

## Sequential associative memory with nonuniformity of the layer sizes

Jun-nosuke Teramae\* and Tomoki Fukai†

Laboratory for Neural Circuit Theory, RIKEN Brain Science Institute, Hirosawa 2-1, Wako, Saitama 351-0198, Japan

(Received 17 May 2006; revised manuscript received 1 October 2006; published 11 January 2007)

Sequence retrieval has a fundamental importance in information processing by the brain, and has extensively been studied in neural network models. Most of the previous sequential associative memory embedded sequences of memory patterns have nearly equal sizes. It was recently shown that local cortical networks display many diverse yet repeatable precise temporal sequences of neuronal activities, termed “neuronal avalanches.” Interestingly, these avalanches displayed size and lifetime distributions that obey power laws. Inspired by these experimental findings, here we consider an associative memory model of binary neurons that stores sequences of memory patterns with highly variable sizes. Our analysis includes the case where the statistics of these size variations obey the above-mentioned power laws. We study the retrieval dynamics of such memory systems by analytically deriving the equations that govern the time evolution of macroscopic order parameters. We calculate the critical sequence length beyond which the network cannot retrieve memory sequences correctly. As an application of the analysis, we show how the present variability in sequential memory patterns degrades the power-law lifetime distribution of retrieved neural activities.

DOI: 10.1103/PhysRevE.75.011910

PACS number(s): 87.19.La, 84.35.+i, 87.18.Sn

### I. INTRODUCTION

Many models have been proposed for learning and generating sequences of neural activities. A well-studied class of such models includes associative memory models with asymmetric recurrent synaptic connections [1–4]. The mechanism studied in these models appears to play an active role for storing temporal sequences of visual memories in the inferotemporal cortex [5]. Another well-studied model of sequence generation is “synfire chain” that represents a feed-forward neural network to propagate synchronous spike packets [6,7]. This class of neural networks was originally proposed to explain precise spike sequences in monkey brains, and was recently revisited by other experimental studies [8,9]. From the viewpoint of circuit structure, the discrimination between the two model classes is not rigorous when recurrent connections are sparse in associative memory models.

Here, we consider an associative memory model of binary neurons that stores sequences of memory patterns with highly variable sizes. The motivation of this study comes from the recent experimental finding of “neuronal avalanches,” which revealed that cortical networks have the ability to produce a diverse repertoire of sequences of synchronous activities. Interestingly, the size and time length (i.e., lifetime) of these activities are distributed obeying power laws of exponents  $-1.5$  and  $-2$ , respectively [10,11]. These results suggest that neuronal avalanches may reflect a critical branching process in neural dynamics, since the process shows the same power laws as avalanches [12,13]. Alternatively, neural avalanches may reflect the existence of a diverse ensemble of cell assemblies rather than the critical state in neural dynamics. To obtain the observed power laws, the size and length of each subnetwork of these cell assem-

blies should obey the same power-law distributions as the size and lifetime of neuronal avalanches. We have proposed a stochastic neuronal wiring rule to embed such an ensemble of layered subnetworks in a large pool of excitatory and inhibitory neurons [14].

In this paper, we analyze a binary network analogy of this spiking-neuron network model to study the properties of sequence retrieval numerically and analytically. Recurrent or feed-forward neural networks for sequence memory recall have been studied extensively [15,16]. For instance, we may define the storage capacity as the upper bound for the total length of sequences beyond which the system cannot retrieve memory patterns [17–20]. Besides the storage capacity, we are particularly interested in how the nonuniform construction of memory sequences might affect the retrieval dynamics in this model. In fact, the evolution equations of the macroscopic order parameters become stochastic rather than deterministic due to the variability in memory patterns. Interestingly, the stochastic nature of the macroscopic dynamics induces an exponential cutoff in the power-law size distribution of retrieved sequences, with the cutoff position depending on the memory load. Namely, the variable memory sizes destabilize sequence retrieval. Throughout the paper, we use “length” and “lifetime” to refer to the physical length of each embedded sequence and the length of the actually retrieved one, respectively.

### II. ASSOCIATIVE MEMORY FOR SEQUENCE RECALL WITH HIGHLY VARIABLE PATTERN SIZES

Population of  $N$  identical neurons updated at every discrete time step are described as,

$$S_i(t+1) = F[h_i(t)] = F\left[\sum_{j=1}^N J_{ij}S_j(t)\right], \quad i = 1, \dots, N \quad (1)$$

where  $S_i$  and  $h_i$  are the state variable and internal potential of the  $i$ th neuron, respectively. We update all neurons synchro-

\*Electronic address: teramae@brain.riken.jp

†Electronic address: tfukai@brain.riken.jp

nously to mimic synchronous propagations of neural activities such as neuronal avalanches. For simplicity, we restrict response function  $F$  to a Heaviside function  $F(x)=\Theta(x-\theta)$  with threshold  $\theta$ :  $\Theta(x)=1$  if  $x>0$ , or 0 otherwise. We, however, can easily extend the results of our analysis to the case with a general class of nonlinear response functions. Matrix element  $J_{ij}$  represents coupling strength from neuron  $j$  to neuron  $i$ . Memory pattern  $\{\tilde{\xi}_i^\mu\}$  defines a set of neurons belonging to the  $\mu$ th layer or cell assembly of a memorized sequence:  $\tilde{\xi}_i^\mu=1$  if the layer includes neuron  $i$ , or  $\tilde{\xi}_i^\mu=0$  otherwise. In many previous models of associative memory, the value of  $\tilde{\xi}_i^\mu$  was determined according to the following rule:

$$\tilde{\xi}_i^\mu = \begin{cases} 0, & \text{with Prob } 1-a \\ 1, & \text{with Prob } a \end{cases}, \quad (2)$$

where  $a$  is the average size of each memory pattern normalized by the system size  $N$ . By contrast, in this study we treat sequences of such memory patterns that possess variable sizes. Thus, we modify the above rule such that successive memory patterns may have different sizes,

$$\tilde{\xi}_i^\mu = \begin{cases} 0, & \text{with Prob } 1-a_\mu \\ 1, & \text{with Prob } a_\mu \end{cases}. \quad (3)$$

We later determine  $\{a_\mu\}$  according to a stochastic process.

Below, we primarily investigate the retrieval process of the network storing a single sequence of length  $P$ , although we can easily extend our analysis to memory retrieval with extensively many sequences of finite length. The sequence length is defined as the number of memory patterns stored in the neural network in proportion to the size of the network,  $P=\alpha N$ . To avoid possible finite size effects, we impose a periodic boundary condition on the memory sequence as

$$\{\tilde{\xi}_i^{\mu=P+1}\} = \{\tilde{\xi}_i^{\mu=1}\}. \quad (4)$$

In terms of these memory patterns, we construct the following synaptic coupling matrix:

$$\begin{aligned} J_{ij} &= \sum_{\mu} \frac{1}{Na_{\mu}(1-a_{\mu})} (\tilde{\xi}_i^{\mu+1} - a_{\mu+1})(\tilde{\xi}_j^{\mu} - a_{\mu}) \\ &= \sum_{\mu} \frac{1}{Na_{\mu}(1-a_{\mu})} \xi_i^{\mu+1} \xi_j^{\mu}, \end{aligned} \quad (5)$$

where  $\xi_i^{\mu}(\equiv \tilde{\xi}_i^{\mu} - a_{\mu})$  satisfies  $\langle \xi_i^{\mu} \rangle = 0$  and  $\langle (\xi_i^{\mu})^2 \rangle = a_{\mu}(1-a_{\mu})$ .

### III. ORDER-PARAMETER EQUATIONS DESCRIBING RETRIEVAL PROCESS

When the system size  $N$  is sufficiently large, we can describe the retrieval dynamics in terms of evolutionary equations for macroscopic order parameters. We can derive such equations analytically in the limit of  $N \rightarrow \infty$ , keeping  $\{a_{\mu}\}$  constant. The derivation goes in parallel with the previous analyses [19,20] that were conducted based on a version of the self-consistent signal-to-noise analysis (SCSNA) [21] extended to the nonequilibrium case. This method can rather

accurately describe the retrieval dynamics in the Hopfield-type associative memory models [19]. Moreover, the order-parameter equations obtained at the equilibrium coincide with those obtained by replica calculations [21]. We introduce a pattern overlap to measure the similarity between the population neural activity and the  $\mu$ th memorized pattern in the sequence,

$$m_{\mu}(t) = \frac{1}{Na_{\mu}(1-a_{\mu})} \sum_i \xi_i^{\mu} S_i(t), \quad (6)$$

which is normalized such that the value takes unity when the population activity retrieves the  $\mu$ th pattern correctly. We measure the strength of the population activity by

$$y(t) = \frac{1}{N} \sum_i S_i(t), \quad (7)$$

which takes a value of unity when all neurons are activated simultaneously.

We can assume, without loss of generality, that only the  $t$ th memory pattern has a sizable overlap at time step  $t$ , i.e.,  $m_{\mu=t}(t) \sim O(1)$ . Overlaps with the other memory patterns are significantly small at this time step, that is,  $m_{\mu \neq t}(t) \sim O(1/N)$ , since the pattern vectors are mutually independent. Thus, we may divide  $h_i(t)$  into a signal term having a magnitude of  $O(1)$  and the residual terms as follows:

$$\begin{aligned} h_i(t) &= \sum_{\mu} \xi_i^{\mu+1} m_{\mu}(t) = \xi_i^{t+1} m_t(t) + \sum_{\mu \neq t} \xi_i^{\mu+1} m_{\mu}(t) \equiv \xi_i^{t+1} m_t(t) \\ &\quad + Z_i(t). \end{aligned} \quad (8)$$

We can treat the crosstalk term  $Z_i(t)$  as independent random variables obeying a Gaussian distribution whose mean and variance are 0 and  $\sigma^2(t)$ , respectively. To derive the evolution equation for  $\sigma^2(t)$ , we have to consider

$$\sigma^2(t+1) = \langle Z_i^2(t+1) \rangle = \left\langle \left( \sum_{\mu \neq t+1} \xi_i^{\mu+1} m_{\mu}(t+1) \right)^2 \right\rangle, \quad (9)$$

where

$$m_{\mu}(t+1) = \frac{1}{Na_{\mu}(1-a_{\mu})} \sum_j \xi_j^{\mu} F \left( \sum_{\nu=1}^{\alpha N} \xi_j^{\nu+1} m_{\nu}(t) \right), \quad (\mu \neq t+1) \quad (10)$$

are overlaps with the nonretrieved patterns. Note that  $S_j$  is correlated with  $\xi_j^{\mu \neq t+1}$  since  $S_j$  depends on it through a recurrent synaptic input. Therefore, we separate the term involving  $\xi_j^{\mu \neq t+1}$  from the remaining terms and expand  $F$  up to the first order of this term as

$$\begin{aligned} &F \left( \xi_j^{\mu} m_{\mu-1}(t) + \sum_{\nu \neq \mu} \xi_j^{\nu} m_{\nu-1}(t) \right) \\ &= F \left( \sum_{\nu \neq \mu} \xi_j^{\nu} m_{\nu-1}(t) \right) + \xi_j^{\mu} m^{\mu-1}(t) F' \left( \sum_{\nu \neq \mu} \xi_j^{\nu} m_{\nu-1}(t) \right) \\ &\equiv S_j^{(\mu)}(t) + \xi_j^{\mu} m_{\mu-1}(t) S_j^{(\mu)'}(t), \quad (\mu \neq t+1) \end{aligned} \quad (11)$$

where thus introduced  $S_j^{(\mu)}(t)$  and  $S_j^{(\mu)'}(t)$  do not explicitly

depend on  $\xi_j^\mu$ . Substituting Eq. (11) for  $F(\sum_j \xi_j^{\mu+1} m_\nu(t))$  in the right-hand side of Eq. (10) gives

$$\begin{aligned} m_\mu(t+1) &= \frac{1}{Na_\mu(1-a_\mu)} \sum_j \xi_j^\mu S_j^{(\mu)}(t+1) \\ &\quad + \frac{1}{Na_\mu(1-a_\mu)} \sum_j (\xi_j^\mu)^2 S_j^{(\mu)'}(t+1) m_{\mu-1}(t) \\ &\equiv m^{(\mu)}(t+1) + U^\mu(t) m_{\mu-1}(t), \quad (\mu \neq t+1) \end{aligned} \quad (12)$$

where  $U^\mu(t)$ , the coefficient of  $m_{\mu-1}(t)$ , does not explicitly depend on  $\xi_j^\mu$ . Since  $\mu-n \neq t+1-n$  for an arbitrary integer  $n$ , we can repeatedly apply the above calculation to Eq. (12) and obtain the following equation for  $Z$ :

$$\begin{aligned} Z_i(t+1) &= \sum_{\mu \neq t+1} \xi_i^{\mu+1} (m^{(\mu)}(t+1) + U^\mu(t) m^{(\mu-1)}(t) \\ &\quad + U^\mu(t) U^{\mu-1}(t-1) m^{(\mu-2)}(t-1) \\ &\quad + U^\mu(t) U^{\mu-1}(t-1) U^{\mu-2}(t-2) m^{(\mu-3)}(t-2) + \dots). \end{aligned} \quad (13)$$

We have to average the square of  $Z_i$  over different realizations of memory sequence to calculate the variance given in Eq. (9). As  $\langle \xi_i^\mu \rangle = 0$ , the products of different memory patterns do not contribute to this averaging, and we need to evaluate only the self-squared terms appearing in the square of Eq. (13). The first term, for example, gives

$$\begin{aligned} &\left\langle \sum_{\mu \neq t+1} (\xi_i^{\mu+1})^2 (m^{(\mu)}(t+1))^2 \right\rangle \\ &= \left\langle \sum_{\mu \neq t+1, j} (\xi_i^{\mu+1})^2 (\xi_j^\mu)^2 (S_j^{(\mu)}(t+1))^2 \left( \frac{1}{Na_\mu(1-a_\mu)} \right)^2 \right\rangle \\ &= \left\langle \sum_{\mu \neq t+1} (\xi_i^{\mu+1})^2 \frac{1}{Na_\mu(1-a_\mu)} \right\rangle y(t+1) \\ &= \left\{ N\alpha \sum_{a,b} \text{Prob}(a_\mu = a) \text{Prob}(a_{\mu+1} = b | a_\mu = a) \right. \\ &\quad \left. = a \frac{b(1-b)}{Na(1-a)} \right\} y(t+1) = (c\alpha) y(t+1), \end{aligned} \quad (14)$$

where

$$\begin{aligned} c &\equiv \sum_{a,b} \text{Prob}(a_\mu = a) \text{Prob}(a_{\mu+1} = b | a_\mu = a) \frac{b(1-b)}{a(1-a)} \\ &= \frac{1}{\alpha N} \sum_\mu \frac{a_{\mu+1}(1-a_{\mu+1})}{a_\mu(1-a_\mu)} \end{aligned} \quad (15)$$

arises from the nonuniformity of the size of memory patterns. As shown in Eq. (13), the effect of the factor  $c$  is to rescale the memory storage load as  $\alpha \rightarrow c\alpha$ . We can calculate the other terms in a similar fashion to obtain a recursive relationship,

$$\begin{aligned} \sigma^2(t+1) &= \alpha c (y(t+1) \\ &\quad + U^2(t) y(t) + U^2(t) U^2(t-1) y(t-1) + \dots) \\ &= \alpha c \left( y(t+1) + U^2(t) \left\{ \frac{1}{\alpha c} \sigma^2(t) \right\} \right) \\ &= (\alpha c) y(t+1) + U^2(t) \sigma^2(t). \end{aligned} \quad (16)$$

Consequently, the macroscopic equations are summarized as

$$\begin{aligned} m(t+1) &= \frac{1}{a_{t+1}(1-a_{t+1})} \int Dz \langle \xi^{t+1} F[\xi^{t+1} m(t) + \sigma(t) z] \rangle_{\xi^{t+1}}, \\ \sigma^2(t+1) &= (\alpha c) y(t+1) + U^2(t) \sigma^2(t), \\ y(t+1) &= \int Dz \langle F[\xi^{t+1} m(t) + \sigma(t) z] \rangle_{\xi^{t+1}}, \\ U(t) &= \int Dz \langle F'[\xi^{t+1} m(t) + \sigma(t) z] \rangle_{\xi^{t+1}}, \end{aligned} \quad (17)$$

where  $Dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$ . In particular, when  $F$  is a Heaviside function, we can further rewrite the above equations as

$$m(t+1) = 1 - \frac{1}{2} [\text{erfc}(\phi_{t+1}^1) + \text{erfc}(\phi_{t+1}^0)],$$

$$y(t+1) = a_{t+1} - \frac{1}{2} [a_{t+1} \text{erfc}(\phi_{t+1}^1) - (1-a_{t+1}) \text{erfc}(\phi_{t+1}^0)],$$

$$\begin{aligned} \sigma^2(t+1) &= (\alpha c) y(t+1) + \frac{1}{2\pi} \{ a_{t+1} \exp[-(\phi_{t+1}^1)^2] \\ &\quad + (1-a_{t+1}) \exp[-(\phi_{t+1}^0)^2] \}^2, \end{aligned} \quad (18)$$

in terms of normalized variables,

$$\begin{aligned} \phi_{t+1}^1 &= \frac{(1-a_{t+1})m(t) - \theta}{\sqrt{2}\sigma(t)} \\ \phi_{t+1}^0 &= \frac{a_{t+1}m(t) + \theta}{\sqrt{2}\sigma(t)}, \end{aligned} \quad (19)$$

where  $\text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du$  is the complementary error function.

When the successive memory patterns have an equal size,  $a_\mu = a$ , the equations coincide with the conventional evolution equations of the order parameters. The equations accurately predicted the memory retrieval dynamics and the storage capacity in many models. To confirm the validity of our analyses, we numerically solved the evolution equations with constant  $a_\mu$  for various initial overlaps. The numerical solutions agreed well with the trajectories of the pattern overlap that were obtained by simulations of the original network model (Fig. 1).

#### IV. REDUCTION OF STORAGE CAPACITY DUE TO STRUCTURAL NONUNIFORMITY

The nonuniformity of the pattern size gives rise to two different effects on the order parameter equations: the scaling

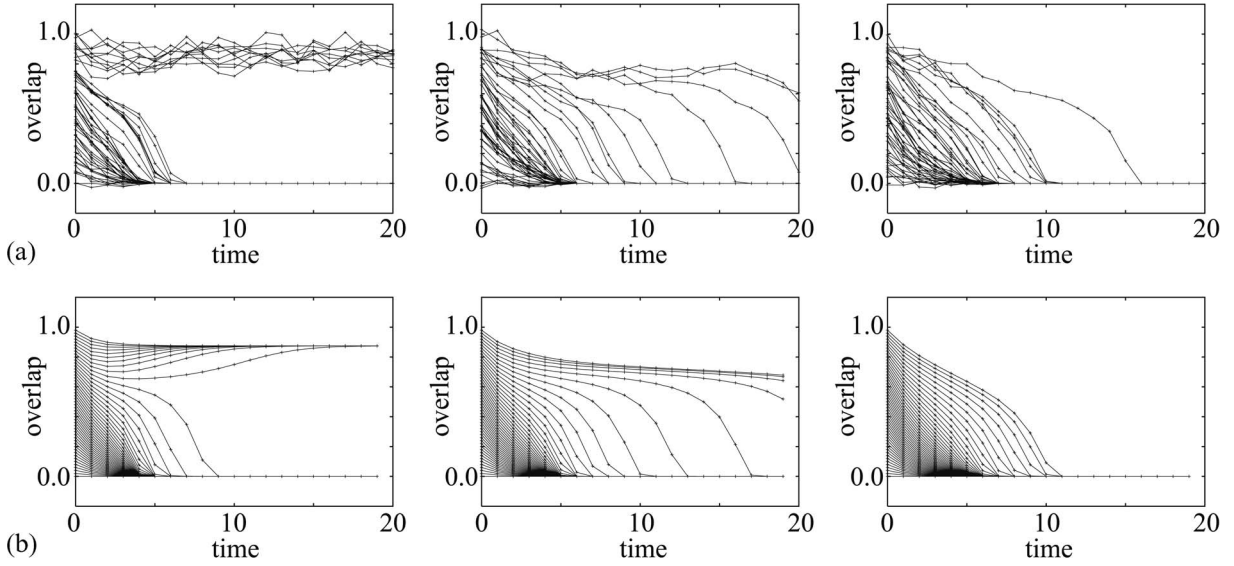


FIG. 1. Time evolution of the pattern overlap during memory retrieval. A network model embeds an infinitely long sequence of memory patterns with a constant memory size,  $a=0.1$ . The pattern overlap was obtained from (a) numerical simulations of a 4000-neuron network or (b) numerical solutions to the order-parameter equations. The memory load, or equivalently the sequence length, was set as  $\alpha=0.6$  (left),  $0.7$  (middle), and  $0.8$  (right).

factor  $c$  and the temporal fluctuation due to  $a_i$  explicitly included in the equations. Since the value of the scaling factor is generally greater than unity,

$$c = \frac{1}{P} \sum_{\mu} \frac{a_{\mu+1}(1-a_{\mu+1})}{a_{\mu}(1-a_{\mu})} \geq \prod_{\mu} \frac{a_{\mu+1}(1-a_{\mu+1})}{a_{\mu}(1-a_{\mu})} = 1, \quad (20)$$

where the equality holds when all  $a_{\mu}$ 's have the same value. This implies that the present neural network behaves as if an effective number of memory patterns were larger than the actual one. Therefore, the variations in the memory pattern size reduce the storage capacity.

This situation degrades the effect of sparse coding on improving the storage capacity. When all  $a_{\mu}$ 's have an equal value, the storage capacity is known to be a monotonically decreasing function of the sparseness of network activity [20,22,23]. Therefore, we can improve the storage capacity by decreasing that value. If, however,  $a_{\mu}$ 's have different values and only some of them are decreased, the value of  $c$  is increased and the effect of the increased nonuniformity may overwhelm the benefit of small  $a_{\mu}$ . Thus, the storage capacity is determined through the tradeoff between the two competing effects. To see this more clearly, we consider a simple example in which  $a_{\mu}$ 's take only two values,  $b_1$  and  $b_2 (< b_1)$  according to the following probabilistic rule:

$$a_{\mu} = \begin{cases} b_1, & \text{with Prob } 1-p \\ b_2, & \text{with Prob } p \end{cases}. \quad (21)$$

The average of  $a_{\mu}$ 's over all memorized patterns is a decreasing function of  $p$  since  $b_2 < b_1$ . The storage capacity, therefore, would increase monotonically with an increase in  $p$ , if we neglect the scaling factor that is a nonmonotonic function of  $p$  (Fig. 2, dashed line). The solid curves in Fig. 2 display the values of the storage capacity which were calculated from the order-parameter Eqs. (18) and (19) with  $\{a_{\mu}\}$  given

by Eq. (21). The storage capacity increases monotonically if we neglect the scaling factor  $c$  (gray). It is, however, not monotonous if we consider this factor (black). As we increase  $p$ , the storage capacity initially decreases and takes a minimum at a finite value of  $p$ .

## V. STOCHASTIC FLUCTUATIONS IN ORDER-PARAMETER DYNAMICS

When the sizes of successive memory patterns are determined as random variables, the order-parameter equations

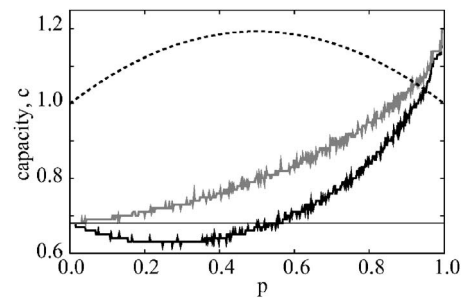


FIG. 2. Storage capacity of a network model embedding a sequence of memory patterns. Their normalized sizes  $\{a_{\mu}\}$  were chosen from a binomial distribution at  $b_1=0.1$  (with Prob  $1-p$ ) and  $b_2=0.04$  (with Prob  $p$ ). The capacity was calculated by numerically solving the order-parameter equations. The scaling factor can be calculated as  $c=(B_1-B_2)^2 p(1-p)/(B_1 B_2)+1$ , where  $B_1=b_1(1-b_1)$  and  $B_2=b_2(1-b_2)$ , and changes nonmonotonically with  $p$  (dashed curve). If the scaling factor is neglected in the expressions, the storage capacity increases monotonically with increasing  $p$  (gray curve). The scaling factor, however, turns the storage capacity into a nonmonotonic function that takes a minimum value at a finite value of  $p$  (black curve).



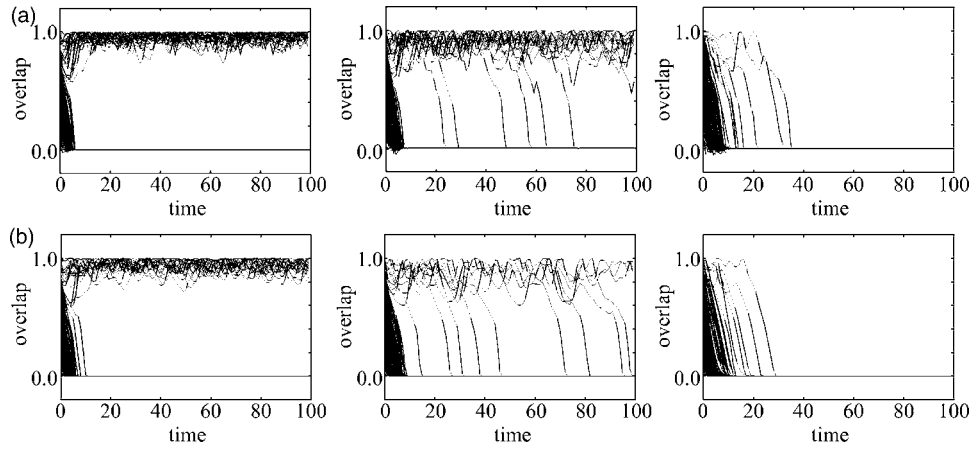


FIG. 3. Time evolution of the pattern overlap during memory retrieval. A network model embeds an infinitely long sequence of memory patterns with highly variable memory sizes. The sizes of memory patterns were chosen from a uniform distribution  $[0.01:0.1]$ . The pattern overlap was obtained from (a) numerical simulations of a 4000-neuron network or (b) numerical solutions to the order-parameter equations. The memory load, or equivalently the sequence length, was set as  $\alpha=0.6$  (left),  $0.7$  (middle), and  $0.8$  (right). Note that the range of the horizontal axes is five times longer than that of Fig. 1 to show the stochastic behavior clearly.

including stochastic variables  $\{a_\mu\}$  describe a set of stochastic equations rather than deterministic ones even in the limit of  $N \rightarrow \infty$ . We solved these stochastic equations in the case where we determined the sizes of individual memory patterns according to independent identical uniform distributions. Figures 3(a) and 3(b) display the time evolution of the pattern overlap obtained from numerical simulations of the original network model and from numerical solutions to the macroscopic equations (18), respectively. We performed these numerical simulations for various initial values of the pattern overlap and the memory load. The results show that the pattern overlap may initially be attracted to  $m=1$ , as in the previous model (Fig. 1). This retrieval state was stable in the previous model, storing memory patterns of a fixed size. The overlaps, however, cannot stay permanently at this value, when the sizes of successive memory patterns vary as in the present case. The overlaps fluctuate beneath  $m=1$  for a while until they suddenly drop from the vicinity of unity to 0. If we use an analogy of the potential function, the state  $m=1$  is a metastable state, from which a random drift force kicks out a “particle” to a global minimum defined by  $m=0$  within a finite lifetime.

To evaluate the lifetime of these metastable states, we averaged the trajectories obtained by numerical simulations

of the original network model over many different realizations of the memory-size fluctuations [Fig. 4(a)]. As shown in the figures, the averaged overlaps decay faster when the network embeds a larger number of memory patterns, indicating the lifetime is a decreasing function of the memory storage. Solving the order-parameter equations gives a reasonable approximation to the behavior of the averaged overlaps at each value of the memory load [Fig. 4(b)]. To quantify the decaying behavior, we plotted time evolution of the averaged overlaps calculated by the order-parameter equations at various memory loads and for the same initial conditions [Fig. 5(a)]. Fitting these curves with exponentially decaying functions provides the decay constant  $\tau$  as a function of the memory load, that is, the length of memory sequence relative to the network size [Fig. 5(b)]. For a relatively small number of memory patterns,  $\tau$  may be practically regarded as infinity, and the network can trace the memory sequence for a sufficiently long time. However,  $\tau$  can be much smaller in networks embedding a larger number of memory patterns, and the network cannot retrieve a memory sequence that is longer than the characteristic time scale.

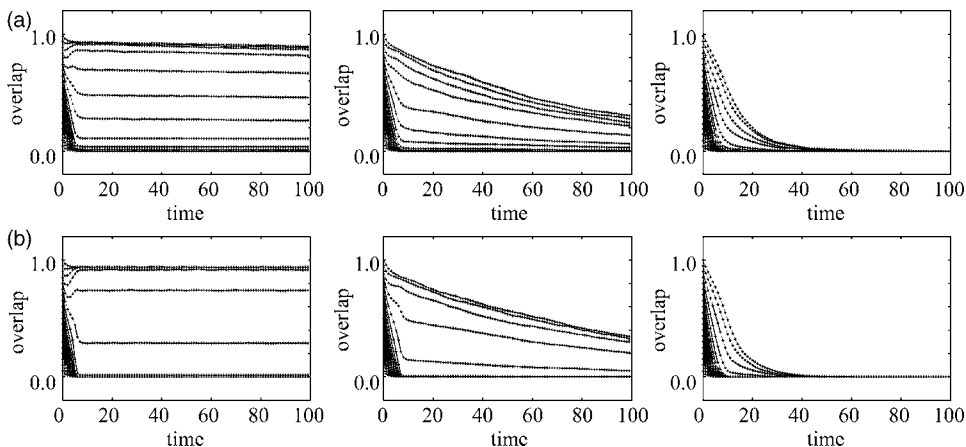


FIG. 4. Time evolution of the average pattern overlap during memory retrieval. The dynamical trajectories were calculated from (a) numerical simulations of a 1000-neuron network similar to that shown in Fig. 3 or (b) numerical solutions to the order-parameter equations. Other parameters were the same as in Fig. 3. The trajectories were averaged over 1000 different realizations of memory patterns with variable sizes.

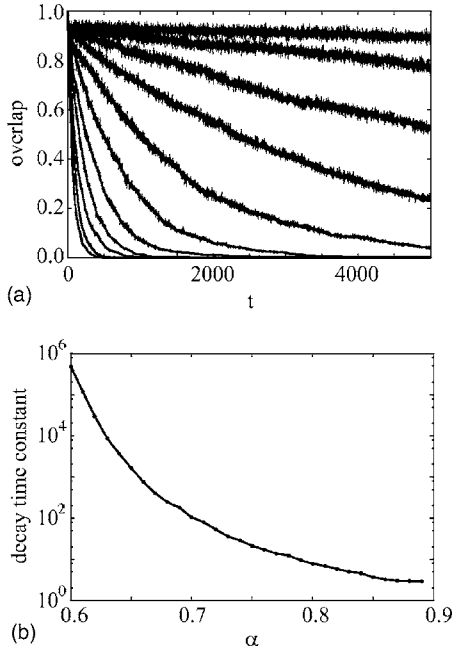


FIG. 5. Estimation of the mean lifetime of the sequence memory retrieval in Fig. 4. The average of the pattern overlap was taken over 1000 different realizations of the memory size fluctuations. (a) From top to bottom, the value of  $\alpha$  was increased from 0.6 to 0.7 with a step size of 0.01. We fixed the values of other parameters at the same values as in Fig. 2(b). (b) Exponential functions of  $\alpha$  could well fit the curves shown above to obtain the decay constant.

## VI. POSSIBLE RELATIONSHIP TO NEURONAL AVALANCHES

We have so far considered the associative memory model loaded with an infinitely long sequence satisfying a periodic boundary condition. The results of the previous analysis are applicable to an associative memory model loaded with extensively many sequences of finite length. Such a network is of particular interest since it can generate a set of retrieval states similar to synchronous activity propagations of neuronal avalanches. Here, as an interesting application of the order-parameter analysis, we study an associative memory that stores multiple memory sequences that obey power-law distributions of the total activity size (defined later) and the length. We consider the same power laws as neuronal avalanches show. We will demonstrate that the stochastic fluctuations in the memory pattern size affect the power-law statistics of retrieved activities. Denoting the net activity (the number of 1s) of the  $k$ th memory pattern in the  $\mu$ th sequence as  $a_{\mu,k}$ , we generate  $a_{\mu,k}$ 's recursively through a Markov stochastic process,  $P(a_{\mu,1}, a_{\mu,2}, a_{\mu,3}, \dots) = P_0(a_{\mu,1})P(a_{\mu,2}|a_{\mu,1})P(a_{\mu,3}|a_{\mu,2})\dots$ . The conditional Gaussian distributions,  $P(a_{\mu,k+1}|a_{\mu,k})$ , satisfy

$$\begin{aligned} \langle a_{\mu,k+1} \rangle_{a_{\mu,k}=a} &= a, \\ \text{Var}(a_{\mu,k+1})_{a_{\mu,k}=a} &= \sigma^2 a, \end{aligned} \quad (k=1,2,3,\dots) \quad (22)$$

where  $\langle \dots \rangle_{a_{\mu,k}=a}$  and  $\text{Var}(\dots)_{a_{\mu,k}=a}$  stand for the average and variance taken under the condition  $a_{\mu,k}=a$ . Here,  $\sigma$  is a parameter characterizing the magnitude of fluctuations in the

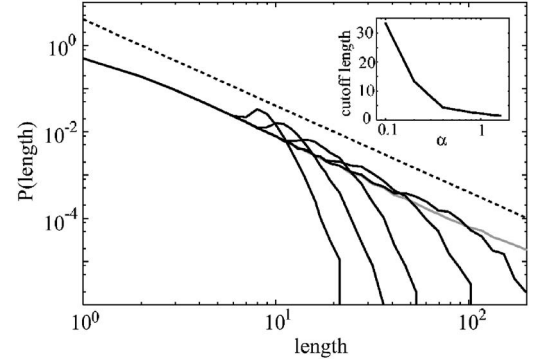


FIG. 6. Memory retrieval in a neural network storing infinitely many sequences of finite size. Power-law distributions of the length of embedded sequences (gray) and the lifetime of retrieved sequences (black) were shown. The dotted line indicates a power law with exponent  $-2$ . The normalized sizes of sequential memory patterns  $a_{\mu,k}$  were chosen recursively from a Gaussian distribution with  $\sigma=0.05$  according to the rules shown in Eq. (22). The range of the pattern size was set as  $a_{\min}=10^{-4}$  and  $a_{\max}=0.5$ . As we increase the number of memory patterns ( $\alpha=0.1, 0.2, 0.4, 0.8$ , and  $1.6$ ), the position at which an exponential cutoff appears in the power-law distributions shifts toward shorter lifetime. The inset shows the cutoff position as a function of  $\alpha$ .

pattern size. If  $a_{\mu,k+1}$  is chosen outside a certain prescribed range,  $a_{\min} \leq a_{\mu,k+1} \leq a_{\max}$ , we terminate this sequence at the  $k$ th memory pattern and start a new sequence as the  $(\mu+1)$ th sequence. Setting the size of initial layers equal to  $a_{\min}$ , we repeat the above process  $P(=\alpha N)$  times to obtain memory sequences with various lengths and sizes. From the obtained sequences, we construct the following synaptic matrix by their superposition:

$$\begin{aligned} J_{ij} &= \sum_{\mu}^{\alpha N} \sum_k^{l_{\mu}} \frac{1}{Na_{\mu,k}(1-a_{\mu,k})} (\hat{\xi}_i^{\mu,k+1} - a_{\mu,k+1})(\hat{\xi}_j^{\mu,k} - a_{\mu,k}) \\ &= \sum_{\mu}^{\alpha N} \sum_k^{l_{\mu}} \frac{1}{Na_{\mu,k}(1-a_{\mu,k})} \xi_i^{\mu,k+1} \xi_j^{\mu,k}, \end{aligned} \quad (23)$$

where  $\{\xi^{\mu,k}\}$  represents pattern vectors defined similarly to Eqs. (3) and (5), and  $l_{\mu}$  is the length of the  $\mu$ th sequence.

The mathematical structure of the above stochastic wiring procedure to embed memory patterns is essentially equivalent to the critical branching process [12,13]. Thus, the size and length distributions of the memory sequences generated by this procedure should obey power laws of exponents  $-3/2$  and  $-2$ , respectively, so the retrieval of the memorized sequences resembles the propagation of synchronous activities in neuronal avalanches. In fact, we can confirm the power-law-obeying tail of extremely long memory sequences in our numerical results (Fig. 6, gray line). However, the present network model cannot retrieve such long sequences perfectly, since the fluctuations in the memory-pattern size destabilize the retrieval dynamics. To see this situation, we constructed the lifetime distribution from the numerical solutions to the order-parameter equations (18) and (19) for the network defined with Eqs. (22) and (23). Here, the lifetime

of each retrieved sequence represents the time step at which the pattern overlap falls off to the failure state  $m=0$ . The lifetime distributions of retrieved sequences well replicate the power law at relatively short lifetimes (Fig. 6, solid curves), meaning that the stochastic size fluctuations do not significantly disturb the retrieval of short sequences. However, the distributions deviate from the power law at longer lifetimes. In fact, each distribution exhibits an exponential cutoff representing a failure in memory retrieval at long lifetimes. As expected from the result shown in Fig. 5(b), increases in the memory load  $\alpha$  shift the position of this cutoff toward a shorter lifetime. Thus, the lifetime of the retrieved activities exhibits a tail that deviates from the power-law distribution.

## VII. DISCUSSION

We have studied associative memory models embedding memory sequences with various sizes and lengths. We have derived the macroscopic order-parameter equations analytically by a standard method to obtain such equations in associative memory models embedding fixed-size memory patterns. The obtained equations well describe the dynamical process of memory retrieval in the present models.

The fluctuations in the memory size create two dynamical effects that have not extensively been studied in literature. First, the nonuniformity of the network structure introduces a numerical factor multiplied by the number of memorized patterns in the order-parameter equations. Since this multiplicative factor is greater than unity, the capacity of sequential memory retrieval is smaller than that obtained without the memory-size fluctuations. Second, the order parameters themselves contain fluctuations since their equations depend explicitly on the sizes of consecutive memory patterns. Therefore, the retrieval process becomes stochastic rather than deterministic, when the pattern size is to obey a stochas-

tic process responsible for the power laws. The value of the pattern overlap falls off quickly from the value corresponding to a successful retrieval, as the random drift force originating from the memory-size fluctuations kicks out a particle wandering around a local minimum. The typical time scale that the state takes in leaving from the metastable retrieval state is in general a monotonically decreasing function of the total number of memorized patterns.

As an intriguing application of the present analysis, we have investigated a sequential associative memory storing extensively many sequences obeying a power-law length distribution. If the network can retrieve all memorized sequences correctly, the lifetime distribution of the retrieved sequences should also obey the same power law as the length distribution of memorized sequences. However, the fluctuations in the size of memory patterns disable the network to retrieve long sequences successfully. Therefore, the resultant distribution of the retrieved sequences deviates from the power law at long lifetimes by showing an exponential cutoff. The lifetime of the retrieved sequences in this model may correspond to the lifetime of synchronous activities propagating during neuronal avalanches. In fact, an exponential cutoff was characteristic of the lifetime distribution obtained from *in vitro* experiments of neuronal avalanches [10]. Such an exponential cutoff is considered to arise from the finite size effect in cortical slices. However, the cutoff may partly reflect the dynamical instability of memory retrieval, as suggested by the present results.

## ACKNOWLEDGMENTS

The present work was partially supported by Grants-in-Aid for Scientific Research of Priority Areas (No. 17022036) and Grant-in-Aid for Young Scientists No. (B) 17700318 from the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

- 
- [1] S. Amari and K. Maginu, *Neural Networks* **1**, 63 (1988).
  - [2] H. Nishimori, T. Nakamura, and M. Shiino, *Phys. Rev. A* **41**, 3346 (1990).
  - [3] A. V. M. Herz, Z. Li, and J. L. van Hemmen, *Phys. Rev. Lett.* **66**, 1370 (1991).
  - [4] M. Morita, *Brain Res. Cognit. Brain Res.* **5**, 137 (1996).
  - [5] N. Matsumoto, M. Okada, Y. Sugase-Miyamoto, and S. Yamane, *J. Comput. Neurosci.* **18**, 85 (2005).
  - [6] M. Abeles, H. Bergman, E. Margalit, and E. Vaadia, *J. Neurophysiol.* **70**, 1629 (1993).
  - [7] M. Diesmann, M. Gewaltig, and A. Aertsen, *Nature* **402**, 529 (1999).
  - [8] R. H. Hahnloser, A. A. Kozhevnikov, and M. S. Fee, *Nature* **419**, 65 (2002); **421**, 294(E) (2003).
  - [9] Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste, *Science* **304**, 559 (2004).
  - [10] J. M. Beggs and D. Plenz, *J. Neurosci.* **23**, 11167 (2003).
  - [11] J. M. Beggs and D. Plenz, *J. Neurosci.* **24**, 5216 (2004).
  - [12] T. E. Harris, *The Theory of Branching Processes* (Dover, New York, 1989).
  - [13] S. Zapperi, K. B. Lauritsen, and H. E. Stanley, *Phys. Rev. Lett.* **75**, 4071 (1995).
  - [14] J. Teramae and T. Fukai, *J. Comput. Neurosci.* (to be published).
  - [15] S. Amari, *IEEE Trans. Comput.* **21**, 1197 (1972).
  - [16] H. Sompolinsky and I. Kanter, *Phys. Rev. Lett.* **57**, 2861 (1986).
  - [17] S. Amari, in *Neural and Synergetic Computers*, edited by H. Haken (Springer-Verlag, Berlin, 1988), p. 85.
  - [18] E. Domany, W. Kinzel, and R. Meir, *J. Phys. A* **22**, 2081 (1989).
  - [19] M. Okada, *Neural Networks* **9**, 1429 (1996).
  - [20] K. Kitano and T. Aoyagi, *J. Phys. A* **31**, L613 (1998).
  - [21] M. Shiino and T. Fukai, *J. Phys. A* **25**, L375 (1992).
  - [22] G. Palm, *Biol. Cybern.* **36**, 19 (1980).
  - [23] M. V. Tsodyks and M. V. Feigl'man, *Europhys. Lett.* **6**, 101 (1988).