

*This paper was presented at a colloquium entitled “Physics: The Opening to Complexity,” organized by Philip W. Anderson, held June 26 and 27, 1994, at the National Academy of Sciences, in Irvine, CA.*

## Rapid local synchronization of action potentials: Toward computation with coupled integrate-and-fire neurons

(neurobiology/phase locking/Lyapunov function/pulse-coupled oscillators)

JOHN J. HOPFIELD\* AND ANDREAS V. M. HERZ†‡

\*Divisions of Biology and Chemistry, California Institute of Technology, Pasadena, CA 91125; and †Beckman Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801

**ABSTRACT** The collective behavior of interconnected spiking nerve cells is investigated. It is shown that a variety of model systems exhibit the same short-time behavior and rapidly converge to (approximately) periodic firing patterns with locally synchronized action potentials. The dynamics of one model can be described by a downhill motion on an abstract energy landscape. Since an energy landscape makes it possible to understand and program computation done by an attractor network, the results will extend our understanding of collective computation from models based on a firing-rate description to biologically more realistic systems with integrate-and-fire neurons.

Most neurons communicate through action potentials—discrete short pulses of electrochemical activity. Action potentials are generated when the membrane potential of a neuron reaches a threshold value. They propagate along the axon of a cell toward synapses with postsynaptic neurons where they initiate ion currents that trigger (or inhibit) action potentials of the postsynaptic cell. Most useful network neuro-modeling, in contrast, is based on model cells whose output is described as a continuous variable that is slowly varying in time. The output variable is usually interpreted as a short-time average of the rate of production of action potentials.

While it may frequently be the case that mean firing rates are an adequate description of neural information, there are many instances where the detailed timing and organization of action potentials matter. The sonar processing of bats (1), sound localization of owls (2), and electrosensation of weakly electric fish (3) are based on time-comparison circuits that discriminate signals with a resolution in the submillisecond range. Specific stimulus-dependent synchronization of action potentials has been observed in the olfactory system of the locust (4) and in the primary visual cortex of cats (5, 6) and has been suggested to be of “neurocomputational” use in binding features together (7). Experiments have also shown that neurons in frontal cortex exhibit pronounced stimulus-dependent temporal correlations (8). All these different experimental results give strong impetus to studying how to do computations based on action potential timing.

The modeling using action potentials has been of two natures. Analytical work has been done chiefly on networks in which each neuron has equal synaptic coupling strength to all other neurons (9–15). Although these networks exhibit interesting dynamical features, their connectivity is too simple to support biologically relevant computation. At the other extreme, numerical studies of much more elaborate networks, with multicompartment cells and axonal propagation delays,

have been carried out (16–21). The simulations display rich behavior, but the networks are completely intractable from a mathematical point of view. Thus our ability to understand and generalize from such simulations is rather limited.

In this paper we analyze the dynamics of networks of integrate-and-fire neurons with local or global interactions. We prove that various models reach a complex but cyclic attractor in a short time. One model has a particularly simple mathematical structure that admits a Lyapunov function. Its convergence to periodic limit cycles may thus be understood as a downhill march in an abstract energy landscape. Simulations of the other integrate-and-fire models display almost identical dynamical behavior in the short-time regime, while deviating at longer times. However, only the short-term behavior of the models seems biologically relevant, since computational decisions must be taken rapidly, and in any event the assumption of constant input from other areas implicit in all models breaks down at longer times. By focusing on short-time aspects (and by not making any mean-field approximation), our approach differs from other mathematical studies of networks of integrate-and-fire neurons with nonuniform interactions (22–26).

The existence of a Lyapunov function for a simple structure of coupled model neurons has previously allowed an understanding of computation by attractors and of associative memory (27). We believe it will similarly be possible to understand some aspects of collective computation based on action potentials through the Lyapunov function of the integrate-and-fire model and that the similarity of the short-term behavior of the other models will extend this understanding to a broad class of networks.

### The Model Systems

We consider networks of interconnected integrate-and-fire model neurons. Below the firing threshold, each neuron  $i$  operates as a leaky integrator,

$$C(du_i/dt) = -[u_i(t) - u_0]/R + I_i(t). \quad [1]$$

In the absence of the input current  $I_i(t)$ , the cell potential  $u_i(t)$  drifts to its rest value  $u_0$ . The term  $u_0/R$  can be absorbed in  $I_i(t)$ , and we will, therefore, focus on the case where  $u_0 = 0$ . By rescaling time, the capacitance  $C$  and the resistance  $R$  of the cell membrane can be taken as unity except for the limiting cases  $C \rightarrow 0$  or  $R \rightarrow \infty$ . The latter describes a perfectly integrating cell,

$$du_i/dt = I_i(t). \quad [2]$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡Present address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, England.

When the potential of cell  $i$  reaches the threshold  $u_{\text{thresh}}$ , the cell produces an action potential of negligible duration and resets its potential to  $u_{\text{reset}}$ . For convenience, units can be chosen such that  $u_{\text{thresh}} = 1$  and  $u_{\text{reset}} = 0$ .

When an action potential arrives at a synapse from cell  $j$  to cell  $i$ , a synaptic current briefly flows. This current is sometimes modeled as a step followed by an exponential decay or by an  $\alpha$  function. In either case, the integral of the current flow is finite, and the current rapidly decays. Within the present description, the duration of this current is set to zero. By assuming a linear summation of the inputs and vanishing signal delays, the current  $I_i(t)$  into cell  $i$ ,  $1 \leq i \leq N$ , is then given by

$$I_i(t) = \sum_j T_{ij} f_j(t) + I_i^{\text{ext}}(t), \quad [3]$$

where the instantaneous firing rate  $f_j(t)$  is a sum of Dirac  $\delta$  functions

$$f_j(t) = \sum_n \delta(t - t_j^n). \quad [4]$$

$t_j^n$  represents the times at which neuron  $j$  generates an action potential,  $T_{ij}$  represents the strength of the synaptic connection from neuron  $j$  to neuron  $i$ , and  $I_i^{\text{ext}}(t)$  denotes an external input current that is assumed to be constant in time unless otherwise stated.

The general behavior of the system is now as follows. While none of the neurons is producing an action potential, Eq. 1 can be integrated and yields (for  $C = R = 1$ )

$$u_i(t) = [u_i(t_0^+) - I_i^{\text{ext}}] e^{-(t-t_0)} + I_i^{\text{ext}} \quad \text{for } t \geq t_0, \quad [5]$$

where  $t_0$  denotes the last firing time. When the potential  $u_j$  of neuron  $j$  reaches 1 (the threshold), it drops instantaneously to 0. At the same time, the potential  $u_i$  of each neuron  $i$  to which neuron  $j$  makes a synapse is increased by  $T_{ij}$ .

Because the duration of action potentials and of synaptic currents has been set equal to zero, the description given so far contains an ambiguity. To which value should neuron  $i$  be reset if at time  $t$  an action potential is produced by cell  $j$ ,  $T_{ij} > 0$ , and  $u_i(t^-) > 1 - T_{ij}$ ? For in this case, the action potential will raise  $u_i$  above 1, and cell  $i$  should generate its action potential during the flow of synaptic current produced by the synapse  $T_{ij}$ . When synaptic (and dendritic) time constants are longer than the duration of action potentials, what should actually happen is that cell  $j$  should fire when its potential reaches  $u_{\text{thresh}} = 1$ , and the synaptic current from synapse  $T_{ij}$  that arrives after cell  $i$  fires should be integrated to yield a positive potential (relative to  $u_{\text{reset}}$ ) afterward. Thus, if cell  $j$  fires first and at time  $t$  and that event evokes a firing of neuron  $i$ , then after both action potentials have been generated, the two membrane potentials should be

$$u_i(t^+) = T_{ji} \quad [6]$$

and

$$u_i(t^+) = u_i(t^-) + T_{ij} - 1. \quad [7]$$

The first equation represents the fact that neuron  $j$  fired first when  $u_j = 1$  and was reset to 0, and when subsequently neuron  $i$  generated its action potential, this changed the potential of neuron  $j$  to  $T_{ji}$ . The second equation represents the fact that neuron  $i$  fired second and reduced its potential by 1 when it did so but received the synaptic current  $T_{ij}$  when neuron  $j$  fired.

The updating rule can be generalized to a large network of neurons by the following algorithm. As the potentials all increase with time, a first neuron  $j$  reaches  $u_j = 1$ . Reset that potential to zero, and then change the potential of each neuron  $i$  by  $T_{ij}$ . If, by following this procedure, some of the potentials become  $\geq 1$ , pick the neuron with the largest potential, say

neuron  $k$ , decrease its potential by 1, and then change the potential of each neuron  $l$  by  $T_{lk}$ . Continue the procedure until no membrane potential is  $\geq 1$  and then "resume the flow of time" and again let each potential  $u_i$  increase according to Eq. 5.

This deterministic algorithm preserves the essence of the idea that firing an action potential carries a neuron from  $u_{\text{thresh}}$  to  $u_{\text{reset}}$ , and effectively apportions the synaptic current into a part that is necessary to reach threshold and a part that raises the potential again afterward. Because the firing of one neuron can set off the instantaneous firing of others, this model can generate events in which many neurons are active simultaneously. Notice that throughout what follows, terms such as "synchronized neurons" always refer to the time of spike generation. According to this definition, a periodic network state (also called a "phase-locked solution") may or may not be "globally synchronized."

When synaptic (and dendritic) time constants are shorter than the duration of an action potential, all contributions from the synaptic current that arrive during spike generation are lost, and Eq. 7 should be replaced by  $u_i(t^+) = 0$ . Distinguishing between models with zero and nonzero leakage currents, we are led to four different scenarios:

Model A, the first leaky integrate-and-fire model, characterized by Eqs. 1, 3, 4, 6, and 7;

Model B, like model A but whenever an action potential is generated by neuron  $j$  at time  $t$ ,  $u_j$  is reset to 0, whether or not  $u_j(t^-) = 1$  or  $u_j(t^-) > 1$ ;

Model C, like model A but in the limit  $R \rightarrow \infty$  where Eq. 1 is replaced by Eq. 2;

Model D, like model C but with  $u_j$  reset to 0 as in model B.

To analyze the dynamic role of action potentials with fixed size, we include a fifth related model of geophysical origin (28): model E, like model D but with "action potentials" whose size is proportional to the membrane potential of the spiking unit so that if  $j$  is active at time  $t$ ,  $u_i$  is changed by  $T_{ij}u_j(t)$ .

The order in which the neurons are updated in an event in which several neurons fire at once is described above and has been used in all variants. However, for models A and C, the order does not matter as long as  $T_{ij} \geq 0$ . For these cases, any procedure for choosing the updating sequence of the neurons at or above threshold will yield the same result because the reset is by a fixed negative amount (herein:  $-1$ ) regardless of whether immediately prior to reset  $u = 1$  or  $u > 1$ .

In model C, the cumulative effects of action potentials and slow membrane dynamics commute if  $T_{ij} \geq 0$ . This makes the model equivalent to a class of "Abelian avalanche" models (29, 30). Closely related sandpile-like models relax to a critical state with fluctuations on all length scales, a phenomenon that has been coined "self-organized criticality" (31). The apparent similarity between the microscopic dynamics of such models and networks with integrate-and-fire neurons has led to speculations about a possible biological role of the stationary self-organized critical state (19, 25, 26). However, whereas for earthquakes, avalanches, and sandpiles, the main interest is in the properties of the stationary state, for neural computation, it is the convergence process itself that does the computation and is thus of particular interest (32).

### Simulations of Locally Connected Networks

The models allow for countless realizations. To perform numerical experiments and for thinking about the biological relevance, a specific connection pattern must be described. We choose to study simple planar networks with local connections, in crude correspondence to the structure of subareas of the neocortex. In neurobiology, the output of one neocortical area will provide time-varying inputs to another, but for simplicity,

the effects of this changing feedback will be omitted. Thus, when thinking about the dynamics of one patch of neurons, the behavior of the model after perhaps 200 ms is irrelevant from a biological point of view because the inputs from other areas would probably have changed on such a time scale. Equivalently, in the dynamical behavior of the system, we should be interested in phenomena that occur within the time necessary for a typical neuron to generate 20 spikes or less.

Within this conceptualization, information coming from the periphery or other areas of the brain can be thought of as either providing initial values for the  $u_i$  or defining the  $I_i^{\text{ext}}(t)$ . In the simulations and theoretical analysis, the first paradigm was used, and the external input was taken to be constant in time and equal for all neurons,  $I_i^{\text{ext}}(t) = I$ . This choice corresponds to the experimental situation of "stimulus-induced oscillations" (5), whereas the alternative scheme would model "stimulus-locked oscillations."

Square arrays with excitatory interactions and various boundary conditions were used to experimentally explore the similarities and differences between the variant models. For simplicity, the initial values of the  $u_i$  were chosen from a uniform random distribution between 0 and 1. Synapses of fixed strength  $\alpha$  joined each neuron with its four nearest neighbors. For such connection schemes with few interactions per neuron, even large networks with up to  $10^6$  neurons can efficiently be simulated on a workstation. (By using appropriate variable transformations, a fast algorithm proposed in ref. 34 can be extended to all realizations of models A–E.)

Fig. 1 A–C presents typical results of simulations with periodic boundary conditions. The abscissae denote the discrete times at which action potentials take place; the ordinates show the number of simultaneous action potentials. When viewed in space, large events are due to large connected ensembles of synchronized neurons. To keep the number of data points in a reasonable range, only results from relatively small networks with  $40 \times 40$  cells are shown.

All models exhibit rapid convergence to (approximately) periodic solutions with locally synchronized neural ensembles that are phase-locked with respect to each other. In models A and C, and for the parameters used in the simulations, this convergence process is completed after roughly five firing cycles. Due to the larger average reset, the transient is slightly longer in models B, D, and E. Further simulations show that the convergence time is almost independent of the system size and decreases quickly for decreasing  $\alpha$ .

What takes place on a larger time scale, however, is different for the variant models. Model A slowly reorganizes toward global synchronization. During this relaxation process, clusters of synchronized neurons usually merge to form larger clusters. However, as shown in Fig. 1A *Inset*, synchronized groups may also split. Model B reorganizes toward a reduced set of phase-locked solutions. In general, this process terminates before the state of global synchrony is reached (Fig. 1B *Inset*). Both reorganization processes are governed by local adaptations at the boundaries between synchronized clusters. Thus the time needed to reach a stationary solution increases with system size.

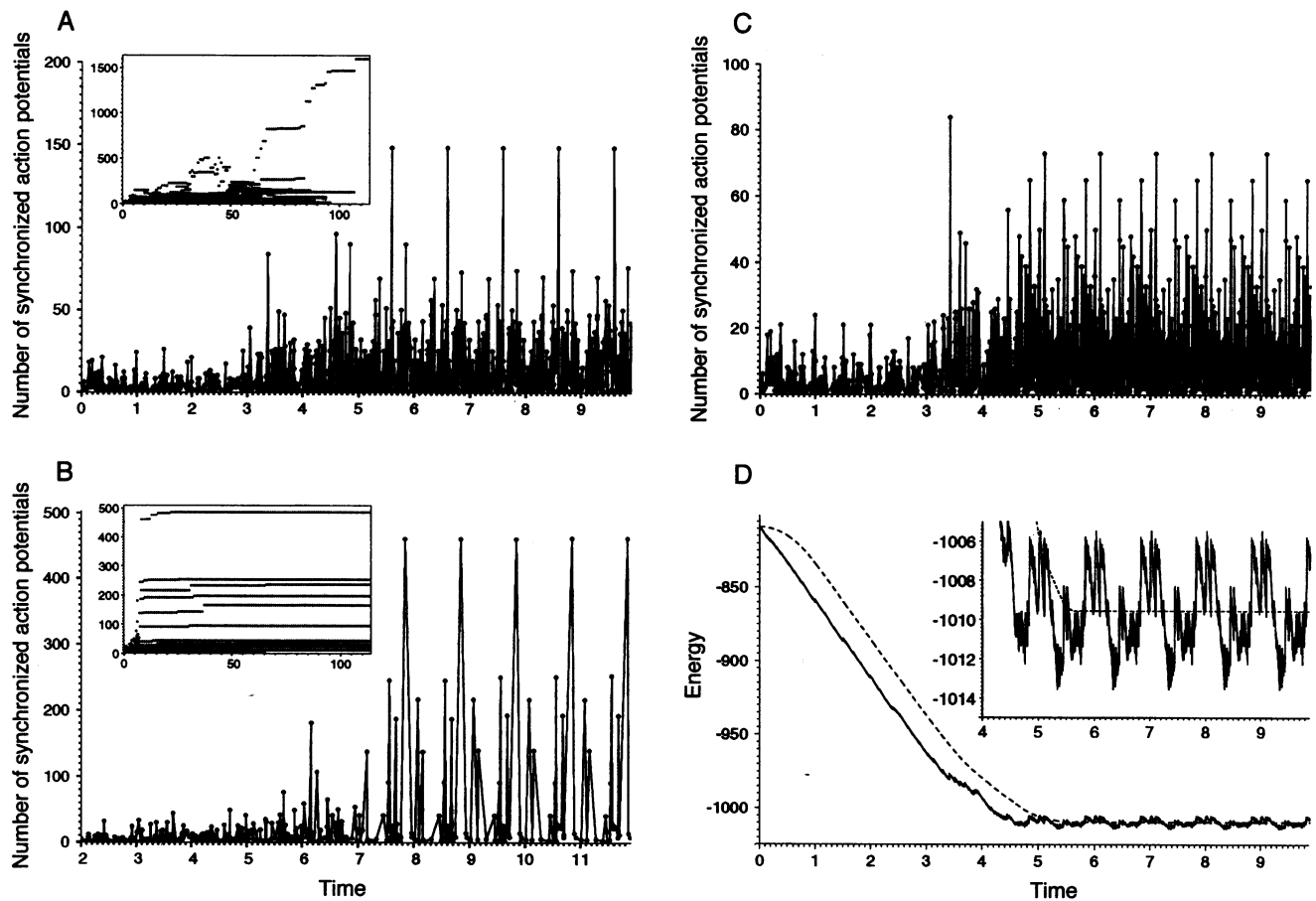


FIG. 1. Synchronization of action potentials. Results from numerical simulations of models A, B, and C with  $40 \times 40$  neurons, periodic boundary conditions,  $I = 10$ , and nearest-neighbor interactions of strength  $\alpha = 0.24$  are shown in A, B, and C, respectively. Each dot represents the number of simultaneous action potentials as a function of (normalized) time. All models exhibit rapid local synchronization of action potentials, followed by a slow convergence to globally synchronized solutions in model A (A *Inset*), and a slow reorganization to a reduced set of phase-locked solutions in model B (B *Inset*). Model C does not exhibit slow relaxation phenomena (C). This is also visible from the time evolution of the Lyapunov function  $E$  (solid line) and the Lyapunov functional  $L$  (dashed line) shown in D. As required, the latter approaches a constant value (D *Inset*).

Models C and D do not show any slow reorganization after the initial convergence to phase-locked oscillations (Fig. 1C). During one period of the oscillation, every neuron fires exactly once. For model C, this implies that the period is  $P = (1 - 4\alpha)/I$ : when neuron  $i$  fires at time  $t$ ,  $u_i$  is reduced by 1. Each of the four incoming action potentials increases  $u_i$  by  $\alpha$ . Between the firings,  $u_i$  increases at a rate  $I$ . At time  $t + P$ , the original value of  $u_i$  is reached again and the cycle is completed. Model D shows oscillations with the same period. For solutions without multiple spikes, this follows from the above argument. However, general limit cycles with period  $P$  and one spike per neuron and cycle can only occur if the membrane potentials do not exceed the threshold. Otherwise, they would decrease by more than one during the firing of the action potential but still only increase by one over the entire cycle. It follows that in events with multiple neurons firing "at the same time," the potentials become fine-tuned such that if neuron  $i$  is triggered by neuron  $j$  at time  $t$ ,  $u_i(t^-) = 1 - \alpha$ . (The same argument applies to  $P$ -periodic solutions of model E.) This implies that although every periodic firing sequence of model C can be realized in model D, the volume of all attractors is greatly reduced when measured in the space of the dynamical variables  $u_i$ .

In contrast to models A–D, that is, to systems with action potentials of fixed size, model E typically exhibits a few transient bursts with many neurons firing in synchrony, after which the system relaxes to simple periodic solutions with no or only a few synchronized clusters (34–36). A heuristic understanding of the difference in the collective behavior is readily obtained. In all models, a neuron  $i$  is absorbed into an adjacent synchronized cluster if its membrane potential  $u_i$  at the firing time  $t$  of the cluster is close enough to the firing threshold. In general,  $u_i(t)$  will be larger than the threshold and differ from cycle to cycle during the formation of the cluster. In model E, a fixed fraction of  $u_i(t)$  is transmitted to all neighbors, including cells that belong to the synchronized cluster. Changes in the pulse size may cause some cells to fire out of synchrony and thus result in a break-up of the cluster. In systems with fixed action potentials, this phenomenon cannot occur because the only information transmitted to neighboring cells is in the firing time. We hypothesize that it is this reduction of information that allows the dynamic stabilization of synchronized ensembles. According to this view, action potentials not only represent a static "yes–no" decision of an isolated neuron but also play an important dynamic role in networks of interconnected neurons.

To test the robustness of the synchronization process, various alterations of the network dynamics and architecture were studied. Time-dependent noise was included in Eq. 4 to model the stochastic nature of neurotransmitter release. Quenched noise was added to the reset value of individual neurons to allow different spontaneous firing rates. The simulations reveal that synchrony (as defined by the exact firing times) is not a robust feature of the system dynamics. We attribute this result in part to the strict measure of coherence used; interspike-interval distributions show only a gradual loss of synchrony with increasing noise level.

As an example of structural inhomogeneities, networks with open boundary conditions were studied. Neurons at the edges (and corners) of the simulated sheets receive less input and cannot sustain the same firing rate as neurons inside the network. The resulting dynamical perturbations are clearly visible in the long-time behavior. In models A and C, the size of synchronized clusters gradually changes since new neurons constantly enter and leave a cluster. In models B and D, however, the fine tuning described earlier is able to act as a counterbalance. This leads to an intermittent behavior where clusters stay unchanged over many iterations before they suddenly merge or break apart. (For a model of mechanical stick–slip friction that is equivalent to model D with  $\alpha = 1/4$ ,

this behavior has also been reported in ref. 37.) Another signature of open boundary conditions is a size-independent change of the dominant firing frequency in model D—from  $I/(1 - 4\alpha)$  to  $I/(1 - 3\alpha)$ —demonstrating the entrainment of the entire system by its boundary (38). Model E, finally, exhibits a power-law distribution of the cluster sizes with strong temporal fluctuations (28, 34–36). The results show again that pulses of fixed size are essential to stabilize locally synchronized activity patterns.

### Convergence in the Two Models Without Leakage

The rapid phase locking of models C and D observed in the numerical experiments can be verified analytically (38).

**THEOREM 1.** *Let all synapses be excitatory,*

$$T_{ij} \geq 0, \quad [8]$$

*let all neurons have the same sum  $A$  of incoming synaptic strengths,*

$$\sum_j T_{ij} = A < 1, \quad [9]$$

*and let the external input current be the same for all neurons,*

$$I_i^{\text{ext}} = I. \quad [10]$$

*Then for arbitrary initial condition  $(u_1, u_2, \dots, u_N)$ , networks C and D converge in finite time to a cyclic solution with period*

$$P = (I - A)/I. \quad [11]$$

*The attractor is reached as soon as every neuron has fired once. Depending on the initial condition, the periodic solution can be very complex and contain events in which one neuron fires alone and others in which many fire in synchrony. On the attractor, each neuron fires exactly once in a period.*

*Proof:* We first show that under the conditions of the theorem, no neuron fires more than once in any interval of length  $P$ .

**LEMMA.** *Let  $n_i(t, t')$  denote the number of times neuron  $i$  fires in the interval  $[t, t')$ . If Eqs. 8–10 hold, then  $n_i(t, t + P) \leq 1$ .*

*Proof of the Lemma:* Starting at time  $t$ , let  $i$  denote the first neuron that fires twice and let  $t'$  denote the time when it does so. The total change in  $u_i$  from  $t$  to  $t'$  due to synaptic currents and the external input must be greater than or equal to 1,

$$(t' - t) \frac{(1 - A)}{P} + \sum_j T_{ij} n_j(t, t') \geq 1. \quad [12]$$

Since neuron  $i$  is the first to fire twice, the number of firings of each of the other neurons up to  $t'$  is less than or equal to 1. For  $T_{ij}$  nonnegative, the lefthand side of Eq. 12 is, therefore, less than or equal to  $(t' - t)(1 - A)/P + A$ . This implies that Eq. 12 can only be satisfied if  $t' - t \geq P$ .

Returning to the proof of *Theorem 1*, let  $t_{\text{max}}$  denote the first time where every neuron has fired at least once. (Some cells may have fired repeatedly before  $t_{\text{max}}$ , depending on the parameter values and initial conditions.) For all neurons  $i$ , let  $t_i$  denote the last firing time before  $t_{\text{max}}$ ,  $t_{\text{min}}$  denote the minimum of all these times  $t_i$ , and  $k$  denote a cell that fires (without being triggered by other cells) at  $t_{\text{min}}$  for the last time.

By definition, every cell discharges at least once in the interval  $[t_{\text{min}}, t_{\text{max}}]$ . This implies in particular that every neuron  $j$  from which cell  $k$  receives synaptic input emits one or more action potentials in that interval. Each spike adds  $T_{kj}$  to  $u_k$ . The total change of  $u_k$  in  $[t_{\text{min}}, t_{\text{max}}]$  is thus equal or greater than  $A + I(t_{\text{max}} - t_{\text{min}})$ . This number has to be smaller than 1 because otherwise neuron  $k$  would fire a second time in the interval

$[t_{\min}, t_{\max}]$ , in contradiction to the assumption. It follows that  $t_{\max} - t_{\min} < P$ .

Combined with the *Lemma* (evaluated at time  $t = t_{\max} - P$ ), this means that every cell fires exactly once in  $[t_{\min}, t_{\max}]$  and no cell fires in  $(t_{\max} - P, t_{\min})$ . Since  $t_{\max} \leq I^{-1}$ , the last result proves that in finite time  $t_{\max} - P$ , a limit cycle is approached in the sense that  $u_i(t) = u_i(t + P)$  for  $t > t_{\max} - P$ . The argument also shows that the attractor is reached as soon as every neuron has fired once.

The proof does not depend in any way on the reset mechanism. This implies that it covers not only models C and D but also all intermediate schemes where a neuron  $i$  firing at time  $t$  is relaxed to some value between 0 and  $u_i(t^-) - 1$ . Perhaps surprisingly, this includes stochastic updating rules.

### Lyapunov Function for Model C

Model C admits a second unrelated convergence proof.<sup>§</sup> Consider the quantity

$$E = - \sum_i u_i. \quad [13]$$

**THEOREM 2.** *In addition to the conditions of Theorem 1, let all neurons also have the same sum A of outgoing synaptic strengths,*

$$\sum_i T_{ij} = A < I. \quad [14]$$

Then

$$E(t + P) \leq E(t). \quad [15]$$

On the attractor,  $E$  is cyclic with period  $P$ , though it can be a very complicated function within one cycle.

*Remark:* Notice that as in *Theorem 1*, symmetry of the synaptic connections,  $T_{ij} = T_{ji}$ , is not required.

*Proof:* The change of  $E$  in the interval  $P$  is

$$\begin{aligned} E(t + P) - E(t) &= -(1 - A)N - \sum_{i,j} T_{ij} n_j(t, t + P) + \sum_i n_i(t, t + P). \quad [16] \end{aligned}$$

The first term comes from the constant input current, the second term comes from the effect of the firing of other neurons, and the third term comes from neuron  $i$  itself firing. By using the sum rule (14), we find

$$E(t + P) - E(t) = -(1 - A)[N - \sum_i n_i(t, t + P)]. \quad [17]$$

Since due to the *Lemma*, the change of  $E$  in each time interval  $P$  is nonpositive and since  $E$  is bounded for all solutions, the system performs a (discrete) downhill march in the energy landscape generated by  $E$ —if  $E(t)$  is measured after time steps of fixed length  $P$  (for an illustration, see Fig. 1D). Independent from *Theorem 1*, the result implies that after a finite time each  $u_i(t)$  is periodic with period  $P$  and that every neuron fires once in any interval of length  $P$ .

To avoid the somewhat unfamiliar evaluation of the Lyapunov function  $E$  at the discrete times  $t + kP$ , one may alternatively use the functional  $L = \int_{t-P}^t E(s) ds$ . Along solutions of model C,  $L$  is differentiable with  $dL(t)/dt = E(t) - E(t - P)$  for all  $t \geq P$ , and the same conclusions are reached.

<sup>§</sup>A similar proof has been given by Gabrielov (29, 30).

### Behavior of the Two Models with Leakage

For concreteness, the mathematical analysis of models A and B will be restricted to the paradigm of Fig. 1: planar sheets of neurons with periodic boundary conditions, nearest-neighbor interactions of strength  $\alpha$ , and constant uniform external input  $I_i^{\text{ext}}(t) = I$ . Multiple instantaneous firings of the same neuron are excluded by taking  $4\alpha < 1$ , and by choosing  $I > 1$ , neurons cannot relax to a permanently quiescent state with subthreshold membrane potentials.

The most simple stationary solutions are periodic oscillations where all neurons fire in unison. At some time  $t_0$ , a first neuron reaches the firing threshold, resets to zero, and triggers (directly or indirectly) all four neighbors. The period  $P_L$  of the oscillation is then determined by the conditions  $u_i(t_0^+) = 4\alpha$  and  $u_i(t_0^- + P_L) = 1$ . Solving Eq. 5 for  $P_L$  gives

$$P_L = \ln(I - 4\alpha) - \ln(I - 1). \quad [18]$$

What can be said about the stability of this solution? First, it can be shown that in the absence of external perturbations, a triggered neuron cannot fall behind the wave of excitation.<sup>¶</sup> Second, if a small perturbation  $\varepsilon$  in the firing time of the first neuron still results in a globally synchronized firing pattern, the network activity will simply be shifted in time (by  $\varepsilon$ ). Third, it holds in general that if a neuron receives a fixed number of action potentials between two of its firings, the interval between the two firing times is maximal if the action potentials arrive synchronized and immediately after the first firing. This follows from the fact that the graph of  $u(t)$  is concave down between firings (see Eq. 5) and suggests that if due to some small external perturbation, a neuron fails to stay within the synchronized pulse, the system will eventually return to global synchrony even if during this process, other neurons become desynchronized.

For model A, a heuristic argument implies that the fully synchronized oscillation is also globally attractive. The argument is based on a comparison between the dynamics generated by model A and model C. Viewed in the phase space spanned by the membrane potentials, the time evolution of model C between firings is represented by straight lines parallel to the diagonal  $u_1 = u_2 = \dots = u_N$ . In model A, the trajectories between firings are straight lines toward the point  $u_1 = u_2 = \dots = u_N = I$ , the equilibrium in the absence of spike generation. In other words, the solutions of model A approach each other between firings, whereas the solutions of model C stay equidistant. The resets due to action potentials are the same for both models. A solution of model A thus gradually shifts between the periodic solutions of model C and approaches the only orbit common to both models, the globally synchronized state. The picture also explains why solutions of model A exhibit a rapid (approximate) phase locking with a “period” close to that of model C.

In a comparison between model B and model D, the above argument cannot be used to prove convergence to global synchronization: Because the membrane potential of a firing neuron is reset to zero in these models, nearby but different states may be mapped into one single state in an elementary firing event. This restoration phenomenon can balance the

<sup>¶</sup>The membrane potential of a neuron triggered by  $k$  neighbors at time  $t_0$  must satisfy  $1 - k\alpha \leq u(t_0^-) < 1 - (k - 1)\alpha$ . In model B, it is reset to zero and then receives input from the remaining  $(4 - k)$  neighbors, that is,  $u(t_0^+) = (4 - k)\alpha$ . By using Eq. 5, one obtains  $u(t_0^- + P_L) = 1 - k\alpha \exp(-P_L)$ , or  $1 - k\alpha < u(t_0^- + P_L) < 1$ . This implies that a neuron may only move forward in the firing order and never surpasses the first neuron. In model A, one has  $u(t_0^+) = u(t_0^-) + 4\alpha - 1$  independent of  $k$  (see Eq. 7). This gives  $u(t_0^- + P_L) = 1 + [u(t_0^-) - 1]\exp(-P_L)$  or  $u(t_0^-) < u(t_0^- + P_L) < 1$ , leading to the same conclusion as above. In addition, it follows that for  $t \rightarrow \infty$ , the time evolution of all membrane potentials in model A becomes equal.

contracting process between firings and admits spatially non-uniform periodic solutions (Fig. 1B). Numerical simulations reveal that their period is the same as that of the globally synchronized state and that every neuron fires once during one period of the oscillation. This unexpected result implies that there is a minimal size for any locally synchronized cluster of neurons: The cluster has to contain the triggering neuron and all neurons this cell is receiving input from. Otherwise, the triggering neuron would receive input at some intermediate time of its firing cycle, resulting in a decreased interspike interval due to the convexity of  $u(t)$ .

For a two-dimensional sheet of neurons with nearest-neighbor interactions, the minimal cluster is a cross-shaped pattern with five neurons. It is shown in the *Appendix* that an (attractive) set of stable spatiotemporal spike patterns exists where a discharge of these five neurons alternates with a uniform firing of the “background” neurons. This disproves a conjecture that networks of integrate-and-fire neurons with uniform input and translationary invariant connection scheme always approach the state of global synchronization (9). [It should be noted that the mathematical proofs in ref. 9 refer to a variant of model B where Eq. 6 is replaced by  $u_j(t^+) = 0$ .] The present proof is readily extended to other activity patterns and demonstrates that (leaky) integrate-and-fire neurons with short-range interactions exhibit much richer collective phenomena than systems with all-to-all couplings.

So far, only excitatory connections have been discussed. There is, however, strong numerical evidence that inhibition does not change the overall picture. We have studied networks under conditions that exclude run-away solutions,<sup>||</sup> guaranteed that no neuron relaxes permanently to a subthreshold state, and satisfied Eqs. 9 and 10. All simulations approached periodic limit cycles whose period is given by Eq. 11 or 18, with  $4\alpha$  replaced by  $A$ . This observation gives hope that further analytical understanding of the models discussed in this paper is possible, although we believe that the mathematical situation is more complicated. A convergence proof based on Lyapunov functions such as Eq. 13 is possible because every periodic solution of model C has the same period. This is not the case for models A and B, as shown by the following counterexample. Consider a spatiotemporal “checkerboard” pattern, where the “black” sites fire at even multiples of  $\Delta/2$ , the “white” sites fire at odd multiples of  $\Delta/2$ . A self-consistent calculation of the firing pattern leads to an implicit equation for  $\Delta$ ,  $4\alpha \exp(-\Delta/2) + I[1 - \exp(-\Delta)] = 1$ . Comparison with Eq. 18 shows that  $\Delta \neq P_L$ . Although this firing pattern is unstable, its mere existence indicates that it will be difficult to find Lyapunov functions for models A and B.

### Toward Collective Computation with Action Potentials

The response of the networks to spatially structured stimuli such as grey-valued (visual) patterns is best understood if one first focuses on initial conditions drawn from a uniform random distribution with mean  $\bar{u}$  and width  $w$ . A simple natural image may then be thought of as an array of noisy grey-valued

plateaus, separated by sharp brightness variations at object boundaries.

For nearest-neighbor interactions of strength  $\alpha$ , the system behavior is governed by two parameters, one characterizing the dependence upon the internal dynamics ( $\alpha$ ), the other describing the dependence upon the statistics of the initial conditions ( $w$ ). For simplicity, we focus on two-dimensional realizations of model C with periodic boundaries and  $I_i^{\text{ext}} = I$ .

If  $w \leq \alpha$ , the first action potential of a cell activates all neighboring neurons, which in turn trigger their neighbors, and results in a global synchronization. For small  $\alpha$  and large enough  $w$ , such an avalanche is impossible. On the other hand, there is also not enough time for gradual long-range ordering to set in because the convergence takes place in finite time. Thus the probability distribution of events of large size must decrease exponentially with size for small  $\alpha$  and large  $w$ . Simulations show that for large systems the two dynamic regimes are separated by a critical line where the probability to obtain a cluster of a certain size is given by a power law (38).

When operating in the present mode of a stimulus-induced oscillation, the computational capabilities of the network are then as follows. With a fixed value for  $\alpha$ , the network always relaxes to a periodic oscillation of period  $P = 1 - 4\alpha$ , independent of the structure of the initial condition. The spatiotemporal characteristics of the emergent oscillation, however, strongly reflect the statistics of the initial pattern. Regions with small variability are smoothed out and represented by locally synchronized clusters of neurons whose firing times encode the stimulus quality. Regions with high variability give rise to spatially uncorrelated firing patterns. Integrate-and-fire networks thus operate like spatiotemporal extensions of resistive grids with line-breaking elements (33, 39).

The picture emerging from these findings suggests that even simple locally coupled integrate-and-fire neurons are able to encode objects—herein defined as regions of similar grey value—by synchronized firing patterns that are held in a dynamic short-time memory through neural reverberations. More general network architectures that include longer-range connections and inhibitory synapses are able to perform specific computations as demonstrated in Fig. 2. In a next step, feature-sensitive neurons could be included and might help to obtain a quantitative comprehension of the computations involved in cortical object recognition.

### Discussion

A “neurodynamics” including action potentials displays much richer collective phenomena than do models that represent action potentials only by their statistical average effect. Previously, this richness has prevented a thorough understanding of how to compute in a useful fashion with a feedback network of such (model) neurons. On the other hand, the notion of attractors and a Lyapunov function has provided powerful methods to analyze neural networks based on a firing-rate description. The convergence proof (*Theorem 1*) and the existence of a Lyapunov function (*Theorem 2*) for special cases of networks of spiking neurons leads us to hope that it will be possible to achieve a quantitative understanding of the relationship between connections, initial conditions, and computation in these systems as well.

Model C, which has a Lyapunov function, applies to patterns of connectivity much more varied than have been previously studied. In fact, all models with the same input current and total incoming synaptic strength exhibit short-term convergence to a periodic state whose nature is similar to that of the model with a Lyapunov function. Furthermore, rapid (approximate) phase locking of action potentials has been observed in experiments (5, 6) and is also exhibited by a number of more elaborate modeling approaches (13, 15, 20, 21). Thus it seems appropriate to focus on this particular representative.

<sup>||</sup>Sufficient conditions for bounded  $u_i$  follow readily from a worst-case analysis. Let  $T_i^+ = \sum_j \theta(T_{ij})$  denote the total incoming excitatory synaptic strength of neuron  $i$ ,  $T_i^- = \sum_j \theta(-T_{ij})$  denote the total inhibitory strength,  $T^+ = \max_i \{T_i^+\}$ , and  $T^- = \max_i \{T_i^-\}$ . Let us focus on model C with uniform input,  $I_i^{\text{ext}} = I$ . Multiple firings of one neuron during a single event cannot occur if the total current of all neurons that excite a given neuron does not drive that neuron above threshold if it has just fired. For  $T^+ < 1$ , the  $u_i(t)$  are thus bounded from above. Each  $u_i$  is bounded from below if in any time interval  $u_i$  decreases less due to inhibitory input than it increases due to the current  $I$ . The shortest possible time interval between two spikes of the same neuron is  $(1 - T^+)/I$ . In that time, the potential of a subthreshold neuron increases by at least  $1 - T^+$  and decreases by at most  $T^-$ . This defines the condition  $1 - (T^+ + T^-) > 0$ .

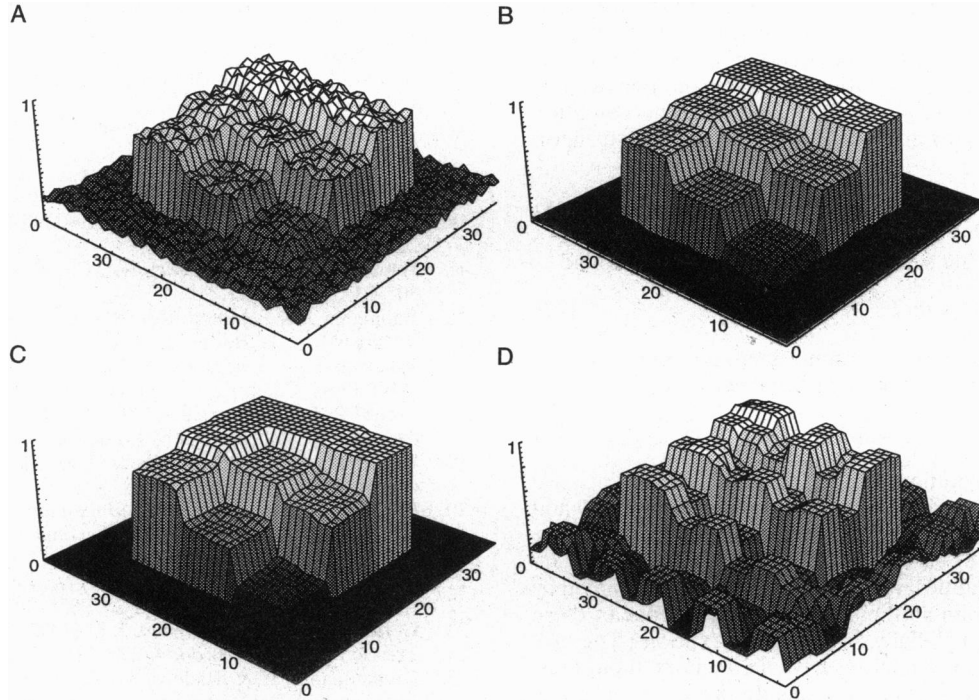


FIG. 2. Collective computation with action potentials. Stimulus pattern (A) and response of three networks of type C with  $40 \times 40$  neurons ( $x$  and  $y$  axis) and periodic boundary conditions (B–D). The initial values for the membrane potentials ( $z$  axis in A) are noisy plateaus with means  $\bar{u} = 0.9, 0.85, 0.8; 0.8, 0.7, 0.6; 0.7, 0.5, 0.3$ ; and a background of 0.2. The noise is uniform with width  $w = 0.1$  (A). The output of the network is represented by the (normalized) time elapsed since the last firing of a neuron ( $z$  axis in B–D). (B and C) Noise reduction through moderate local excitation; the output patterns closely resemble the original (noiseless) image. (B) Each neuron is connected to its four nearest neighbors with  $\alpha = 0.06$ . (C) Additional couplings to the four next nearest neighbors (with strength  $\beta = 0.03$ ) result in an overfitting of areas with little brightness variation. Contrast enhancement is possible through long-range inhibition as shown in D where in addition to local excitation ( $\alpha = 0.05$  and  $\beta = 0.02$ ), inhibitory long-range connections were included. They are of strength  $\gamma = -0.03$  and join each neuron  $i$  with 16 cells (out of 32) along the boundary of a square of size  $9 \times 9$ , centered at  $i$ .

To stay within the present class of models, incoming information has to be thought of as setting initial conditions for a set of neurons, which then are allowed to follow their own intrinsic dynamics. However, the presumption that all neurons have the same steady input (or leakage) current is inadequate for biology. Sufficiently different input currents would seem to necessarily decouple neurons from the periodic behavior. Undoubtedly, inputs are also constantly changing. Nevertheless, the initial-condition approach is similar to the way that the mathematics of associative memory was described, and we can similarly expect some useful insight into the computational abilities of integrate-and-fire neurons from this approach.

In closing, let us point out that many natural phenomena arise from the interaction of coupled threshold elements. Networks of spiking nerve cells are just one example in a large class of systems that includes swarms of flashing fireflies (9), avalanches (31), and earthquakes (28). The corresponding models show a high degree of similarity on the microscopic level. For example, the time evolution of single earthquake faults has been described by model E (28). For open boundary conditions, this system approaches a self-organized critical state with fluctuations on all length scales. It has been argued that this result explains the power-law behavior observed in the size-frequency relation of earthquakes. Similarly, the periodic behavior of the present models provides a plausible explanation of the quasi-cyclic behavior seen in the recurrence pattern of some large earthquakes (38).

In view of the hope of some physicists for a “theory of everything” relating diverse complex systems, the generality and specificity of the results in this area of modeling might be noted. A range of models were shown to have similar short-time behavior. However, the asymptotic behavior of the different models depends sensitively on model details, such as

boundary conditions, and ranges from self-organized criticality to system-wide synchronization. The utility of a “theory of everything” under such circumstances depends entirely on what one hopes to learn.

#### Appendix: Nontrivial Oscillations in Model B

Let us focus on a cross-shaped pattern, the minimal synchronized cluster possible for excitatory nearest-neighbor interactions. Let us further assume that the background fires at times  $kP_L, k \in \mathbb{N}$ , and that the cross-shaped pattern is active at times  $t = \Delta + kP_L$ . To analyze the existence and stability of this solution, three particular classes of cells have to be studied: The four nearest neighbors of the center of the cross will be called  $a$  cells. Further away from the center and part of the background are the four  $b$  and four  $c$  cells. The latter are the diagonal neighbors of the center cell; the former are two lattice sites away from the center, bordering one  $a$  cell and three other background cells.

At time 0,  $a$  cells receive three action potentials (two from  $c$  cells and one from a  $b$  cell), so that  $u(0^+) = u(0^-) + 3\alpha$ . Just before the center neuron fires, Eq. 5 gives  $u(\Delta^-) = [u(0^+) - I] \exp(-\Delta) + I$ ; right after this event,  $u$  reaches the value  $u(\Delta^-) + \alpha$  and is then reset to  $u(\Delta^+) = 0$ . Finally, at time  $P_L$ , the initial state is reached again,  $u(0^-) = u(P_L^-) = I[1 - \exp(\Delta - P_L)]$ . By using the last expression to eliminate  $u(0^-)$ , one verifies that  $a$  cells are below threshold at  $t = \Delta^-$  if

$$\Delta > P_L - \ln(4/3), \quad [19]$$

and at or above threshold at  $t = \Delta$  if

$$\Delta \leq -\ln[\exp(-P_L) - 1/4] - \ln(4/3). \quad [20]$$

Depending on their potential at  $t = 0^-$ ,  $b$  cells may be triggered by  $k = 1, 2$ , or 3 cells at  $t = 0$ . This leads to three possible activity bands, labeled by  $k$  and defined by  $1 - k\alpha \leq u(0^-) < 1 - (k - 1)\alpha$ . At time  $P_L^-$ , solutions of each band are given by  $u(P_L^-) = 1 + \alpha \exp(-P_L)[\exp(\Delta) - k - 1]$ . The condition  $u(P_L^-) = u(0^-)$  leads to three constraints for  $\Delta$ . The strongest condition is

$$\Delta < \ln[4 - 2 \exp(P_L)]. \quad [21]$$

Applied to  $c$  cells, the same argument provides the bound

$$\Delta < \ln[2 - 1/2 \exp(P_L)]. \quad [22]$$

A final constraint is simply the requirement that the cross fires before the background neurons become active again,

$$\Delta < P_L. \quad [23]$$

A self-consistent solution may exist if all constraints are compatible. Eqs. 19 and 20 are always compatible, Eqs. 19 and 21 are compatible if  $5I > 16 - 44\alpha$ , and Eqs. 19 and 22 are compatible if  $3I > 8 - 20\alpha$ . The example  $\alpha = 0.2$  and  $I = 2.0$  satisfies both inequalities. In this case,  $\Delta$  is only constrained by Eq. 23. Numerical simulations confirm the existence of these phase-locked but not globally synchronized solutions. They are stable because triggered neurons have no memory about their previous state and triggering neurons do not experience any restoring forces when slightly perturbed.

This work was stimulated by a series of most helpful discussions with John Rundle about the connections and differences between earthquake models and networks of spiking neurons. We are also grateful to W. Gerstner and T. Heskes for critical comments and suggestions and to K. Schulten for providing powerful computing facilities that allowed fast interactive large-scale simulations. A.V.M.H. was supported by a Beckman Institute Fellowship; J.J.H. was supported in part by National Science Foundation Grant BIR 9207487 and by the Ron and Maxine Linde Venture Fund.

1. Suga, N., O'Neill, W. E., Kujirai, K. & Manabe, T. (1983) *J. Neurophysiol.* **49**, 1573–1626.
2. Konishi, M. (1992) *Harvey Lect.* **86**, 47–64.
3. Carr, C. E., Heiligenberg, W. & Rose, G. J. (1986) *J. Neurosci.* **6**, 107–119.
4. Laurent, G. & Davidowitz, H. (1994) *Science* **265**, 1872–1875.
5. Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H. J. (1988) *Biol. Cybern.* **60**, 121–130.
6. Gray, C. M. & Singer, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1698–1702.
7. von der Malsburg, C. (1981) *The Correlation Theory of Brain Function* (MPI for Biophys. Chem., Göttingen, Germany).
8. Abeles, M. (1991) *Corticonics* (Cambridge Univ. Press, Cambridge, U.K.).
9. Mirollo, R. E. & Strogatz, S. H. (1990) *SIAM J. Appl. Math.* **50**, 1645–1662.
10. Kuramoto, Y. (1991) *Physica D* **50**, 15–30.
11. Abbott, L. F. & van Vreeswijk, C. (1993) *Phys. Rev. E* **48**, 1483–1490.
12. Tsodyks, M., Mitkov, I. & Sompolinsky, H. (1993) *Phys. Rev. Lett.* **71**, 1280–1283.
13. Usher, M., Schuster, H. & Niebur, E. (1993) *Neural Comput.* **5**, 570–586.
14. van Vreeswijk, C. & Abbott, L. F. (1993) *SIAM J. Appl. Math.* **53**, 253–264.
15. Hansel, D., Mato, G. & Meunier, C. (1995) *Neural Comput.* **7**, 307–337.
16. Buhmann, J. & Schulten, K. (1986) *Biol. Cybern.* **54**, 319–335.
17. Wilson, M. A. & Bower, J. M. (1989) in *Methods in Neuronal Modeling: From Synapses to Networks*, eds. Koch, C. & Segev, I. (MIT Press, Cambridge, MA), pp. 291–334.
18. Traub, R. D. & Miles, R. (1991) *Neural Networks of the Hippocampus* (Cambridge Univ. Press, Cambridge, U.K.).
19. Usher, M., Stemmler, M., Koch, C. & Olami, Z. (1994) *Neural Comput.* **6**, 795–836.
20. Bush, P. & Sejnowski, T. (1994) *Inhibition Synchronizes Sparsely Connected Cortical Neurons Within and Between Columns in Realistic Network Models*, preprint.
21. Tsodyks, M. & Sejnowski, T. (1994) *Network* **6**, 111–124.
22. Abbott, L. F. (1990) *J. Phys. A* **23**, 3835–3859.
23. Gerstner, W. & van Hemmen, J. L. (1992) *Network* **3**, 139–164.
24. Treves, A. (1993) *Network* **4**, 259–284.
25. Dunkelmann, S. & Radons, G. (1994) in *Proceedings of the International Conference on Artificial Neural Networks*, eds. Marimnaro, M. & Morasso, P. G. (Springer, London), pp. 867–871.
26. Kentridge, R. W. (1994) in *Computation and Neural Systems*, eds. Eeckman, F. H. & Bower, J. M. (Kluwer, Dordrecht, The Netherlands), pp. 531–535.
27. Hopfield, J. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558.
28. Olami, Z., Feder, H. J. S. & Christensen, K. (1992) *Phys. Rev. Lett.* **68**, 1244–1247.
29. Gabrielov, A. (1993) *Physica A* **195**, 253–274.
30. Gabrielov, A., Newman, W. I. & Knopoff, L. (1994) *Phys. Rev. E* **50**, 188–196.
31. Bak, P., Tang, C. & Wiesenfeld, K. (1987) *Phys. Rev. Lett.* **59**, 381–384.
32. Hopfield, J. J. (1994) *Phys. Today* **46**, 40–46.
33. Geman, S. & Geman, D. (1984) *IEEE Trans. Pattern Analysis Machine Intelligence* **6**, 721–741.
34. Grassberger, P. (1994) *Phys. Rev. E* **49**, 2436–2444.
35. Socolar, J. E. S., Grinstein, G. & Jayaprakash, C. (1994) *Phys. Rev. E* **47**, 2366–2376.
36. Middleton, A. A. & Tang, C. (1995) *Phys. Rev. Lett.* **74**, 742–745.
37. Feder, H. J. S. & Feder, J. (1991) *Phys. Rev. Lett.* **66**, 2669–2672.
38. Herz, A. V. M. & Hopfield, J. J. (1994) *Earthquake Cycles and Neural Reverberations: Collective Oscillations in Systems with Coupled Threshold Elements*, preprint.
39. Poggio, T., Torre, V. & Koch, C. (1985) *Nature (London)* **317**, 314–319.