

Cooperation of spike timing-dependent and heterosynaptic plasticities in neural networks: A Fokker-Planck approach

Liqiang Zhu

School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing, China and Department of Electrical Engineering, Arizona State University, Tempe, Arizona 85287

Ying-Cheng Lai

Department of Electrical Engineering, Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287

Frank C. Hoppensteadt

Courant Institute, New York University, New York, New York 10012

Jiping He

Department of Bioengineering, Arizona State University, Tempe, Arizona 85287

(Received 15 September 2005; accepted 6 March 2006; published online 11 May 2006)

It is believed that both Hebbian and homeostatic mechanisms are essential in neural learning. While Hebbian plasticity selectively modifies synaptic connectivity according to activity experienced, homeostatic plasticity constrains this change so that neural activity is always within reasonable physiological limits. Recent experiments reveal spike timing-dependent plasticity (STDP) as a new type of Hebbian learning with high time precision and heterosynaptic plasticity (HSP) as a new homeostatic mechanism acting directly on synapses. Here, we study the effect of STDP and HSP on randomly connected neural networks. Despite the reported successes of STDP to account for neural activities at the single-cell level, we find that, surprisingly, at the network level, networks trained using STDP alone cannot seem to generate realistic neural activities. For instance, STDP would stipulate that past sensory experience be maintained forever if it is no longer activated. To overcome this difficulty, motivated by the fact that HSP can induce strong competition between sensory experiences, we propose a biophysically plausible learning rule by combining STDP and HSP. Based on the Fokker-Planck theory and extensive numerical computations, we demonstrate that HSP and STDP operated on different time scales can complement each other, resulting in more realistic network activities. Our finding may provide fresh insight into the learning mechanism of the brain. © 2006 American Institute of Physics. [DOI: [10.1063/1.2189969](https://doi.org/10.1063/1.2189969)]

Learning and adaptation are key to natural selection and evolution, as the successful survival of an individual species depends strongly on its abilities to learn new experiences and to adapt the acquired skills to the changing environment. The ability to alter behavior is a result of changes in the nervous system. How learning and adaptation are carried out in the brain has been among the most fundamental issues in neuroscience. Studies on learning range from molecules and cells through neural networks, to animal behavior and psychology, in which both experimental and theoretical disciplines play important roles. Motivated by the recent experimental observations on spike timing-dependent and heterosynaptic plasticities, we address how neural networks change during learning and adaptation. By utilizing a physical theory based on the Fokker-Planck equation and extensive numerical computations, we establish a biophysically plausible learning rule incorporating both types of neural plasticity and show that it can result in realistic neural activities at the network level. This work thus represents an example of how tools from statistical and nonlinear

physics can be applied to addressing interesting problems in biological sciences.

I. INTRODUCTION

Synaptic plasticity, changes in the synaptic conductance in response to learning and adaptation, is fundamental to memory and the development of neural circuits. The classical Hebbian rule,¹ which has been the foundation for research on the role of synaptic plasticity, is based on the intuition that if, in the process of learning or adaptation, input from one neuron results in the firing of another neuron, then the synaptic connection between those two neurons is potentiated. This postulate has received support from many experiments showing that synapses can go through not only long-term potentiation (LTP), but also long-term depression (LTD), depending on the pattern of the neural activity. Motivated by these observations, several forms of Hebbian learning have been proposed to induce LTP/LTD in terms of pre- and/or postsynaptic firing rates, the most representative being the BCM rule, named after Bienenstock, Copper, and

Munro.² According to the BCM rule, for instance, LTP (LTD) arises if synaptic inputs result in a postsynaptic firing rate above (below) a threshold that depends on the postsynaptic activity. While the BCM rule can indeed account for important neural phenomena,³ it is basically a *rate-based rule*, i.e., it ignores entirely the timing information of the synaptic activity. Recent experiments have indicated that this timing information can play a critical role in determining the synaptic changes in both sign and magnitude.⁴ In particular, LTP can be observed if the presynaptic action potential is *followed* by a postsynaptic one, whereas LTD occurs if the temporal order of the action potentials is reversed. This type of plasticity, named *spike timing-dependent plasticity* (STDP), is believed to be important in shaping the various synaptic plasticities required for learning and adaptation.⁴

In general, Hebbian learning contains a positive-feedback mechanism. That is, once a synapse is potentiated, it becomes easier for the presynaptic neuron to make the postsynaptic neuron fire, promoting further potentiation of the synapse. The stability of the Hebbian learning rule is thus important. While modifications such as incorporating a postsynaptic-dependent threshold in BCM or introducing dependence on the initial synaptic size in STDP can make Hebbian learning stable, a wealth of evidence suggests that homeostatic mechanisms are also employed by neural networks to maintain stability. In particular, a neuron or a network of neurons has some preset activity level that is dynamically maintained. For example, chronically reducing inhibition in cortical networks initially raises firing rates, but over a period of days firing rates can return to the control level.⁵ Similarly, as shown in Ref. 6, interactions among neurons in the primary motor cortex can be strengthened at the beginning of adaptation, but usually return to the original level after a few days. Another example is the heterosynaptic plasticity (HSP) recently reported in Ref. 7, in which LTP or LTD introduced in one synapse lead to opposite changes in other synapses on the same postsynaptic neuron, and during the process the total synaptic conductance changes little. In this sense, HSP can prevent the phenomenon of runaway synapses introduced by Hebbian learning, and thus represents a homeostatic mechanism stabilizing neural activities. At present, however, the fundamental feature(s) of a neuron or a network of neurons being dynamically maintained is still not clear. Furthermore, how this mechanism interacts with other regulatory mechanisms is not well understood either.

This study is motivated by the two recent experimental discoveries on Hebbian and homeostatic mechanisms: STDP and HSP. A question of interest is, if both STDP and HSP take place in a neural circuit, how do they interact with and possibly complement each other so as to modify the neural circuit for learning and adaptation? To be as general as possible, we shall investigate the effects of STDP and HSP on randomly connected neural networks. At the present, however, the molecular mechanism underlying LTP/LTD is still not well understood, hindering a universal formulation of the STDP rule. To overcome this difficulty, we shall employ a theoretical analysis based on stochastically dynamic processes and the Fokker-Planck paradigm to propose an implementation of STDP that is compatible with rate-based BCM

and is also able to generate realistic synaptic distributions. Our study shows that (1) STDP alone produces broadened unimodal or bimodal weight distributions, weak competition between recurrent connections in networks, and activity-induced learning; (2) Networks trained using STDP alone cannot produce realistic activities, e.g., past sensory experience is maintained forever if it is no longer activated; (3) HSP can induce strong competition between sensory experiences; (4) HSP and STDP operating on different time scales can complement each other to generate more realistic network activities.

In Sec. II, we detail our physical theory based on the Fokker-Planck paradigm to model STDP and HSP. Section III presents results of extensive numerical computations using random neural networks. A discussion is given in Sec. IV.

II. PHYSICAL THEORY OF SYNAPTIC PLASTICITY

The learning rule studied in this paper consists of two components—STDP and HSP. STDP is local and homosynaptic in the sense that only the synapses experiencing the spiking activities are modified, which can be modeled as

$$\Delta g = G(\Delta t|g), \quad (1)$$

where Δg is the percentage change in synaptic conductance g due to a pair of pre- and postsynaptic spikes separated by time Δt (positive Δt implies that presynaptic spike precedes postsynaptic spike and negative one for the reverse order). On the other hand, HSP is nonlocal and heterosynaptic in the sense that modification of one synapse may be accompanied by changes in other synapses on the same neuron. Based on the observation in Ref. 7, our idealized HSP rule reads

$$\tau_{\text{HSP}} \frac{d\bar{g}}{dt} = -\bar{g} + g_{\text{goal}}, \quad (2)$$

where τ_{HSP} is the time constant of HSP, \bar{g} is the sum of presynaptic conductances of one neuron, and g_{goal} is the desired value of the total conductance. In reality, STDP and HSP operate on different time scales: minutes and hours, respectively.

A. Theoretical formulation

Consider one typical synapse connecting two neurons in a network and its weight changes according to STDP/HSP and the pre-/postsynaptic spike timings. Since inputs from thousands of presynaptic neurons cause the firings of the postsynaptic neuron, weight changes of the synapse under consideration can be modeled as a random walk with other synaptic inputs as background noise. In the limit of small $|\Delta g|$, which is the case of STDP/HSP, the probability of observing a synaptic weight g at time t , or $P(g, t)$, can be described by the following Fokker-Planck equation:⁸⁻¹⁰

$$\frac{\partial P(g, t)}{\partial t} = -\frac{\partial}{\partial g}[A(g)P(g, t)] + \frac{1}{2} \frac{\partial^2}{\partial g^2}[B(g)P(g, t)], \quad (3)$$

where the drift $A(g)$ and the diffusion $B(g)$ are

$$A(g) = \int_{-\infty}^{\infty} \Delta g P_g(\Delta g|g) d\Delta g, \quad (4)$$

$$B(g) = \int_{-\infty}^{\infty} \Delta g^2 P_g(\Delta g|g) d\Delta g,$$

and $P_g(\Delta g|g)$ is the probability for the event $g \rightarrow g + \Delta g$. Note that $P(g, t = \infty)$ is confined to the region $[0, g_{\text{goal}}]$ due to the approximate reflecting boundary conditions imposed. Because the amplitude of Δg is small, the diffusion term $B(g)$ is small. As a result, the final $P(g, t = \infty)$ will concentrate near where the drift term $A(g)$ vanishes.

To find $A(g)$, one can write $\Delta g = \Delta g_1 + \Delta g_2$, where Δg_1 is caused by STDP and Δg_2 is caused by HSP. The interaction between these two drift forces is weak due to the fact that they operate on different time scales. Thus, it is natural to assume these two drift forces are independent (this is true for the asymptotic solutions). We have

$$A(g) = \int \int (\Delta g_1 + \Delta g_2) P_g(\Delta g_1, \Delta g_2|g) d\Delta g_1 d\Delta g_2$$

$$= \int \Delta g_1 P_g(\Delta g_1|g) d\Delta g_1 + \int \Delta g_2 P_g(\Delta g_2|g) d\Delta g_2$$

$$\equiv A_1(g) + A_2(g), \quad (5)$$

where $A_1(g)$ and $A_2(g)$ are the drift terms due to STDP and HSP, respectively.

B. $A_1(g)$ for different implementations of STDP

Substituting Eq. (1) into the expression of $A_1(g)$ in Eq. (5), we obtain

$$A_1(g) = \int_{-\infty}^{\infty} G(\Delta t|g) P_t(\Delta t|g) d\Delta t, \quad (6)$$

where $P_t(\Delta t|g) = P_g(\Delta g|g) G'(\Delta t|g)$ is the probability that the effective pair of pre- and postsynaptic spikes is separated by time Δt at a synapse with conductance g . Two basic questions that determine $G(\Delta t|g)$ and $P_t(\Delta t|g)$ are (1) whether relative change in synaptic weight due to a pair of spikes depends on the weight itself^{8,9,11} and (2) how pre- and postsynaptic spikes in long spike trains *pair together* to induce the change.^{12,13} It is possible, however, to address these two questions by investigating the functional consequences of all possible answers and examining how well these consequences are consistent with existing knowledge, e.g., broad conductance distributions [Fig. 1(e) in Ref. 14 and Fig. 1(b) in Ref. 15] and dependence on firing rates (the BCM rule²).

For the first question, two types of STDP learning rules have been proposed in the literature. The first type is weight-independent (or additive) STDP, which can be modeled as

$$\Delta g_1 = G(\Delta t|g) = \begin{cases} c_p e^{-\Delta t/\tau_p}, & \Delta t > 0, \\ -c_d e^{\Delta t/\tau_d}, & \Delta t < 0, \end{cases} \quad (7)$$

where c_p , c_d , τ_p , and τ_d are constants. If g is outside the region of $[0, g_{\text{max}}]$ due to the learning process, g is forced to be boundary values. The second type is weight-dependent (or multiplicative) STDP, mathematically represented by

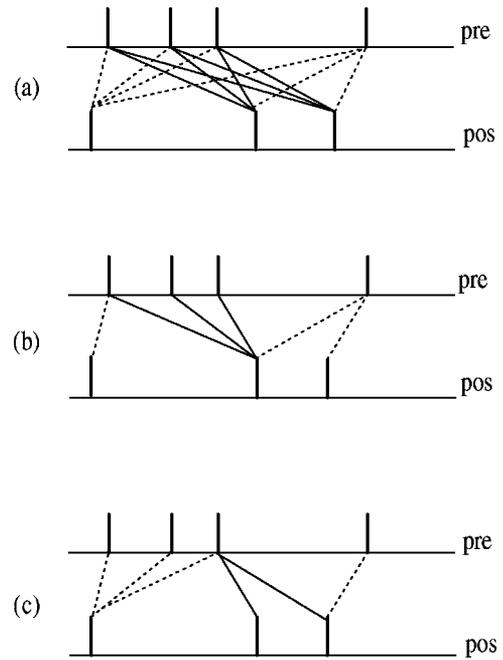


FIG. 1. Pairing rules. (a) All-to-all. (b) Nearest neighbor. (c) Latest neighbor. The solid line represents LTP pairs and the dotted lines are for LTD pairs.

$$\Delta g_1 = G(\Delta t|g) = \begin{cases} c_p e^{-\Delta t/\tau_p} (1-g), & \Delta t > 0, \\ -c_d g e^{\Delta t/\tau_d}, & \Delta t < 0, \end{cases} \quad (8)$$

where g is the normalized value with respect to a constant g_{max} ; c_p , c_d , τ_p , and τ_d are constants. The following parameter set is used in numerical simulations: $c_p = 0.001$, $c_d = 0.003$, and $\tau_p = \tau_d = 20$ ms.

Many pairing rules have been proposed to address the second question. Among them, all-to-all and nearest-neighbor pairing rules have been widely used. The all-to-all rule assumes every presynaptic spike interacts with every postsynaptic spike, as shown in Fig. 1(a). An alternative to the above simple scheme is a nearest-neighbor interaction, in which only the first presynaptic event after a given postsynaptic event can produce depression, and only the first postsynaptic spike after a given presynaptic event can produce potentiation, as shown in Fig. 1(b). There could also be other forms of pairing rules, such as postsynaptic-centric or presynaptic-centric rules. In this study, we propose a latest-neighbor pairing rule, which is motivated by a dynamical model of long-term synaptic plasticity.¹⁶ In particular, building on the current understanding of the molecular mechanism of synaptic plasticity, Abarbanel *et al.* proposed in Ref. 16 that synaptic plasticity is the result of interactions between two processes: one due to presynaptic activities and another due to postsynaptic activities. The prediction of this model has been shown to be in good agreement with many experiments. If we assume that the idea of having two processes is correct, and also assume the presynaptic (postsynaptic) process will be reset by any new presynaptic (postsynaptic) event, then at any time instant only the latest-neighbor pairs of pre- and postsynaptic spikes contribute to the plasticity, because the current states of the pre- and postsynaptic

processes are determined totally by these two spikes, respectively. This pairing rule is illustrated in Fig. 1(c).

Next, we will consider six possible implementations of STDP based on two possible $G(\Delta t|g)$ and three possible $P_i(\Delta t|g)$ due to the three pairing rules discussed above. The analysis of weight-dependent $G(\Delta t|g)$ with latest-neighbor pairing rule will be given in detail, while other implementations will be discussed briefly.

1. Weight-dependent STDP with latest-neighbor pairing rule

Two factors contribute to $P_i(\Delta t|g)$. One is the causal contribution due to the direct and indirect interactions between the presynaptic and postsynaptic neurons, which depend on the details of the network model. Another is noncausal due to the random coincidence of the two spike trains. For the latest-neighbor pairing rule, the noncausal factor can be found as follows. Suppose the interspike intervals (ISIs) of pre- and postsynaptic spike trains are independent of each other and follow the distribution $\Phi_i(T)$ and $\Phi_o(T)$, respectively. As shown in Fig. 1(c), all pairs of pre- and postsynaptic spikes that lead to LTP can be found by looking at each interspike interval (ISI) in the presynaptic spike train: the pairs connecting the leading presynaptic spike with each of the postsynaptic spikes within the ISI are the LTP pairs. Similarly, all pairs of pre- and postsynaptic spikes that lead to LTD can be located by examining each ISI in the postsynaptic spike train, i.e., the pairs connecting the leading postsynaptic spike with each of the presynaptic spikes within the ISI are the LTD pairs. Thus, to find the probability of observing a latest-neighbor pair with positive Δt , we first consider an ISI with duration T in the presynaptic spike train. Because the pre- and postsynaptic spikes are not correlated (since we are examining the noncausal factor), all pairs with positive $\Delta t < T$ have equal probability that is proportional to the product of the pre- and postsynaptic mean firing rates. Pairs with $\Delta t > T$ will have zero probability. Thus, we have

$$\text{prob}(\Delta t|0 < \Delta t < T) = \lambda_i \lambda_o, \quad (9)$$

$$\text{and } \text{prob}(\Delta t|\Delta t \geq T) = 0,$$

where λ_i and λ_o are the mean firing rates of pre- and postsynaptic neurons, respectively. Averaging over all possible ISIs in the presynaptic spike train yields

$$\begin{aligned} \text{prob}(\Delta t|\Delta t > 0) &= \int_0^{\Delta t} 0 \cdot \Phi_i(T) dT + \int_{\Delta t}^{\infty} \lambda_i \lambda_o \Phi_i(T) dT \\ &= \lambda_i \lambda_o \int_{\Delta t}^{\infty} \Phi_i(T) dT. \end{aligned} \quad (10)$$

The probability for the case of negative Δt can be obtained similarly,

$$\text{prob}(\Delta t|\Delta t < 0) = \lambda_i \lambda_o \int_{|\Delta t|}^{\infty} \Phi_o(T) dT. \quad (11)$$

Figure 2 illustrates this result for the case of Poisson spike trains.

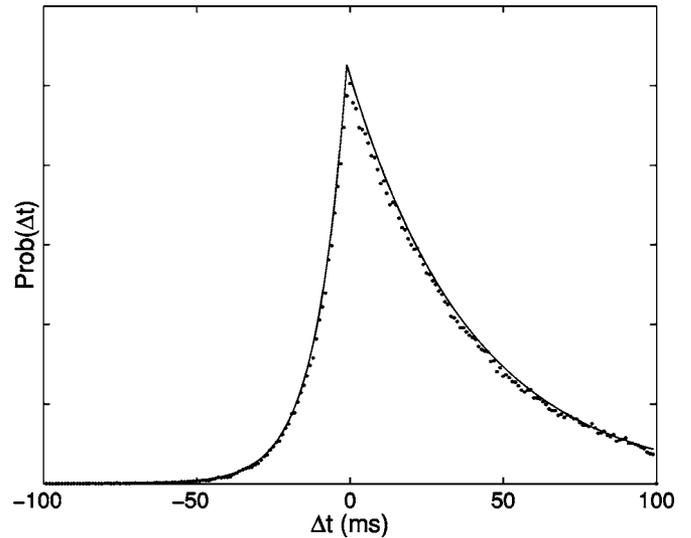


FIG. 2. Probability $\text{prob}(\Delta t)$. Points are obtained from simulation with two Poisson spike trains ($\lambda_i=25$ Hz, $\lambda_o=100$ Hz). The solid curve is the theoretical prediction.

For STDP with latest-neighbor pairing rule, we thus have

$$P_i(\Delta t|g) \approx C(\Delta t|g) + \lambda_i \lambda_o \int_{|\Delta t|}^{\infty} \Phi(T) dT, \quad (12)$$

where $\Phi(T) = \Phi_{i(o)}(T)$ if $\Delta t > 0 (< 0)$; the term $C(\Delta t|g)$ represents the causal factor, and the second term the noncausal.

If both $C(\Delta t|g)$ and $\Phi(T)$ are known, we could solve for $P(g, t)$ from Eq. (3). However, $C(\Delta t|g)$ and $\Phi(T)$ also depend on the weight distribution $P(g, t)$. Given an initial weight distribution $P(g, t=0)$, synaptic weights in a large network may evolve in a complicated way and, hence, Eq. (3) has to be solved *self-consistently*. To gain insight into the problem, we assume, with time-invariant external inputs to the network, the random process can achieve an asymptotic steady state, at which $C(\Delta t|g)$ and $\Phi(T)$ for each network neuron are invariant (Assumption I). Furthermore, we assume spike trains are Poisson (Assumption II),

$$\Phi(T) = \lambda e^{-\lambda T}, \quad (13)$$

where λ is the firing rate of the pre- or the postsynaptic neuron. To obtain a simple form for $C(\Delta t|g)$, we also assume there is only one-way connection and no other indirect connections (e.g., common inputs) between two neurons exist (Assumption III). We will relax these three assumptions and discuss the corresponding consequences.

Under Assumption III, we have $C(\Delta t|g)=0$ if $\Delta t < 0$, since the pre- and postsynaptic events are not correlated, and $C(\Delta t|g) \neq 0$ if $\Delta t > 0$, representing the effect that a presynaptic spike always enhances the probability of postsynaptic firing. Although the shape of $C(\Delta t|g)$ depends on the details of the system model, in general, $C(\Delta t|g)$ is small and can be assumed to be proportional to g and λ_i .¹⁰ The integration of $C(\Delta t|g)$ inside the STDP learning window is

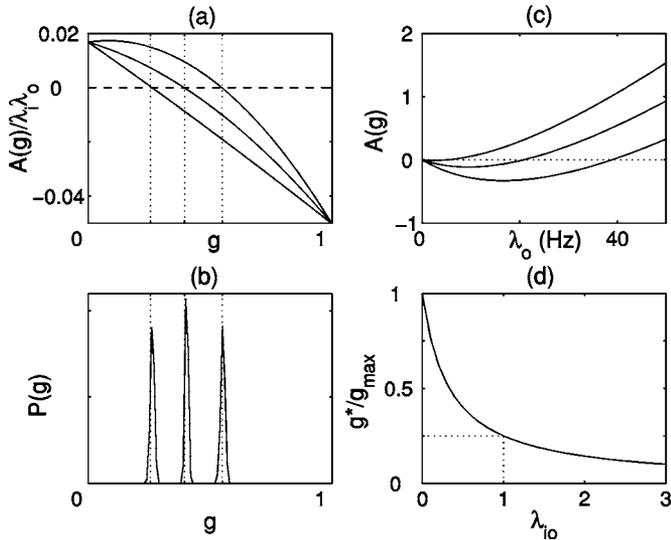


FIG. 3. (a) $A_1(g)$ for $\alpha=0.02, 0.2, 0.4$, respectively (from left to right). Input and output firing rates are 5 Hz. (b) Final weight distribution $P(g)$ due to $A_1(g)$ in (a) estimated from 200 runs. (c) Qualitative consistency with BCM. The drift term $A_1(g)$, or the expected value of Δg , depends on output firing rate λ_o . Three curves correspond to different initial synaptic weights (from top to bottom, $g=0.25, 0.3, 0.35$). (d) The dependence of g^* on $\lambda_{io}=(\lambda_i + 1/\tau_p)/(\lambda_o + 1/\tau_d)$.

$$\int C(\Delta t|g)e^{\Delta t/\tau} dt \approx \lambda_i \alpha g, \quad (14)$$

where $\tau=\tau_p=\tau_d$, and α is a small positive constant proportional to the correlation between the input and the output spike trains, or equivalently, to the correlation among all inputs. This is reasonable considering that dozens of coincident presynaptic spikes are needed to evoke a postsynaptic firing and more correlated inputs result in higher probability of postsynaptic firing and hence higher correlation between pre- and postsynaptic activities. These considerations lead to

$$\frac{A_1(g)}{\lambda_i \lambda_o} = c_p(1-g) \left(\frac{1}{\lambda_i + 1/\tau_p} + \frac{\alpha g}{\lambda_o} \right) - \frac{c_d g}{\lambda_o + 1/\tau_d}, \quad (15)$$

where the first term on the right-hand side of the equation represents the strength of LTP, and the second the LTD.

The dependence of $A_1(g)$ on input correlation is illustrated in Fig. 3(a) for three values of α . Let g^* be the point where LTP and LTD are balanced, i.e., $A_1(g^*)=0$. We see that the position of g^* moves to the right as α increases; so does the final weight distribution [Fig. 3(b)]. This indicates that the correlation between inputs can be encoded into the synaptic strength due to the STDP learning rule, similar to other STDP implementations.^{8,9} If the constraint on connections is removed, e.g., network neurons can have recurrent connections and/or common inputs, $C(\Delta t|g)$ will become more complicated. For example, $C(\Delta t|g) \neq 0$ even if $\Delta t < 0$. As a consequence, the integration of $C(\Delta t|g)$ inside the STDP learning window can be positive or negative for different neurons in the network, which results in different LTP/LTD-balancing positions for different synapses. Especially for the case of two neurons with recurrent connections, competition between the two synapses can be induced. As we will show

later in a simulation example, this effect becomes more and more evident as the correlation among external inputs increases.

To find the dependence of $A(g)$ on input and output firing rates, we assume $\alpha=0$ in Eq. (15) and obtain

$$g^* = \left[1 + \frac{c_d(\lambda_i + 1/\tau_p)}{c_p(\lambda_o + 1/\tau_d)} \right]^{-1}, \quad (16)$$

which suggests that synaptic weights are also sensitive to input and output firing rates, as shown in Fig. 3(d). If the distribution of neuron firing rate is broad, the weight distribution in the network must be broad, too [unlike the distribution in Fig. 3(b)]. Although the diversity in cross correlations between pre- and postsynaptic neurons can broaden the weight distribution too, it is not likely to be the major cause because neurons in reality are only weakly correlated. The diversity of the firing rate is partially caused by induced competition between neurons, which can be understood as follows. If the firing rate of the postsynaptic neuron (excitatory) increases, for example, g will increase according to Eq. (16) and at the same time local inhibition will be strengthened due to the increased excitation to the local inhibitory neurons. The increased local inhibition will lower the firing rates of local excitatory neurons. The lost excitation of postsynaptic neuron will be less than that of the presynaptic neuron because of increased g , causing the ratio λ_i/λ_o to decrease and g to increase further. As a consequence of this chain reaction, one neuron becomes less active because another become more active, naturally introducing neural competition. It is also important to note that the STDP with latest-neighbor pairing rule is compatible¹³ with rate-based Hebbian learning rules. For example, the BCM rule² has the form

$$\tau_g \frac{dg}{dt} = \lambda_i \lambda_o (\lambda_o - \theta), \quad (17)$$

where τ_g is a time constant, and θ is a variable threshold used to induce and stabilize LTP/LTD. In general, it is assumed that θ depends on the activity of the postsynaptic neuron and increases faster than the postsynaptic firing rate to prevent g from having unrealistically large values. To compare STDP with BCM, we can rewrite Eq. (15) as

$$\frac{dg}{dt} = \lambda_i \lambda_o F(\lambda_o, \theta), \quad (18)$$

$$\theta = \frac{c_d g}{c_p(1-g)} (\lambda_i + 1/\tau_p) - 1/\tau_d,$$

where $F(x, y) > 0$ for $x > y$, and $F(x, y) < 0$ for $x < y$. If λ_i is fixed, the threshold θ depends on g . As shown in Fig. 3(c), θ increases with g . Because increasing g can cause only a slight increase in λ_o , θ will increase much faster than λ_o . Note that the constraint g_{\max} and the term $(1-g)$ in Eq. (8) are not essential for weight-dependent STDP, but they ensure a faster convergence. It can thus be seen that this learning rule is able to induce LTP/LTD and stabilize synaptic modifications as BCM, but with higher temporal accuracy. Although spike trains in reality are not perfectly Poissonian

(e.g., having refractory phase), the above general conclusions should still hold qualitatively.

Assumption I in our model is that an asymptotic steady state can be achieved under invariant external inputs to the network. It is not clear how good this assumption is. Thus, we have carried out numerical simulations using a random network model, with results in good agreement with our theoretical analysis. The simulation results will be shown in Sec. III.

2. $A_1(g)$ from alternative implementations of STDP

In the following, the correlation between pre- and postsynaptic neurons is assumed to be zero since it is small in reality. Similar to the derivation of Eq. (10), the probability that pre- and postsynaptic spikes are separated by Δt for the all-to-all rule is

$$\text{prob}(\Delta t) = \lambda_i \lambda_o. \quad (19)$$

For the nearest-neighbor rule, the probability is the reverse of the one for the latest neighbor,

$$\begin{aligned} \text{prob}(\Delta t | \Delta t > 0) &= \lambda_i \lambda_o \int_{|\Delta t|}^{\infty} \Phi_o(T) dT, \\ \text{prob}(\Delta t | \Delta t < 0) &= \lambda_i \lambda_o \int_{|\Delta t|}^{\infty} \Phi_i(T) dT. \end{aligned} \quad (20)$$

For weight-independent STDP paired with the all-to-all rule, we have

$$\frac{A_1(g)}{\lambda_i \lambda_o} = c_p \tau_p - c_d \tau_d, \quad (21)$$

which implies that all synapses are either depressed to zero or potentiated to a maximum value depending on the sign of $c_p \tau_p - c_d \tau_d$. This appears unrealistic.

For weight-independent STDP paired with the nearest-neighbor rule, we have

$$\frac{A_1(g)}{\lambda_i \lambda_o} = \frac{c_p}{\lambda_o + 1/\tau_p} - \frac{c_d}{\lambda_i + 1/\tau_d}, \quad (22)$$

which also implies an unrealistic bimodal distribution.

For weight-independent STDP paired with the latest-neighbor rule, we obtain

$$\frac{A_1(g)}{\lambda_i \lambda_o} = \frac{c_p}{\lambda_i + 1/\tau_p} - \frac{c_d}{\lambda_o + 1/\tau_d}, \quad (23)$$

which also implies an unrealistic bimodal distribution.

For weight-dependent STDP paired with the all-to-all rule, we find

$$g^* = \left[1 + \frac{c_d \tau_p}{c_p \tau_d} \right]^{-1}, \quad (24)$$

which does not depend on the firing rates. So, this implementation is not compatible with rate-based Hebbian rules such as the BCM rule.

For weight-dependent STDP paired with the nearest-neighbor rule, we have

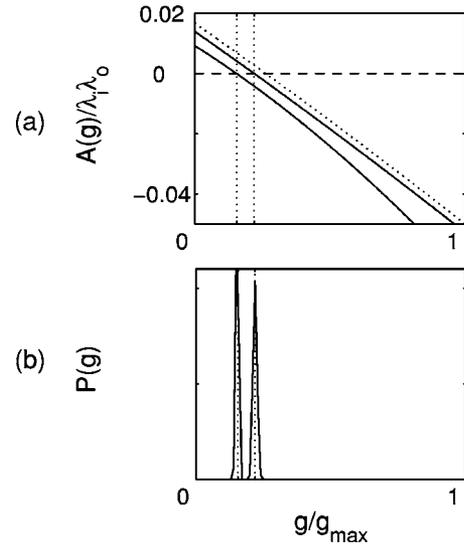


FIG. 4. Influence of HSP on $A(g)$. (a) The drift term $A(g)$ for $\lambda_i = \lambda_o = 1$ Hz, 5 Hz, respectively. $\alpha = 0.02$, $\tau_{\text{HSP}} = 1000$ s, $g_{\text{goal}}/N = 0.15$. The dotted curve is $A_1(g)$ (STDP) with $\lambda_{i,o} = 5$ Hz. (b) $P(g)$ due to $A(g)$ in (a) is estimated from 200 runs.

$$g^* = \left[1 + \frac{c_d(\lambda_o + 1/\tau_p)}{c_p(\lambda_i + 1/\tau_d)} \right]^{-1}, \quad (25)$$

which implies that the larger λ_o is, the smaller g^* . This contradicts the BCM rule.

There can be many other possible implementations of STDP. Here, we only wish to emphasize that any meaningful implementation should be consistent with existing knowledge. It can be seen that some implementations, such as ones discussed above, are not realistic. In the rest of this paper, only weight-dependent STDP with latest-neighbor pairing rule will be considered.

C. $A_2(g)$ and $A(g)$

The drift force caused by HSP can be written as

$$A_2(g) = \frac{1}{\tau_{\text{HSP}}} \left[\frac{g_{\text{goal}}}{N} - \int_{-\infty}^{\infty} g P(g, t = \infty) dg \right], \quad (26)$$

where N is the total number of input synapses. We see that $A_2(g)$ does not depend on g in the asymptotic state since the second term on the right-hand side is the mean, which does not depend on g . For simplicity, we write $A_2 \equiv A_2(g)$.

Under the three assumptions made in Sec. II B 1, $A_1(g)$ is given by Eq. (15). Finally, we have

$$\frac{A(g)}{\lambda_i \lambda_o} = c_p (1 - g) \left(\frac{1}{\lambda_i + \tau_p} + \frac{\alpha g}{\lambda_o} \right) - \frac{c_d g}{\lambda_o + \tau_d} + \frac{A_2}{\lambda_i \lambda_o}. \quad (27)$$

Comparing Eqs. (15) and (27), we see that $A(g)$ is approximately a shifted version of $A_1(g)$, as shown in Fig. 4(a). The amplitude of this shift is proportional to $A_2/\lambda_i \lambda_o$. If A_2 is fixed and $\lambda_i \lambda_o$ is large, i.e., sensory inputs are activated, the influence of HSP can be negligible. If $\lambda_i \lambda_o$ is small, i.e., the neuron undergoes spontaneous activities, the influence of HSP cannot be neglected and the mean of the weight distribution is shifted.

Another important feature that HSP introduces is the competition between inputs. Suppose there are two groups of inputs: one is strong and another is weak. The relative strength may be induced by different firing rates, the frequencies of activation, correlations, and so on. According to the relative strength of their drift forces produced by STDP, the additional drift force produced by HSP will be allocated to the two groups. That is, the two groups of inputs have to compete for the total available weight or the control of the neuron. Thus, we see that HSP naturally introduces competition among neurons.

III. NUMERICAL SIMULATIONS

Our model network consists of 240 excitatory:60 inhibitory neurons. The excitatory:inhibitory ratio is set to be 4:1, representing the statistics of neurons in local neural networks of brain. The network receives inputs from 100 excitatory neurons that generate Poisson spike trains. The probability that a neuron is connected with another through a synapse is 0.3, which is a little higher than values obtained by many experimental studies. For example, it was shown in Ref. 17 that the connection rate among thick tufted layer 5 neurons in rat visual cortex is about 11.6%. Although the connection rate of 0.3 might be higher than reality, the equivalent connection rate after learning should be much lower because many synapses during the learning process could be largely depressed. Each network neuron is modeled by the following set of ordinary differential equations:¹⁸

$$\begin{aligned} \frac{dV}{dt} &= -(17.81 + 47.58V + 33.8V^2)(V - 0.48) \\ &\quad - 26R(V + 0.95) + I_s + I, \\ \frac{dR}{dt} &= (1/\tau_R)[-R + 1.29V + 0.79 + 3.3(V + 0.38)^2], \end{aligned} \quad (28)$$

where V is the membrane potential, R is the recovery variable with time constant $\tau_R=5.6$ ms for excitatory neurons and 2.1 ms for inhibitory neurons, and current I_s is due to the synaptic inputs from neurons modeled in the network. The background current I to each neuron is modeled as synaptic current due to uncorrelated Poissonian spike trains. In simulation, I results in a firing rate of about 1 Hz (7 Hz) for excitatory (inhibitory) neurons when $I_s=0$. These equations are a simplified version of the Hodgkin-Huxley equations for mammalian cortical neurons, being able to produce a good approximation to spike shapes, firing rates, and bursting behavior throughout the physiological range. Each synapse is modeled by the following set of ordinary differential equations:¹⁸

$$\frac{df}{dt} = (1/\tau_{\text{syn}})[-f + \text{Hvs}(V_{\text{pre}} - \Omega)], \quad (29)$$

$$\frac{dS}{dt} = (1/\tau_{\text{syn}})(-S + f),$$

where f is an intermediate variable for the synaptic potential S , τ_{syn} is the time delay of the synapse, $\text{Hvs}(x)$ is the Heaviside step function,

$$\text{Hvs}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases} \quad (30)$$

V_{pre} is the membrane potential of the presynaptic neuron, and Ω is the threshold for synaptic conductance change. The current into the postsynaptic neuron is $I_s = -gS(V - E_{\text{syn}})$, where E_{syn} is the synaptic reversal potential, g controls the synaptic conductance, and V is the postsynaptic membrane potential. In our simulations, $g_{\text{max}}=1.0$ (i.e., 1000 pS). For excitatory synapses, we set $\tau_{\text{syn}}=2.0$ ms, $\Omega=-0.3$ (i.e., -30 mV), and $E_{\text{syn}}=0$. For inhibitory synapses, we use $\tau_{\text{syn}}=0.5$ ms, $\Omega=-0.4$ (i.e., -40 mV), and $E_{\text{syn}}=-0.75$ (i.e., -75 mV). More details about the parameter selection can be found in Ref. 18. Using this parameter set of neurons and synapses, the total current into a postsynaptic neuron due to one excitatory presynaptic action potential (AP) is 1.5 times of the one due to inhibitory AP. The networks are in random state with averaged firing rate for excitatory neurons of 1 Hz and inhibitory ones of 7 Hz when the input neurons are silent; $E-E$ (excitatory-excitatory) synapses are initialized by a uniform distribution between 0 and g_{max} , all other types of synapses are g_{max} and are not plastic.

While there are well-established learning rules for $E-E$ synapses, learning rules involving inhibitory neurons have not been well established. Thus, in this study, only $E-E$ synapses are plastic; all others are assumed to be constant. In reality, STDP and HSP operate on different time scales. To make simulation feasible, however, we assume a smaller time scale, $\tau_{\text{HSP}}=10$ s, for HSP in Eq. (2).

A. Self-organization

Although sensory experience is important to refining neural circuits, spontaneous activities are believed to be critical to the circuit formation during early development.¹⁹ So, we first study how STDP learning rules sculpt the neural network from spontaneous activities. In this simulation, $E-E$ synapses are initialized by a uniform distribution between 0 and g_{max} , as shown in Fig. 5(a). All other types of synapses are g_{max} . Then, the synapses between excitatory network neurons are allowed to undergo self-organization due to the spontaneous activities while input neurons are kept silent. The firing rates of network neurons become approximately invariant after 60 s. Simulation runs until 1000 s, at which steady state is assumed to be achieved. As shown in Fig. 5(b), the final distribution of network synaptic conductances has a peak at $g=0.25$, which is consistent with the prediction on the single postsynaptic neuron model by the Fokker-Planck theory with zero correlation [Eq. (16)]. This can be understood since network neurons are weakly

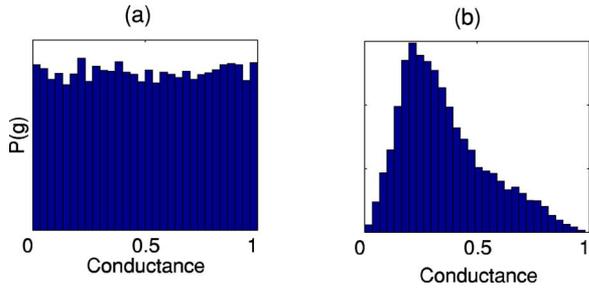


FIG. 5. Self-organization of network excitatory synapses through STDP. (a) The initial distribution of $E-E$ synaptic conductance. (b) The final distribution of $E-E$ synaptic conductance with a peak at 0.25.

correlated and inputs to each neuron should be Poissonian during spontaneous activities. This result is quite similar to the one based on the single postsynaptic neuron model,⁸ in which multiplicative noise is added in the STDP learning rules to mimic experimental observations.

B. Correlation-based learning, dependence on firing rate, and neural competition

In this simulation, $E-E$ synapses are initialized by a uniform distribution between 0 and g_{max} , and all other types of synapses are g_{max} . After the first 100 s, during which the network undergoes spontaneous activities with input neurons at silence, the input neurons are activated at 30 Hz. Firing rates of network neurons become approximately invariant after 1000 s. The simulation runs until 2000 s, at which we assume the learning process achieves the asymptotic steady state.

Figure 6(a) shows the firing rate distribution of the excitatory network neurons with independent inputs, or $C=0$. Roughly speaking, network neurons can be divided into two groups according to firing rates, higher or lower than 1 Hz. The total weights of afferent synapses on neurons from different groups are also different, corresponding to the two groups shown in Fig. 6(b). This is the result of competition between neurons. The dependence of g^* on firing rate is shown in Fig. 7 (dots). A closer examination shows that the data points, obtained from neurons with firing rate higher than 1 Hz, appear close to the theoretical prediction [Eq. (16)] (not shown in this figure). The small deviations are mainly caused by the fact that the ISI distributions are not

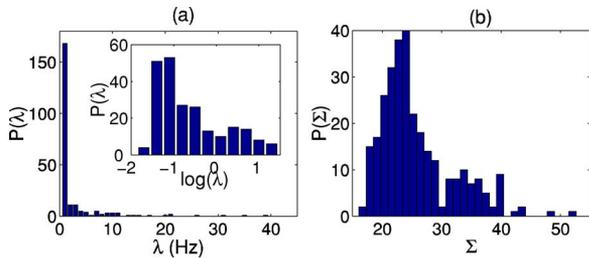


FIG. 6. Simulation 2. Distributions of firing rate and summation of weights. (a) Histogram of firing rates of excitatory network neurons. The inset plot is on a semilogarithmic scale. (b) Histogram of Σ , the total $E-E$ synaptic weights converging on one excitatory network neuron. In simulation, inputs are independent with firing rate 30 Hz.

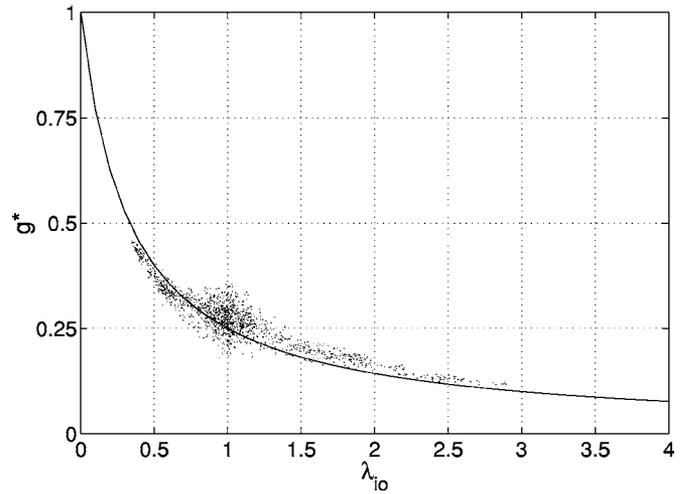


FIG. 7. Dependence of g^* on firing rate. The solid curve is given by Eq. (16) with $\lambda_{io}=(\lambda_i+1/\tau_p)/(\lambda_o+1/\tau_d)$. The dots are obtained in a numerical simulation.

perfectly exponential (with refractory period), while Eq. (16) assumes they are perfectly exponential. Figure 8(a) shows the results of weight distributions for uncorrelated inputs. The broad weight distributions are due to the broad firing rate distribution. The two peaks in the weight distribution of network synapses signify weak competition between network synapses. This may not be surprising if one realizes the existence of reciprocal connections between a pair of neurons. If the forward synapse is strengthened (or weakened), the backward synapse must be weakened (or strengthened) at the same time since changes in both synapses are induced by the same pair of spikes. Repeating such experiences can finally result in different balancing positions for the reciprocal connections. That is, the strength ratio of LTP and LTD in Eq. (15) is different for the pair of synapses, resulting in different values for g^* . Figures 8(b) and 8(c) show the results for inputs with increasing correlations. Apparently, the mean of input conductances is proportional to the correlation between inputs, as predicted by the theory [Eq. (15)]. For network synapses, although the shapes of their distributions are

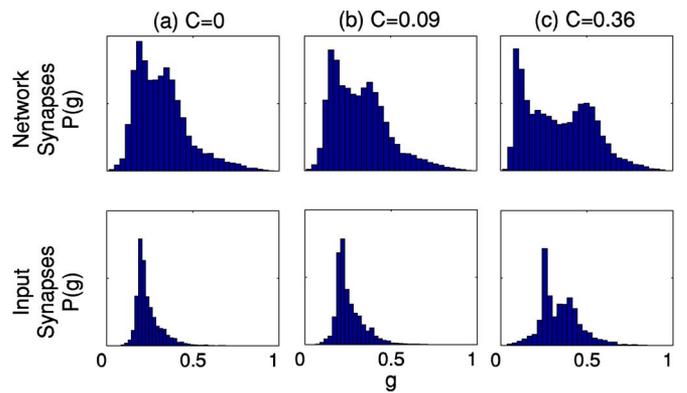


FIG. 8. Learning input correlations. Duration of the simulation is 2000 s. The firing rate of inputs is 30 Hz. (a) Correlation among inputs is $C=0$. Upper panel: the final distribution of $E-E$ network synaptic conductance. Lower panel: the final distribution of $E-E$ input synaptic conductance. (b) $C=0.09$. (c) $C=0.36$.

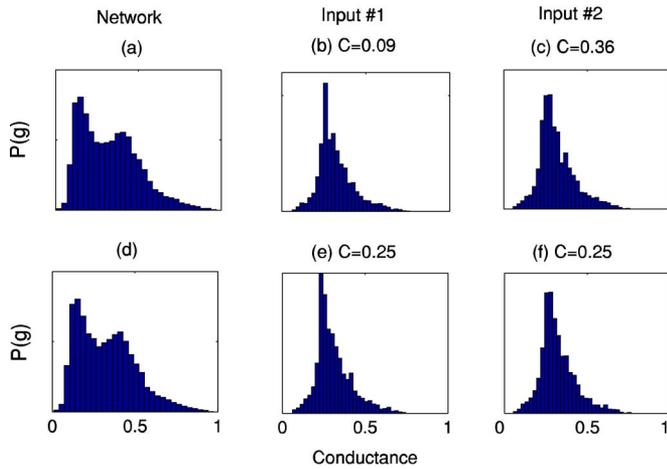


FIG. 9. Cooperative learning. Upper panels correspond to the case where there are two independent groups of external inputs with correlations $C = 0.09$ and $C = 0.36$, respectively. Lower panels are for the case where two independent groups of external inputs have $C = 0.25$ and $C = 0.25$, respectively. They have the same firing rate of 30 Hz. (a) The final distribution of $E-E$ network synaptic conductance. The final distribution of $E-E$ input synaptic conductance in group 1 (b) and group 2 (c).

slightly different from the one due to spontaneous activities, it can still be seen that the mean is proportional to the input correlation. The distance between the two peaks also increases as input correlation increases as a result of increased competition.

C. Cooperative learning

Synapses are initialized by using distributions obtained in simulation 1. Duration of the simulation is 120 s. Figure 9 shows the result for the cases where there are two independent input groups. In the first case, the cross correlation is $c = 0.09$ for group 1 and $c = 0.36$ for group 2. For the second case, the cross correlation inside each group is the same, $c = 0.25$. The firing rate of input spike trains is 30 Hz. Comparing with Fig. 8, it can be seen that there is cooperative learning between the two groups. For example, without cooperative learning, the two groups in the first case (upper panels) should have different means. In particular, the mean value of group 1 in the first case would be smaller than that of the group in the second case since it has weaker correlation ($0.09 < 0.25$).

D. Synaptic competition

Synapses are initialized by using distributions obtained in simulation 1. In simulations, the inputs are divided into two groups. The first group is activated during the first 40 s, while the second is activated during the second 40 s. The simulation results for the network trained by STDP only are shown in the middle of Fig. 10. It can be seen that the means of the final distributions of two groups are the same, similar to the case in Fig. 9. The distribution of the first group is unchanged during the second 40 s because its firing rate is zero and STDP never updates the weights. This is not real-

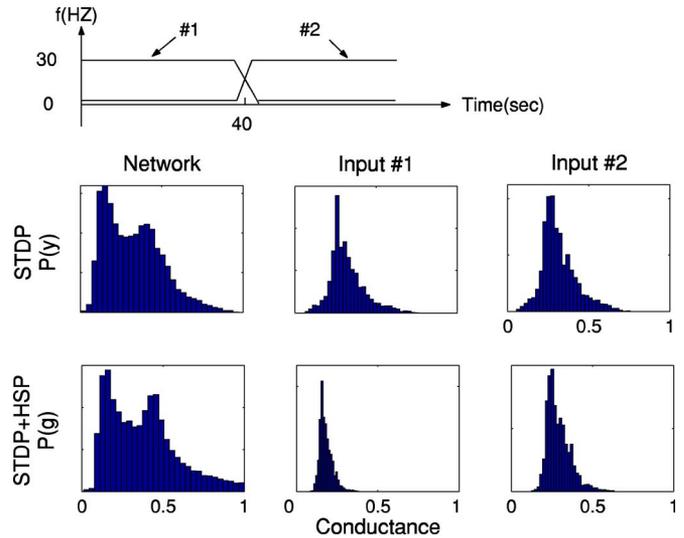


FIG. 10. Competition between inputs. Top: the network receives two groups of inputs. Group 1 is on during the first 40 s, while group 2 is on during the second 40 s. Upper three panels: final distributions using STDP only. Lower three panels: final distributions using both STDP and HSP.

istic since memory can be forgotten and new sensory experience can overwrite the old one. Thus, competitive learning is lacking in STDP.

The simulation results for the network trained by STDP and HSP are shown in the bottom of Fig. 10, where $g_{\text{goal}} = 30$ is used. We see that the shapes of the distributions are similar to the ones trained by STDP only. The only differences are the pronounced double peaks in network synapses and the suppression of the first group of input synapses, as expected due to the competition of limited total weights. Apparently, HSP enhances the competition between network and input synapses, regardless of whether they are activated or not. At the same time, HSP does not prevent STDP learning because of the different time scale of these two learning rules.

IV. DISCUSSION

Because of our limited understanding of the mechanisms for LTP/LTD, many possible implementations of STDP have been proposed. For example, STDP has been assumed to be weight dependent, weight independent, or a combination of both;²⁰ the integration of STDP may use all-to-all pairing, nearest-neighbor pairing, or latest-neighbor pairing, etc. Our theoretical and numerical studies at the network level reveal that weight-dependent STDP with the latest-neighbor pairing rule can generate stable and more realistic distributions of synaptic conductance, induce correlation-based learning and strong competition among network neurons, and is compatible with rate-based Hebbian rules.

The major difficulty in the current theoretical analysis using the Fokker-Planck theory lies in the fact that the exact form of causal factor $C(\Delta t|g)$ in Eq. (12) is unknown for recurrent networks. To overcome this difficulty, usually some restrictions have to be made upon the network models. For example, Refs. 8–10 use a nonrecurrent network model, which has a simple form of $C(\Delta t|g)$, similar to this paper.

Reference 21 studies a recurrent network in synchronized state, e.g., 3-cycle state, which also has a simple form of $C(\Delta t|g)$ with values at a few discrete positions. In this paper, we start from nonrecurrent networks and solve the Fokker-Planck equation in a way similar to Refs. 8–10. Based on this closed-form solution for nonrecurrent networks, we discuss, qualitatively, how recurrent connections would change the form of $C(\Delta t|g)$ and weight distributions, and then provide supporting evidence from numerical simulations.

The closed-form solution for nonrecurrent networks also helps us in studying the consequences of different STDP pairing rules and how compatible they are with existing knowledge, similar to Ref. 13. Here, it should be noted that the nearest-neighbor rule suggested in Ref. 13 is different from the latest-neighbor rule proposed in this paper (and also different from the nearest-neighbor rule discussed in this paper). The nearest-neighbor rule in Ref. 13 assumes there is a calcium saturation in postsynaptic neuron, so the first succeeding postsynaptic spike overrides the effect of subsequent spikes. For example, there is no second LTP pair in Fig. 1(c) according to this rule. If this saturation does exist, then the question is when the dynamics of a postsynaptic calcium channel recovers from the saturation. Implicitly assumed by the rule of Ref. 13, it recovers suddenly and completely right after the next presynaptic spike. In fact, as suggested in Ref. 22, saturations exist in both pre- and postsynaptic parts and the dynamics recover from saturations exponentially with different time constants. So, in this sense, the latest-neighbor pairing rule with exponentially recovered saturations at both pre- and postsynaptic sides is more realistic. As argued in Ref. 13, on the other hand, the exponentially recovered saturation has little effect on the properties of plasticity rules. So, for simplicity, we have proposed a plain latest-neighbor pairing rule for STDP in this paper. Further discussion about the STDP pairing rules can be found in the literature, e.g., Refs. 23–25.

As a homosynaptic rule, STDP assumes that events happening in one synapse will not influence directly the other synapses targeting on the same postsynaptic neuron. This homosynaptic nature results in the lack of competitive learning between synapses converging on one neuron. In situations where strong competition is needed, e.g., developing selectivity, some heterosynaptic mechanism is necessary.⁷ HSP operates on a time scale different from STDP, which allows the coexistence of the activity-induced learning and competitive learning. Our results demonstrate that combining STDP and HSP can produce a biophysically plausible learning rule that better characterizes the learning mechanism of the brain.

Apparently, the plasticity model studied in this paper is simplified and idealized. In reality, there are not only much more complicated heterosynaptic interactions at work,^{7,26} but also complicated homosynaptic dynamics. For example, STDP may require multiple pairings and have complicated pairing rules, take place with delays, and take different properties at different locations on the dendritic tree.²⁷ Furthermore, the details of STDP largely depend on the system being studied. While some of these assumptions made in this study will change the weight distributions, such as pairing

rules, some will not. For example, an action delay of STDP much shorter than the mean interspike intervals would not change the learning process. Even when they are comparable, its effect should be only a delay on the learning process and have little influence on the asymptotic weight distributions. It will be interesting to investigate the effect of network structure on weight distributions.

ACKNOWLEDGMENTS

This work is sponsored by DARPA under Grant No. MDA972-00-1-0027. Y.C.L. was also supported by the NSF under Grant No. ITR-0312131. We also want to thank the anonymous reviewers for their comments and helpful suggestions.

¹D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Wiley, New York, 1949).

²E. L. Bienenstock, L. N. Copper, and P. W. Munro, *J. Neurosci.* **2**, 1854 (1982).

³A. Artola, S. Bröcher, and W. Singer, *Nature (London)* **347**, 69 (1990); A. Artola and W. Singer, *Trends Neurosci.* **16**, 480 (1993); A. Ngezahayo, M. Schachner and A. Artola, *J. Neurosci.* **20**, 2451 (2000).

⁴H. Markram, J. Lubke, M. Frotscher, and B. Sakmann, *Science* **275**, 213 (1997); L. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, and M.-M. Poo, *Nature (London)* **395**, 37 (1998); G. Q. Bi and M. M. Poo, *J. Neurosci.* **18**, 10464 (1998); D. Debanne, B. H. Gähwiler, and S. M. Thomson, *J. Physiol. (London)* **507**, 237 (1998); D. E. Feldman, *Neuron* **27**, 45 (2000); P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson, *ibid.* **32**, 1149 (2001).

⁵G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B. Nelson, *Nature (London)* **391**, 892 (1998).

⁶L. Zhu, Y.-C. Lai, F. C. Hoppensteadt, and J. He, *Neural Comput.* **15**, 2359 (2003).

⁷S. Royer and D. Paré, *Nature (London)* **422**, 518 (2003).

⁸M. C. W. van Rossum, G. Q. Bi, and G. G. Turrigiano, *J. Neurosci.* **20**, 8812 (2000).

⁹J. Rubin, D. Lee, and H. Sompolinsky, *Phys. Rev. Lett.* **86**, 364 (2001).

¹⁰H. Câteau and T. Fukai, *Neural Comput.* **15**, 597 (2003).

¹¹L. F. Abbott and K. I. Blum, *Cereb. Cortex* **6**, 406 (1996); W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, *Nature (London)* **384**, 76 (1996); R. Kempter, W. Gerstner, and J. L. van Hemmen, *Phys. Rev. E* **59**, 4498 (1999); S. Song, K. D. Miller, L. F. Abbott, *Nat. Neurosci.* **3**, 919 (2000).

¹²R. C. Froemke and Y. Dan, *Nature (London)* **416**, 433 (2002); G.-Q. Bi, *Biol. Cybern.* **87**, 319 (2002).

¹³E. M. Izhikevich and N. S. Desai, *Neural Comput.* **15**, 1511 (2003).

¹⁴Z. Nusser, S. Cull-Candy, and M. Farrant, *Neuron* **19**, 397 (1997).

¹⁵R. J. O'Brien, S. Kamboj, M. D. Ehlers, K. R. Rosen, G. D. Fischbach, and R. L. Huganir, *Neuron* **21**, 1067 (1998).

¹⁶H. D. I. Abarbanel, R. Huerta, and M. I. Rabinovich, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10132 (2002).

¹⁷S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, *PLoS Biol.* **3**, e68 (2005).

¹⁸H. R. Wilson, *J. Theor. Biol.* **200**, 375 (1999).

¹⁹L. C. Katz and C. J. Shatz, *Science* **274**, 1133 (1996).

²⁰R. Güttig, R. Aharonov, S. Rotter, and H. Sompolinsky, *J. Neurosci.* **23**, 3697 (2003).

²¹N. Levy, D. Horn, I. Meilijson, and E. Ruppín, *Neural Networks* **14**, 815 (2001).

²²R. C. Froemke and Y. Dan, *Nature (London)* **416**, 433 (2002).

²³P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson, *Neuron* **32**, 1149 (2001).

²⁴H. X. Wang, R. C. Gerkin, D. W. Nauen, and G.-Q. Bi, *Nat. Neurosci.* **8**, 187 (2005).

²⁵G.-Q. Bi and J. Rubin, *TINS* **28**, 222 (2005).

²⁶G.-Q. Bi, *Biol. Cybern.* **87**, 319 (2002); Y. Humeau, H. Shaban, S. Bisière, and A. Luthi, *Nature (London)* **426**, 841 (2003).

²⁷A. Saudargiene, B. Porr, and F. Wörgötter, *Neural Comput.* **16**, 595 (2004).