

シナプスに遅延と切断を含む連想記憶モデルの理論

三好 誠司[†] 岡田 真人^{††,†††}

[†] 神戸高専電子工学科 〒654-0154 神戸市西区学園東町 8-3

^{††} 理化学研究所 脳科学総合研究センター脳数理研究チーム 〒351-0198 埼玉県和光市広沢 2-1

^{†††} 科学技術振興事業団 ERATO 川入学習動態脳プロジェクト 〒619-0288 京都府相楽郡精華町光台 2-2-2

E-mail: [†]miyoshi@kobe-kosen.ac.jp, ^{††}okada@brain.riken.go.jp

あらまし 相関学習型の連想記憶モデルにおいて、シナプスの結合率が低下すると記憶容量自体は小さくなるが、シナプス1本あたりの記憶容量であるシナプス効率が増大することが知られている。我々は、遅延シナプスを導入することにより、シナプス総本数を一定に保ったまま結合率を下げることを提案し理論的な解析を行う。統計神経力学と離散フーリエ変換を用いてモデルの定常状態を理論的に扱い、記憶容量を定量的に議論する。その結果、結合率を下げることにより記憶容量が増大・発散することが明らかになる。この結果は、実際の神経系において観測されているシナプスの過剰生成と刈り込みの理論的裏付けとなるものであり、また、脳内においては平衡状態よりも動的なアトラクタを記憶する方が望ましいことを示唆するものである。

キーワード 連想記憶, ニューラルネットワーク, 遅延, シナプス切断, 統計神経力学

A Theory of Associative Memory Model with Synaptic Delay and Pruning

Seiji MIYOSHI[†] and Masato OKADA^{††,†††}

[†] Department of Electronic Engineering, Kobe City College of Technology, Kobe 651-2194, Japan

^{††} Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, 351-0198, Japan

^{†††} Kawato Dynamic Brain Project, ERATO, Japan Science and Technology Corporation, 619-0288, Japan

E-mail: [†]miyoshi@kobe-kosen.ac.jp, ^{††}okada@brain.riken.go.jp

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 whole network then decreases. To overcome this difficulty, we propose that the connecting rate is decreased while keeping the total number of synapses constant by introducing delayed synapses. First, we re-derive the Yanai-Kim theory by employing the statistical neurodynamics. Next, we derive macroscopic steady state equations of the model by using the discrete Fourier transformation. Furthermore, two types of synaptic prunings are treated analytically. 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. These results theoretically support the significance of pruning following over-growth 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.

Key words 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 mem-

ory [1], models with randomly pruned synapses have been discussed in detail [2]–[5]. 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 whole network decreases.

On the other hand, it is universally observed that synapses

are pruned following over-growth in real neural systems [5]–[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 supposed that increasing synapse efficiency in that way gives functional significance to synaptic pruning following over-growth. They used a correlation-type auto-associative memory model to verify this hypothesis from a theoretical standpoint. 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 increases by obtaining storage capacity. In this paper, synaptic pruning as described above is called systematic pruning. Though 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. Therefore, Mimura et al. [4] analyzed this system by using the SCSNA [21], which is the method of statistical mechanics. They showed that systematic pruning increases synapse efficiency by the order of $-\ln(1-R)$ over random pruning at the limit when R approaches unity, where R ($0 \leq R \leq 1$) is the rate of synaptic pruning. The important point in this case is that the storage capacity of the whole network decreases, though synapse efficiency increases by random pruning or systematic pruning.

To overcome this difficulty, we propose to decrease 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], [27] 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 re-derive the Yanai-Kim theory [24], [25] using the statistical neurodynamics [22], involving 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 ex-

remely large [26]. Thus, accounting for the parallel symmetry of time steps, which holds in the steady state of the Yanai-Kim theory, we derive the macroscopic steady state equations by employing the discrete Fourier transformation, where the computational complexity does not formally depend on L . Using the derived steady state equations, storage capacities can be quantitatively discussed even for a large L limit.

Next, synaptic pruning in the delayed network is treated 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; that is, 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 by systematic pruning.

2. Delayed Network

2.1 Model

The structure of a 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

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

$$F(\cdot) = \text{sgn}(\cdot) \quad (2)$$

$$u_i^t = \sum_{l=0}^{L-1} \sum_{j=1}^N J_{ij}^l x_j^{t-l}, \quad (3)$$

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. Here, sgn is the sign function defined as

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

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

Let us consider the storing sequence of αN memory patterns, $\xi^1 \rightarrow \xi^2 \rightarrow \dots \rightarrow \xi^\mu \rightarrow \dots \rightarrow \xi^{\alpha N}$. Here, α and αN are the loading rate and the length of the sequence, respectively. Each component of ξ^μ is assumed to be an independent

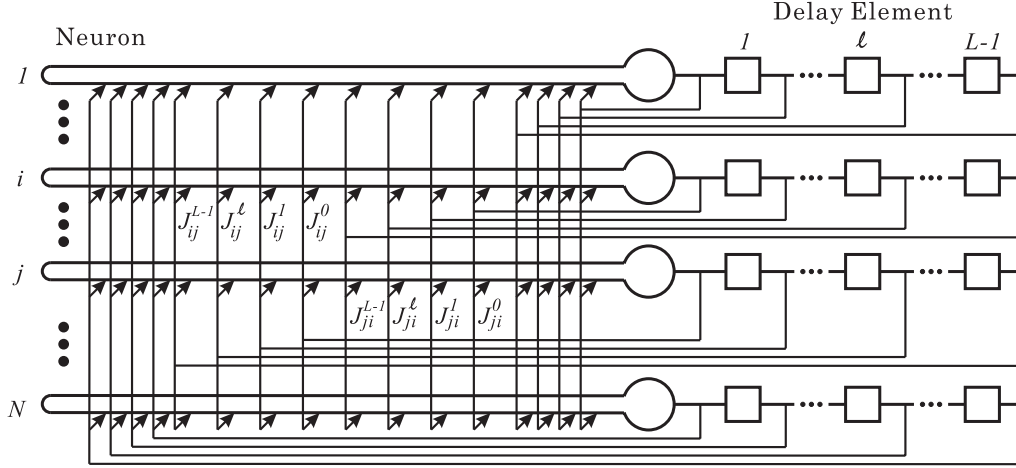


Figure 1 Structure of delayed network.

random variable that takes a value of either $+1$ or -1 according to the following probabilities,

$$\text{Prob}[\xi_i^\mu = \pm 1] = \frac{1}{2}. \quad (5)$$

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

$$J_{ij}^l = \frac{c_l}{N} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^\mu \quad (6)$$

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

Correlation learning is an algorithm based on Hebb rule, and it is inferior to the error correcting learning in terms of storage capacity. However, as seen in eqn (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 some of the patterns stored as a sequence are 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 some 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 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 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 recalling becomes unstable.

We define the overlap, 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

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

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

Amari and Maginu [22] proposed the statistical neurodynamics. This analytical method handles the dynamical behavior of 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 σ_t^2 . They then derived recursive relations for the variance and the overlap.

Using eqns (1)-(7), we can derive the following macrodynamical equations.

$$\sigma_t^2 = \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l, t-l'}, \quad (8)$$

$$v_{t-l, t-l'} = \alpha \delta_{l, l'} + U_{t-l} U_{t-l'} \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1, t-l'-k'-1} + \alpha (c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l}), \quad (9)$$

$$U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{t-1}} \exp\left(-\frac{(s^{t-1})^2}{2\sigma_{t-1}^2}\right), \quad (10)$$

$$s^t = \sum_{l=0}^{L-1} c_l m_{t-l}, \quad (11)$$

$$m_{t+1} = \text{erf}\left(\frac{s^t}{\sqrt{2}\sigma_t}\right), \quad (12)$$

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. In this paper, the initial condition is that the states of all

neurons and all delay elements are set to be the stored in pattern sequences. In this case, $m_l = 1$ ($l = 0, \dots, L - 1$) and $v_{l,t} = \alpha$ ($l = 0, \dots, L - 1$).

We note that the derived macrodynamical equations and the Yanai-Kim theory [24], [25] coincide.

2.3 Macroscopic Steady State Analysis by Discrete Fourier Transformation and Discussion

The Yanai-Kim theory re-derived 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 as shown in eqns (8) and (9), 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, i.e., the asymptotic behavior of the storage capacity in a large L limit. Thus, in this section, we consider the Yanai-Kim theory in a steady state and derive the macroscopic steady state equations of the delayed network. Furthermore, the storage capacity is analyzed for a large L by solving the derived equations numerically [25].

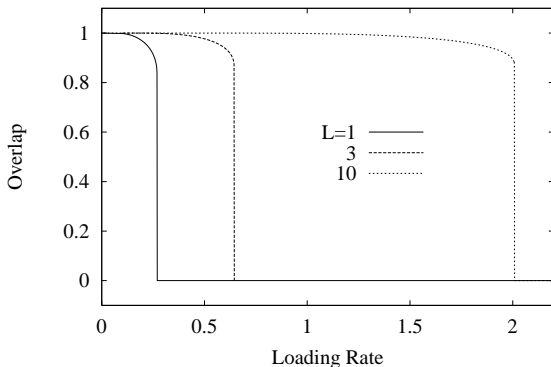


Figure 2 Relationship between loading rate α and overlap m . (theory)

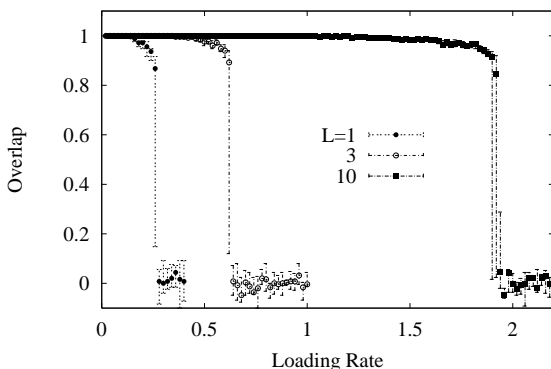


Figure 3 Relationship between loading rate α and overlap m . (computer simulation)

For simplicity, let us assume that $c_l = 1$, $l = 0, \dots, L - 1$. In a steady state, $v_{t-l,t-l'}$ can be expressed as $v_{l-l'}$ because of the parallel symmetry in terms of time step. Using the discrete Fourier transformation and the inverse discrete Fourier transformation, we obtain

$$\sigma^2 = \sum_{n=1-L}^{L-1} (L - |n|) \lim_{T \rightarrow \infty} \frac{1}{2T+1} \sum_{r=-T}^T V(r) e^{j2\pi \frac{rn}{2T+1}}. \quad (13)$$

From eqns (10)-(12) and (13), we can obtain the steady state equations in terms of the network's macroscopic variables as eqns (14)-(17).

$$\sigma^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\alpha [(1-U) \sin(\pi x) + U \sin \{(2L+1) \pi x\}] [1 - \cos(2L\pi x)]}{\sin(\pi x) [2 \sin^2(\pi x) - U^2 \{1 - \cos(2L\pi x)\}]} dx \quad (14)$$

$$U = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp\left(-\frac{s^2}{2\sigma^2}\right) \quad (15)$$

$$s = mL \quad (16)$$

$$m = \operatorname{erf}\left(\frac{s}{\sqrt{2}\sigma}\right) \quad (17)$$

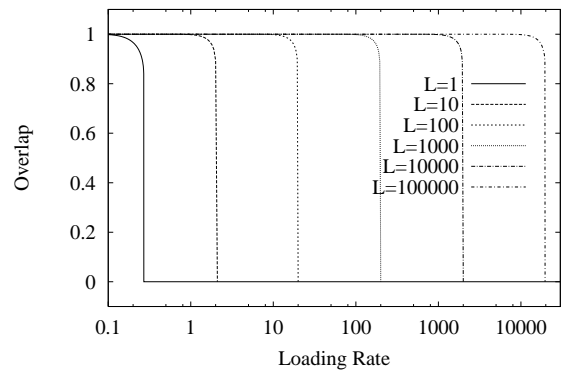


Figure 4 Relationship between loading rate α and overlap m . These lines are obtained by solving steady state equations numerically. Storage capacity is $0.195L$, with a large L limit.

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 calculation 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 $N = 500$, 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 delay L . Data points \bullet , \circ , \blacksquare indicate the medians of the sixth largest values 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.

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. From this figure, 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. 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 over-growth [5]–[15] is observed. Since this pruning following over-growth 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 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]. A couple 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 converges to $\frac{2}{\pi}$ by random pruning, and diverges as $\frac{2}{\pi}(-2 \ln c)$ at the limit where the connecting rate c is extremely small. Here, the relation between connecting rate c and pruning rate R is given by $c = 1 - R$. The important point here is that the storage capacity of the whole network decreases, since the number of synapses decreases.

In the following discussion, a delayed network with synaptic pruning is analyzed on the basis of macrodynamical equations and macroscopic steady state equations 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 mechanisms, therefore, it is important to investigate its effect on the dynamical behavior of pattern recalling and storage capacity.

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

$$J_{ij}^l = \frac{c_l c_{ij}^l}{Nc} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu}, \quad (18)$$

$$\text{Prob}[c_{ij}^l = 1] = 1 - \text{Prob}[c_{ij}^l = 0] = c, \quad (19)$$

where c is the connecting rate.

Modifying eqn (18), we obtain

$$J_{ij}^l = \frac{c_l}{N} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu} + \frac{c_l (c_{ij}^l - c)}{Nc} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu}. \quad (20)$$

We can obtain the macrodynamical equations for random pruning as follows,

$$\tilde{\sigma}_t^2 = \sigma_t^2 + \frac{\alpha(1-c)}{c} \sum_{l=0}^{L-1} c_l^2, \quad (21)$$

$$\sigma_t^2 = \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l, t-l'}, \quad (22)$$

$$\begin{aligned} v_{t-l, t-l'} &= \alpha \delta_{l, l'} \\ &+ U_{t-l} U_{t-l'} \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1, t-l'-k'-1} \\ &+ \alpha (c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l}), \end{aligned} \quad (23)$$

$$U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_{t-1}} \exp\left(-\frac{(s^{t-1})^2}{2\tilde{\sigma}_{t-1}^2}\right), \quad (24)$$

$$s^t = \sum_{l=0}^{L-1} c_l m_{t-l}, \quad (25)$$

$$m_{t+1} = \text{erf}\left(\frac{s^t}{\sqrt{2}\tilde{\sigma}_t}\right), \quad (26)$$

where the initial conditions are the same as in the case of a fully connected network. δ is Kronecker's delta. Equation (21) 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. Finally, as for a fully connected network, the macroscopic steady state equations in the case of random pruning become

$$\tilde{\sigma}^2 = \sigma^2 + \frac{\alpha(1-c)}{c} \sum_{l=0}^{L-1} c_l^2 \quad (27)$$

$$U = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}} \exp\left(-\frac{s^2}{2\tilde{\sigma}^2}\right) \quad (28)$$

$$s = mL \quad (29)$$

$$m = \text{erf}\left(\frac{s}{\sqrt{2}\tilde{\sigma}}\right) \quad (30)$$

where σ^2 is given by eqn (14).

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. That means eqns (27)-(30) are solved numerically and the steady state overlaps m_∞ are investigated using $c = 1/L$, where $c_l = 1$. Figure 5 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 $N = 500$, 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 rates α and length of delays L . Data points \bullet , \circ , \blacksquare , \square , $*$ indicate the medians of the sixth largest values 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.

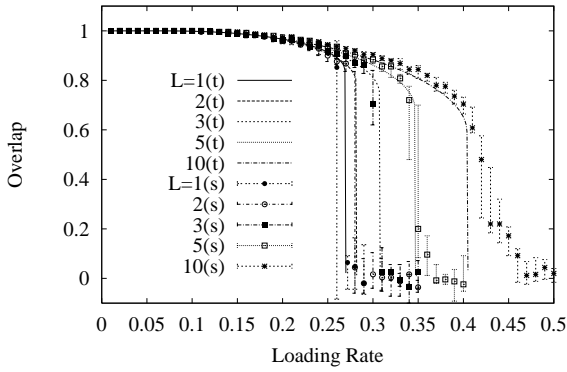


Figure 5 Relationship between loading rate α and overlap m when synapses are randomly pruned. (theory(t) and computer simulation(s))

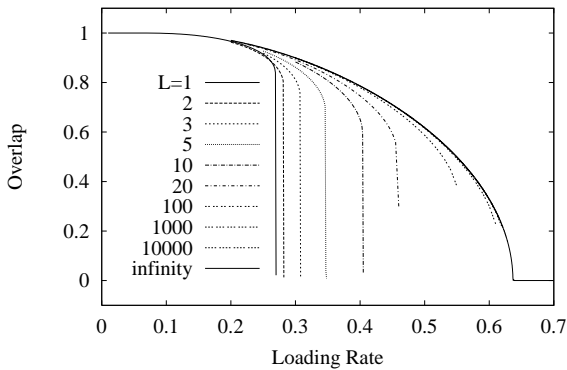


Figure 6 Relationship between loading rate α and overlap m when synapses are randomly pruned. (theory)

Figure 5 displays the following. In the case of $L =$

1 ($c = \frac{1}{L} = 1.0$), that 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 published values [28]–[30]. As the length of delay L increases, storage capacity α_C increases though the total number of synapses is constant. This phenomenon is due to the time lag of synaptic inputs by delays that reduces 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. Figure 6 shows the results.

In Figure 6, the property $L = \infty$ is obtained by setting $U = 0$. The storage capacity in this case is $2/\pi = 0.6366 \dots$. Figure 6 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 over-growth 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.

Systematic pruning like that can be expressed by nonlinear function $f(\cdot)$ shown in Figure 7. Synapses in the range of $-z_{th} < z < +z_{th}$ are pruned by $f(\cdot)$. In this case, the relationship between the connection rate c and z_{th} is given by

$$c = \int_{\{z|f(z) \neq 0\}} Dz = 1 - \operatorname{erf}\left(\frac{z_{th}}{\sqrt{2}}\right), \quad (31)$$

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

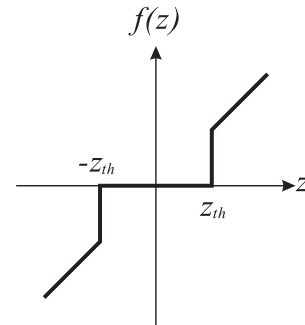


Figure 7 Nonlinear function for systematic pruning.

In the systematic pruning model, synaptic connections are constituted by

$$J_{ij}^l = \frac{c_l \sqrt{\alpha N}}{N} f(T_{ij}^l), \quad (32)$$

$$T_{ij}^l = \frac{1}{\sqrt{\alpha N}} \sum_{\mu} \xi_i^{\mu+1+l} \xi_j^{\mu}. \quad (33)$$

Equation (33) is a stochastic variable that obeys normal distribution $N[0, 1]$.

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

$$J_{ij}^l = \frac{c_l \sqrt{\alpha N}}{N} f(T_{ij}^l) \quad (34)$$

$$= \frac{c_l \sqrt{\alpha N}}{N} \left(J T_{ij}^l + \left(f(T_{ij}^l) - J T_{ij}^l \right) \right), \quad (35)$$

where

$$J = \int D x f'(x) \quad (36)$$

$$= \int D x x f(x). \quad (37)$$

We can obtain the macrodynamical equations for systematic pruning as follows,

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

$$\sigma_t^2 = J^2 \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} c_l c_{l'} v_{t-l, t-l'}, \quad (39)$$

$$\tilde{J}^2 = \int D x (f(x))^2, \quad (40)$$

$$J = \int D x x f(x), \quad (41)$$

$$\begin{aligned} v_{t-l, t-l'} &= \alpha \delta_{l, l'} \\ &+ U_{t-l} U_{t-l'} \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} c_k c_{k'} v_{t-l-k-1, t-l'-k'-1} \\ &+ \alpha (c_{l-l'-1} U_{t-l'} + c_{l'-l-1} U_{t-l}), \end{aligned} \quad (42)$$

$$U_t = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}_{t-1}} \exp\left(-\frac{(s^{t-1})^2}{2\tilde{\sigma}_{t-1}^2}\right), \quad (43)$$

$$s^t = J \sum_{l=0}^{L-1} c_l m_{t-l}, \quad (44)$$

$$m_{t+1} = \operatorname{erf}\left(\frac{s^t}{\sqrt{2}\tilde{\sigma}_t}\right), \quad (45)$$

where the initial conditions are the same as in the case of a fully connected network. δ is Kronecker's delta. Eqn (38) 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. Finally, as in the case of full connections, the macroscopic steady state equations in the case of systematic pruning become

$$\tilde{\sigma}^2 = \sigma^2 + \alpha \left(\tilde{J}^2 - J^2 \right) \sum_{l=0}^{L-1} c_l^2, \quad (46)$$

$$U = \sqrt{\frac{2}{\pi}} \frac{1}{\tilde{\sigma}} \exp\left(-\frac{s^2}{2\tilde{\sigma}^2}\right), \quad (47)$$

$$s = m L, \quad (48)$$

$$m = \operatorname{erf}\left(\frac{s}{\sqrt{2}\tilde{\sigma}}\right), \quad (49)$$

where σ^2 is given by eqn (14).

As for random pruning, the storage capacity α_C is investigated under the condition that $c \times L$ is constant. That means eqns (46)-(49) are solved numerically, and the steady state overlaps m_∞ are investigated using $c = 1/L$, where $c_l = 1$. Figure 8 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 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.

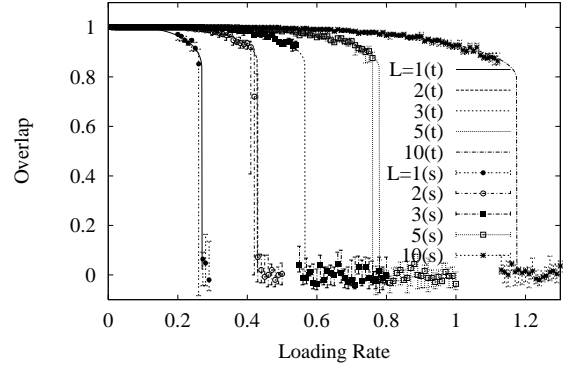


Figure 8 Relationship between loading rate α and overlap m when synapses are systematically pruned. (theory(t) and computer simulation(s))

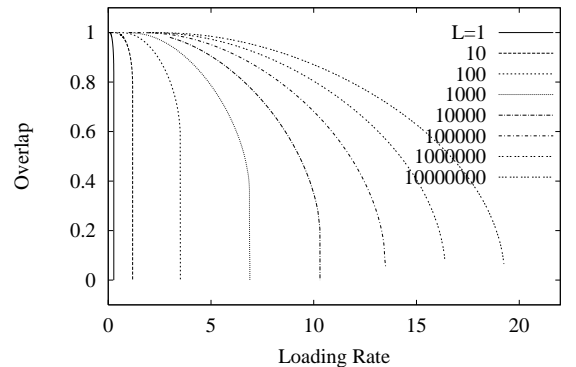


Figure 9 Relationship between loading rate α and overlap m when synapses are systematically pruned. (theory)

Figure 8 shows the following. As the length of delay L increases, storage capacity α_C increases though the total number of synapses is constant. 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. Figure 9 shows the results.

Figure 9 shows the following. As the length of delay L increases, storage capacity α_C increases, though the total number of synapses is constant; the tendency to 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 2.8, i.e., $\alpha_C = 2.8 \log L$. In the 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. However, this fact is also supported by Mimura et al. [4] described in Sec.3.1.

4. Conclusion

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 re-derived the Yanai-Kim theory, which involves macrodynamical equations for the dynamics of the network with serial delay elements. Next, accounting for the parallel symmetry of the re-derived equations, we derived the macroscopic steady state equations of the model by using the discrete Fourier transformation. 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 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. These results theoretically support the significance of pruning following over-growth of synapses in the brain [5]–[15] and strongly suggest that the brain prefers to store dynamic attractors such as sequences or limit cycles, rather than equilibrium states.

In this paper, we treated random patterns, in which each element takes ± 1 with a probability of 0.5. It is interesting future work to treat sparse patterns in which each element takes $+1, 0$.

Acknowledgment

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Japan, Grant-in-Aid for Scientific Research, 13780303 and 14580438.

References

- [1] M. Okada, *Neural Networks*, vol.9, pp.1429–1458, 1996.
- [2] H. Sompolinsky, *Physical Review A*, vol.34, pp.2571–2574, 1986.
- [3] M. Okada, T. Fukai and M. Shiino, *Physical Review E*, vol.57, pp.2095–2103, 1998.
- [4] K. Mimura, T. Kimoto and M. Okada, (cond-mat/0207545)
- [5] G. Chechik, I. Meilijson and E. Ruppin, *Neural Computation*, vol.10, pp.1759–1777, 1998.
- [6] P. R. Huttenlocker, *Brain Res.*, vol.163, pp.195–205, 1979.
- [7] P. R. Huttenlocker, C. De Courten, L. J. Garey, and H. Van der Loos, *Neuroscience Letters*, vol.33, pp.247–252, 1982.
- [8] J. P. Bourgeois and P. Rakic, *J. Neurosci.*, vol.13, pp.2801–2820, 1993.
- [9] J. Takacs and J. Hamori, *J. of Neuroscience Research*, vol.38, pp.515–530, 1994.
- [10] G. M. Innocenti, *Trends Neurosci.*, vol.18, pp.397–402, 1995.
- [11] M. F. Eickenhoff and P. Rakic, *Developmental Brain Research*, vol.64, pp.129–135, 1991.
- [12] P. Rakic, J. P. Bourgeois, and P. S. Goldman-Rakic, *Progress in Brain Research*, vol.102, pp.227–243, 1994.
- [13] M. P. Stryker, *J. of Neuroscience*, vol.6, pp.2117–2133, 1986.
- [14] A. W. Roe, S. L. Pallas, J. O. Hahn, and M. Sur, *Science*, vol.250, pp.818–820, 1990.
- [15] J. R. Wolff, R. Laskawi, W. B. Spatz, and M. Missler, *Behavioural Brain Research*, vol.66, pp.13–20, 1995.
- [16] K. Fukushima, *Kybernetik*, vol.12, pp.58–63, 1973.
- [17] S. Miyoshi and M. Okada, *Transactions of IEICE*, vol.J83-A, pp.1330–1332, 2000, in Japanese.
- [18] J.J. Hopfield, *Proc. National Academy of Sciences*, vol.79, pp.2554–2558, 1982.
- [19] D.J. Amit, H. Gutfreund and H. Sompolinsky, *Physical Review A*, vol.32, pp.1007–1018, 1985.
- [20] D.J. Amit, H. Gutfreund and H. Sompolinsky, *Physical Review Letters*, vol.55, pp.1530–1533, 1985.
- [21] M. Shiino and T. Fukai, *Journal of Physics A: Mathematical and General*, vol.25, L375–L381, 1992.
- [22] S. Amari and K. Maginu, *Neural Networks*, vol.1, pp.63–73, 1988.
- [23] M. Okada, *Neural Networks*, vol.8, no.6, pp.833–838, 1995.
- [24] H.F. Yanai and E.S. Kim, *Technical report of IEICE*, NC92-116, pp.167–174, 1993, in Japanese.
- [25] S. Miyoshi, H.F. Yanai and M. Okada, *Neural Networks*, (submitted).
- [26] S. Miyoshi and M. Okada, *Transactions of IEICE*, vol.J85-A, pp.124–133, 2002, in Japanese.
- [27] S. Miyoshi and K. Nakayama, *Proc. ICNN'95*, pp.1955–1960, 1995.
- [28] M. Kawamura and M. Okada, *Journal of Physics A: Mathematical and General*, vol. 35, pp. 253–266, 2002.
- [29] A. Düring, A.C.C. Coolen and D. Sherrington, *Journal of Physics A: Mathematical and General*, vol. 31, pp. 8607–8621, 1998.
- [30] S. Amari, *Proc. ICNN'88*, pp.633–640, 1988.