Dynamics-based sequential memory: Winnerless competition of patterns

Philip Seliger, Lev S. Tsimring, and Mikhail I. Rabinovich
Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093-0402
(Received 13 May 2002; published 16 January 2003)

We introduce a biologically motivated dynamical principle of sequential memory which is based on winnerless competition (WLC) of event images. This mechanism is implemented in a two-layer neural model of sequential spatial memory. We present the learning dynamics which leads to the formation of a WLC network. After learning, the system is capable of associative retrieval of prerecorded sequences of patterns.

DOI: 10.1103/PhysRevE.67.011905 PACS number(s): 87.19.La, 84.35.+i, 87.18.Sn, 89.75.Kd

The ability to process sequential information has long been seen as one of the most important functions of living and artificial intelligent systems. In spite of the long history of studies of sequential learning and memory, little is known about dynamical principles of learning and remembering of multiple events and their temporal order by neural systems. Here we propose a dynamical principle of *winnerless competition* (WLC) that can be the basic mechanism of the sequential memory. The essence of the idea is that the sequential memory is encoded in a multidimensional dynamical system with a complex heteroclinic trajectory connecting a sequence of saddle points. Each of the saddle points represents an event in a sequence to be remembered. The specific structure of the phase space is such that each saddle point can have many stable directions but only a single unstable direction. All saddle points are unidirectionally connected by these one-dimensional unstable separatrices. Once the state of the system approaches one fixed point representing a certain event, it is drawn along an unstable separatrix toward the next fixed point, and so on. The existence and stability of such heteroclinic structure is determined by specific asymmetric inhibitory connections between neurons within the WLC neural network. These connections are formed by the sensory inputs caused by sequential events in a sequence.

In this paper, we demonstrate this principle in a model of the spatial sequential memory in the hyppocampus. It is well accepted that the hippocampus plays the central role in acquisition and processing of information related to the representation of physical space. The most spectacular manifestation of this role is the existence of so called ''place cells'' which repeatedly fire when an animal is in a certain spatial location [1]. While much effort has been spent on experimental search and modeling of the so called ''cognitive map'' [2] as a paradigm for spatial memory, recent neurophysiological research favors an alternative concept of spatial memory based on a linked collection of stored *episodes* [3]. Each episode comprises a sequence of *events* which, besides spatial locations, may include other features of environment (orientation, odor, sound, etc.). Each distinct event is accompanied by time-locked activity of a certain hippocampal cell. Dynamical modeling of the emerging concept of the episodic memory is of apparent general interest for neuroscience. Several models of associative sequential memory have been proposed in the literature [4]. Most of them are based on the generalization of the Hopfield associative memory network [5] to include asymmetric synaptic connections. Accordingly, they suffer from difficulties typical for Hopfield-type networks: the abundance of spurious attractors (sequences), complex structure of attractor basins, and sensitivity to noise. Furthermore, these models are based on dynamical equations with memory, which is difficult to justify biologically.

A dynamical model of the sequential spatial memory should be based on the following experimental facts. First, there is a clear separation between neurons directly responding to specific stimuli (we call them sensory neurons, SN) and hyppocampal cells in CA1 and CA3 regions (principal neurons, PN). The PNs fire in response to a combined vector of stimuli corresponding to a particular event. Second, while sensory neurons are not directly connected to each other, the PNs are coupled via inhibitory connections controlled by interneurons. Third, the synaptic connections among PNs and between PNs and SNs exhibit Hebbian long-term potentiation [6,7]. Based on these features of the hippocampal network, we propose a two-layer dynamical model of the sequential spatial memory (SSM) that can answer the following key questions. (i) How is a certain event (e.g., an image of environment) recorded in the structure of the synaptic connections between multiple SNs and a single PN during learning? (ii) What kind of the cooperative dynamics forces individual PCs to fire sequentially, which would correspond to a specific route (a sequence of scenes) in the environment? (iii) How complex should this network be to store a certain number of different episodes without mixing different events or storing spurious episodes?

Let us discuss the learning objectives which would lead to formation of the sequential SSM. The first objective is to learn a projection map: as a result of unsupervised learning the image of a particular environment (snapshot) encoded by heightened activity of the group of SNs leads to the heightened activity (firing) of just one PN (see Fig. 1). The second objective is to learn the temporal sequence of images. This can be achieved by modifying inhibitory connections among PNs due to long-term potentiation (see, e.g., Ref. [6]). The resulting structure of the phase space for the PN layer will exhibit features of the winnerless competition [8]. After the learning is completed, the neural network should be able to reproduce a specific route following a starting pattern.

The two-layer structure of the SSM model is reminiscent of the projection network implementation of the *normal form projection algorithm* (NFPA) [9]. In that model, the dynam-

 ©2003 The American Physical Society

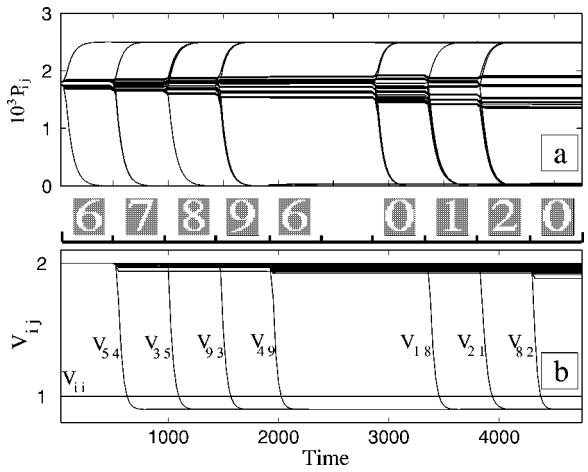

FIG. 1. The strengths of the connection coefficients between the sensory and the principal layers (a) and within the principal layer (b). Parameters of simulations: $N_s=588$, $N_p=10$, $\alpha=1$, $\beta=2.5$, $V_1=0.9$, $\epsilon=0.01$, $\sigma=10^{-4}$, $\tau=480$.

ics of the network is cast in terms of the normal form equations which are written for amplitudes of certain normal forms which correspond to different patterns stored in the system. The normal form dynamics can be chosen to follow certain dynamical rules, for example, in Ref. [9] it was shown that a Hopfield-type network with improved capacity can be built using this approach. Furthermore, in Ref. [9] it was proposed that specific choices of the coupling matrix for the normal form dynamics can lead to multistability among more complex attracting sets than simple fixed points, such as limit cycles or even chaotic attractors. As we will see below, the model of SSM after learning is completed can be viewed as a variant of the NFPA with *m* specific choice of normal form dynamics corresponding to the winnerless competition among different patterns.

Consider a two-level network of N_s SN x_i and N_p principal neurons a_i. Similar to the projection network model [9], we assume that sensory neurons do not have their own dynamics and are slaved to either external stimuli in the learning (or storing) regime, or to the PNs in the retrieval regime. In the learning regime, $x_i=I_i$ where $\{I_i\}$ is a binary input pattern consisting of 0's and 1's. During the retrieval phase, $x_i=\Sigma_{j=1}^{N_p} P_{ij}a_j$, where P_{ij} is the $N_s\times N_p$ projection matrix of connections between SNs and PNs.

The PNs are driven by SNs during the learning phase, but they also have their own dynamics controlled by inhibitory interconnections (see above). After the learning is finished, the direct driving from SNs is disconnected. The equations for the amplitudes of PNs, a_i, read

$$\dot{a}_i=a_i-a_i\sum_{j=1}^{N_p} V_{ij}a_j+\alpha a_i\sum_{j=1}^{N_s} P_{ij}^T x_j+\xi(t), \qquad (1)$$

where $\alpha\neq 0$ in the learning phase, and $\alpha=0$ in the retrieval phase. We use the transposed projection matrix P_{ij}^T assuming that the coupling between SNs and PNs is bidirectional and symmetric. The last term in the rhs of Eq. (1) represents small positive external perturbations which we model as white noise uniformly distributed between 0 and σ, however, in reality it can represent input signals from other parts of the brain which control learning and retrieval dynamics.

After a certain pattern is presented to the model, the sensory stimuli reset the state of the PN layer according to the projection rule $a_i=\Sigma_{j=1}^{N_s} P_{ij}^T x_j$, but then a_i change according to Eq. (1).

In addition to the dynamics of SNs and PNs during learning and retrieval phases, we need to introduce two learning processes: (i) forming the projection matrix P_{ij} which is responsible for connecting a group of sensory neurons of the first layer corresponding to a certain stored pattern to a single principal neuron which represents this pattern at the PN level; (ii) learning of the competition matrix V_{ij} which is responsible for the temporal (logical) ordering of the sequential memory.

Projection matrix. The slow learning dynamics of the projection matrix is controlled by the following equation

$$\dot{P}_{ij}=\epsilon a_i(\beta x_j-P_{ij}). \qquad (2)$$

with $\epsilon\ll 1$. We assume that initially all P_{ij} connections are nearly identical $P_{ij}=1+\eta_{ij}$, where η_{ij} are small random perturbations, $\Sigma_j\eta_{ij}=0$, $\langle\eta_{ij}^2\rangle=\eta_0^2\ll 1$. Additionally, we assume that initially matrix V_{ij} is purely competitive: $V_{ii}=1$ and $V_{ij}=V_0>1$ for $i\neq j$.

Consider a scenario when we want to ''memorize'' a certain pattern **A** in our projection matrix. We apply a set of inputs A_i corresponding to the pattern **A** to the SNs. As before, we assume that external stimuli render the SNs in one of two states: excited ($A_i=1$) and quiescent ($A_i=0$). The initial state of the PN layer is fully excited $[a_i(0)=\Sigma_j P_{ij}A_j]$. According to the competitive nature of interaction of PNs after a short transient, only one of them (neuron **A**) which corresponds to maximum $a_i(0)$ remains excited and others become quiescent (inhibited). Which neuron becomes ''responsible'' for the pattern **A** is actually random, as it depends on the initial projection matrix P_{ij}. As it follows from Eq. (2), for small ϵ ''synapses'' of suppressed PNs do not change, whereas synapses of the (single) excited neuron evolve such that the connections between excited SNs and PNs neurons amplify towards $\beta>1$, and connection between excited PNs and quiescent SNs decay to zero [see Fig. 1(a)]. As a result, the first input pattern will be ''recorded'' in one of the rows of the matrix P_{ij}, while other rows will remain almost unchanged.

Now suppose that we want to record a second pattern different from the first one. We can repeat the procedure described in the preceding paragraph, namely, apply external stimuli (pattern **B**) to the SNs, ''project'' them to the initial state of the PN layer $[a_i(0)=\Sigma_j P_{ij}B_j]$, and let the system evolve. Since synaptic connections from SNs suppressed by the first pattern to neuron **A** have been eliminated, a new set of stimuli corresponding to pattern **B** will excite neuron **A** weaker than most of the others, and competition will lead to selection of one principal neuron **B** *different* from neuron **A**. In such a way we can record as many patterns as there are PNs.

Competition matrix. The sequential order of patterns recorded in the projection network is determined by the competition matrix V_{ij}. Initially it is set to $V_{ij}=V_0>1$ for $i \neq j$ and $V_{ii}=1$ which corresponds to winner-take-all competition. The goal of sequential spatial learning is to record the transition of pattern **A** to pattern **B** in the form of suppressing the competition matrix element V_{BA}. The slow dynamics of the nondiagonal elements of the competition matrix are controlled by the delay-differential equation

$$\dot{V}_{ij}=\epsilon a_i(t)a_j(t-\tau)(V_1-V_{ij}). \tag{3}$$

As seen from Eq. (3), only the matrix elements corresponding to $a_i(t)\neq 0$ and $a_j(t-\tau)\neq 0$, are changing towards the asymptotic value $V_1<1$ corresponding to the desired transition. Since most of the time (except for short transients) only one of the principal neurons is excited, only one of the connections V_{ij} is changing at any time [see Fig. 1(b)]. As a result, an arbitrary (nonrepeating) sequence of patterns can be recorded. If, after a series of nonrepeating patterns, we show the first pattern again, the ''loop'' of heteroclinic connections will be closed and the system will be able to reproduce a repeating sequence of patterns in a cyclic manner.

If the dimension of the secondary layer N_s permits, it is easy to record into the network more than one sequence of patterns. To avoid a spurious connection between the sequences, the time interval between the last pattern of the first sequence and the first pattern of the second sequence should be greater than τ.

In Fig. 1 we show the simulation results for a slow dynamics of weights P_{ij} and V_{ij} during a learning phase in a network with 588 sensory and 10 principal neurons for $\epsilon=0.01$. As stored patterns I_i we take ten digits $0,\ldots,9$ represented as 21×28 pixel dithered images. Two loop sequences of patterns have been presented: ''0,'' ''1,'' ''2,'' and ''6,'' ''7,'' ''8,'' ''9.'' Note that these images are not precisely orthogonal to each other, and yet the system is able to associate them to different PNs. While a certain pattern is presented to the SN layer, certain matrix coefficients P_{ij} decay, some other (connecting excited neurons of the sensory layer and a single excited neuron of the PN layer) approach 2.5 and remaining connections remain almost unchanged. After a switch from one pattern to the next in a sequence the corresponding matrix coefficient V_{ij} decays to a low value $V_1=0.9$.

Now, presenting a test pattern **T** ''resembling'' one of the recorded patterns to the sensory layer $[x_i(0)=T(i),\ a_i(0)=\Sigma_i P_{ij}T_j]$, will initiate a periodic sequence of patterns corresponding to the previously recording sequence recorded in the network. Figure 2(a) shows the behavior of the principal neurons after two different initial patterns have been presented, one resembling digit ''0'' and another resembling digit ''6.'' In both cases, the system quickly settled onto a cyclic generation of patterns associated with a given test pattern. At any given time except for a short transient time between the patterns, only a single principal neuron is ''on,'' which corresponds to a particular pattern. The order in which the principal neurons are turned on is completely determined by the structure of the WLC matrix V_{ij}. The duration of each

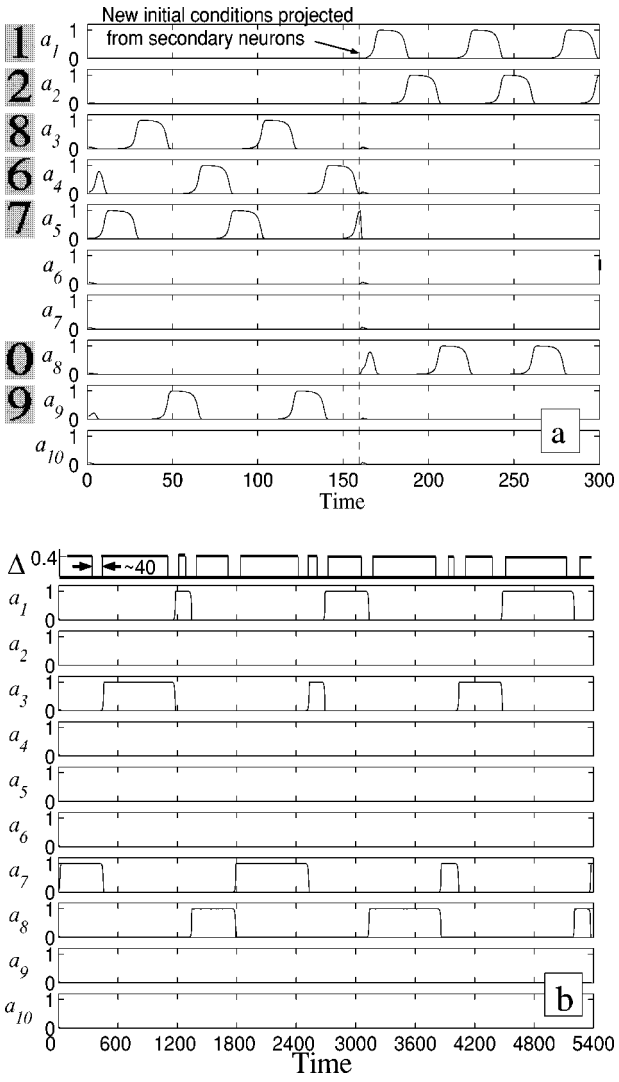

FIG. 2. Amplitudes of principal neurons during the memory retrieval phase, (a) periodic retrieval, two different test patterns presented, (b) aperiodic retrieval with modulated inhibition. Parameters of simulations are the same as in Fig. 1.

''state'' is determined by the magnitude of external perturbations σ. For $\sigma=0$, the system would asymptotically approach the separatrices and so the durations of each state would grow indefinitely. For a finite σ, the duration of each state scales as $-\ln\sigma$.

In the above example, patterns are retrieved from the spatial memory periodically in time. However, it may be desirable for a system to be able to control the duration of individual patterns in the sequence. This can be easily achieved by modulating the overall strength of inhibitory connections $V_{ij}+\Delta(t)$. While $\Delta(t)>1-V_1$, all fixed points are stable nodes, and so a single principle neuron keeps firing. In order to advance to the next pattern, $\Delta(t)$ is suppressed to zero for a short period of time $[O(-\ln\sigma)]$. An example of such nonperiodic retrieval of images is shown in Fig. 2(b).

In conclusion, we introduced a principle of operation for the sequential memory which is based on the winnerless competition and illustrated it in the model of the sequential spatial memory in hippocampus. It is embodied in the two-

layer neuronal structure with the first layer serving as a sensory input for the second layer which performs winnerless competition among representative principal neurons. We introduced the learning rules for the projection and the competition matrices which lead naturally to the desired function of the network. We also demonstrated that external perturbations can influence the timing of the transitions among the stored patterns, however, the sequence of patterns is robust against external perturbations. The model can operate in two regimes: externally timed switching controlled by global modulation of inhibitory connections and spontaneous periodic switching between patterns. The latter can be relevant for route replays during sleep. Of course, our model only describes a generic mechanism of sequential memory, in real biological systems neurons generate nonstationary spike trains and synaptic dynamics is time dependent (see Refs. [7,10]). Moreover, instead of a single PN a given pattern can be represented by a group of neurons which would increase the structural stability of the memory. All these generalizations will be addressed in our future work.

The authors gratefully acknowledge support from the Engineering Research Program of the Office of Basic Energy Sciences at the U.S. Department of Energy, Grant No. DE-FG03-96ER14592, and from NSF Grant No. EIA-013708.

[1] J. O'Keefe and J. Dostrovsky, Brain Res. **34**, 171 (1971); J. O'Keefe and N. Burgess, Nature (London) **381**, 425 (1996); M. Wilson and B.L. McNaughton, Science **261**, 1055 (1993).

[2] J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, New York, 1978); N.A. Scmjuk, A.D. Thime, and H.T. Blair, Hippocampus **3**, 387 (1993); R. Muller, Neuron **17**, 979 (1996); A. Samsonovich and B. McNaughton, J. Neurosci. **17**, 5900 (1997); K.I. Blum and L.F. Abbott, Neural Comput. **8**, 85 (1996); A.D. Redish and D.S. Touretzky, *ibid.* **10**, 73 (1998).

[3] E.R. Wood, P.A. Dudchenko, and H. Eichenbaum, Nature (London) **397**, 613 (1999); H. Eichenbaum and P. Dudchenko, Neuron **23**, 209 (1999).

[4] D. Kleinfeld, Proc. Natl. Acad. Sci. U.S.A. **83**, 9469 (1986); H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861 (1986); S. Amari and H. Yani, in *Associative Neural Memories: Theory and Implementation*, edited by H. Hassoun (Oxford University Press, New York, 1993), p. 170; L.F. Abbott and K.I. Blum, Cereb. Cortex **6**, 406 (1996).

[5] J.J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[6] T.V.P. Bliss and G.L. Collingridge, Nature (London) **361**, 31 (1993).

[7] H. Markram, J. Lubke, J. Froscher, and B. Sakmann, Science **275**, 213 (1997).

[8] M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H.D.I. Abarbanel, and G. Laurent, Phys. Rev. Lett. **87**, 068102 (2001).

[9] B. Baird and F. Eeckman, in *Associative Neural Memories: Theory and Implementation*, Ref. [4], p. 135.

[10] G. Bi and M. Poo, Annu. Rev. Neurosci. **24**, 139 (2001).