Learning-induced synchronization of a globally coupled excitable map system

Yoshinori Hayakawa and Yasuji Sawada
Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan
(Received 7 September 1999)

We propose a pulse-coupled neural network model in which one-dimensional excitable maps connected in a time-delayed network serve as the neural processing units. Although the individual processing unit has simple dynamical properties, the network exhibits collective chaos in the active states. Introducing a Hebbian learning algorithm for synaptic connections enhances the synchronization of excitation timing of the units within a subpopulation. The synchronizing clusters approximately exhibit a power-law size distribution, suggesting a hierarchy of synchronization. After applying a stationary signal to a subpopulation of the units with learning, the network then reproduces the signal. The learnable time range is much longer than the inherent time scale of the processing units, i.e., the synaptic delay time. Also, the network can reproduce periodic signals with time resolution finer than the delay time. Our present network model can be considered as a temporal association device which operates in chaotic states.

PACS number(s): 05.45.Xt, 87.18.Sn

I. INTRODUCTION

Recent studies aimed at understanding higher brain mechanisms have been more concerned with the dynamical aspects of collective behavior of nerve systems. Several experiments suggest that the timing of neuronal impulses may play a crucial role in function and information representation in several regions of the mammalian brain [1–4]. To understand temporal coding processes in the brain by means of temporal coding, we need to understand several fundamental properties of pulse-coupled networks with a large number of neuronal processing elements. Several models of neurons have demonstrated that partial or global synchronization of biological oscillators requires both inhibitory coupling along with transmission delay [5]. On the other hand, while learning induced by interactions with the environment must lead to self-organized spiking [6], we do not understand how synaptic plasticity affects spiking patterns. Furthermore, global reorganization of spike timing caused by input stimuli is indispensable for ''binding'' of information [7].

One of the essential properties of spiking-neurons is their excitability, the threshold gating of emitted pulses. To represent this excitability requires at least two degrees of freedom using ordinary differential equations (ODEs) such as the Fitzhugh-Nagumo equations. On the other hand, a one-dimensional map with a continuous variable can exhibit excitable, oscillatory, and chaotic dynamics with a few parameters.

Since the subject of our present study is the robust properties of collective neuronal dynamics which are less dependent on details of neurons, we do not take biological models of action potential for specific nerve cells, such as the Hodgkin-Huxley equations; a one-dimensional map suffices for the representation of abstract neurons with excitable dynamics. Furthermore, simulating the Hodgkin-Huxley equations and the ODEs derived from them is time-consuming [8]. For both numerical and theoretical analysis, a simpler and easier to handle form would aid a large-scale simulation.

The integrated-and-fire (IF) models would be another representation of excitable spiking neurons in simplified forms. To represent the excitability and refractoriness, however, the IF model requires internal variables, complicating the IR model compared to ours, in which a single element follows a purely one-dimensional dynamics. Therefore, we take a one-dimensional map with excitable dynamics as a model neural processing unit.

In this paper, we propose a coupled one-dimensional map connected in an all to all network as an abstract neuronal network with spiking neurons. After constructing the phase diagram of the global activity of the network for random synaptic connection, we show that a Hebbian local learning rule can induce synchronization of spiking and a synchronized cluster self-organizes into a characteristic cluster size distribution. We also show that incoming stimuli can control very long period bursting of spikes with resolution time finer than the element's characteristic time scale.

II. MODEL

We employ a simplified globally coupled network of excitable neural processing units. To represent neuronal activity $x(t)$ at a discrete time t in a simple manner, we use a one-dimensional iterative map $x(t+1)=f(x(t))$ with the following piecewise linear form [9]:

$$f(x)=\begin{cases} 0 & \text{for } x<-1 \\ ax & \text{for } -1\leqslant x<0.2 \\ bx-0.1 & \text{for } 0.2\leqslant x<0.85 \\ c(x-1) & \text{for } 0.85\leqslant x \end{cases} \tag{1}$$

where $0<a$, $c<1$, $1<b$ are constants (Fig. 1). These parameters determine the shape of the pulse. In typical numerical simulations, we choose $a=0.5$, $b=1.5$, and $c=0.04$. For typical neurons, one iterative step in this model corresponds to milliseconds of time. The absolute magnitude of $x(t)$ is arbitrary.

To avoid numerical instability or divergence in large-scale simulations, we assume that x resets to zero for $x<-1$, which happens when the magnitude of inhibition is

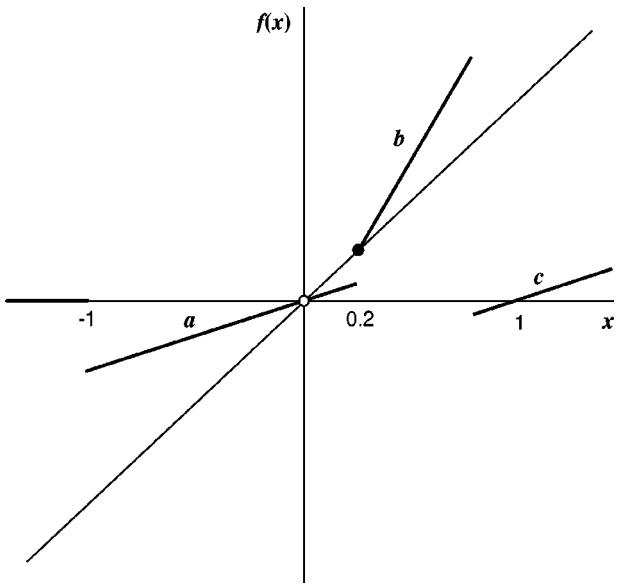

FIG. 1. Piecewise-linear one-dimensional map $f(x)$. The bullet is the unstable fixed point which gives the threshold for excitation, and the circle is the stable fixed point. a, b, and c are the slopes in each segment. Typically, $a=0.5$, $b=1.5$, and $c=0.04$.

too large. However, this resetting is not essential to our results. In most cases, the operation point of x remains always above -1.

This map has a stable fixed point at $x_s=0$ and an unstable one at $x_u=0.2$. If the value of x exceeds x_u, it will ''fire'' a single pulse and enter the resting state of the negative domain of x. Otherwise, x approaches x_s monotonically. The weak negative excitation after a single pulse is analogous to the refractory state in the Hodgkin-Huxley model. Thus while it neglects some biological details, our model qualitatively reproduces the behavior of the Hodgkin-Huxley equations.

We consider N excitable maps connected by transmission lines with delay times τ and synaptic strength w_{ij} from element j to i. The whole dynamics of the system then is

$$x_i(t+1)=f\left(x_i(t)+\sum_j^N w_{ij}x_j(t-\tau)+y_i(t)\right), \qquad (2)$$

where i and j are the indices of the elements, and $y_i(t)$ is an external bias to element i. All elements $(i=1,\ldots,N)$ update synchronously following Eq. (2). For simplicity, we assume a constant, uniform delay τ. We can simulate the dynamics of a large-scale network much faster than the conventional Hodgkin-Huxley model or its reduced forms due to the simplified update rule.

In numerical simulations, we first prepare a set of $\{x_i(t)\}$ for $-\tau\leqslant t\leqslant 0$ using Gaussian distributed random numbers with a mean of 0.2 and a variance of $(0.2)^2$ as initial conditions. This choice of initial values ensures the initial excitation of a large fraction of elements. Otherwise, e.g., starting in the neighborhood of x_s, the whole network eventually enters a resting state. We set the synaptic weight matrix using Gaussian distributed random numbers.

III. CONDITIONS FOR SPONTANEOUS EXCITATIONS

In this section, we consider global network activity in terms of the statistical neuro-mechanics formulated by Amari [10]. We define the activity of the network at time t as

$$X(t)=\sum_{i=1}^N x_i(t). \qquad (3)$$

Since each element is excitable, the states $x_i(t)$ approximate binary valued elements, that is, in the resting state $x_i(t)\simeq 0$ and in the excited state, $x_i(t)\simeq 1$. We assume the transient time between these two states is negligibly small.

In a mean field description of the activity of a neural network with random synaptic weights, if the number of elements is large enough, inputs to a neuron behave as a sum of independent stochastic variables. When the activity is $X(t-\tau)$ at time $t-\tau$, the inputs Ω for an element are

$$\Omega=\sum_j^N w_{ij}x_j(t-\tau)\simeq \sum_k^{X(t-\tau)} \omega_k, \qquad (4)$$

where ω is the stochastic variable which represents the distribution of weights w_{ij}. Here we limit our consideration to the zero-bias case. For simplicity, we assume that the weights have a Gaussian distribution of mean μ and variance σ^2. Thus, Ω also has a Gaussian distribution of mean μX and variance $\sigma^2 X$. We also assume uncorrelated successive values of x_i.

If the slope a near the stable fixed point is small, the time scale around x_s becomes very short, so $x_i(t)\simeq x_s$ at all time for nonfiring elements. The probability for a nonfiring element to be driven beyond the threshold $x^*(=x_u)$ is

$$P_f(X)=\int_{x^*}^{\infty} G(x;\mu X,\sigma\sqrt{X})dx, \qquad (5)$$

where $G(x;\mu,\sigma)$ is a Gaussian probability density function of mean μ and variance σ^2.

$P_f(X)$ maps activity of $X(t-\tau)$ to $X(t)$ as

$$X(t)=NP_f(X(t-\tau)), \qquad (6)$$

where $N\gg 1$ is the total number of elements. Here we assume that the delay parameter, τ, is larger than the spiking period of the map and that the correlation between two successive firing events is negligible. A stable fixed point in Eq. (6) at $0<X^*<N$ is necessary for sustained spontaneous firing of a finite population of the elements.

This property of $P_f(X)$ yields a saddle-node bifurcation line in (μ,σ) space for a finite value of X, i.e., the transition is first order. Numerically, we can evaluate the activity X for any μ and σ. Figure 2 shows typical outputs from a single element for a few σ at fixed μ $(=0)$. For weak coupling, i.e., small σ, excitation induced by the initial $x_i(t\leqslant 0)$ does not last, while elements activate spontaneously above a critical value of σ.

Figure 3 shows the phase diagram obtained by direct numerical simulation of Eqs. (1) and (2) with $N=40$ and the theoretical transition line obtained from Eq. (6). Small a reduces the memory effects of the element, and the transition line obtained from Eq. (6) agrees with the numerical simu-

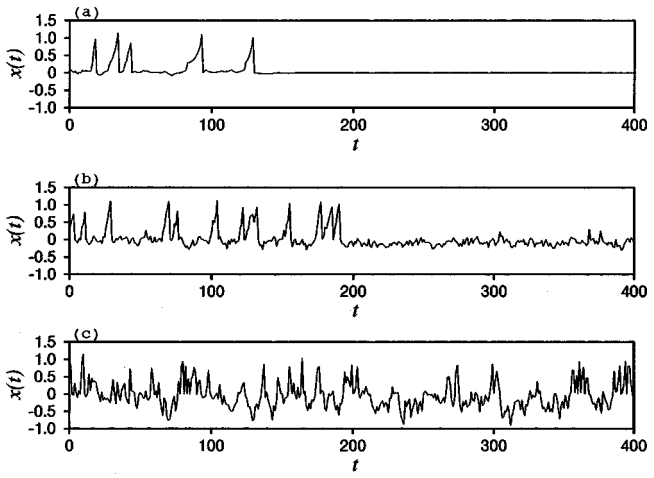

FIG. 2. Behavior of an element in a fully connected network with random synaptic weights. The standard deviations of the synaptic weights are (a) $\sigma=0.05$, (b) $\sigma=0.1$, and (c) $\sigma=0.2$, with average $\mu=0$. The number of elements $N=40$. Other conditions are described in the text.

lations. To account for relaxation near x_s in future work, we will estimate the probability density function of $x_i(t)$ quantitatively. In the following sections, we choose the initial distribution of synaptic weights near the transition line, to enhance the network's sensitivity to changes of learned weights.

IV. DEGREE OF SYNCHRONIZATION

Measuring the activity of elements disregards the correlation of firing events. To measure the mutual correlation among a large population of elements, we need to determine the degree of synchronization in a simple but general form. The correlation between the signals of elements i and j is

$$c_{ij}=\langle[x_i(t)-\langle x_i(t)\rangle][x_j(t)-\langle x_j(t)\rangle]\rangle, \tag{7}$$

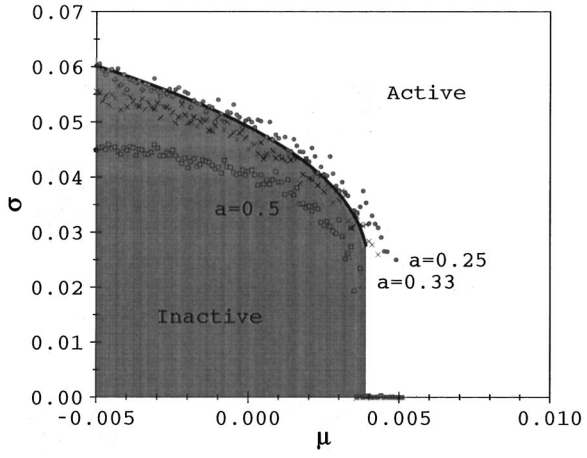

FIG. 3. Phase diagram for random synaptic weights in the parameter space (μ,σ) for $N=100$. Solid curve is the transition line to global activity derived in the text [Eq. (6)]. In the gray region, the network is inactive for the entire initial condition. Circles show the transition points obtained in simulations with $a=0.25$. ×'s represent $a=0.33$, and squares $a=0.5$.

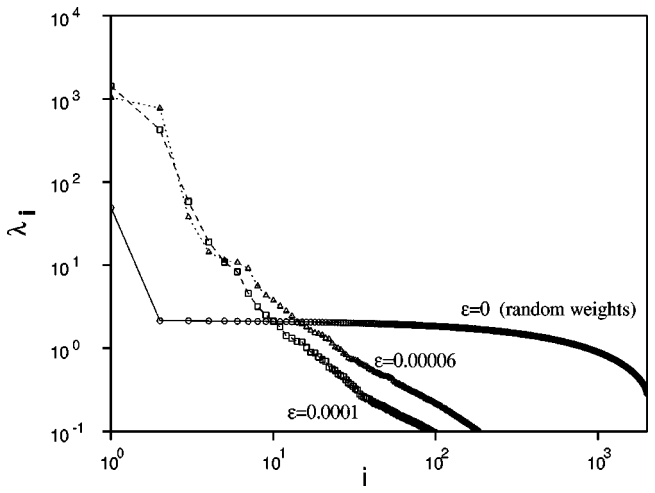

FIG. 4. Distribution of eigenvalues for the covariance matrix R, in descending order as a function of rank, i. $N=2000$. For nonzero learning coefficient ϵ, large eigenvalues emerge: White circles connected by solid lines, $\epsilon=0$; white triangles with dotted lines, $\epsilon=0.000\,06$; and white rectangles with dashed lines, $\epsilon=0.0001$. For all synaptic connections $\tau=11$.

where angular brackets denote a time average. Normalizing the pair correlations,

$$r_{ij}=\frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}, \tag{8}$$

yields the covariance matrix $R=\{r_{ij}\}$. From the definition, $r_{ij}=1$ if $x_i(t)$ changes synchronously with $x_j(t)$, and $r_{ij}=0$ in the absence of correlation.

Since R decomposes into the form $R=YY^T$ using a matrix Y (see the Appendix), R is positive definite so every eigenvalue of R is real and positive. By normalization, all diagonal elements $r_{ii}=1$. Therefore, summation of the eigenvalues $\Sigma\lambda_i$ gives the size of the matrix, i.e., the number of elements. If m correlated elements have no temporal correlation with other groups, the covariance matrix R has an independent block of size m, with corresponding eigenvalues m (once) and zero $(m-1$ times). The eigenvector corresponding to m has m nonzero components, which indicate the locations (indices) of the elements belonging to the group.

Thus the eigenvalues gives the cluster size distribution for mutual synchronization over long periods. We sort the eigenvalues in descending order, and plot them versus their indices in an *eigenspectrum* of synchronization. In Fig. 4 ($\epsilon=0$), we show an example of the eigenspectrum for a randomly connected network with $N=2000$.

When $\{x_i(t)\}$ can be represented by a Gaussian stochastic variable which is independent of any signal $\{x_j(t')\}$ for $j\neq i$ and at $t'\neq t$, we expect the maximum eigenvalue of R to be order 1 as described in the Appendix. Thus the probability that a large cluster of order N appears could be negligible (almost zero probability) in the large N limit, if the $\{x_i(t)\}$ are uncorrelated. In other words, a large eigenvalue in the eigenspectrum indicates that a nontrivial collective excitation mode exists.

V. REINFORCEMENT OF SYNCHRONIZATION

We now introduce a learning algorithm into our model. We divide the neural processing units into a number of subsets $\{S_k | k=1,\ldots,M \ll N\}$, and impose a time-dependent bias signal $\{y_l(t)\}$ on the *visible* units belonging to S_l $(l=1,\ldots,K<M)$, separately. We call the subsets without the external bias, i.e., $y_l(t)=0$, *hidden units*.

We set the internal states of every neural unit $\{x_i(t)\}$ to be random at first, and also set the synaptic weights to be Gaussian random values characterized by (μ_0,σ_0). While the external signals are present, we reinforce the synaptic weights by a Hebbian rule with time delay τ. The amount of reinforcement from element j to i is

$$\delta w_{ij}(t)=\epsilon x_i(t)x_j(t-\tau), \tag{9}$$

where ϵ is a small positive number (the *learning coefficient*). At each step, we update all the synaptic weight synchronously:

$$w_{ij}(t+1)=w_{ij}(t)+\delta w_{ij}(t). \tag{10}$$

Since the state variables allow both signs, the learned w_{ij} can be either excitatory or inhibitory. Due to the finite delay τ, learning will yield asymmetric weight matrices $\{w_{ij}\neq w_{ji}\}$ in general.

Starting from randomly weighted connections, without learning, i.e., in the case $\epsilon=0$, the time series $\{x_i(t)\}$ (see Fig. 2) has no apparent temporal structure. Accordingly, the power spectrum of $\{x_i(t)\}$ is white. Changing the system size, N, from 100 to 4000, the largest eigenvalue $\lambda_1(N)$ of the matrix R increases more slowing than N as shown in Fig. 5(a), indicating that the time series of randomly connected elements without learning behave as independent stochastic events.

Without external signals, all elements are hidden with nonzero learning coefficient ϵ; the timing of spikes self-organizes starting from random Gaussian synaptic weights. Consequently, in late stages of reinforcement given by Eq. (9), the eigenspectrum will reflect the emergence of large clusters. We expect that larger learning coefficients lead to larger size synchronized clusters.

Figure 5(b) shows the dependence of the largest cluster size on the learning coefficient, after 4000 fixed learning steps. We find a critical value of ϵ at which a system-wide cluster emerges. Above the transition, the largest clusters contain almost the same population (around 60–80 %) independent of ϵ. Below the transition, reinforcement of weights is ineffective. Near or above the transition point, the eigenspectrum decays approximately as a power law $\lambda_k \sim k^{-\beta}$ for large eigenvalues, where $\beta\sim 2$ (Fig. 4). This scaling holds within our computational limit $(N<4000)$, suggesting the emergence of a hierarchy of synchronization clusters during learning in very large-scale networks. Most elements belongs to the largest cluster because of large β.

In a recent experiment on the cultured network of cortical neurons of a rat, clustering of neuronal activity is observed using an optical measurement of intracellular Ca^{2+} concentration [11]. By a similar quantitative analysis of synchronization for the neurons, Sano *et al.* found the emergence of synchronizing neurons after several days of network growth.

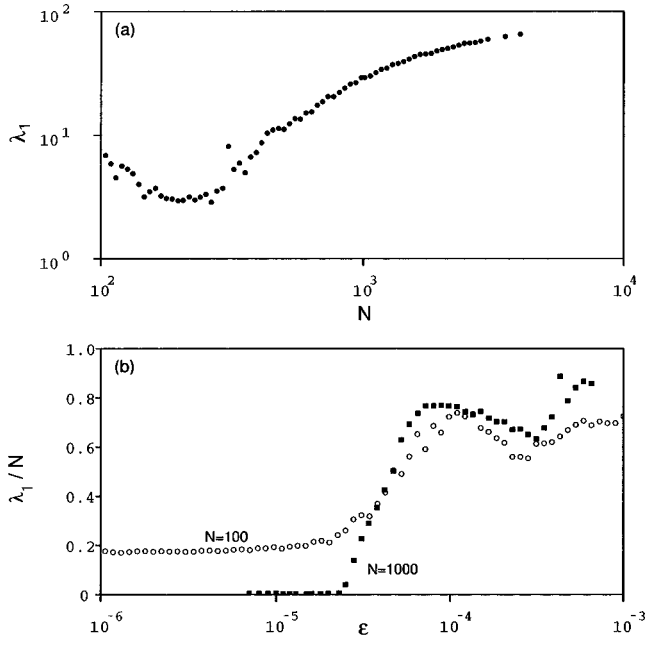

FIG. 5. Maximum eigenvalues, λ_1, for the covariance matrix R. (a) System size dependency. Synaptic weights are random Gaussian random numbers ($\mu=0$, $\sigma=0.1$), with $\tau=11$. The matrix R is calculated from 10 000 steps of numerical simulation at each system size N. (b) Learning parameter ϵ dependency for $N=100$ (circles) and $N=1000$ (squares). Maximum eigenvalues are normalized by N.

In their experiment, there are many synchronizing clusters among the neurons with different cell population. Unfortunately, statistical properties such as the cluster size distribution are not clear so far for the cultured neurons. We may be able to make quantitative comparison to the *in vitro* experiment in future work.

VI. TEMPORAL LEARNING OF A STATIONARY SIGNAL

Learning causes elements to form synchronously spiking clusters, but the spiking is irregular in time. What happens if we apply external signals to a subpopulation of elements? We choose $\frac{1}{4}$ of the elements to be visible.

The external signal is a sine wave of period T_i:

$$y_v(t)=A\sin(2\pi t/T_i), \tag{11}$$

where v represents the indices of all visible elements. The amplitude A is order 1, typically 0.5. Our numerical simulations employ T_i's ranging from unity to over ten times the synaptic delay τ.

We set the initial synaptic weights and internal states to randomly distributed Gaussian as was done before and present the external signal during the first N_L time steps. We call this period the *learning phase*. In the *recall phase*, we turn off the external signal and set the learning coefficient ϵ zero. The network continues to produce the time series autonomously with learned weights. For typical simulations, $N_L=4000$ and $\epsilon=10^{-4}$. We analyze about 10^5 points of data during the recall phase.

During learning, there is visible alternate bursting and resting with period T_i due to the bias signal $y_v(t)$. As the

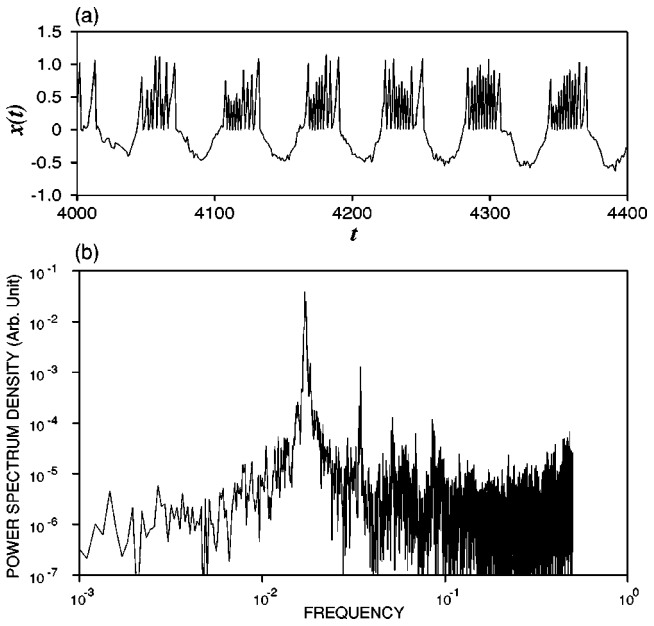

FIG. 6. (a) Periodic bursting of a visible element after learning a sinusoidal external signal, and (b) power spectrum of the collective activity of the visible elements. $N=40$, and 10 elements experienced a periodic sinusoidal bias of period $T_i=50$ during the learning phase of 4000 steps with learning parameter $\epsilon=0.0002$. Thus, 80 cycles of the signal were presented to the network. For initial synaptic weights, $\mu=0$, $\sigma=0.04$, and $\tau=11$.

reorganization of w_{ij} proceeds, this bursting and resting cycle gradually appears in the hidden units as well, while the phase and/or frequency of the cycle may differ from that of the visible elements.

Figure 6(a) shows that, during the recall phase, if the learning parameters are appropriate, the excitation of elements persists even without the external bias signal. Remarkable aspects of the time series, which cannot be observed without the external signal $y_v(t)$, include (i) autonomous quasiperiodic bursting excitation, (ii) very long and accurate intervals between bursts, and (iii) self-organization of synchronized clusters having different bursting times in the hidden elements. Furthermore, the burst interval T_o found in the visible elements is close to that of the external signal imposed during learning.

We evaluated the periodicity of bursting by Fourier analysis of the time series. The power spectrum shows the bursting interval T_o as a lower frequency peak [see Fig. 6(b)]. In Fig. 7, we plot T_o as a function of T_i. In this case, the delay parameter τ, which is the longest time scale explicitly assumed at the element level, is 11. The period of the reproduced signal is much longer than τ, ranging from 10 to 100 of τ, so many units must contribute to the generation of the long-term coherence; excitation pathways must be long enough to maintain the coherence over a period $T_o \gg \tau$.

At shorter time scales, T_i and T_o are clearly proportional with slope close to unity. Therefore, the network can reproduce the timing of stimuli without special design. The resolution of the timing of reproduced signals is finer than the delay time τ which we expect to be the ''quantum'' of timing. In other words, the whole network dynamics produces the timing, not just the delayed transmission of pulses.

The reproduction of stimuli is robust over a wide range of

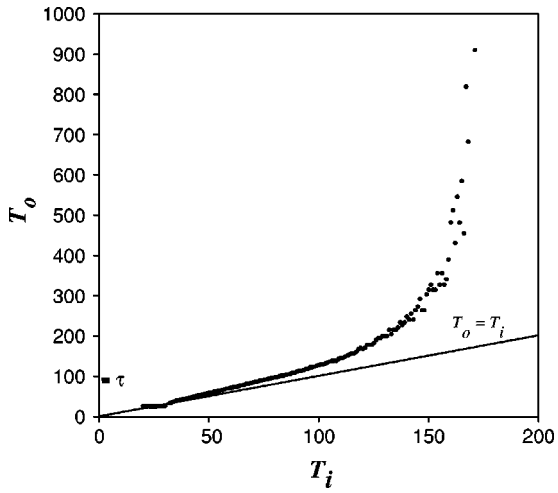

FIG. 7. Period T_o of learning-induced bursting as a function of the period T_i of the sinusoidal bias. $N=40$. Ten visible elements are subject to the sinusoidal bias signal, the rest are hidden elements. The plot shows the status of one visible element after 4000 learning steps with initial synaptic weights Gaussian random ($\mu=0$, $\sigma=0.04$) and delay time $\tau=11$. The periodicity of the output signal is evaluated from the peak of the power spectrum for 8192 data points taken during recall.

parameters ϵ and N_L, and the bursting period seems independent of those parameters. With the pulse rate representation used in conventional neural network models, synaptic weights directly affect the characteristic time scale of neuronal dynamics. On the contrary, with pulse timing coding, the magnitude of the weights affects less the temporal behavior of the network.

VII. COLLECTIVE CHAOS

We use the maximum Lyapunov exponent for the network to characterize the degree of temporal disorder of the signals given by Eqs. (1) and (2) [12]. The time-delayed interactions prevent analytic calculation, so we first construct an $(N \times D)$-dimensional state vector,

$$\boldsymbol{s}(t)=(x_1(t),\ldots,x_N(t),\ x_1(t-1),\ldots,x_N(t-1),\ldots, x_1(t-D),\ldots,x_N(t-D)x), \tag{12}$$

where the integer $D>0$ is large enough that the results are independent of D. By solving Eq. (2) from an initial conditions at $\boldsymbol{s}_0(t)$, and from a perturbed one at $\boldsymbol{s}_1(t)=\boldsymbol{s}_0(t)+\boldsymbol{\delta}$, simultaneously, we estimate the maximum Lyapunov exponent Λ:

$$\Lambda=\lim_{T\to\infty}\frac{1}{T}\ln\left(\frac{|\boldsymbol{s}_1(t+T)-\boldsymbol{s}_0(t+T)|}{|\boldsymbol{\delta}|}\right). \tag{13}$$

In the following discussion, we use $T>10^4$ and $D>100$ to enforce convergence of Λ.

Consider the transition to spontaneous firing in the parameter space (μ,σ) (Fig. 3). Since we expect the transition to spontaneous activity to be first order, a finite fraction of the elements change their operational points beyond the unstable fixed point x_s at the transition. Thus the maximum Lyapunov

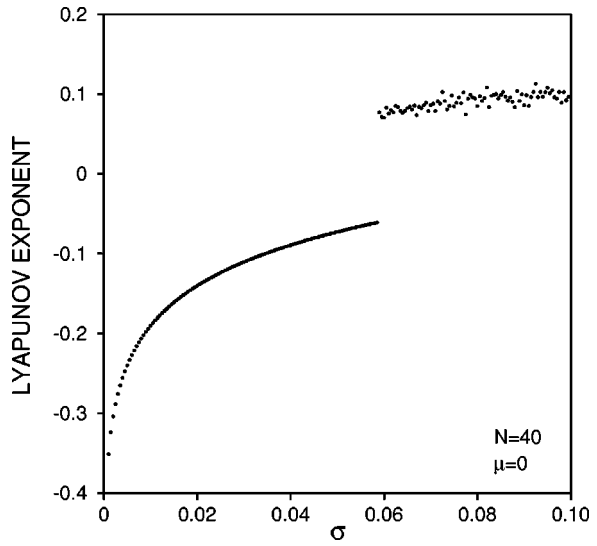

FIG. 8. Maximum Lyapunov exponent of the network with random synaptic weights as the standard deviation σ of the weight distribution. $\mu=0$. At the inactive-active transition, the maximum Lyapunov number becomes positive discontinuously. $N=40$.

exponent, which describes the most unstable mode, becomes positive discontinuously at the transition as seen in Fig. 8.

In the parameter regime of active states, the dynamics is deterministically chaotic due to the excitable elements. Even when recalling a learned periodic bursting signal, the maximum Lyapunov exponent remains positive. Accordingly, the power spectrum of the repetitive bursting has a broad peak around the bursting frequency [Fig. 6(b)]. Numerical simulations show no neutral states, thus we conjecture that almost everywhere in parameter space active states are chaotic.

VIII. DISCUSSION

Our present model and learning algorithm should be able to learn more complex time sequences. It can recall the absolute time length of input signals without any prior knowledge of the time scale due to the almost linear response of the bursting period as shown in Fig. 7. On the contrary, conventional associative memories can barely control the timing of recalled patterns, i.e., the time scale of the neuronal elements often determines the transition time between successive patterns during association [13,14].

As an example of the learning of more complex signals, Fig. 9 shows the learning and recalling for three different periodic signals. For stable learning and recall, we slightly modified our asymmetric reinforcement rule:

$$\delta w_{ij}=\epsilon(x_i(t)-\hat{x})\theta(x_j(t-\tau)-\hat{x}), \tag{14}$$

where $\hat{x}$ is a constant and $\theta(x)$ is the Heaviside function. In Eq. (14), we assume that only output signals larger than $\hat{x}$ affect the synaptic weights, i.e., only the positive pulses transmit mutual information which may be closer to real synapses than Eq. (9). The qualitative behavior of the network under Eq. (14) seems to be the same as that under Eq. (9), while avoiding spurious pulse transmission during the resting states by choosing $\hat{x}\simeq x_u$, which makes numerical behavior

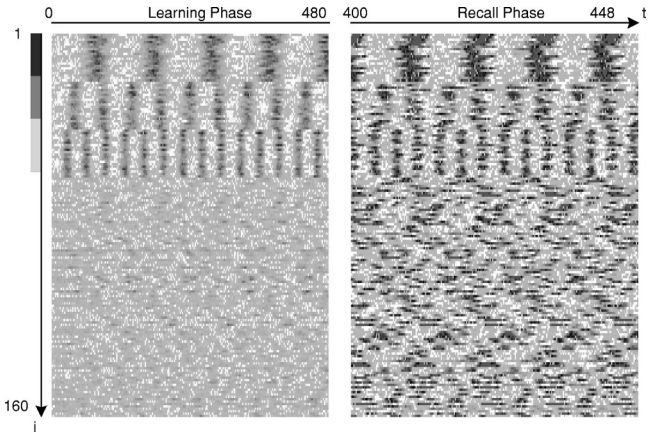

FIG. 9. Activity of the elements during learning of three sinusoidal signals of different periods. Light shade shows positive internal states, and dark shadow negative. Only the first 480 steps of learning and recall are plotted. $N=160$ with 60 visible elements. From the top of the plot, the first 20 elements have sinusoidal bias of a period of 100 during learning (black bar), the second 20 have a period of 50 (gray bar), and the third 20 have a period of 100/3 (light gray bar), respectively. The synaptic delay τ is randomly distributed with mean 20 and variance 10. Learning algorithm follows Eq. (14), $\epsilon=0.0001$ and $\hat{x}=0.2$.

more stable in larger networks. Figure 9 shows that the network can recall more than one signal simultaneously with correct relative phases.

More complex or nonstationary signals seem not to be reproduced as effectively as simple periodic ones. For instance, complex but periodic signals tend to be recalled as a simpler periodic bursting with a typical time period of the original. On the other hand, when we give short-time nonstationary signals to our network with learning, unexpected periodic bursting often takes place in the recall phase. However, we have no data on the network's efficiency or capacity for temporal association for general patterns.

Although the dynamics of the present model is deterministic, timing of the resultant signals is quite irregular due to chaos. The chaotic dynamics of the network may help generate timing among different neurons; despite the lack of randomness, synapses experience varied combinations of input pulse timings, and they can reinforce desirable pulses by Hebbian learning. Thus, hidden elements are driving learning as a reservoir of timing. Without hidden elements, learning of long period signals fails in simulations. During the learning of stationary signals, selection of timing occurs even in the hidden elements, and finally periodic bursting is also seen in hidden elements (see Fig. 9 in its recall phase). In this context, the scaling property of synchronization (Sec. V) may play an important role to generate diversity of timing in the hidden units.

Contrary to the chaotic internal dynamics, after the learning of a periodic signal, bursting with a very precise interval occurs in the recall phase. This bursting seems to be quite stable against initial perturbations and the bursting interval holds finer resolution than the characteristic time scale of the units. In other words, the present network is able to acquire stable dynamics in long time scale using chaos by learning.

In our model, the strength of chaos relates directly to the slope b above the unstable fixed point in the piecewise-linear

map f. As long as N is large, at any moment multiple excited elements contribute $\ln|b|$ to the Lyapunov exponent. Thus, when the average activity of the network is $X>0$, from the definition, the maximum Lyapunov exponent is approximately

$$\Lambda \simeq \frac{X}{N}\ln|b|, \tag{15}$$

where we assume that the elements stay on the unstable branch of the map with probability X/N on average. Numerical simulations agree with Eq. (15). The slope b may affect the dynamical complexity of our model, but the dependence is beyond our present scope.

Finally, we compare our model to the integrate and fire (IF) model. In the IF model, the membrane potential V_i follows the equation.

$$\frac{dV_i}{dt} = -V_i + I_i, \tag{16}$$

where I_i is the synapse current into the neuron. When V_i exceeds a threshold value V^*, depolarization takes place producing a single spike. Then, V_i resets to a constant during a finite refractory period. Although the IF model seems simple, representing the excitability and refractoriness requires hidden variables, complicating the IR model compared to ours in which a single element follows a purely one-dimensional dynamics. Tuning the slopes b and c, we can adjust the shape and timing of the spike to resemble IF behavior. Using a nonlinear function g, such as $g(x)=\theta(x-\hat{x})$ as in Eq. (14), would allow only positive spikes to propagate, as assumed in the IF model.

ACKNOWLEDGMENTS

We would like to thank Dr. Masaki Sano and Yoshihiro Hayakawa for helpful discussions and comments. We also would like to thank Dr. Kazuyuki Aihara for helpful information and comments about related works. We are grateful to Dr. James A. Glazier for a careful reading of the manuscript and illuminating comments. This research is financed by the Sound Technology Promotion Foundation, Japan.

APPENDIX: MAXIMUM EIGENVALUE OF COVARIANCE MATRIX FOR RANDOM TIME SERIES

Suppose that every $x_i(t)$ is a Gaussian stochastic variable independent of $x_j(t')$ for $i\neq j$ and $t\neq t'$ $(1\leqslant i,j\leqslant N)$. Without loss of generality, we assume that the average of $\{x_i(t)\}$ is zero and its second moment is V^2.

The elements of the covariance matrix R then becomes

$$r_{ij} = \frac{\sum_{t=0}^{T} x_i(t)x_j(t)}{TV^2}, \tag{A1}$$

where T is the averaging time. Note that $x_i(t)$ and $x_j(t)$ are also mutually independent.

From the definition of R, we can always decompose R into $R=YY^T$, where $Y=\{y_{ij}\}$ is the $T\times N$ matrix with element $y_{ij}=x_i(j)/(T^{1/2}V)$. We assume that the size of the matrix N and the averaging time T are the same order, i.e., $T=N\gg 1$. Since Y and Y^T have the same eigenvalues and eigenvectors, the eigenvalues $\{\lambda_i\}$ of R can be represented as $\lambda_i=\xi_i^2$ using the eigenvalues $\{\xi_i\}$ of Y.

The density of eigenvalues of a random matrix obeys the ''semicircle law'' [15]. Thus we can expect that the eigenvalue density $z(\xi)$ of the random matrix Y is

$$z(\xi) = \begin{cases} \dfrac{N^{1/2}}{2\pi}(2-\xi^2)^{1/2}, & \xi^2<2, \\ 0, & \xi^2>2. \end{cases} \tag{A2}$$

Using Eq. (A2), we can obtain the eigenvalue density $\rho(\lambda)$ of R as

$$\rho(\lambda) = \frac{1}{2\lambda^{1/2}} z(\lambda^{1/2}) \tag{A3}$$

for $\lambda>0$. Therefore, the largest eigenvalue of R is

$$\lambda_1 = 2 \tag{A4}$$

in the limit $N(\sim T)\to\infty$.

[1] *Neural Representation of Temporal Patterns*, edited by E. Covey, H. L. Hawkins, and R. F. Port (Plenum, New York, 1995).
[2] C. M. Gray, P. König, A. K. Engel, and W. Singer, Nature (London) **338**, 334 (1989).
[3] M. Ahissar, M. Ahissar, and H. Bergman, J. Neurophysiol. **67**, 203 (1992).
[4] A. K. Engel, A. K. Kreiter, and P. Konig, Proc. Natl. Acad. Sci. USA **88**, 6048 (1991).
[5] U. Ernst, K. Pawelzik, and T. Geisel, Phys. Rev. E **57**, 2150 (1998); U. Ernst, K. Pawelzik, and T. Geisel, Phys. Rev. Lett. **74**, 1570 (1995).
[6] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[7] See, e.g., *Large-scale Neuronal Theories of the Brain*, edited by C. Koch and J. L. Davis (MIT Press, Cambridge, MA, 1994).
[8] H. C. Tuckwell, *Introduction to Theoretical Neurobiology Vol. 1 & 2* (Cambridge University Press, Cambridge, 1988).
[9] A similar one-dimensional map was proposed as a neuronal processing element by E. Labos, in *Chaos in Biological Systems*, edited by H. Degn, A. V. Holden, and L. F. Olsen (Plenum Press, New York, 1987), p. 195.
[10] S. Amari, Proc. IEEE **59**, 35 (1971).
[11] M. Sano and A. Yoshida (private communication).
[12] W. D. Dechert and R. Gencay, Physica D **90**, 40 (1996).
[13] H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861 (1986).
[14] M. H. Hassoun, *Fundamentals of Artificial Neural Network* (MIT Press, Cambridge, MA, 1995).
[15] M. L. Metha, *Random Matrices and the Statistical Theory of Energy Levels* (Academic Press, New York, 1967).