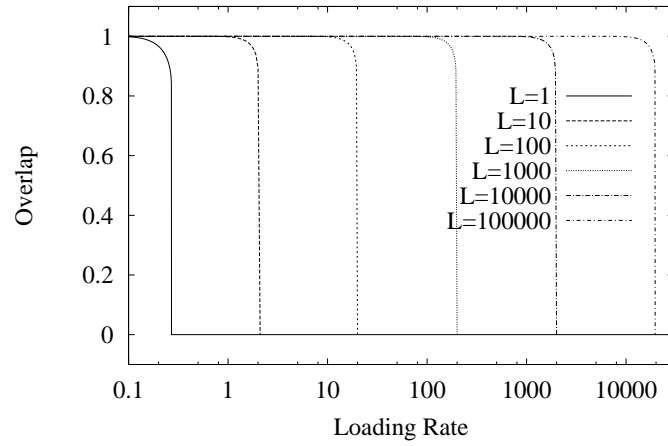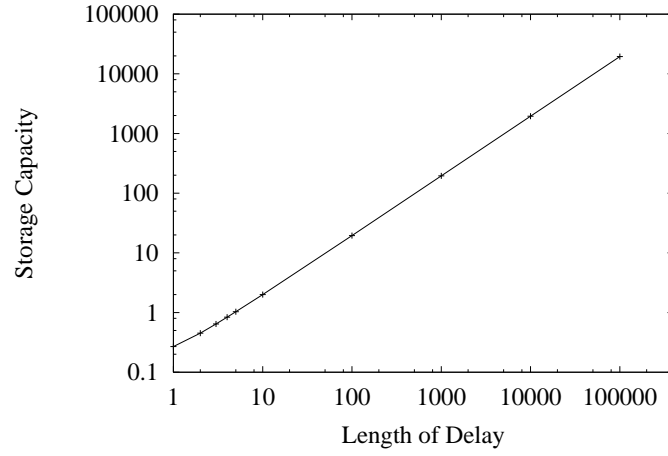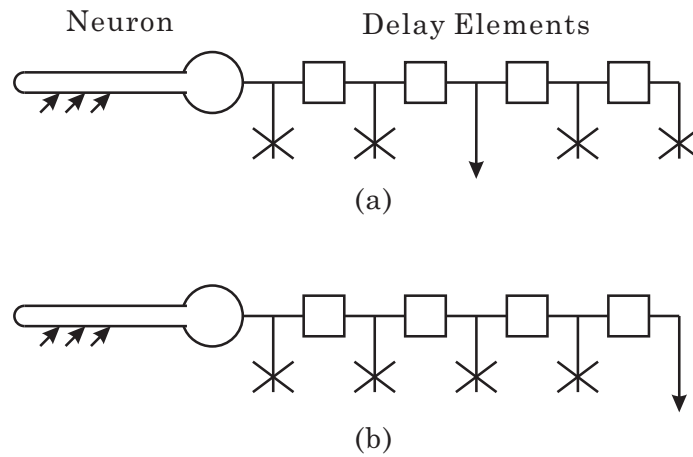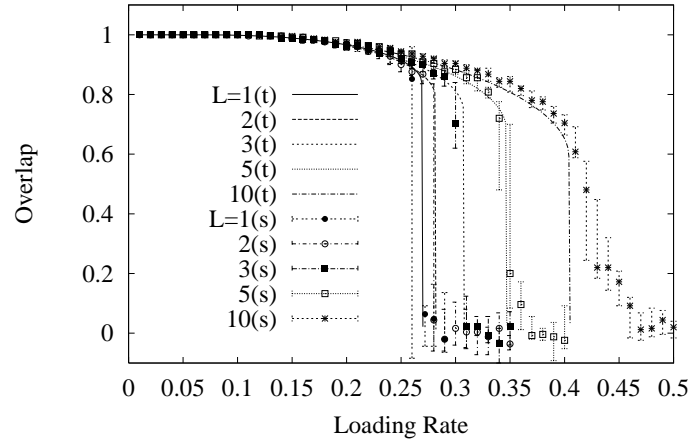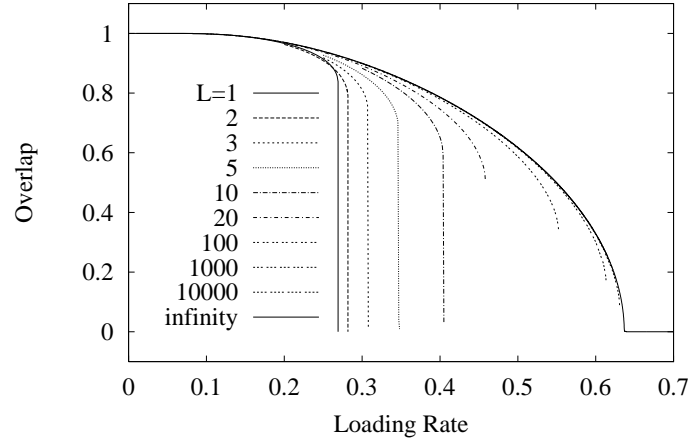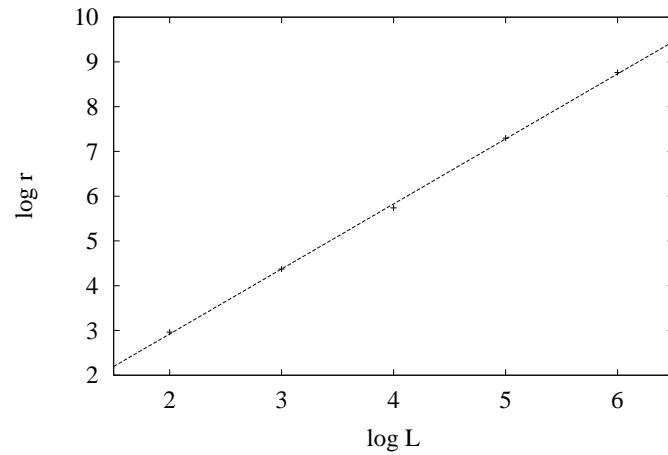

Figure 4: Relationship between loading rate α and overlap m. These lines are obtained by solving steady state equations numerically.

Figure 5: Relationship between length of delay L and storage capacity α_C. This line is obtained by solving the steady state equations numerically. Storage capacity is $0.195L$, with a large L limit.

Figure 6: Representation of delayed synapses by pruning: (a) length of delay is three, and (b) length of delay is five.

Figure 7: Relationship between loading rate α and overlap m when synapses are randomly pruned (theory(t) and computer simulation(s)).

Figure 8: Relationship between loading rate α and overlap m when synapses are randomly pruned (theory).

Figure 9: Relationship between $\log L$ and $\log r$ when synapses are randomly pruned.
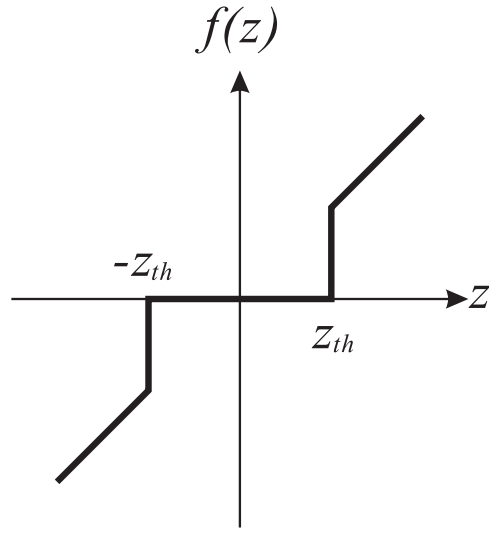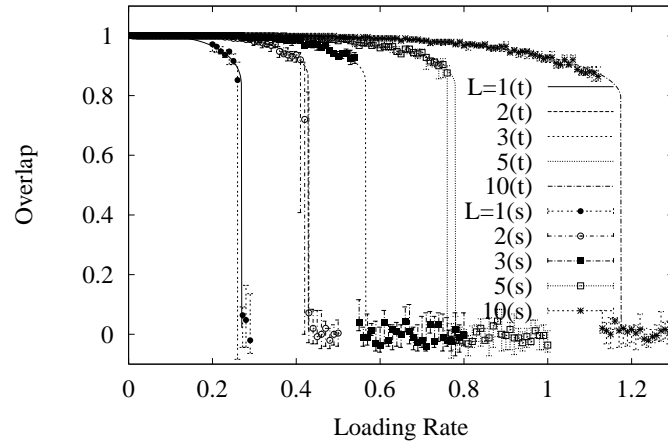
Figure 10: Nonlinear function for systematic pruning.

Figure 11: Relationship between loading rate α and overlap m when synapses are systematically pruned (theory(t) and computer simulation(s)).
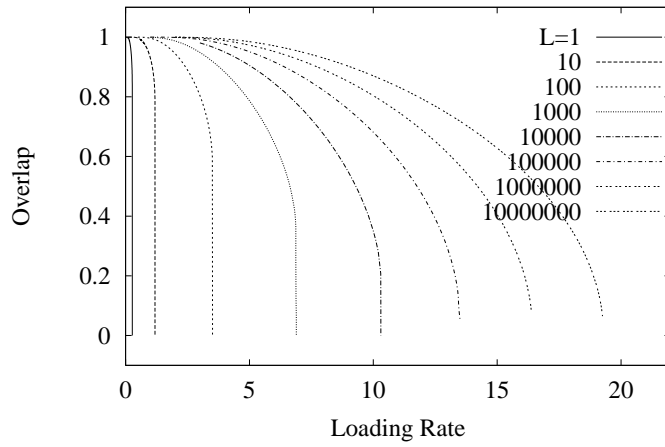
Figure 12: Relationship between loading rate α and overlap m when synapses are systematically pruned (theory).
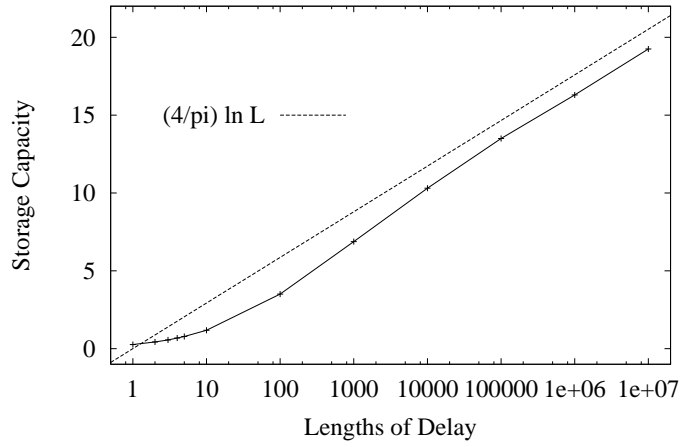

Figure 13: Relationship between length of delay L and storage capacity α_C when synapses are systematically pruned.
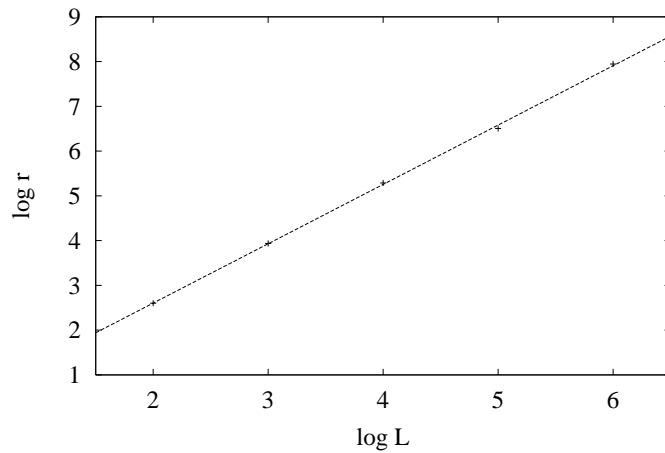

Figure 14: Relationship between $\log L$ and $\log \sigma^2$ when synapses are systematically pruned.