

## Dynamical Encoding by Networks of Competing Neuron Groups: Winnerless Competition

M. Rabinovich,<sup>1</sup> A. Volkovskii,<sup>1</sup> P. Lecanda,<sup>2,3</sup> R. Huerta,<sup>1,2</sup> H.D.I. Abarbanel,<sup>1,4</sup> and G. Laurent<sup>5</sup>

<sup>1</sup>*Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093-0402*

<sup>2</sup>*GNB, E.T.S. de Ingeniería Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain*

<sup>3</sup>*Instituto de Ciencia de Materiales de Madrid, CSIC Cantoblanco, 28049 Madrid, Spain*

<sup>4</sup>*Department of Physics and Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 93093-0402*

<sup>5</sup>*California Institute of Technology, Division of Biology, MC 139-74, Pasadena, California 91125*

(Received 29 December 2000; published 20 July 2001)

Following studies of olfactory processing in insects and fish, we investigate neural networks whose dynamics in phase space is represented by orbits near the heteroclinic connections between saddle regions (fixed points or limit cycles). These networks encode input information as trajectories along the heteroclinic connections. If there are  $N$  neurons in the network, the capacity is approximately  $e(N-1)!$ , i.e., much larger than that of most traditional network structures. We show that a small winnerless competition network composed of FitzHugh-Nagumo spiking neurons efficiently transforms input information into a spatiotemporal output.

DOI: 10.1103/PhysRevLett.87.068102

PACS numbers: 87.10.+e, 05.45.-a, 87.18.Bb

Information about the environment is generally encoded into spike sequences by neurons in animal sensory nervous systems [1]. There is a growing body of evidence [2–6] that, in some systems, the representation of information comes through both identity and temporal encoding: each stimulus is characterized by a specific and reproducible sequence of firing across specific neurons [7]. To build a reasonable dynamical theory of such an encoding, we must first understand the principles on which the dynamics of these sensory networks is based and predict some advantages that such stimulus representation has for further processing.

We explore a class of dynamical systems that we call competitive networks or winnerless competition (WLC) networks. These produce identity-temporal or spatiotemporal coding in the form of deterministic trajectories moving along heteroclinic orbits that connect saddle fixed points or saddle limit cycles in the system's state space. These saddle states correspond to the activity of specific neurons or groups of neurons, and the separatrices connecting these states correspond to sequential switching from one state to another.

We use observed features of olfactory processing networks [2] as a guide to our study of computation using competitive networks. Figure 1 shows the simultaneously recorded activity of three different projection neurons (PNs) in the locust antennal lobe (AL) evoked by two different odors: despite similar PN activities before the stimulus onset (the result of the action of noise) each odor evokes a specific spatiotemporal activity pattern that results from interactions between these and other neurons in the network [2,8].

Using this experimental data and knowledge about the anatomy and physiology of the AL we hypothesize that such olfactory networks form, store, and recognize patterns using a WLC strategy. The experiments on which our

ideas are based indicate the following features of neural encoding: the representation of input (sensory) information (i) uses both identity (or "space") and time, (ii) sensitively depends on the stimulus, (iii) is deterministic and reproducible, and (iv) is robust against noise. These observations suggest (a) that a dynamical system which possesses

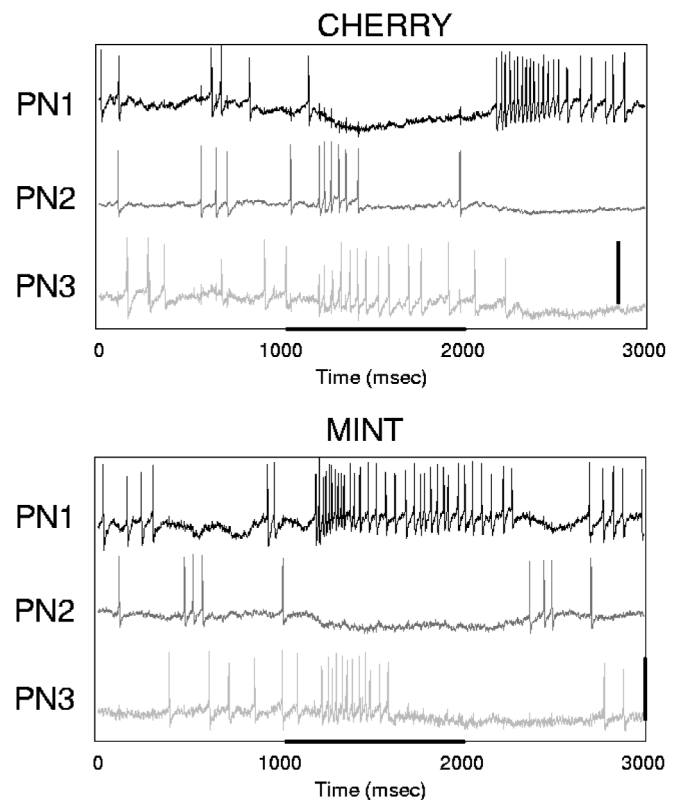


FIG. 1. The temporal patterns produced by three simultaneously sampled PNs in the locust antennal lobe when two different odors are presented. The horizontal bar indicates the time interval when the stimulus was presented (see [8] for details).

these characteristics should be strongly dissipative so that orbits rapidly “forget” the state of the system when the stimulus is turned on and (b) that the system represents information by transient trajectories, rather than attractors (regular or strange), of the unstimulated system. The system cannot have multistability because noise could drive the system to an inappropriate “representation,” possibly that of a different environmental stimulus. Our competitive systems possess the desired characteristics.

We illustrate our ideas with examples from networks of  $N$  neurons with dynamical variables  $\mathbf{y}_i(t) = [y_i^1(t), y_i^2(t), \dots, y_i^m(t)]$ ;  $i = 1, 2, \dots, N$ .  $y_j^1(t)$  is the neuron membrane potential. We consider

$$\frac{d\mathbf{y}_i(t)}{dt} = F[\mathbf{y}_i(t)] - \sum_{j=1}^N G_{ij}(\mathbf{S}) \cdot [y_j^1(t), y_j^1(t)] + \tilde{S}_i(t), \quad (1)$$

where  $F(\mathbf{y}_i)$  is the nonlinear function that describes the dynamics of individual neurons.  $G_{ij}(\mathbf{S}) \cdot [\dots]$  is a nonlinear operator describing inhibitory action of the  $j$ th neuron onto the  $i$ th neuron.  $\mathbf{S}(t) = \{S_i(t)\}$  and  $\tilde{\mathbf{S}}(t)$  are the vectors representing stimuli to the network.

From experiments [2] we infer that a stimulus acts in two principal ways: (i) it excites a subset of neurons through the additive  $\tilde{\mathbf{S}}(t)$  in (1); (ii) it modifies the effective inhibitory connections between the neurons through  $G_{ij}(\mathbf{S})$  in (1). In the insect antennal lobe, for example, this modification occurs as a result of activation of the inhibitory interneurons that connect different PNs, e.g., different  $\mathbf{y}_i$ . The intrinsic dynamics of these neurons is governed by many variables corresponding to ion channels and intracellular processes. Such a detailed description, however, is not needed to illustrate the principle of “coding with separatrices.” We need only capture the “firing” or “not-firing” state of the component neurons. We thus simplify our model to an equation for the firing rate  $a_i(t) > 0$  of neural activity:

$$\frac{da_i(t)}{dt} = a_i(t) \left[ \sigma_i(\mathbf{S}) - \left( a_i + \sum_{j \neq i}^N \rho_{ij}(\mathbf{S}) a_j(t) \right) \right] + \tilde{S}_i(t). \quad (2)$$

We used here the nonlinear inhibitory operator  $G$ :  $G(\mathbf{S}) \cdot [a_i, a_j] = \rho_{ij}(\mathbf{S}) a_i a_j$ , where  $\rho_{ij}(\mathbf{S})$  is the strength of inhibition of neuron  $j$  onto  $i$ .  $\sigma_i(\tilde{\mathbf{S}}) = -1$  when there is no stimulus, and  $+1$  when the stimulus has a component at neuron  $i$ . When  $\sigma_i = -1$ , the quiet resting state  $a_i = 0$  is stable. When a stimulus is applied and  $\sigma_i = +1$ , the system moves away from this quiet state onto a sequence of heteroclinic trajectories. This instability triggers the system into rapid action, provides robustness against noise, and allows a response independent of the state at stimulus onset. We neglect the dynamics of the synaptic connections because we do not account for spike dynamics. For the “average” description of neurons, synapses can be considered as “nonlinear triggers.” [Equation (2) is a

Lotka-Volterra equation.] These have been analyzed in detail for  $N = 3$  and  $\mathbf{S} = \tilde{\mathbf{S}} = 0$  in [9–11].

To begin, we ignore the additive sensory input in (2). This tells us how the network operates and allows an estimation of its encoding capacity. Information about the input resides solely in the couplings  $\rho_{ij}(\mathbf{S})$ . If the inhibitory connections for a specific stimulus are symmetric  $\rho_{ij} = \rho_{ji}$  and identical  $\rho_{i \neq j} = \rho$ ,  $\rho_{ii} = 1$ , the dynamics is very simple. For weak coupling,  $\rho < 1$ , the system has a global attractor: the stable fixed point  $a_i = \frac{1}{1 + \rho(N-1)}$  corresponding to simultaneous activity of all neurons. If  $\rho > 1$  the system is multistable: depending on the initial conditions, one neuron becomes active and the others are quiet. When the inhibitory connections are not symmetric, the system with  $N$  competitive neurons has different closed heteroclinic orbits that consist of saddle points and one-dimensional separatrices connecting them. Such heteroclinic orbits are global attractors in phase space and are found in various regimes of the  $\rho_{ij}(\mathbf{S})$ . This implies that, if the stimulus is changed, another orbit in the vicinity of the heteroclinic orbit becomes a global attractor for this stimulus.  $C$ , the capacity of the network, indicates the number of different items which the network may thus encode through its activity. To have a closed heteroclinic orbit,  $N \geq 3$ . Each saddle fixed point on the closed heteroclinic orbit must have one positive eigenvalue. The other  $N - 1$  are negative so that all other directions are attracted to the saddle. Movement from saddle point to saddle point results. The sequence from fixed point  $i$  to fixed point  $j$  must occur when  $\rho_{ii} = 1$ ,  $\rho_{ij} > 1$ , and  $\rho_{ji} < 1$  [12]. The only positive eigenvalue of the Jacobian at the fixed point is  $J = 1 - \rho_{ji}$ .

$C$  can be estimated as follows. If we satisfy the conditions for one heteroclinic orbit to exist, we can build another from it by permuting the indices of the  $a_i$  and of the matrix  $\rho_{ij}$ . There are  $N!$  permutations of the indices. Some of these generate the same heteroclinic orbit: firing as  $(1, 2, 3, 4, 5)$  or  $(2, 3, 4, 5, 1)$  is equivalent. For a given permutation there are  $N - 1$  permutations that are cyclically equivalent. The number of heteroclinic orbits involving all  $N$  neurons is  $(N - 1)!$ . There are still more heteroclinic orbits. These are associated with the  $N - 1$ ,  $N - 2, \dots$  dimensional subspaces which can be selected by eliminating one saddle point at a time from allowed orbits. The total number of these is the capacity  $C$ ,

$$C = \sum_{k=3}^N \binom{N}{k} (k - 1)! = N! \sum_{k=3}^N \frac{1}{k(N - k)!}$$

so  $C > N! \frac{1}{N} \sum_{k=0}^{N-3} \frac{1}{k!}$  and  $C < N! \frac{1}{3} \sum_{k=0}^{N-3} \frac{1}{k!}$ . For large  $N$ ,

$$\left( 1 - \frac{1}{e(N - 2)!} \right) < \frac{C}{e(N - 1)!} < \frac{N}{3} \left( 1 - \frac{1}{e(N - 2)!} \right).$$

A WLC network’s rich behavior can be illustrated when

we connect spiking FitzHugh-Nagumo (FN) [13] model neurons with the inhibitory interactions of our general framework. The network we studied is described by ( $i = 1, 2, \dots, 9$ ):

$$\begin{aligned} \tau_1 \frac{dx_i(t)}{dt} &= f[x_i(t)] - y_i(t) - z_i(t)[x_i(t) - \nu] \\ &\quad + 0.35 + S_i, \\ \frac{dy_i(t)}{dt} &= x_i(t) - by_i(t) + a, \\ \tau_2 \frac{dz_i(t)}{dt} &= \sum_j g_{ji} G[x_j(t)] - z_i(t). \end{aligned} \quad (3)$$

Here we use a dynamical model of inhibition:  $z_i(t)$  is a synaptic current modeled by first-order kinetics. The variable  $x_i(t)$  denotes the membrane potential,  $y_i(t)$  is a recovery variable, and  $f(x) = x - \frac{1}{3}x^3$  is the internal FN nonlinearity. The stimulus is taken as a constant. We use a step function for  $G(x) = 0, x \leq 0$ , and  $G(x) = 1, x > 0$ , as the synaptic connection.  $S_i$  is the stimulus, and  $g_{ji}$  is the strength of synaptic inhibition:  $g_{ji} = 2$  if the  $j$ th neuron inhibits the  $i$ th; 0 otherwise. The other parameters are  $a = 0.7, b = 0.8, \tau_1 = 0.08, \tau_2 = 3.1$ , and  $\nu = -1.5$ .

Our numerical simulations show that the network produces different spatiotemporal patterns in response to different stimuli. Figure 2 presents examples of these activities corresponding to two different stimuli. The system was in the resting state  $x_i \approx -1.2, y_i \approx -0.62, z_i = 0$  before the stimulus began at  $t = 0$ . As one can see, the patterns are considerably different and distinguishable.

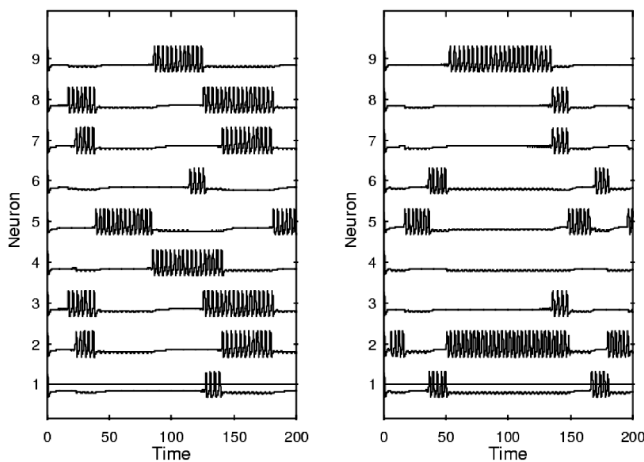


FIG. 2. The spatiotemporal patterns generated by a network of nine FitzHugh-Nagumo neurons with inhibitory connections. We used the external stimuli:  $S_1 = 0.1, S_2 = 0.15, S_3 = 0, S_4 = 0, S_5 = 0.15, S_6 = 0.1, S_7 = 0, S_8 = 0, S_9 = 0, \tau_2 = 3.1$  (left), and  $S_1 = 0.01, S_2 = 0.03, S_3 = 0.05, S_4 = 0.04, S_5 = 0.06, S_6 = 0.02, S_7 = 0.03, S_8 = 0.05, S_9 = 0.04, \tau_2 = 4.1$  (right). The coefficients  $g_{15}, g_{52}, g_{21}, g_{24}, g_{45}, g_{52}, g_{65}, g_{26}, g_{36}, g_{53}, g_{74}, g_{57}, g_{84}, g_{58}, g_{86}, g_{89}$ , and  $g_{95}$  are equal to 2 and the other  $g_{ji}$  are equal to 0. We plot  $x_i(t)$  versus time.

To characterize the network as a processing device, we ask how well it carries information from input to output and ask if temporal coding is needed to make this communication effective. We first specify how we represent input and output information; we must define the input and output coding spaces. In our nine neuron model (3) each element may receive an excitatory input  $S_i$ . The input signal is binary: each element is excited or not. So we can represent the net input as a nine-digit binary number  $d = d_1, \dots, d_9$ :  $d_i = 1$ , if  $S_i > 0$ , and  $d_i = 0$ , if  $S_i = 0$ , for  $i = 1, \dots, 9$ . The output of each element  $x_i(t)$  is also considered to be binary. At any time the outputs of all nine elements can be given as a nine-digit binary number  $b = b_1, \dots, b_9$  with  $b_i = 1$ , if  $x_i > 0$ , and  $b_i = 0$ , if  $x_i \leq 0$ . The time-dependent output over a time  $KL$  is represented as  $K$  sequences of binary numbers of length  $L$ :  $Q_k^L = \{b^1, b^2, \dots, b^L\}, k = 1, 2, \dots, K$ .

To investigate the robustness and reproducibility of the representation of the stimulus we generated  $D$  random input signals  $d^1, \dots, d^D$ , and then evaluated the output from Eq. (3) for each input  $d^i$  using different network initial conditions selected randomly from inside a sphere in state space with radius  $r$ , centered at zero. A total of  $K$  different output sequences  $Q_k^L, k = 1, \dots, K$  were recorded corresponding to all inputs  $d^1, \dots, d^D$  and all initial conditions. The calculations were performed for two different radii  $r$  of initial conditions to examine reliability against noise. To illustrate the dependence of information transfer on the length  $L$  of the output sequence, namely, the role of time in the output encoding, we determined the average mutual information  $I(L)$  between input and output sequences

$$I(L) = \sum_{i=1}^D \sum_{k=1}^K P(d^i, Q_k^L) \log \left\{ \frac{P(d^i, Q_k^L)}{P(d^i)P(Q_k^L)} \right\},$$

where  $P(d^i, Q_k^L)$  is the joint probability distribution of inputs and outputs,  $P(d^i)$  is the distribution of inputs, and  $P(Q_k^L)$  is the probability of output sequences. We drew the  $d^i$  from a uniform distribution:  $P(d^i) = 1/D$ .  $I(L)$  answers the question: On average over all inputs and all outputs, how much information about the input sequences  $d^i$  do we learn by observing the output sequences  $Q_k^L$ ?  $I(L)$ , as a function of the output sequence length  $L$  with different radii  $r$  of initial conditions, is presented in Fig. 3. One can see that  $I(L)$  reaches its maximum allowed value, the entropy of the input signal, when  $L \geq 4$ .  $D = 10$  in these calculations. From these results we see that full information about the inputs (sensory stimuli) can be found in the outputs (heteroclinic orbits connecting the clusters of FN models) for output sequences larger than 3. Further, this occurs in a manner largely insensitive to the system noise level provided one decodes this sequence of length  $L > 3$ . We conclude that temporal encoding, namely,  $L > 1$ , is required to achieve a good representation of the input data.

We have investigated a class of neural network models whose stimulus-dependent dynamics reproduces the rich

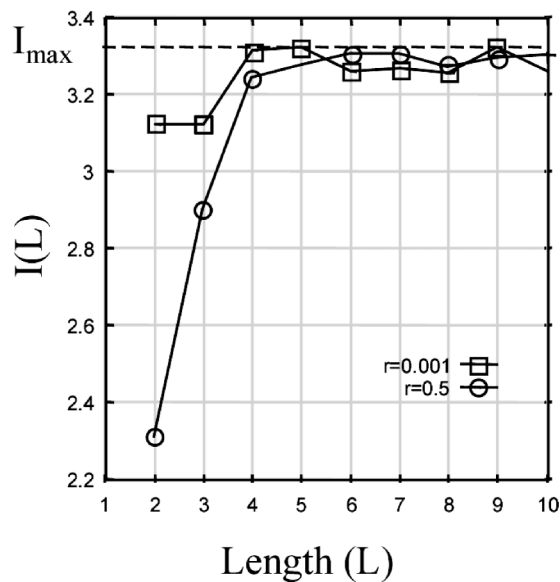


FIG. 3. Average mutual information  $I(L)$  as a function of output sequence length  $L$  and noise radius  $r$ .

spatiotemporal features observed in insect olfactory systems [2,7]. The principle of such a representation relies on the fact that specific trajectories act as global attractors in state space. This behavior corresponds to experimental data, showing that activity proceeds across parts of the insect olfactory processing network by the sequential activation and deactivation of subgroups of neurons. This “activation path” within the network is the global attractor which appears under the action of the stimulus in the neighborhood of the heteroclinic orbit. A stimulus can thus be thought of as an informational signal that reorganizes the global attractor in a stimulus-specific manner, forcing the system to evolve through state space along a transient, but deterministic, path joining unstable “saddle states.” Once the stimulus ceases, each active neuron returns to its baseline activity, controlled by intrinsic properties, basal connection strengths, and noise. The properties of these simplified models can also emerge from more realistic networks having the same dominance of asymmetrical inhibitory connections. We conjecture that a large network with sparse, random connections will effectively exhibit the same stimulus-dependent sequential activation and deactivation of subgroups of neurons [8]. WLC encoding is very sensitive to the forcing stimulus. This is because the heteroclinic linking of a specific set of saddle points or cycles is always unique. Two like stimuli, activating greatly overlapping subsets of the network, may thus

become easily separated and recognized at the next step of the processing because small initial differences will be amplified in time. This is verified experimentally in the fish olfactory bulb [14]. Our central idea does not depend on the nature of the stimulus. It may thus apply to brain circuits other than olfactory processing systems. It may perhaps underlie interesting experimental observations such as the flipping between quasistationary states of activity seen in a monkey cortex [15]. Beyond the biological observations which suggested these investigations, WLC networks provide an attractive model for computation because of their large capacity as well as their robustness to noise contamination.

We thank Mark Stopfer for experimental data. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Engineering and Geosciences, under Grants No. DE-FG03-90ER14138 and No. DE-FG03-96ER14592, by NIDCD (G. L.), and by M. Ciencia y Tecnología BF12000-0157 (R. H.).

- 
- [1] F. Rieke *et al.*, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1997).
  - [2] G. Laurent, M. Stopfer, R. W. Friedrich, M. Rabinovich, A. Volkovskii, and H. D. I. Abarbanel, *Annu. Rev. Neurosci.* **24**, 263 (2001).
  - [3] A. J. Hudspeth and N. K. Leggett, *Curr. Opin. Neurobiol.* **10**, 631 (2000).
  - [4] G. Buzsaki and J. J. Chrobak, *Curr. Opin. Neurobiol.* **5**, 504 (1995).
  - [5] W. Singer and C. M. Gray, *Annu. Rev. Neurosci.* **18**, 555 (1995).
  - [6] P. Dayan and L. F. Abbott, <http://play.ccs.brandeis.edu/~abbott/book/TOC.html>
  - [7] M. Wehr and G. Laurent, *Nature (London)* **384**, 162 (1996).
  - [8] M. Bazhenov, M. Stopfer, M. Rabinovich, H. D. I. Abarbanel, T. Sejnowski, and G. Laurent, *Neuron* **30**, 569 (2001).
  - [9] R. M. May and W. I. Leonard, *SIAM J. Appl. Math.* **29**, 243 (1975).
  - [10] S. Grossberg, *J. Theor. Biol.* **73**, 101 (1978).
  - [11] T. Fukai and S. Tanaka, *Neural Comput.* **9**, 77 (1977).
  - [12] V. Afraimovich and M. I. Rabinovich (to be published).
  - [13] R. FitzHugh, *Biophys. J.* **1**, 445 (1961); J. Nagumo, S. Arimoto, and S. Yoshizawa, *Proc. IRL* **50**, 2061 (1960).
  - [14] R. W. Friedrich and G. Laurent, *Science* **291**, 889 (2001).
  - [15] M. Abeles, H. Bergman, I. Gat, E. Seideman, N. Tishby, and E. Vaadia, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8616 (1995).