To appear in *Neural Computation*, vol.14, no.12, pp.2883–2902, 2002

Self-Regulation Mechanism of Temporally Asymmetric Hebbian Plasticity

Narihisa Matsumoto[†,*] and Masato Okada[*]

[†]Graduate School of Science and Engineering,

Saitama University, Saitama 338-8570, Japan

[*]RIKEN Brain Science Institute, Saitama 351-0198, Japan

E-mail: {xmatumo,okada}@brain.riken.go.jp

abstract

Recent biological experimental findings have shown that the synaptic plasticity depends on the relative timing of the pre- and post-synaptic spikes. This determines whether Long Term Potentiation (LTP) or Long Term Depression (LTD) is induced. This synaptic plasticity has been called "Temporally Asymmetric Hebbian plasticity (TAH)". Many authors have numerically demonstrated that neural networks are capable of storing spatio-temporal patterns. However, the mathematical mechanism of the storage of spatio-temporal patterns is still unknown, and the effect of LTD is particularly unknown. In this paper, we employ a simple

neural network model and show that interference between LTP and LTD disappears in a sparse coding scheme. On the other hand, the covariance learning rule is known to be indispensable for the storage of sparse patterns. We also show that TAH has the same qualitative effect as the covariance rule when spatio-temporal patterns are embedded in the network.

1 Introduction

Recent biological experimental findings have indicated that the synaptic plasticity depends on the relative timing of the pre- and post- synaptic spikes. This determines whether Long Term Potentiation (LTP) or Long Term Depression (LTD) is induced (Bi and Poo, 1998; Markram et al., 1997; Zhang et al., 1998). LTP occurs when a presynaptic firing precedes a postsynaptic firing by no more than about 20 ms. In contrast, LTD occurs when the presynaptic firing follows the postsynaptic firing. The transition between LTP and LTD takes place within a few ms. A learning rule of this type is called "Temporally Asymmetric Hebbian learning (TAH)" (Abbott and Song, 1999; Rubin et al., 2001) or "Spike Timing Dependent synaptic Plasticity (STDP)" (Song et al., 2000). Many authors have numerically demonstrated that neural networks are capable of storing spatio-temporal patterns (Gerstner et al., 1996; Kempter et al., 1999; Munro and Hernandez, 2000; Song et al., 2000; Rao and Sejnowski, 2000; Yoshioka, 2002). Song et al. have used TAH in discussing the variability of spike generation in a network that consists of spiking neurons (Song et al., 2000). They found that the area of LTD has to be slightly larger than that of LTP for stability. Namely, a precise balance between

LTP and LTD is crucial. Yoshioka has used TAH in discussing an associative memory network that consists of spiking neurons (Yoshioka, 2002). He found that stable retrieval requires that the areas of LTP and LTD be equal. Munro and Hernandez have numerically shown that LTD makes the retrieval of spatio-temporal patterns from a network possible even in a noisy environment (Munro and Hernandez, 2000). However, they did not discuss why TAH is effective in terms of the storage and retrieval of spatio-temporal patterns. Since both LTP and LTD are effects of TAH, the interference between LTP and LTD may prevent the retrieval of patterns. To investigate the unknown mathematical mechanism for retrieval, we employ an associative memory network that consists of binary neurons. Simplifying the dynamics of the internal potential enables us to analyze the details of the retrieval process. We define a learning rule that is similar to the formulations used in previous works. We show the mechanism which makes the retrieval of spatio-temporal patterns possible in this network.

The storage of spatio-temporal patterns in associative memory networks by the covariance learning rule has been the subject of a lot of research (Chechik et al., 1999; Kitano and Aoyagi, 1998). A number of biological findings have implied that sparse coding schemes may be used in the brain (Miyashita, 1988). The covariance rule is well-known to be indispensable when the sparse patterns are embedded in a network as attractors (Amari, 1989; Okada, 1996). The information on the firing rate for the stored patterns is *not* indispensable to TAH, although it is indispensable to the covariance rule. We theoretically show that TAH has the same qualitative effect as the covariance rule when the spatio-temporal patterns are embedded in the network. This means that differences between spike times induce LTP or LTD, and this spike time difference is capable of canceling out the effect of the firing rate information. We conclude that this is why information on the firing rate is not required for the storage of patterns by TAH.

We introduce our model for the storage of spatio-temporal patterns in section 2. In section 3, we discuss the mechanism by which the spatio-temporal patterns are stored and retrieved, and discuss the basis for TAH having the same effect as the covariance learning rule. In section 4, we investigate the properties of our model. In section 5, we discuss our model. The Appendix presents a detailed derivation of the statistical neurodynamical equations of the model.

2 Model

We investigate a network which consists of N binary neurons that are mutually connected (Figure 1). In this paper, we consider the case where $N \to \infty$. We employ a neuronal model with binary state $\{0, 1\}$. We define discrete time steps and the following rule for synchronous updating:

$$u_i(t) = \sum_{j=1}^{N} J_{ij} x_j(t), \tag{1}$$

$$x_i(t+1) = \Theta(u_i(t) - \theta), \tag{2}$$

$$\Theta(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0, \end{cases} \tag{3}$$

where $x_i(t)$ is the state of the i-th neuron at time t, $u_i(t)$ is the internal potential of that neuron, and θ is a uniform threshold. If the i-th neuron fires at t, its state is $x_i(t) = 1$; otherwise, $x_i(t) = 0$. The specific value of the threshold is discussed later. J_{ij} is the synaptic weight from the j-th neuron to the i-th neuron.

Each element ξ_i^μ of the μ-th memory pattern $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \xi_2^\mu, \cdots, \xi_N^\mu)$ is generated independently by

$$\mathrm{Prob}[\xi_i^\mu = 1] = 1 - \mathrm{Prob}[\xi_i^\mu = 0] = f. \tag{4}$$

The expectation of $\boldsymbol{\xi}^\mu$ is $\mathrm{E}[\xi_i^\mu] = f$, and thus f is considered as the mean firing rate of the memory pattern. The memory pattern is "sparse" when $f \to 0$, and this coding scheme is called "sparse coding". The synaptic weight J_{ij} follows the form of the synaptic plasticity, which depends on the difference in spike times between the i-th (post-) and j-th (pre-) neurons. The difference determines whether LTP or LTD is induced. A learning rule of this type is called "Temporally Asymmetric Hebbian learning (TAH)" or "Spike Timing Dependent synaptic Plasticity (STDP)". The biological experimental findings indicate that LTP or LTD is induced when the difference in the pre- and post-synaptic spike times falls within about 20 ms (Zhang et al., 1998) (Figure 2(a)). We define a single time step in equations (1)–(3) as 20 ms for Figure 2(a), and durations within 20 ms are ignored (Figure 2(b)). Figure 2(b) shows that LTP is induced when the j-th neuron fires one time step before the i-th neuron, $\xi_i^{\mu+1} = \xi_j^\mu = 1$, while LTD is induced when the j-th neuron fires one time step after the i-th neuron, $\xi_i^{\mu-1} = \xi_j^\mu = 1$. Previous work has indicated the significance of the balance between LTP and LTD (Song et al., 2000). Therefore, we define that the area of LTP is the same as that of LTD, and that the amplitude of LTP is also the same as that of LTD. On the basis of these definitions, we employ the following learning rule:

$$J_{ij} = \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} (\xi_i^{\mu+1}\xi_j^\mu - \xi_i^{\mu-1}\xi_j^\mu). \tag{5}$$

The number of memory patterns is $p = \alpha N$, where α is defined as the "loading rate". There is a critical value α_C of the loading rate, so that a loading rate larger than α_C makes retrieval of the pattern sequence unstable. α_C is called the "storage capacity". Previous works have shown that the learning method described by equation (5) is capable of storing spatio-temporal patterns, that is, pattern sequences (Munro and Hernandez, 2000; Rao and Sejnowski, 2000). We show that a sequence of p memory patterns is retrieved, i.e. $\boldsymbol{\xi}^1 \rightarrow \boldsymbol{\xi}^2 \rightarrow \cdots \rightarrow \boldsymbol{\xi}^p \rightarrow \boldsymbol{\xi}^1 \rightarrow \cdots$ (Figure 3). In other words, $\boldsymbol{\xi}^1$ is retrieved at $t = 1$, $\boldsymbol{\xi}^2$ is retrieved at $t = 2$, and $\boldsymbol{\xi}^1$ is retrieved at $t = p + 1$.

Here, we discuss the value of the threshold θ. The threshold value is well-known to require time-dependent control according to the progress of the retrieval process (Amari, 1989; Okada, 1996). One candidate for the algorithm to control the threshold value is to maintain the mean firing rate of the network at that of memory pattern, f, as follows:

$$f = \frac{1}{N} \sum_{i=1}^{N} x_i(t) = \frac{1}{N} \sum_{i=1}^{N} \Theta(u_i(t) - \theta(t)). \tag{6}$$

The threshold value obtained is well-known to be nearly optimal, since it gives an approximate maximal value for the storage capacity (Okada, 1996).

3 Theory

Many neural network models for the storage and retrieval of sequential patterns by TAH have been discussed by many authors (Gerstner et al., 1996; Kempter et al., 1999; Munro and Hernandez, 2000; Rao and Sejnowski, 2000). They have numerically shown that TAH is effective for storing the patterns. For example, Munro and Hernandez showed that their model provides the retrieval of stored patterns even in noisy environments (Munro and Hernandez, 2000). However, reason for the effectiveness of TAH has not been mentioned in previous works. Exploring this mechanism is the main purpose of our paper.

Here, we discuss the mechanisms that the network learned by TAH can store and retrieve sequential patterns. Before providing details of this retrieval process, we discuss the simple situation where there are very few memory patterns relative to the number of neurons, i.e., $p \sim O(1)$. Let the state at time t be the equivalent to the t-th memory pattern: $\boldsymbol{x}(t) = \boldsymbol{\xi}^t$. The internal potential $u_i(t)$ of equation (1) is then given by

$$u_i(t) = \xi_i^{t+1} - \xi_i^{t-1}. \tag{7}$$

$u_i(t)$ depends on two independent random variables, ξ_i^{t+1} and ξ_i^{t-1}, according to the equation (7). The first term ξ_i^{t+1} is the signal term for the recall of the pattern $\boldsymbol{\xi}^{t+1}$, which is intended to be retrieved at time $t+1$, and the second term ξ_i^{t-1} may interfere with the retrieval of $\boldsymbol{\xi}^{t+1}$. According to the equation (7), $u_i(t)$ takes a value of 0, -1 or $+1$: see Table 1.

$\xi_i^{t+1} \backslash \xi_i^{t-1}$	0	1
0	0	-1
1	+1	0

Table 1: When $(\xi_i^{t+1}, \xi_i^{t-1}) = (1, 0)$, $u_i(t) = +1$, and when $(\xi_i^{t+1}, \xi_i^{t-1}) = (0, 1)$, $u_i(t) = -1$. Other than in these conditions, $u_i(t) = 0$. $\xi_i^{t-1} = 1$ is the case where interference of LTD appears.

As stated above, $\xi_i^{t-1} = 1$ is the case where interference of LTD appears. If the threshold $\theta(t)$ is set between 0 and +1, $\xi_i^{t+1} = 0$ is not affected by the interference of $\xi_i^{t-1} = 1$. When $\xi_i^{t+1} = 1$ and $\xi_i^{t-1} = 1$, the interference *does* affect the retrieval of $\boldsymbol{\xi}^{t+1}$. We consider the probability distribution of the internal potential $u_i(t)$ to examine how the interference of LTD affects the retrieval of $\boldsymbol{\xi}^{t+1}$. The probability of $\xi_i^{t+1} = 1$ and $\xi_i^{t-1} = 1$ is f^2, that of $\xi_i^{t+1} = 1$ and $\xi_i^{t-1} = 0$ is $f - f^2$, that of $\xi_i^{t+1} = 0$ and $\xi_i^{t-1} = 1$ is $f - f^2$, and that of $\xi_i^{t+1} = 0$ and $\xi_i^{t-1} = 0$ is $(1-f)^2$. The probability distribution of $u_i(t)$ is given by Figure 4(a), i.e.,

$$\mathrm{Prob}(u_i(t)) = (f-f^2)\delta(u_i(t)-1)+(1-2f+2f^2)\delta(u_i(t))+(f-f^2)\delta(u_i(t)+1). \tag{8}$$

If the threshold $\theta(t)$ is set between 0 and +1, the state $x_i(t+1)$ is 1 with a probability of $f - f^2$ and 0 with a probability of $1 - f + f^2$. The overlap between the state $\boldsymbol{x}(t+1)$ and the memory pattern $\boldsymbol{\xi}^{t+1}$ is given by

$$m^{t+1}(t+1) = \frac{1}{Nf(1-f)} \sum_{i=1}^{N} (\xi_i^{t+1} - f) x_i(t+1) = 1 - f. \tag{9}$$

In a sparse limit, $f \to 0$, the probability of $\xi_i^{t+1} = 1$ and $\xi_i^{t-1} = 1$ approaches 0. This means that the interference of the LTD disappears in this case, and the model is able to retrieve the next pattern $\boldsymbol{\xi}^{t+1}$. The overlap $m^{t+1}(t+1)$ then approaches 1.

We now discuss whether information on the firing rate is or is not indispensable to TAH. To investigate this, we consider the case where there is a large number of memory patterns, i.e., $p \sim O(N)$. Using equation (9), the internal potential $u_i(t)$ of the i-th neuron at time t is represented as

$$u_i(t) = (\xi_i^{t+1} - \xi_i^{t-1}) m^t(t) + z_i(t), \tag{10}$$

$$z_i(t) = \sum_{\mu \neq t}^{p} (\xi_i^{\mu+1} - \xi_i^{\mu-1}) m^{\mu}(t). \tag{11}$$

$z_i(t)$ is called the "cross-talk noise", and represents contributions of non-target patterns other than $\boldsymbol{\xi}^{t-1}$, and prevents the retrieval of the target pattern, $\boldsymbol{\xi}^{t+1}$. This term disappears in the finite loading case, $p \sim O(1)$.

The covariance learning rule, i.e.,

$$J_{ij} = \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} (\xi_i^{\mu+1} - f)(\xi_j^{\mu} - f) \tag{12}$$

is indispensable when sparse patterns are embedded in a network as attractors (Amari, 1989; Okada, 1996). Under sparse coding schemes, unless the covariance rule is employed, the cross-talk noise *does* diverge in the large N limit. Storage of the patterns thus becomes impossible. The information on the firing rate of the stored patterns is *not* indispensable to TAH, while it is indispensable to the covariance rule. We used the method of the "statistical neurodynamics" (Amari and Maginu, 1988; Okada, 1995) to examine whether the variance of the cross-talk noise diverges or not. If it is possible to store a pattern sequence, the cross-talk noise follows a Gaussian distribution with a mean of 0 and time-dependent variance $\sigma^2(t)$. Otherwise, $\sigma^2(t)$ diverges. The probability distribution of $u_i(t)$ is given by Figure 4(b). Since $\sigma^2(t)$ changes over time, it is necessary to control the threshold to an appropriate value at each time step (Amari, 1989; Okada, 1996). We then apply the statistical neurodynamics to obtain recursive equations for the overlap $m^t(t)$ between the network state $\boldsymbol{x}(t)$ and the target pattern $\boldsymbol{\xi}^t$ and the variance $\sigma^2(t)$ of cross-talk noise. The details of the derivation are given in the Appendix. Here, we show the recursive equations for $m^t(t)$ and $\sigma^2(t)$.

$$m^t(t) = \frac{1-2f}{2} \mathrm{erf}(\phi_0) - \frac{1-f}{2} \mathrm{erf}(\phi_1) + \frac{f}{2} \mathrm{erf}(\phi_2), \tag{13}$$

$$\sigma^2(t) = \sum_{a=0}^{t} {}_{2(a+1)}\mathrm{C}_{(a+1)}\alpha q(t-a) \prod_{b=1}^{a} U^2(t-b+1), \tag{14}$$

$$U(t) = \frac{1}{\sqrt{2\pi}\sigma(t-1)}\{(1-2f+2f^2)e^{-\phi_0^2} + f(1-f)(e^{-\phi_1^2} + e^{-\phi_2^2})\}, \tag{15}$$

$$q(t) = \frac{1}{2}\left(1 - (1-2f+2f^2)\mathrm{erf}(\phi_0) - f(1-f)(\mathrm{erf}(\phi_1) + \mathrm{erf}(\phi_2))\right), \tag{16}$$

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y \exp(-u^2)du, \quad {}_b\mathrm{C}_a = \frac{b!}{a!(b-a)!}, \quad a! = a \times (a-1) \times \cdots \times 1,$$

$$\phi_0 = \frac{\theta(t-1)}{\sqrt{2}\sigma(t-1)}, \phi_1 = \frac{-m^{t-1}(t-1) + \theta(t-1)}{\sqrt{2}\sigma(t-1)}, \phi_2 = \frac{m^{t-1}(t-1) + \theta(t-1)}{\sqrt{2}\sigma(t-1)}.$$

These equations reveal that the variance $\sigma^2(t)$ of cross-talk noise does *not* diverge, as long as a pattern sequence is retrievable. This result means that TAH has the same qualitative effect as the covariance rule.

Next, we discuss the mechanism that the variance of the cross-talk noise does not diverge. Let us consider equation (5). The synaptic weight J_{ij} from the j-th neuron to the i-th neuron is also derived in the following way.

$$J_{ij} = \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} (\xi_i^{\mu+1}\xi_j^{\mu} - \xi_i^{\mu-1}\xi_j^{\mu}) \tag{17}$$

$$= \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} (\xi_i^{\mu}\xi_j^{\mu-1} - \xi_i^{\mu}\xi_j^{\mu+1}) \tag{18}$$

$$= \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} \xi_i^{\mu}(\xi_j^{\mu-1} - \xi_j^{\mu+1}) \tag{19}$$

$$= \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} \xi_i^{\mu}\left\{(\xi_j^{\mu-1} - f) - (\xi_j^{\mu+1} - f)\right\} \tag{20}$$

This equation implies that TAH includes the information on the firing rate of the memory patterns when spatio-temporal patterns are embedded in a network. Therefore, the variance of the cross-talk noise does not diverge, and this is another factor that enables the network to store and retrieve a pattern sequence. We conclude that the differences in spike timing induce LTP or LTD, and that the effect of the firing rate information may be canceled out by the differences in spike timing.

4 Results

We used the statistical neurodynamics and computer simulation to investigate the properties of our model and examine its behavior under the following two conditions: a fixed threshold and a time-dependent threshold.

Figure 5 shows the dependence of the overlap $m^t(t)$ and the mean firing rate of the network, $\bar{x}(t) = \frac{1}{N}\sum_i^N x_i(t)$ on the loading rate α when the mean firing rate of the memory pattern is $f = 0.1$ and the threshold is $\theta = 0.52$, which is optimized to maximize the storage capacity. The upper line denotes the steady-state values of the overlap $m^t(t)$ in retrieval of the pattern sequence. $m^t(t)$ is obtained by setting the initial state of the network at the first memory pattern: $\boldsymbol{x}(1) = \boldsymbol{\xi}^1$. In this case, the storage capacity is $\alpha_C = 0.27$. The lower line indicates the dependence of the critical initial overlap m_C on α. m_C is obtained by setting the initial state of the network at $\boldsymbol{\xi}^1$ with additional noise. The method to add noise is that $100s\%$ of the minority components ($x_i(1) = 1$) are flipped, while the same number of majority components ($x_i(1) = 0$) are also flipped (Okada, 1996). Thus the mean firing rate of the network is kept equal to that of the memory pattern, which is f. The initial overlap $m^1(1)$ is given as $1 - \frac{2s}{1-f}$. The stored pattern sequence is retrievable when the initial overlap $m^1(1)$ is greater than the critical value m_C. Therefore, the lower line represents the basin of attraction for the retrieval of the pattern sequence. The dashed line shows the steady-state values of the normalized mean firing rate of the network, $\bar{x}(t)/f$, for the pattern sequence. The data points and error bars indicate the results of computer simulation of 11 trials with 5000 neurons: $N = 5000$, with the former indicating median values and the latter indicating 1/4 and 3/4 deviations. A discrepancy between the values of m_C obtained by the computer simulations and the theory exists. To investigate the cause of the discrepancy, we perform computer simulations with 50000 neurons and show the results in Figure 6. The data points and error bars indicate the results of computer simulations of 11 trials. The values of parameters are set at the same values as $N = 5000$ case of Figure 5. The lines of the theoretical result are identical with that of Figure 5. The storage capacity of the simulations approached to the analytical result, compared with that of Figure 5. This fact implies that the discrepancy between the storage capacity of the computer simulations and the analytical result is generated by the finite size effect of the computer simulations. Figure 7 shows the probability distribution of the internal potentials in the steady state for $\alpha = 0.1$. The histogram is obtained by computer simulation with $N = 5000$, and the solid line denotes the theoretical result of the statistical neurodynamics. Since the results for computer simulation and the statistical neurodynamics coincide, we give only the results as obtained by the statistical neurodynamics.

Next, we examine the application of the threshold control scheme given as equation (6), where the threshold is controlled to maintain the mean firing rate of the network at f. $q(t)$ in equation (16) is equal to the mean firing rate because $q(t) = \frac{1}{N}\sum_{i=1}^{N}(x_i(t))^2 = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$ under the condition $x_i(t) = \{0, 1\}$. Thus, the threshold is adjusted to satisfy the following equation:

$$f = q(t) = \frac{1}{2}\left(1 - (1 - 2f + 2f^2)\mathrm{erf}(\phi_0) - f(1-f)(\mathrm{erf}(\phi_1) + \mathrm{erf}(\phi_2))\right). \qquad (21)$$

Another candidate for the threshold control scheme is to maintain the activity at $f - f^2$, which is the firing rate of the signal term in equation (10): $\mathrm{Prob}(u_i(t) = 1) = f - f^2$. In this case, the left-hand side f of equation (21) is simply replaced by $f - f^2$. Figure 8 shows the overlap $m^t(t)$ as a function of loading rate α for both cases; (a) is to maintain activity at f and (b) is to maintain at $f - f^2$ with $f = 0.1$. The respective storage capacities are $\alpha_C = 0.234$ in (a) and $\alpha_C = 0.2597$ in (b). The storage capacity under condition (a) is larger than under condition (b), and both basins become larger than under the fixed threshold condition, i.e., $\theta = 0.52$ (Figure 5). The network is thus made robust against noise. This means that even if the initial state $\boldsymbol{x}(1)$ is different from the first memory pattern $\boldsymbol{\xi}^1$, that is, the state involves a lot of noise, the pattern sequence is retrievable.

Finally, we discuss the dependence of the storage capacity on the firing rate f of the memory pattern. The storage capacity is known to diverge as $\frac{1}{f|\log f|}$ in a sparse limit, $f \to 0$ (Tsodyks and Feigle'man, 1988; Perez-Vicente and Amit, 1989). Therefore, we investigate the asymptotic property of the storage capacity in the sparse limit. Figure 9 shows the dependence of the storage capacity on the firing rate when the threshold is controlled to maintain the network activity at f (symbol $\circ$) and at $f - f^2$ (symbol $+$). As f decreases, the storage capacity in the case where activity is maintained at f approaches that for activity maintained at $f - f^2$. The storage capacity diverges as $\frac{1}{f|\log f|}$ in a sparse limit.

5 Discussion

We have used a simple neural network model to examine mechanism by which TAH makes it possible for a pattern sequence to be stored and retrieved. Firstly, we showed that the interference between LTP and LTD disappeared in a sparse coding scheme. This is a factor that enables the network to store and retrieve a pattern sequence. Next, we showed the mechanism that TAH had the same qualitative effect as the covariance learning rule by analyzing the stability of the stored pattern sequence and the retrieval process by means of the statistical neurodynamics. As a consequence of this, the variance of cross-talk noise didn't diverge, and this is another factor which makes the network store and retrieve a pattern sequence. We conclude that the differences in spike timing induce LTP or LTD, and that the effect of the firing rate information may be canceled out by the differences in spike timing. We investigated the properties of our model. To improve the retrieval property of the basin of attraction, we introduced a threshold control algorithm which adjusts a threshold value to maintain the mean firing rate of the network at that of the memory pattern. We found that this scheme enlarged the basin of attraction and made the network more robust against noise. We also found that the loading rate diverged as $\frac{1}{f|\log f|}$ in a sparse limit, $f \to 0$.

Here, we compare the storage capacity of our model with that of the model where the covariance learning rule of equation (12) is used. A pattern sequence which consists of p memory patterns was stored according to this rule. The dynamical equations for this model have been derived by Kitano and Aoyagi (Kitano and Aoyagi, 1998). We calculate the storage capacity α_C^{COV} from their dynamical equations and compare it with that of our model, α_C^{TAH} by obtaining the ratio $\frac{\alpha_C^{TAH}}{\alpha_C^{COV}}$. The threshold control method is described earlier in this paper. As f decreases, the ratio approaches 0.5. The near-halving of the storage capacity for our model is because of the contribution of LTD. Therefore, in terms of storage capacity, the covariance learning rule is better than TAH. However, as stated previously, the information on the firing rate is indispensable to the covariance rule. In biological systems, it seems to be difficult to get the information on the firing rate. Therefore, TAH is more biologically plausible than the covariance rule.

In our analysis, TAH was formulated as equation (5) to realize that the area of LTP was the same as that of LTD. Next, we investigated the balance between LTP and LTD. We considered the case where LTP was induced for the same duration as LTD but the amplitudes of LTP and LTD fluctuated. In our model, it is not possible to cancel out the information on the firing rate when the balance between LTP and LTD is not maintained. Accordingly, the cross-talk noise must diverge. However, if the amplitudes of LTP and LTD are the same on average in a learning process, it is possible to stably retrieve the memory patterns. In other words, the balance between LTP and LTD need not be maintained precisely, but must simply be maintained on average. We intend to discuss this in qualitative terms in a future work.

Finally, we must briefly mention the connection between our model and a biological experimental finding which is called a "synfire chain": a volley of activity is propagated from one pool of neurons to the next pool (Abeles, 1991). There are two candidates to act as a model for a synfire chain: a recurrent network with connections learned by Hebbian learning (Herrmann et al., 1995) and a feed-forward network with uniform connections (Diesmann et al., 1999). Our model is a recurrent network in which a single sequence is embedded by TAH. Since our model is based on recent biological findings, it is more biologically plausible than that of Herrmann et al.. Our model is easily extended to the storage and retrieval of several pattern sequences. When the connectivity is sparse, our model may be regarded as a feed-forward network. Our model is thus a candidate to act as a model for a synfire chain.

Appendix. Derivation of the dynamical equations by the statistical neurodynamics

In this appendix, we derive recursive equations for the overlap $m^t(t)$ and the variance $\sigma^2(t)$. The internal potential $u_i(t)$ at time t is derived from equation (5) and the periodic boundary condition of $\xi_i^{p+1} = \xi_i^1$ and $\xi_i^0 = \xi_i^p$ in the following way:

$$\begin{aligned} u_i(t) &= \frac{1}{Nf(1-f)} \sum_{j=1}^{N} \sum_{\mu=1}^{p} (\xi_i^{\mu+1}\xi_j^{\mu} - \xi_i^{\mu-1}\xi_j^{\mu}) x_j(t) && (22) \\ &= \sum_{j=1}^{N} \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} \left((\xi_i^{\mu+1} - f) - (\xi_i^{\mu-1} - f) \right) (\xi_j^{\mu} - f) x_j(t) && (23) \\ &= \sum_{\mu=1}^{p} (\xi_i^{\mu+1} - \xi_i^{\mu-1}) m^{\mu}(t) && (24) \\ &= (\xi_i^{t+1} - \xi_i^{t-1}) m^t(t) + z_i(t). && (25) \end{aligned}$$

If it is possible to store a pattern sequence, the cross-talk noise will follow a Gaussian distribution with a mean of 0 and time-dependent variance $\sigma^2(t)$: $\mathrm{E}[z_i(t)] = 0, \mathrm{E}[(z_i(t))^2] = \sigma^2(t)$. The overlap is $m^t(t) = \frac{1}{Nf(1-f)} \sum_{i=1}^{N} (\xi_i^t - f) x_i(t)$. We use a

Taylor expansion to derive the following expression for the i-th neuronal state at time $t+1$.

$$
\begin{aligned}
x_i(t+1) &= F(\sum_{\nu=1}^{p}(\xi_i^{\nu+1}-\xi_i^{\nu-1})m^\nu(t)-\theta(t-1)) && (26)\\
&= F(\sum_{\nu\neq\mu}^{p}(\xi_i^{\nu+1}-\xi_i^{\nu-1})m^\nu(t)-\theta(t-1)) \\
&\quad +(\xi_i^{\mu+1}-\xi_i^{\mu-1})m^\mu(t)F'(\sum_{\nu\neq\mu}^{p}(\xi_i^{\nu+1}-\xi_i^{\nu-1})m^\nu(t)-\theta(t-1)) && (27)\\
&= x_i^{(\mu)}(t+1)+(\xi_i^{\mu+1}-\xi_i^{\mu-1})m^\mu(t)x_i'^{(\mu)}(t+1). && (28)
\end{aligned}
$$

Using this relationship, the cross-talk noise at time t, $z_i(t)$, is derived below.

$$
\begin{aligned}
z_i(t) &= \sum_{\mu\neq t}^{p}(\xi_i^{\mu+1}-\xi_i^{\mu-1})m^\mu(t) && (29)\\
&= \frac{1}{Nf(1-f)}\sum_{\mu\neq t}^{p}\sum_{j=1}^{N}\{(\xi_i^{\mu+1}-f)-(\xi_i^{\mu-1}-f)\}(\xi_j^\mu-f)x_j(t) && (30)\\
&= \frac{1}{Nf(1-f)}\sum_{\mu\neq t}^{p}\sum_{j=1}^{N}(\bar{\xi}_i^{\mu+1}-\bar{\xi}_i^{\mu-1})\bar{\xi}_j^\mu x_j^{(\mu)}(t) \\
&\quad +\sum_{\nu\neq t}^{p}U(t)(\bar{\xi}_i^{\nu+1}m^{\nu-1}(t-1)-2\bar{\xi}_i^{\nu-1}m^{\nu-1}(t-1)+\bar{\xi}_i^{\nu-1}m^{\nu+1}(t-1)), && (31)
\end{aligned}
$$

where $U(t)=\mathrm{E}[x_i'^{(\mu)}(t)]$ and $\bar{\xi}_i^\nu=\xi_i^\nu-f$. $x_i'^{(\mu)}(t)$ is independent of $x_i'^{\mu}(t)$ and is the differential of $x_i^{(\mu)}(t)$. Therefore, the square of $z_i(t)$ is given by

$$
\begin{aligned}
(z_i(t))^2 &= \left(\frac{1}{Nf(1-f)}\right)^2\sum_{\mu\neq t}^{p}\sum_{j=1}^{N}(\bar{\xi}_i^{\mu+1}-\bar{\xi}_i^{\mu-1})^2(\bar{\xi}_j^\mu)^2(x_j^{(\mu)}(t))^2 \\
&\quad +\sum_{\nu\neq t}^{p}U(t)^2(\bar{\xi}_i^{\nu+1}m^{\nu-1}(t-1)-2\bar{\xi}_i^{\nu-1}m^{\nu-1}(t-1) \\
&\quad +\bar{\xi}_i^{\nu-1}m^{\nu+1}(t-1))^2 && (32)\\
&= (1^2+(-1)^2)\alpha q(t)+(1^2+(-2)^2+1^2)\alpha q(t-1)U^2(t) \\
&\quad +(1^2+(-3)^2+3^2+(-1)^2)\alpha q(t-2)U^2(t)U^2(t-1)+\cdots && (33)
\end{aligned}
$$

$$= \sum_{a=0}^{t} {}_{2(a+1)}\mathrm{C}_{(a+1)} \alpha q(t-a) \prod_{b=1}^{a} U^2(t-b+1), \tag{34}$$

where $p = \alpha N$, $q(t) = \frac{1}{N}\sum_{i=1}^{N}(x_i^{(\mu)}(t))^2$, ${}_b\mathrm{C}_a = \frac{b!}{a!(b-a)!}$ and $a!$ is the factorial with positive integer a. We applied the relationship $\sum_{a=0}^{b}({}_b\mathrm{C}_a)^2 = {}_{2b}\mathrm{C}_b$ in this derivation. The expectation of $z_i(t)$ is assumed to be 0, so the variance $\sigma^2(t)$ is equal to $\mathrm{E}[(z_i(t))^2]$. We then get a recursive equation for $\sigma^2(t)$:

$$\sigma^2(t) = \sum_{a=0}^{t} {}_{2(a+1)}\mathrm{C}_{(a+1)} \alpha q(t-a) \prod_{b=1}^{a} U^2(t-b+1). \tag{35}$$

Replacing the average over the sites with the average over the Gaussian noise term, a recursive equation for the overlap $m^t(t)$ between the network state $\boldsymbol{x}(t)$ and the retrieval pattern $\boldsymbol{\xi}^t$ is obtained as follows:

$$m^t(t) = \frac{1}{Nf(1-f)} \sum_{i=1}^{N} (\xi_i^t - f) x_i(t) \tag{36}$$

$$= \frac{1}{f(1-f)} \langle\langle \int Dz(\xi^t - f)x(t) \rangle\rangle \tag{37}$$

$$= \frac{1}{f(1-f)} \langle\langle \int Dz(\xi^t - f)F((\xi^t - \xi^{t-2})m^{t-1}(t-1) + \sigma(t-1)z$$

$$-\theta(t-1))\rangle\rangle \tag{38}$$

$$= \frac{1-2f}{2}\mathrm{erf}(\phi_0) - \frac{1-f}{2}\mathrm{erf}(\phi_1) + \frac{f}{2}\mathrm{erf}(\phi_2), \tag{39}$$

where $\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y \exp(-u^2)du$, and $\phi_0 = \frac{\theta(t-1)}{\sqrt{2}\sigma(t-1)}$, $\phi_1 = \frac{-m^{t-1}(t-1)+\theta(t-1)}{\sqrt{2}\sigma(t-1)}$, $\phi_2 = \frac{m^{t-1}(t-1)+\theta(t-1)}{\sqrt{2}\sigma(t-1)}$. $U(t)$ is derived in the following way:

$$U(t) = \frac{1}{N}\sum_{i=1}^{N} x_i'^{(\mu)}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i'(t) = \mathrm{E}[x'(t)] \tag{40}$$

$$= \mathrm{E}[F'((\xi^t - \xi^{t-2})m^{t-1}(t-1) + \sigma(t-1)z - \theta(t-1))] \tag{41}$$

$$= \int_{-\infty}^{\infty} Dz \langle\langle F'((\xi^t - \xi^{t-2})m^{t-1}(t-1) + \sigma(t-1)z - \theta(t-1))\rangle\rangle \tag{42}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz e^{-\frac{z^2}{2}} \langle\langle F'((\xi^t - \xi^{t-2})m^{t-1}(t-1) + \sigma(t-1)z$$

$$-\theta(t-1))\rangle\rangle \tag{43}$$

$$= \frac{1}{\sqrt{2\pi}\sigma(t-1)}\{(1-2f+2f^2)e^{-\phi_0^2} + f(1-f)(e^{-\phi_1^2} + e^{-\phi_2^2})\}. \tag{44}$$

Finally, we derive $q(t)$.

$$q(t) = \frac{1}{N}\sum_{i=1}^{N}(x_i^{(\mu)}(t))^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i(t))^2 = \mathrm{E}[(x(t))^2] \tag{45}$$

$$= \mathrm{E}[F^2((\xi^t - \xi^{t-2})m^{t-1}(t-1) + \sigma(t-1)z - \theta(t-1))] \tag{46}$$

$$= \int_{-\infty}^{\infty} Dz\langle\langle F^2((\xi^t - \xi^{t-2})m^{t-1}(t-1) + \sigma(t-1)z - \theta(t-1))\rangle\rangle \tag{47}$$

$$= \frac{1}{2}\left(1 - (1-2f+2f^2)\mathrm{erf}(\phi_0) - f(1-f)(\mathrm{erf}(\phi_1) + \mathrm{erf}(\phi_2))\right). \tag{48}$$

References

Abbott, L. F. and Song, S. (1999). Temporally asymmetric hebbian learning, spike timing and neuronal response variability. In *Advances in Neural Information Processing Systems 11*, pages 69–75. MIT Press.

Abeles, M. (1991). *Corticonics*. Cambridge Univ. Press.

Amari, S. (1989). Characteristics of sparsely encoded associative memory. *Neural Networks*, 2:1007–1018.

Amari, S. and Maginu, K. (1988). Statistical neurodynamics of various versions of correlation associative memory. *Neural Networks*, 1:63–73.

Bi, G. Q. and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18:10464–10472.

Chechik, G., Meilijson, I., and Ruppin, E. (1999). Effective learning requires neuronal remodeling of hebbian synapses. In *Advances in Neural Information Processing Systems 11*, pages 96–102. MIT Press.

Diesmann, M., Gewaltig, M., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402:529–533.

Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383:76–78.

Herrmann, M., Hertz, J. A., and Prügel-Bennett, A. (1995). Analysis of synfire chains. *Network: Computation in Neural Systems*, 6:403–414.

Kempter, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E*, 59:4498–4514.

Kitano, K. and Aoyagi, T. (1998). Retrieval dynamics of neural networks for sparsely coded sequential patterns. *Journal of Physics A: Mathematical and General*, 31:L613–L620.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275:213–215.

Miyashita, M. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335:817–820.

Munro, P. and Hernandez, G. (2000). LTD facilitates learning in a noisy environment. In *Advances in Neural Information Processing Systems 12*, pages 150–156. MIT Press.

Okada, M. (1995). A hierarchy of macrodynamical equations for associative memory. *Neural Networks*, 8:833–838.

Okada, M. (1996). Notions of associative memory and sparse coding. *Neural Networks*, 9:1429–1458.

Perez-Vicente, C. J. and Amit, D. J. (1989). Optimized network for sparsely coded patterns. *Journal of Physics A: Mathematical and General*, 22:559–569.

Rao, R. P. N. and Sejnowski, T. J. (2000). Predictive sequence learning in recurrent neocortical circuits. In *Advances in Neural Information Processing Systems 12*, pages 164–170. MIT Press.

Rubin, J., Lee, D. D., and Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical Review Letters*, 86:364–367.

Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3:919–926.

Tsodyks, M. V. and Feigle'man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters*, 6:101–105.

Yoshioka, M. (2002). Spike-timing-dependent learning rule to encode spatiotemporal patterns in a network of spiking neurons. *Physical Review E*, 65:011903.

Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., and Poo, M. M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395:37–44.

Figure captions

Figure 1: An associative memory network which stores spatio-temporal patterns. $\boldsymbol{x}(t)$ and $\boldsymbol{x}(t+1)$ are network state at time t and $t+1$ respectively. J_{ij} is synaptic weight from j-th neuron to i-th neuron and is independent of time. The network consists of N binary neurons, that is, each state is $\{0,1\}$.

Figure 2: Temporally Asymmetric Hebbian plasticity: (a) is the result of biological findings (Zhang et al., 1998) and (b) is the learning rule in our model. LTP is induced when the j-th neuron fires one time step before the i-th neuron. LTD is induced when the j-th neuron fires one time step after the i-th one. Synaptic weight J_{ij} follows this form of TAH.

Figure 3: The stored pattern sequence. Each pattern is retrieved periodically. The patterns retrieved are $\boldsymbol{\xi}^1$ at time 1, $\boldsymbol{\xi}^2$ at time 2, and $\boldsymbol{\xi}^1$ at time $t+1$. In short, $\xi_i^{p+1} = \xi_i^1$ and $\xi_i^0 = \xi_i^p$.

Figure 4: The probability distribution of the i-th neuronal potential $u_i(t)$ at time t. (a) is in the case where $p \sim O(1)$ and (b) is $p \sim O(N)$. In (a), there are three delta functional distributions. If the threshold $\theta(t)$ is set between 0 and $+1$ at time t, the pattern $\boldsymbol{\xi}^{t+1}$ is retrievable. In (b), each distribution follows a Gaussian distribution; the respective means are -1, 0 and $+1$, and the variance of each is $\sigma^2(t)$. Adjusting the threshold at an appropriate value is thus essential.

Figure 5: The critical overlap (the lower line) and the overlap in the steady state (the upper line). The dashed line shows the normalized firing rate of the network when the firing rate of stored patterns is 0.1. The threshold is 0.52 and the number of neurons is 5000. The data points indicate the median values and both ends of the error bars indicate 1/4 and 3/4 derivations, respectively. They are obtained in 11 computer simulation trials. The storage capacity is 0.27.

Figure 6: The storage capacity for $N = 50000$. The values of parameters are set at the same values as $N = 5000$ case of Figure 5. The lines of the theoretical results are identical with these of Figure 5. The storage capacity of the simulations approached to the analytical result, compared with that of $N = 5000$ of Figure 5. This fact implies that the discrepancy between the storage capacity of the computer simulations and the analytical result are generated by the finite size effect of the computer simulations.

Figure 7: The probability distribution of internal potential in the steady state of $\alpha = 0.1$. The histogram is obtained by computer simulation with 5000 neurons, while the solid line is obtained by the statistical neurodynamics. Each probabilistic distribution follows a Gaussian distribution.

Figure 8: The critical overlap (the lower line) and the overlap in the steady state (the upper line) when the threshold is varied over time to maintain the mean firing rate of the network at (a) f and (b) $f - f^2$. The dashed lines in both figures show the normalized firing rate of the network when the firing rate of stored patterns is 0.1. The respective storage capacities are $\alpha_C = 0.234$ in (a) and $\alpha_C = 0.2597$ in (b). Both basins of attraction are larger than under the fixed threshold condition, i.e., in Figure 5.

Figure 9: The storage capacity α_C as a function of f in the cases where activity is maintained at f (symbol $\circ$) and at $f - f^2$ (symbol $+$). As f decreases, the storage capacity for activity maintained at f approaches that for activity maintained at $f - f^2$. The capacity diverges as $\frac{1}{f|\log f|}$ in a sparse limit.

Figure 10: A comparison of the storage capacities for our model, α_C^{TAH}, and for the model based on the covariance learning, α_C^{COV}. The ratio $\frac{\alpha_C^{TAH}}{\alpha_C^{COV}}$ is plotted as a function of f. As f decreases, the ratio approaches 0.5. The near-halving of the storage capacity for our model is because of the contribution of LTD.

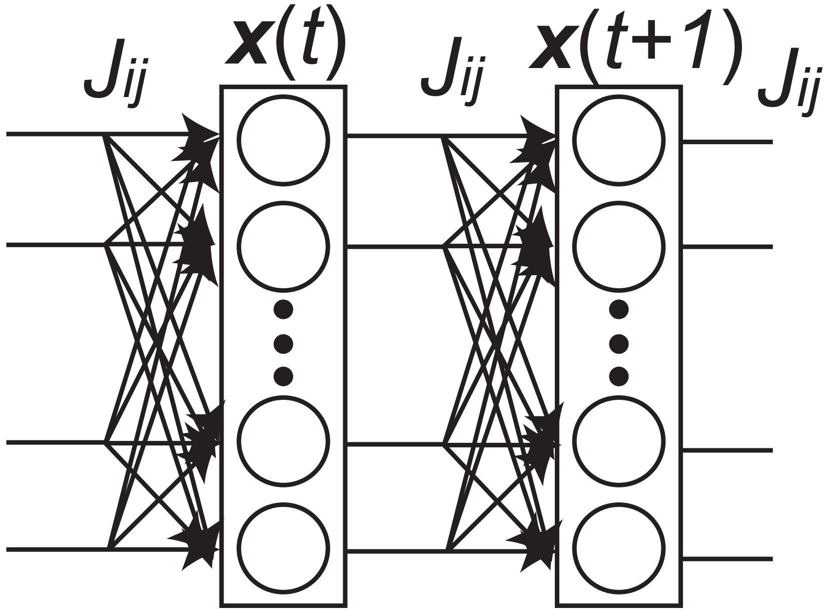


Figure 1

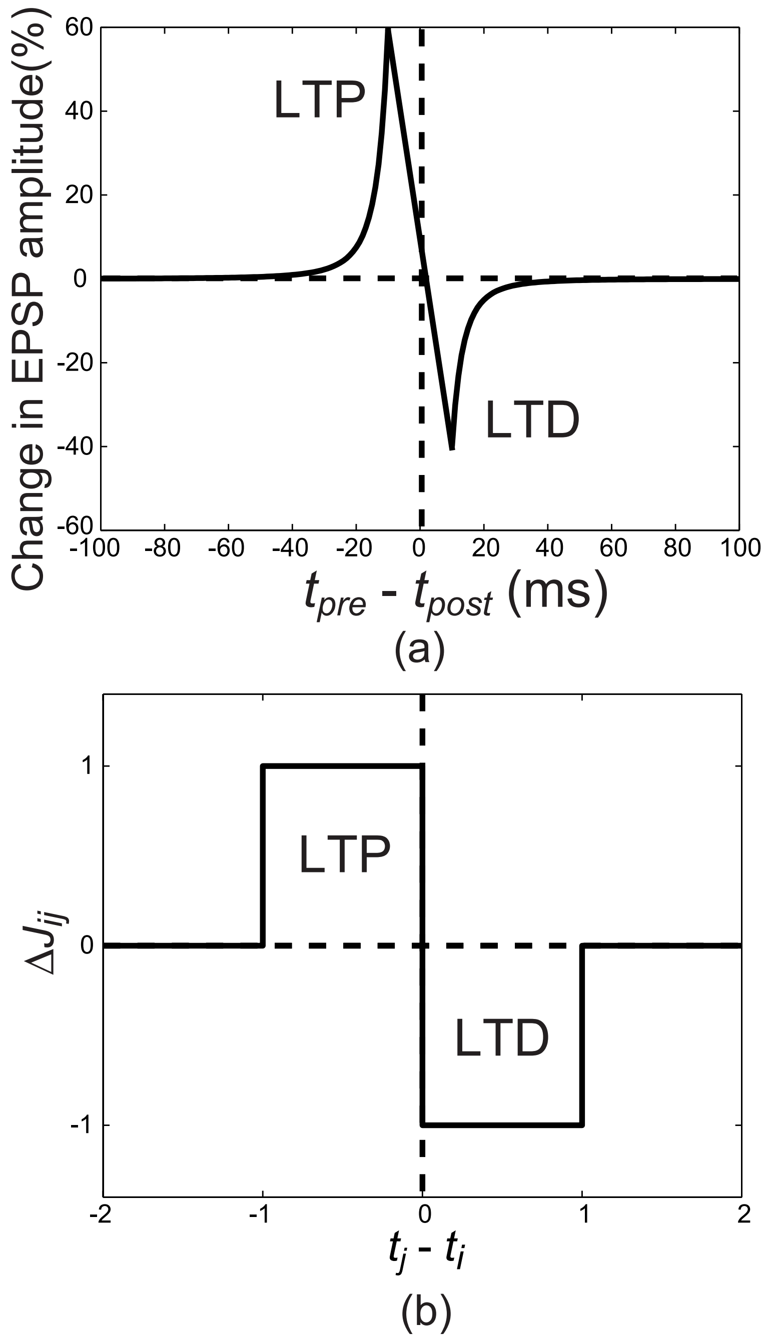

LTP
LTD
Change in EPSP amplitude(%)
t_{pre} - t_{post} (ms)
(a)
LTP
LTD
ΔJ_{ij}
t_j - t_i
(b)

Figure 2

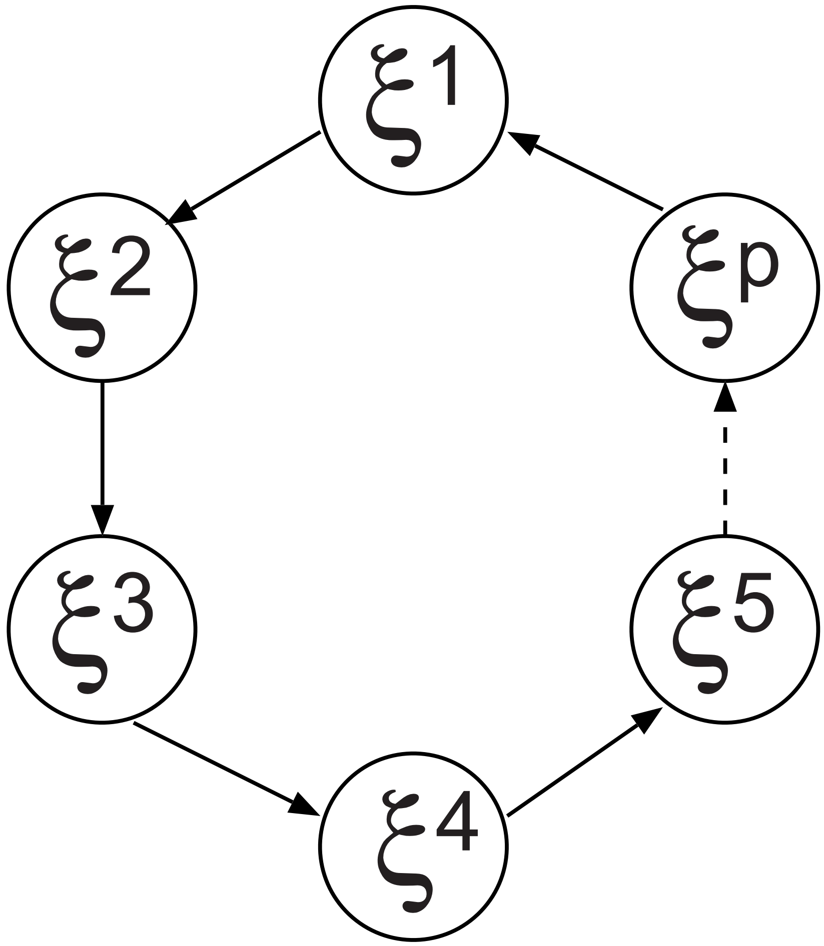


Figure 3

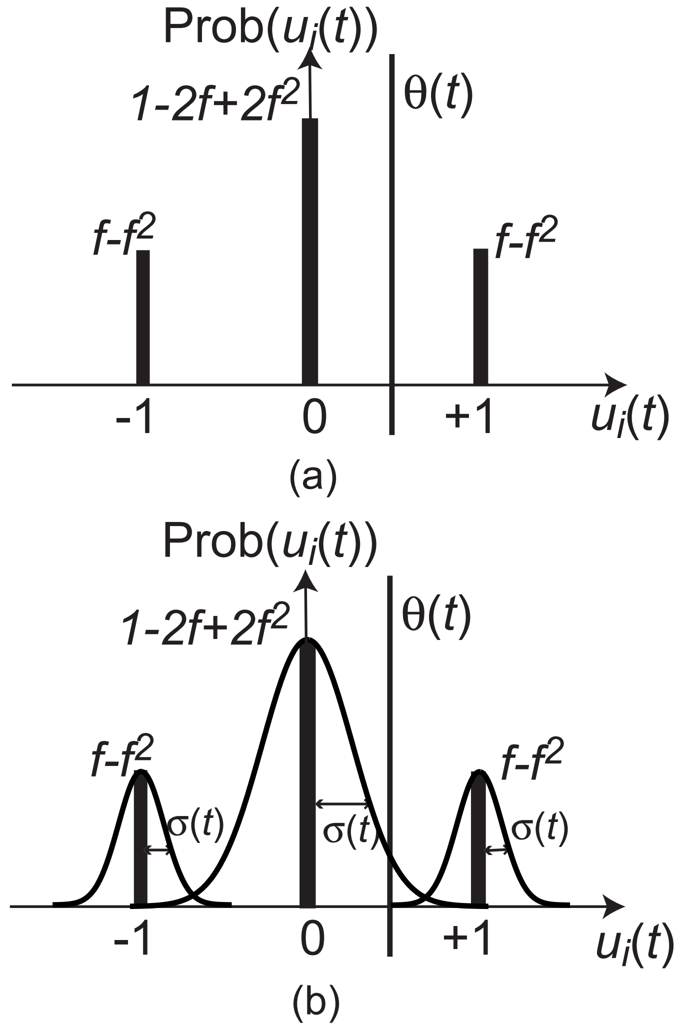

Prob($u_i(t)$)
$1-2f+2f^2$
$\theta(t)$
$f-f^2$
$f-f^2$
-1
0
+1
$u_i(t)$
(a)
Prob($u_i(t)$)
$1-2f+2f^2$
$\theta(t)$
$f-f^2$
$f-f^2$
$\sigma(t)$
$\sigma(t)$
$\sigma(t)$
-1
0
+1
$u_i(t)$
(b)

Figure 4

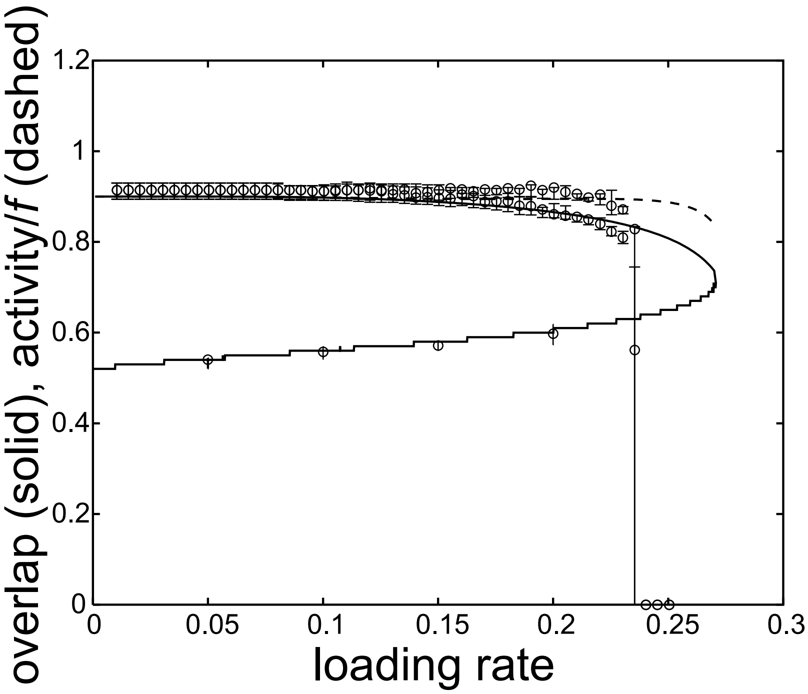

overlap (solid), activity/f (dashed)
1.2
1
0.8
0.6
0.4
0.2
0
0
0.05
0.1
0.15
0.2
0.25
0.3
loading rate

Figure 5

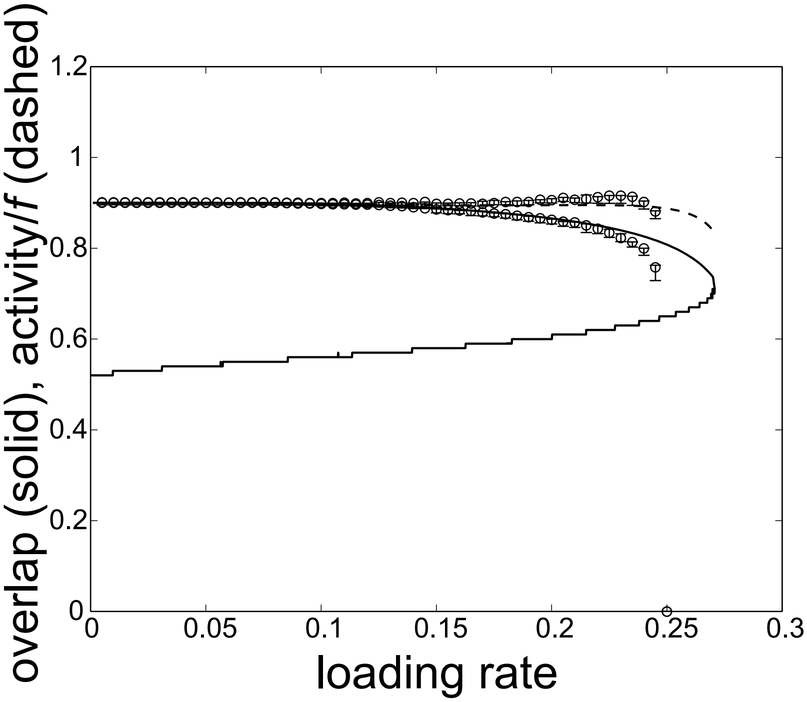
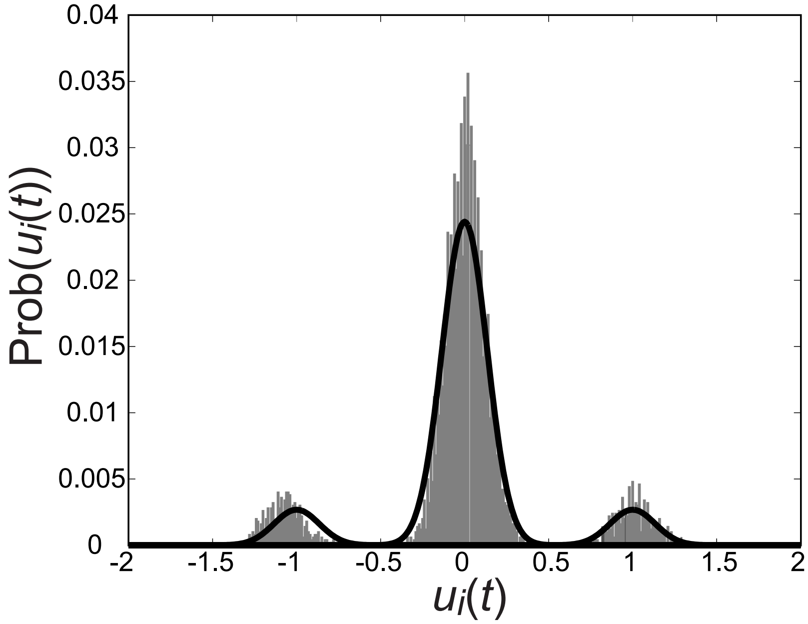

overlap (solid), activity/f (dashed)
1.2
1
0.8
0.6
0.4
0.2
0
0
0.05
0.1
0.15
0.2
0.25
0.3
loading rate

Figure 6



0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Prob($u_i(t)$)
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
$u_i(t)$

Figure 7

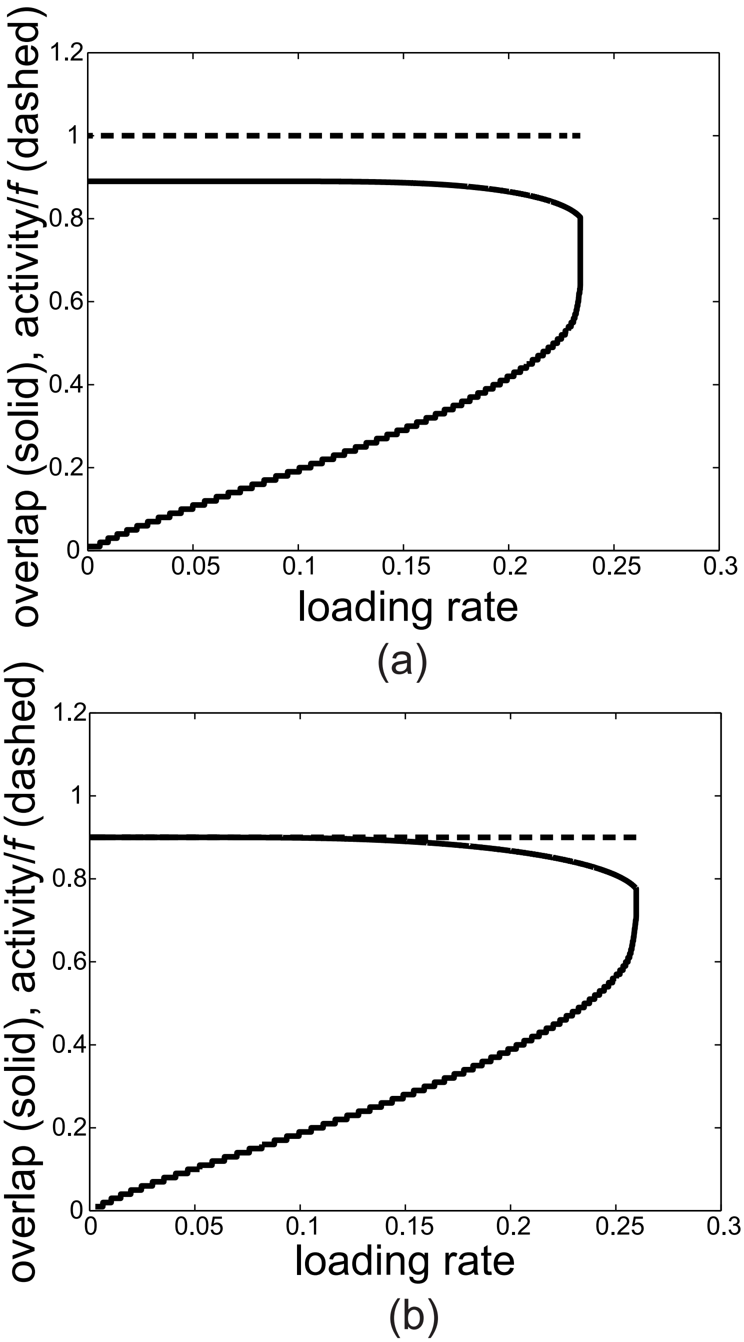
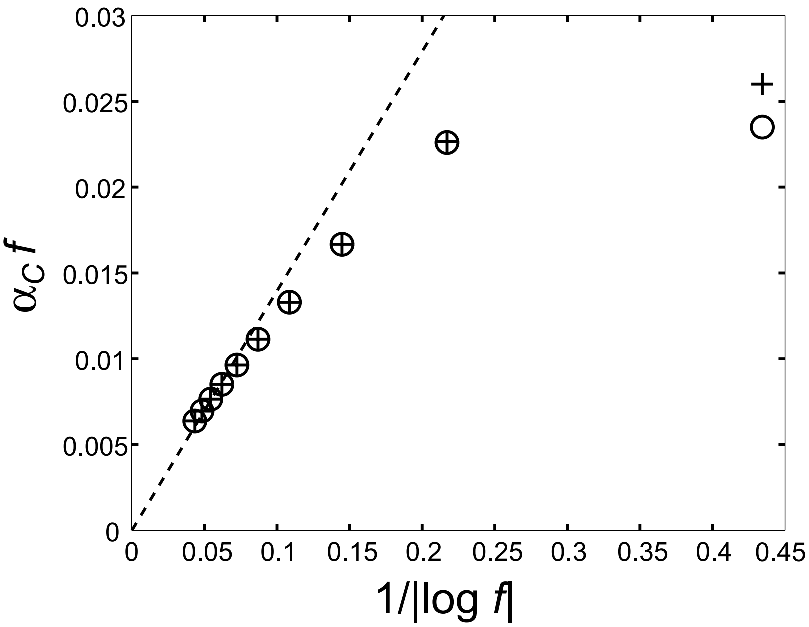

overlap (solid), activity/f (dashed)
loading rate
(a)
overlap (solid), activity/f (dashed)
loading rate
(b)

Figure 8



0.03
0.025
0.02
0.015
0.01
0.005
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
$\alpha_C f$
1/|log f|

Figure 9

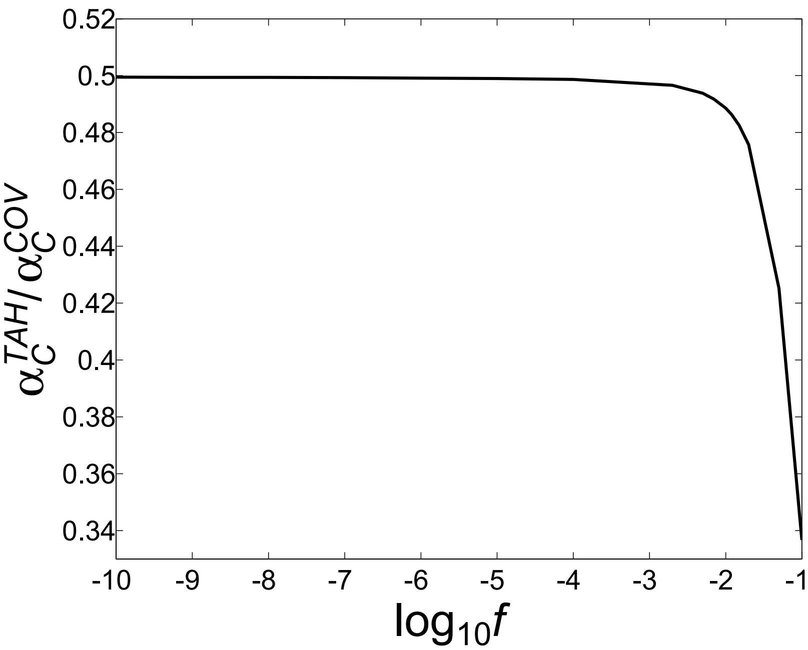

0.52
0.5
0.48
0.46
0.44
0.42
0.4
0.38
0.36
0.34
$\alpha_C^{TAH}/\alpha_C^{COV}$
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
$\log_{10} f$

Figure 10