Mutual information in a dilute, asymmetric neural network model

Elliot Greenfield
Department of Physics, University of California at Berkeley, Berkeley, California 94720

Harold Lecar
Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720

(Received 30 May 2000; revised manuscript received 25 October 2000; published 23 March 2001)

Neural networks with asymmetric synaptic connections ($w_{ij} \neq w_{ji}$) display a broad range of dynamical behavior including fixed point, periodic, and ''chaotic'' trajectories. Previous work has shown that such networks undergo an order-chaos phase transition as various network parameters, such as the connectivity or the degree of asymmetry, are changed. Here, using an information theoretic approach, we present results which suggest that neurons are able to communicate information to each other most effectively in networks that are near the order-chaos transition. We then extend the model to incorporate some biologically relevant features.

DOI: 10.1103/PhysRevE.63.041905 PACS number(s): 87.18.Sn, 87.19.La, 87.18.Bb

I. INTRODUCTION

Much progress has been made in understanding emergent computation in model neural networks. Networks of binary neurons provide a general framework for simulating cognitive processes such as pattern recognition, associative memory, and learning [1]. Neurobiologists have made successful models of small functional networks of known neuroanatomy (e.g., oscillating networks that control rhythmic activities such as digestion [2] and simple locomotion [3]), but the principles of generalized computation remain an open question. In this paper, we study the computational properties of a neural network consisting of binary neurons with dilute asymmetric synaptic connections. This simple model allows us to simulate large networks which can reflect more of the architecture and dynamics of real neural networks.

Our main goal is to determine the dynamical behavior that maximizes the network's ability to perform computations. To this end, we apply information theory, measuring the average mutual information between pairs of pre- and postsynaptic neurons. In order for a network to perform a collective computation, information must be communicated across synaptic connections. Previous workers have demonstrated that neural networks with asymmetric connections undergo a transition from ordered to chaotic behavior as certain network parameters, such as the connectivity, are changed. We find that the average mutual information has a peak near the order-chaos transition. This implies that the network can most efficiently communicate information between cells in a narrow range of connectivities. We derive analytical predictions of the network activity as well as the location of the phase transition. The mutual information peak becomes increasingly pronounced when the basic model is extended to incorporate more biologically realistic features, such as a variable threshold and nonlinear summation of inputs. In general, the peak in the mutual information near the phase transition is a robust feature of the system for a wide range of assumptions about postsynaptic integration.

The network consists of N binary neurons, each receiving input from K other randomly chosen neurons. The K synaptic inputs are a mixture of excitatory and inhibitory synaptic connections. The state of each neuron is represented by the variable $\sigma_i \in \{0,1\}$, where $i=1,\ldots,N$. The $\sigma=0$ state represents a neuron that is at rest (off) and the $\sigma=1$ state represents a neuron that is firing at its maximum rate (on). At each time step, all the neurons in the network are updated in parallel according to the rule

$$\sigma_i(t+1) = \Theta\left[\sum_j w_{ij}\sigma_j(t) - T_i\right], \tag{1}$$

where w_{ij} is the strength of the synaptic input from cell j to cell i, T_i is the threshold of cell i, and Θ is the Heaviside step function. If there is a synaptic connection from cell j to cell i, then the weight w_{ij} is a random variables chosen from a distribution $\rho(w_{ij})$; otherwise, it is zero. Different distributions of synaptic strengths, $\rho(w_{ij})$, can be chosen to reflect different neuronal architectures. Since a physiologically realistic distribution would likely be multimodal, comprised of separate subpopulations of excitatory and inhibitory synaptic weights, an appropriate simplification is a relatively uniform distribution. Similarly the T_i can be chosen to represent the known physiology of spike initiation (see Sec. II C 2).

The order-chaos phase transition has been studied in random Boolean networks (RBNs) by Kauffman [4–6] and in dilute asymmetric neural networks by Kurten [7]. Kurten's model uses $(-1,1)$ neurons and zero threshold. We must use (0,1) neurons in order to investigate the effects of nonzero neuronal thresholds and for the update function to correspond to realistic postsynaptic integration. Appendix A explains why these two representations are not generally equivalent. We show that making this change of variables in the update rule, Eq. (1), generates an extra thresholdlike term that is different for each neuron and significantly alters the network dynamics. In particular, the distribution of activity levels is different, strongly affecting the computed information and mutual information discussed in Sec. II C.

Other recurrent asymmetric neural network models that have been studied include a fully connected network where the synaptic weights consist of a symmetric part and an antisymmetric part: $w_{ij} = w_{ij}^{S} + \lambda w_{ij}^{A}$. This system exhibits a phase transition as λ is varied [8]. Crisanti *et al.* examined

the behavior of entropic quantities at this transition [9] and in a globally coupled logistic map [10]. Another model derives its asymmetry from the fact that a certain fraction of the connections of a symmetric network are cut [11]. For a fully connected network of neurons with a continuous range of firing rates, Sompolinsky *et al.* showed a transition to chaos as the degree of nonlinearity of the update function was increased [12]. We chose completely asymmetric synaptic weights and dilute connections to more closely model typical biological networks.

The Hopfield neural network [13,14], which is the prototypical model of associative memory, employs the simplifying assumption of symmetric synaptic weights, $w_{ij}=w_{ji}$. Though biologically unrealistic, this assumption allows a Lyapunov function to be defined:

$$E=-\frac{1}{2}\sum_{ij} w_{ij}\sigma_i\sigma_j. \tag{2}$$

Accordingly, the system relaxes from its initial state into the closest energy minimum, which is identified as a memory. Networks with asymmetric connections have no Lyapunov function, and thus can have limit cycles in addition to fixed points. Large asymmetric systems can manifest seemingly aperiodic behavior resembling deterministic chaos.

Kauffman showed that a random network of Boolean automata undergoes a transition from ordered to chaotic behavior as a function of K, the number of inputs to each cell [4–6]. For $K\leqslant 2$, the network is in the ''frozen'' phase, with most of the elements fixed in one state or the other and only small, isolated pockets of activity. For $K>2$, the active regions percolate throughout the network, and the network is said to be in the ''chaotic'' phase. Later workers showed that this transition carries over to dilute asymmetric neural networks [11,15,16] at similar values of K, depending on model assumptions. This range of critical connectivities is far smaller than would be expected for a real neural network. The low values of K are not intrinsic to this type of neural network, but result from subsidiary assumptions such as the zero threshold criterion. The transition in our neural network model also occurs at $K=2$ for zero threshold, but it increases significantly for more realistic threshold values.

The ''chaotic'' behavior referred to above is not chaos in the sense of a positive Lyapunov exponent. A Lyapunov exponent cannot be defined for a two-state element. Since the dynamics are deterministic and the state space is finite (2^N possible network states), all trajectories must be periodic. However, the sensitivity of a trajectory to a small perturbation can be generalized by looking at the spread of ''damage'' between two nearby states.

The damage can be defined as the normalized Hamming distance between two network states, $\vec{\sigma}(t)$ and $\vec{\sigma}'(t)$,

$$D(t)\equiv\frac{1}{N}\sum_i |\sigma_i'(t)-\sigma_i(t)|. \tag{3}$$

If the configurations are initially very close, so that $D(0)\ll 1$, then in the ordered phase $D(t)$ will tend to decay to zero, whereas in the chaotic phase $D(t)$ will tend to grow larger and remain finite [11]. Derrida and Pomeau successfully modeled this feature with their ''annealed approximation'' [17]. Earlier it was shown [4] that the average period grows exponentially with the system size N in the chaotic phase, and it is either constant or increases as a power of N in the ordered phase. Both findings correspond to the explanation of the phase transition as resulting from the percolation of active regions.

How does the ability of the network to perform computations depend on its dynamics? Computational ability can be measured directly by training the network to perform a particular task and rating its success. However, one drawback to this approach is that the results would likely depend upon the task that is chosen and the learning algorithm that is employed. It can also be difficult to compare such results with experimental data.

Computational ability can also be measured indirectly by looking at quantities that are believed to be necessary for or indicative of computation, such as information transfer, storage, and processing [18]. The indirect method requires no specification of what type of computation is being performed and therefore relates more closely to our level of understanding of how higher level computation is performed in the brain. For these reasons, we take the indirect approach in this work, and the quantity we examine is the average amount of information communicated between cells, as measured by the mutual information.

In the language of computational theory, the direct measurement quantifies ''useful computation'' and the indirect measurement ''intrinsic computation.'' Intrinsic computation places an upper bound on useful computation [19]. In the present context, the fact that a high level of information is being communicated between cells does not imply that the network is performing a complex computation. However, the converse is true—computation does require the communication of information.

Several workers have studied the relation between computation and order-chaos transitions in a variety of systems. Wolfram found that the behavior of rule-based cellular automata could be classified as steady state, periodic, chaotic, or ''complex.'' Cellular automata with complex dynamics exhibit behavior that is intermediate between periodic and chaotic. Wolfram hypothesized that this class of automata may be capable of universal computation [20,21]. Crutchfield quantified the behavioral complexity of continuous dynamical systems using his ''statistical complexity'' measure, which is the size of the minimal binary automata network needed to reproduce the system's dynamics [22]. He found that as a system moves from ordered to chaotic behavior, for example, the period doubling route to chaos of the logistic map, its statistical complexity peaks at the transition point. Similar ideas about phase transitions and computation have been applied to the study of biological systems, including ant colonies [23] and the immune system [24,25]. Recently, Paczuski *et al.* have demonstrated that a network of competing Boolean automata evolves by an evolutionary process to the border of order and chaos [26].

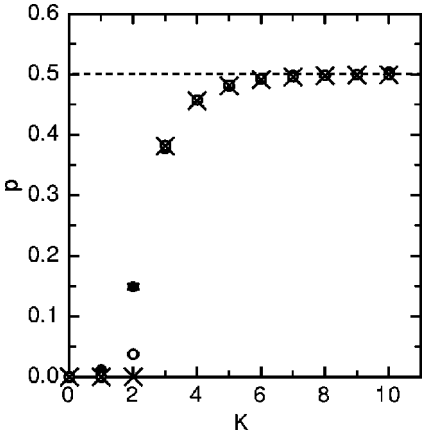

FIG. 1. Average activity as a function of K: $\times$, solution to Eq. (4); •, simulation data for $N=80$ (average of 10^3 networks); ○, simulation data for $N=10\,000$ (average of 250 networks).

II. RESULTS

A. Average activity

The most general feature of the network's dynamics is the average fraction of neurons that are in the $\sigma=1$ state (or, equivalently, the average likelihood that any one neuron is in this state). Unlike the RBN, where there is complete symmetry between the 0 and 1 states, here there is a bias toward the 0 state (even with zero threshold) since a neuron with all of its inputs off will also be off. The average probability that a neuron is on at time $t+1$, $p(t+1)$, can be written in terms of $p(t)$, the average probability that its inputs were on at time t. We show in Appendix B that the average activity follows the equation

$$p(t+1)=\frac{1-[1-p(t)]^K}{2}. \tag{4}$$

The steady state average activity p is the fixed point of Eq. (4). The solutions to Eq. (4) for different K values are shown in Fig. 1. The agreement with simulation data is nearly exact and becomes increasingly good as N increases. These data correspond well to Flyvbjerg's idea of the ''stable core'' [27]. The stable core refers to the fraction of frozen elements, which is 1 in the ordered regime and less than 1 in the chaotic regime in the limit $N\to\infty$. Here, in the ordered regime, the frozen elements are all in the $\sigma=0$ state. Flyvbjerg points out that this quantity plays the role of an order parameter for the transition in RBNs.

B. The phase transition

1. Average period

The location of the phase transition is determined by two methods. We examine the behavior of the average period as a function of N, and we also calculate the evolution of the distance between nearby trajectories using the annealed approximation. Figure 2 shows the dependence of the average period on the number of neurons for $K=1$, 2, and 6. In the ordered phase the average period is either constant or increases as a power of N ($\langle T\rangle\sim N^x$), reflecting the distribution of cycle lengths in the isolated active regions. In the chaotic phase it increases exponentially in N ($\langle T\rangle\sim e^{\alpha(K)N}$) [6]. The average denoted by $\langle\cdots\rangle$ is the average over trials, where each trial has a different set of synaptic weights and initial conditions. This shows that as the active regions percolate, the network can wander through a macroscopic fraction of its state space. For a completely random system (the-''random map''), the dependence of the average period on system size can be solved exactly and is found to be $\langle T\rangle=\sqrt{\pi/8}\,2^{N/2}$ [28].

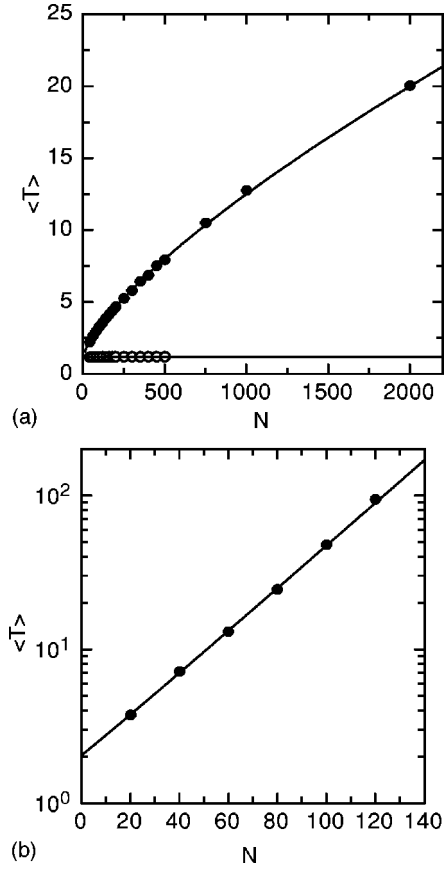

FIG. 2. Average period $\langle T\rangle$ in units of time steps, as a function of system size N: (a) ○, $K=1$ and •, $K=2$ and (b) •, $K=6$ (simulation data, average of 10^5 networks). Solid line indicates power law fit in (a) and exponential fit in (b).

2. Annealed approximation

Although analytic solutions exist for the dynamics of the RBN for the special cases of $K=1$ [29] and $K=N$ [28], approximations must be made for all intermediate values of K. Derrida and Pomeau's ''annealed approximation'' [17] allows an equation to be written for the time evolution of the distance between two network configurations. This approximation effectively ignores correlations that build up between the state of a neuron and the synaptic weights and states of its inputs by averaging over them at each time step. They showed that the annealed approximation closely matches simulation data for a network with ''quenched'' disorder.

The equation is a little more complicated for the case of (0,1) neurons than for Boolean elements or $(-1,1)$ elements since $p \neq \frac{1}{2}$ for all K. Adapting the derivation from [15] we arrive at Eq. (5), which predicts the average distance between two configurations at time $t+1$ as a function of this distance at time t,

$$\langle D(t+1)\rangle = \sum_{n=1}^{K} D(t)^n [1-D(t)]^{K-n} \binom{K}{n} \sum_{\alpha=0}^{K-n} p^\alpha (1-p)^{K-n-\alpha} \binom{K-n}{\alpha} \times \sum_{\beta=0}^{n} \left(\frac{1}{2}\right)^n \binom{n}{\beta} \int dx_1 \rho(x_1) \cdots dx_\alpha \rho(x_\alpha) \int dy_1 \rho(y_1) \cdots dy_\beta \rho(y_\beta) \times \int dz_1 \rho(z_1) \cdots dz_{n-\beta} \rho(z_{n-\beta}) \left| \Theta\left(\sum_{i=1}^{\alpha} x_i + \sum_{i=1}^{\beta} y_i \right) - \Theta\left(\sum_{i=1}^{\alpha} x_i + \sum_{i=1}^{n-\beta} z_i \right) \right|. \quad (5)$$

We use solutions to Eq. (4) to find the value of p to be used in Eq. (5). Equation (5) is the same for an annealed and a quenched network for one time step for two initial uncorrelated network states. Iterating the equation further in time is the annealed approximation.

The chaotic phase is defined by an increase in $D(t)$ for initially close configurations,

$$\left. \frac{dD(t+1)}{dD(t)} \right|_{D(t)=0} > 1. \quad (6)$$

We choose a uniform distribution for $\rho(w_{ij})$ for mathematical convenience and because a physiologically realistic distribution would be a complicated multimodal distribution comprised of separate distributions of weights for excitatory and inhibitory synapses. We find empirically that the results do not vary significantly for different distributions. For a uniform distribution for $\rho(w_{ij})$, the derivative in Eq. (6) can be expressed as

$$\left. \frac{dD(t+1)}{dD(t)} \right|_{D(t)=0} = \frac{K}{2}(1-p)^{K-1} + K \times \sum_{l=1}^{K-1} p^l (1-p)^{K-1-l} \binom{K-1}{l} \frac{1}{2^l (l+1)} \times \sum_{n=0}^{n_{max}} (-1)^n \frac{1}{n!(l-n)!} [(l-2n-1)^{l+1} + (l-2n)^l (2n+1)], \quad (7)$$

where $n_{max} = \mathrm{Int}[(l-1)/2]$. If Eq. (6) is satisfied then the fixed point of Eq. (5) will be nonzero. Note that the long-term behavior of the damage under the annealed approximation is different from that in the quenched system, where $D(t)$ must be periodic. Figure 3 shows the fixed points of Eq. (5), $D^\star$, obtained by setting $D(t+1) = D(t)$. In the ordered phase, the only fixed point is given by $D^\star = 0$. In the chaotic phase, $D^\star = 0$ becomes unstable, as Eq. (6) ensures, and the nonzero fixed point is stable. This prediction for the location of the phase transition agrees with that of the behavior of the average period as described in the previous section.

C. Mutual information

1. Mutual information peak

Having examined the basic dynamical behavior of the network and ascertained the location of the order-chaos transition, we now relate the network's dynamics to its ability to perform computations. Mutual information [30] is chosen as the diagnostic because an essential requirement for collective computation is the communication of information between cells. Let the mutual information of neuron B and one of its input neurons, A, be denoted by $I(A,B)$. Mutual information is defined as

$$I(A,B) = H(A) + H(B) - H(A,B) \quad (8a)$$

$$= H(A) - H(A|B) \quad (8b)$$

$$= H(B) - H(B|A), \quad (8c)$$

where $H(A)$ is the Shannon information of cell A. This has the definition

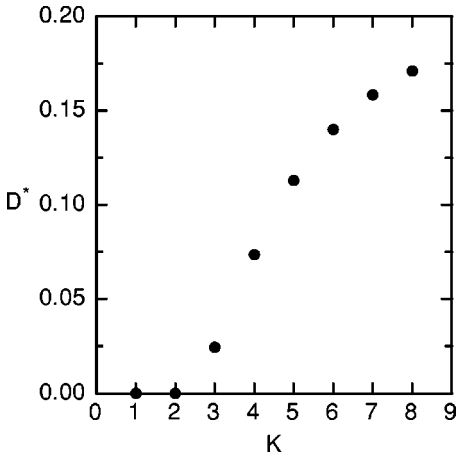

FIG. 3. The fixed points of Eq. (5), $D^\star$, as a function of K.

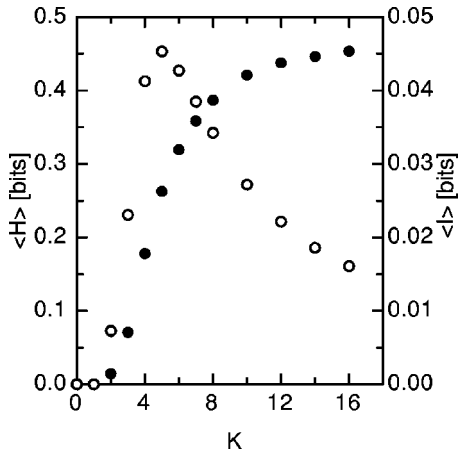

FIG. 4. Average information per cell $\langle H\rangle$, •, left axis, and average mutual information per connection $\langle I\rangle$, ○, right axis, as a function of K (simulation data for $N=50\,000$; average of 100 networks).

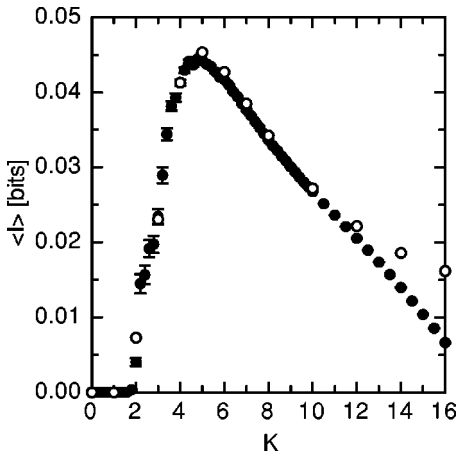

FIG. 5. Average mutual information per connection $\langle I\rangle$: •, Poisson distribution of inputs with mean K; and ○, uniform connectivity K (simulation data for $N=50\,000$; average of 100 networks).

$$H(A)=-\sum_{\sigma_A=0,1}P(\sigma_A)\log_2 P(\sigma_A). \tag{9}$$

The probabilities in Eq. (9) are obtained by averaging over time for each neuron: $P(\sigma_A=1)=p_A=\overline{\sigma_A}$ and $P(\sigma_A=0)=1-p_A$. Thus,

$$H(A)=-p_A\log_2 p_A-(1-p_A)\log_2(1-p_A). \tag{10}$$

The joint entropy $H(A,B)$ and the conditional entropy $H(A|B)$ have the usual definitions,

$$H(A,B)=-\sum_{\sigma_A=0,1}\sum_{\sigma_B=0,1}P(\sigma_A,\sigma_B)\log_2 P(\sigma_A,\sigma_B), \tag{11}$$

$$H(A|B)=-\sum_{\sigma_A=0,1}\sum_{\sigma_B=0,1}P(\sigma_A,\sigma_B)\log_2 P(\sigma_A|\sigma_B). \tag{12}$$

Since we are interested in the amount of information communicated to a neuron from one of its inputs, we measure the mutual information of a neuron at time t and its input at the previous time step $t-1$.

The average information per neuron $\langle H\rangle$ and the average mutual information per connection $\langle I\rangle$ are plotted versus K in Fig. 4. In the ordered phase, $\langle H\rangle$ and $\langle I\rangle$ are zero in the limit $N\to\infty$. As described in Sec. II A, the fraction of neurons frozen in the $\sigma=0$ state approaches 1. In the chaotic phase, $\langle H\rangle$ increases monotonically while $\langle I\rangle$ reaches a peak near the transition and then decreases as the number of inputs becomes larger. This is a key result because it says that, while more information is available in a more chaotic network, the most information can be communicated in a network that is close to the transition. Thus, mutual information increases as the system crosses the transition and information percolates through the network, and then it decreases as the system dynamics become more chaotic and the neural activity becomes progressively decorrelated. The network never becomes completely random, as the RBN does as $K\to N$. This is evidenced by the fact that the average information does not approach 1 bit, as it does in the random map. The reason for this is that the update rule, Eq. (1), can produce only a small fraction of the 2^{2^K} possible Boolean functions of K inputs. These tend to be the least sensitive to changes in the input states [31].

Despite the inherent discreteness of K, we can get a more detailed look at the transition by looking at a network with a distribution of K values. This more closely reflects biological neural networks since all neurons do not have the same number of inputs. The quantities $\langle H\rangle$ and $\langle I\rangle$ are plotted again in Fig. 5 for a Poisson distribution of K values. There is little difference between this case and uniform connectivity, except that the mutual information decays slightly faster for large K. This result is useful because it allows us to see that the transition at $K=2$ is sharp. It also suggests that results obtained with uniform connectivity can be generalized to the more realistic case of a distribution of inputs.

2. Threshold

In order to show that the peak in the mutual information near the transition is robust we investigate a network of neurons with nonzero thresholds. The effect of the threshold is to inhibit the percolation of active regions, thus ordering the system and increasing the critical value of K. Since increasing T necessarily reduces the activity, a more informative measure of information transfer is $\langle I\rangle/p$, which is plotted in Fig. 6 for several threshold values. This quantity is closer to the ''bits/spike'' that is often reported [32].

The scale for the thresholds is set by the distribution of weights $\rho(w_{ij})$, which is chosen here to be a uniform distribution ranging from -1 to $+1$. As expected, the transition occurs at higher values of K as the neural threshold increases. In each case, the mutual information peaks near the transition. The locations of the transitions agree with the predictions of both the annealed approximation and the dependence of the average period on system size.

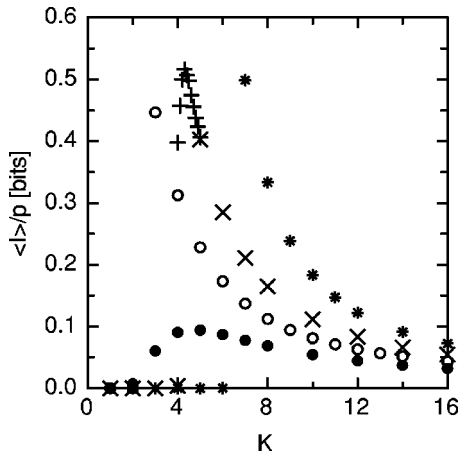

FIG. 6. Average mutual information per connection normalized by activity, $\langle I\rangle/p$, versus K for various thresholds: ● $T=0$; ○, $T=0.3$; ×, $T=0.5$; *, $T=0.7$. Average mutual information per connection normalized by activity, $\langle I\rangle/p$, versus K for various thresholds: ●, $T=0$; ○, $T=0.3$; ×, $T=0.5$; *, $T=0.7$; +, Poisson distribution of inputs with $T=0.5$ (simulation data for $N=50\,000$; average of 100 networks).

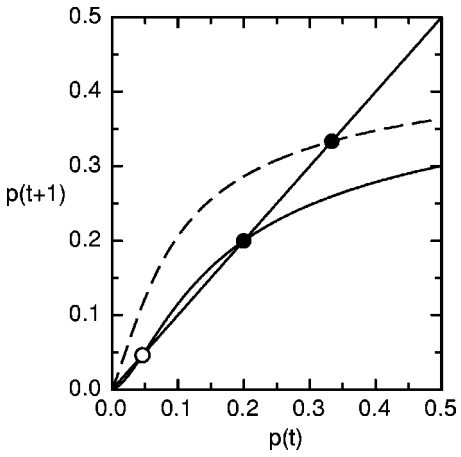

FIG. 7. Average activity $p(t+1)$ as a function of $p(t)$ for $K=50$: dashed line, $T=1$; solid line, $T=1.5$; dotted line, $T=2$. The filled circles mark the stable nonzero fixed points; the open circle marks the unstable nonzero fixed point.

It may seem at first counterintuitive that the average mutual information is larger for small values of the threshold than for zero threshold, since the threshold lowers the average activity. However, this can be understood by examining the distributions of activities for various thresholds. For zero threshold the distribution has peaks at 0 and 1, corresponding to those neurons that are fixed in the $\sigma=0$ and $\sigma=1$ states, respectively. For a relatively small value of the threshold, the peak at 1 is removed and the distribution becomes more uniform, increasing the average information.

3. Higher connectivity

One apparent problem in the foregoing discussion is that the critical connectivity of the model networks is so low. Physiological networks generally have much higher connectivity. Some neurons have as many as 10^4 synapses, and yet they are not in perpetual chaos. In fact, many areas of the cortex are fairly quiet most of the time. Resolving this problem requires the addition of several physiologically derived features. We will indicate some of these features in the discussion. For now, we wish only to examine how the critical connectivity increases with larger values of the threshold.

Making the correspondence between the threshold variable T and a biological threshold voltage is somewhat difficult. Some cortical neurons need to integrate many synaptic inputs in order to cross threshold. The ratio of the size of the threshold to that of an average synaptic input defines the realistic value of T.

For higher threshold, the simulation time becomes impractical. This is because the critical connectivity becomes much larger, and N has to grow accordingly to keep the connectivity dilute. Since the peak in the mutual information follows the phase transition, we can use the annealed approximation to see how the location of the phase transition varies with threshold and connectivity. In the annealed approximation for $T\gg 1$ and $K\gg 1$, Eq. (6) becomes

$$\left.\frac{dD(t+1)}{dD(t)}\right|_{D(t)=0} = K\frac{1}{2\sqrt{2\pi pK\sigma^2}}\exp\left(-\frac{T^2}{2pK\sigma^2}\right), \tag{13}$$

where σ^2 is the variance of $\rho(w_{ij})$. The critical value of K,K_e, is obtained by setting Eq. (13) equal to 1, which yields

$$p\sigma^2 K_c \log_2\left(\frac{K_e}{8\pi p\sigma^2}\right) = T^2. \tag{14}$$

The average activity p is found by solving an equation similar to Eq. (4) that takes into account a nonzero threshold. For $K\gg 1$, this equation can be written as

$$p(t+1) = \frac{1}{2}\sum_{n=1}^{K} p(t)^n[1-p(t)]^{K-n}\binom{K}{n}\mathrm{erfc}\left(\frac{T}{\sqrt{2n\sigma^2}}\right), \tag{15}$$

where $\mathrm{erfc}(\cdots)$ is the complementary error function. Depending on the values of K and T/σ, Eq. (15) may have a stable nonzero solution. Figure 7 is a plot of $p(t+1)$ versus $p(t)$ for three values of T for $K=50$. For larger thresholds, the only stable solution is $p=0$, i.e., the network has no spontaneous activity. It must be assumed that activity arises from external stimulation of the network. In Fig. 8, K_c is plotted as a function of T for several values of p. These data show that as T is increased, K_c increases rapidly to values in the range of connectivities seen in real neural networks.

4. Nonlinear summation

A criticism of some neural network models is that they rely on linear summation of inputs [33]. This is in contrast to some biological neurons, whose dendrites have been shown to add synaptic inputs in highly nonlinear ways [34]. Synaptic interaction can occur in both passive and active dendrites, allowing neurons to implement a fuller range of Boolean functions. To see how nonlinear summation affects the re-

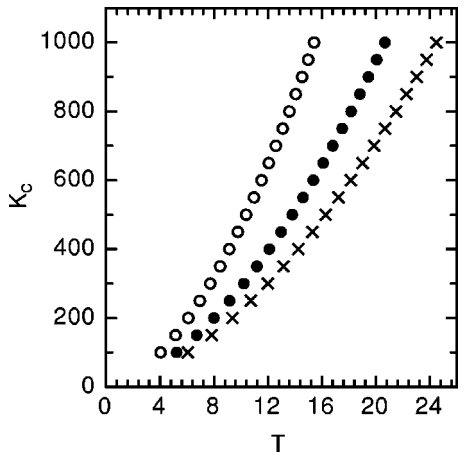

FIG. 8. The critical connectivity K_c as a function of threshold T according to Eq. (13): ○, $p=0.1$; •, $p=0.2$; ×, $p=0.3$.

sults of this paper, we measure the average mutual information in random Boolean networks, the extreme case of nonlinear summation. Figure 9 shows the mutual information as a function of connectivity for a Boolean network. The peak near the transition is clearly present. It is sharper and taller than for the neural network with linear summation. Mutual information provides a method for quantifying the computational benefit of various degrees of nonlinear summation.

III. DISCUSSION

In this paper, we use the concept of mutual information to investigate the relationship between a network's dynamics and its ability to perform computations. To determine the location of the order-chaos transition, we examined the dependence of the average period on system size and applied the annealed approximation to determine the time evolution of the distance between nearby trajectories. We showed how the average information per cell and the average mutual information per connection vary with the number of inputs. We found that, although $\langle H \rangle$ increases in the chaotic phase, $\langle I \rangle$ reaches a maximum near the transition.

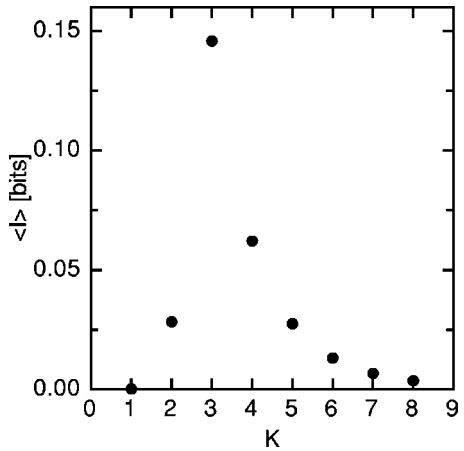

FIG. 9. Average mutual information per connection $\langle I \rangle$ versus K for a random Boolean network (simulation data for $N=10\,000$; average of 100 networks).

We demonstrated the robustness of this result by extending the model to incorporate the biologically relevant cases of nonzero threshold and nonlinear summation. We showed that, while the location of the transition changes with the threshold, mutual information always peaks near the transition. For larger thresholds, the annealed approximation predicts that K_c increases rapidly to values that are closer to the connectivities observed in biology. We also showed that the results carry over to the random Boolean network—the extreme case of nonlinear summation.

While the present model includes such physiological features as nonzero thresholds and dilute, asymmetric connections, it is clearly still quite removed from biological reality. The two-state neuron is a coarse graining of the average firing rate, which is itself an approximation of the neural spike train. However, spiking neurons are also binary on a different time scale, and the model could easily be adjusted to account for this with the introduction of a second time scale—one for the refractory period and one for the time over which a neuron integrates inputs. In fact, the present model can represent a spiking network for the special case where the integration time is equal to the refractory period. In this case, $\sigma=1$ would represent a spike within a small time bin on the scale of the refractory period and $\sigma=0$ would represent the absence of a spike. Such a change would be unlikely to eliminate the phase transition or the peak in mutual information. It would increase the average mutual information, though, because some additional information could be transmitted in the individual spike timing. A spiking model would require a source of noise which would make the model nondeterministic. In the present model, noise in the spike train is time averaged to produce the binary state.

The model also lacks spatial dimensionality since a connection between any two neurons is equally probable. However, it has already been shown that the phase transition also occurs in two- and three-dimensional RBNs. Thus, it is likely that the results of this paper will carry over as well. In a three-dimensional network, different patterns of connectivity could be examined.

These simplifications are reflected in the somewhat low level of mutual information and the small number of inputs for which the transition occurs. However, our results suggest that there is a direct relation between the dynamics of the system and the peak in mutual information. Generally, changes in the details of the underlying system will alter the location of the phase transition and the height of the mutual information peak, but will not eliminate them. The key component appears to be the transition from isolated clusters of activity to the percolation of activity throughout the network. Future work will incorporate more of the physiological features mentioned above, bringing the model even closer to biology and allowing comparison with experimental data.

ACKNOWLEDGMENTS

E.G. has been supported by the National Institutes of Health (NIMH) under Grant No. 1F31MH12286-01. E.G. thanks D. Rokhsar for helpful discussions.

APPENDIX A: (0,1) NEURONS VERSUS (−1,+1) NEURONS

The choice of representation depends on which aspects of neural networks are the primary focus of study. Neural network pioneers such as McCulloch and Pitts used (0,1) neurons to study the network as a logical device analogous to a digital computer [35]. Later workers, such as Hopfield [13] and Amit *et al.* [14], used $(-1,+1)$ neurons to take advantage of the similarity of their models to spin-glass systems and analyze them with the tools of statistical mechanics. In this paper, we choose (0,1) neurons because we are interested in studying physiologically realistic neurons and the influence of the neuronal threshold on network dynamics. In the case of $(-1,+1)$ neurons, an inhibitory connection $(w_{ij}<0)$ coming from an inactive neuron $(S_j=-1)$ acts as an excitation. Similarly, an excitatory connection from an inactive neuron is the same as an inhibition. This leads to dynamics that are difficult to reconcile with physiological networks. For example, a neuron with all of its inputs inactive would have a 50% chance of being active at the next time step. A real neuron would be inactive in such a situation. It is only in the (0,1) representation that the synaptic inputs maintain their excitatory and inhibitory character. Only with this invariance can the update function represent a sum of voltages.

At first glance, the two representations may seem mathematically equivalent. Although switching representations amounts to a simple change of variables it is easy to show that this change of variables will change the distribution of thresholds in potentially important ways. If we make the change of variables, $S_i=2\sigma_i-1$, so that $S_i\in\{-1,1\}$, the update function

$$\sigma_i(t+1)=\Theta\left[\sum_j w_{ij}\sigma_j(t)-T_i\right] \tag{A1}$$

becomes

$$S_i(t+1)=\mathrm{sgn}\left[\sum_j w_{ij}S_j(t)+\sum_j w_{ij}-2T_i\right]. \tag{A2}$$

The transformation generates an extra thresholdlike term, $\Sigma_j w_{ij}$, which is the sum over K random variables chosen from the distribution $\rho(w_{ij})$. Let $\rho(w_{ij})$ have a mean of w_o and a variance of σ^2. For a highly connected network ($K\gg 1$), the central limit theorem is applicable, and the term $\Sigma_j w_{ij}$ is a random variable from a Gaussian distribution with a mean of $K\times w_o$ and a variance of $K\times\sigma^2$. If $\rho(w_{ij})$ is symmetric, then $w_o=0$, and thus the average value of the extra term is zero. But, since the variance is large, the extra threshold term can be very large. For a weakly connected (dilute) network such as the one described in this paper, the extra term is the sum of a small number of random variables. In general, this term is nonzero and different for each neuron. There is no possible distribution $\rho(w_{ij})$ that will make the extra term always zero. Essentially, zero threshold in one representation corresponds to a broad distribution of thresholds in the other. The inequivalence of the two representations has been pointed out previously (for example, [1,36]).

APPENDIX B: DERIVATION OF THE AVERAGE ACTIVITY

The distribution of activities $f(p)$ is a complicated function for which no analytic solution is known. Following [37], it can be expressed as the solution to the integral equation

$$\begin{aligned} f(p)=&\int dw_{i1}\rho(w_{i1})\cdots dw_{iK}\rho(w_{iK})\\ &\times\int dp_1 f(p_1)\cdots dp_K f(p_K)\\ &\times\delta\Bigg[p-\sum_{\sigma_1=0,1}P(\sigma_1|p_1)\cdots\sum_{\sigma_K=0,1}P(\sigma_K|p_K)\\ &\times\Theta\left(\sum_j w_{ij}\sigma_j\right)\Bigg]. \end{aligned} \tag{B1}$$

Using Eq. (B1) the average activity can written as

$$\begin{aligned} p=&\int dp' p'\int dw_{i1}\rho(w_{i1})\cdots dw_{iK}\rho(w_{iK})\\ &\times\int dp_1 f(p_1)\cdots dp_K f(p_K)\\ &\times\delta\Bigg[p'-\sum_{\sigma_1=0,1}P(\sigma_1|p_1)\cdots\sum_{\sigma_K=0,1}P(\sigma_K|p_K)\\ &\times\Theta\left(\sum_j w_{ij}\sigma_j\right)\Bigg]. \end{aligned} \tag{B2}$$

Integrating over p' gives:

$$\begin{aligned} p=&\int dw_{i1}\rho(w_{i1})\cdots dw_{iK}\rho(w_{iK})\\ &\times\int dp_1 f(p_1)\cdots dp_K f(p_K)\\ &\times\sum_{\sigma_1=0,1}P(\sigma_1|p_1)\cdots\sum_{\sigma_K=0,1}P(\sigma_K|p_K)\Theta\left(\sum_j w_{ij}\sigma_j\right). \end{aligned} \tag{B3}$$

Integrating over the weights gives the following for any distribution of weights $\rho(w_{ij})$ that is symmetric about zero:

$$\begin{aligned} p=&\int dp_1 f(p_1)\cdots dp_K f(p_K)\\ &\times\sum_{\sigma_1=0,1}P(\sigma_1|p_1)\cdots\sum_{\sigma_K=0,1}P(\sigma_K|p_K)\\ &\times\frac{1}{2}(1-\delta_{\sigma_1,0}\cdots\delta_{\sigma_K,0}). \end{aligned} \tag{B4}$$

Evaluating the sums over all input states:

$$p=\frac{1}{2}-\frac{1}{2}\left[\int dp_1 f(p_1)(1-p_1)\right]^K \tag{B5}$$

$$=\frac{1-(1-p)^K}{2}. \tag{B6}$$

APPENDIX C: FINITE-SIZE EFFECTS

In studying this system, significant finite-size effects were encountered. Such effects are common in percolation problems (see [38]) and are expected to be larger in this system since there is variation in the percolation parameter for a given value of K. One finite-size effect occurs in the measurement of the average activity of a network in the ordered phase. This quantity is zero in the limit $N\to\infty$ but is nonzero for finite N because of small localized periodic attractors that can exist. For $K=2$, in the range of $N=40$ to $N=20\,000$, the average activity p goes to zero as power law in N: $p(K=2)\sim N^{-0.3}$.

A more important finite-size effect is that which alters the location of the peak in the mutual information. The average mutual information $\langle I\rangle$ is plotted versus K for several values of N in Fig. 10. For $N=80$, the peak is at $K=3$, the lowest value of K in the chaotic phase. As N increases, this peak decreases and a new peak appears at $K=5$. By $N=50\,000$, the curve appears to reach a limiting form. Examination of Fig. 10 reveals that the finite-size effect displayed here is not a gradual shift in peak location with increasing N. Rather, there are two peaks—one at $K=3$ that decreases and one at $K=5$ that increases slightly.

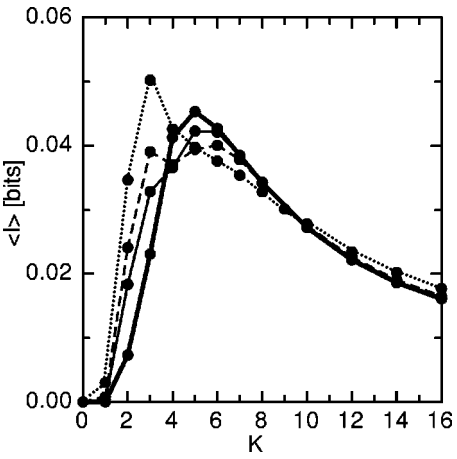

FIG. 10. Average mutual information per connection $\langle I\rangle$ versus K for various system sizes from computer simulations: dotted line, $N=200$; dashed line, $N=800$; thin solid line, $N=2000$; thick solid line, $N=50\,000$.

The origin of both of these finite-size effects can be understood by looking at the distribution of activities for various system sizes. At smaller values of N, there are peaks at multiples of the reciprocals of small integers ($\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{3}{4}$, etc.), which result from local attractors with short periods. These attractors are due to small feedback loops that occur in smaller networks but become increasingly rare as $N\to\infty$. In models that are three dimensional and have local connections, these types of attractors may play a more important role since they would be likely to occur in networks of all sizes.

[1] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, MA, 1991).
[2] R. Harris-Warrick, E. Marder, A. Selverston, and M. Moulins. *Dynamic Biological Networks, The Stomatogastric Nervous System* (MIT Press, Cambridge, MA, 1992).
[3] S. Grillner *et al.*, Trends Neurosci. **18**, 270 (1995).
[4] S. Kauffman, J. Theor. Biol. **22**, 437 (1969).
[5] S. Kauffman, Physica D **10**, 145 (1984).
[6] S. Kauffman, *The Origins of Order* (Oxford, New York, 1993).
[7] K. Kurten, Phys. Lett. A **129**, 157 (1988).
[8] H. Gutfreund, J. Reger, and A. Young, J. Phys. A **21**, 2775 (1988).
[9] A. Crisanti, M. Falcioni, and A. Vulpiani, J. Phys. A **26**, 3441 (1993).
[10] A. Crisanti, M. Falcioni, and A. Vulpiani, Phys. Rev. Lett. **76**, 612 (1996).
[11] B. Derrida, J. Phys. A **20**, L721 (1987).
[12] H. Sompolinsky, A. Crisanti, and H. Sommers, Phys. Rev. Lett. **61**, 259 (1988).
[13] J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[14] D. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[15] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[16] K. Kurten, J. Phys. A **21**, L615 (1988).
[17] B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986).
[18] M. Mitchell, P. Hraber, and J. Crutchfield, Complex Syst. **7**, 89 (1993).
[19] J. Crutchfield, in *Towards the Harnessing of Chaos*, edited by M. Yamaguti (Elsevier Science, Amsterdam, 1994).
[20] S. Wolfram, Physica D **10**, 1 (1984).
[21] J. Crutchfield and M. Mitchell, Proc. Natl. Acad. Sci. U.S.A. **92**, 10 742 (1995).
[22] J. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).
[23] E. Bonabeau, Acta Biotheor. **45**, 29 (1997).
[24] G. Weisbuch, in *Theories of Immune Networks*, edited by H. Atlan and I. Cohen (Springer-Verlag, New York, 1989).
[25] A. Bernardes and R. dos Santos, J. Theor. Biol. **186**, 173 (1997).
[26] M. Paczuski, K. Bassler, and A. Corral, Phys. Rev. Lett. **84**, 3185 (2000).
[27] H. Flyvbjerg, J. Phys. A **21**, L955 (1988).

[28] B. Derrida and H. Flyvbjerg, J. Phys. (Paris) **45**, 971 (1987).
[29] H. Flyvbjerg and N. Kjaer, J. Phys. A **21**, 195 (1988).
[30] C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1998).
[31] As K increases, Eq. (1) can produce a rapidly decreasing fraction of the 2^{2^K} possible Boolean functions. This fraction is 2/4 for $K=1$, 6/16 for $K=2$, and 32/256 for $K=3$. The sensitivity of these functions to changes in the input states can be quantified by measuring the probability that a change of state of one of the inputs causes a change in the output state.
[32] A. Borst and F. Theunissen, Nature Neurosci. **2**, 947 (1999).
[33] C. Koch and G. Laurent, Science **284**, 96 (1999).
[34] C. Koch, *Biophyiscs of Computation* (Oxford University Press, New York, 1999), Chap. 14.
[35] W. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).
[36] M. Abeles, *Corticonics* (Cambridge University Press, Cambridge, 1991).
[37] B. Derrida and H. Flyvbjerg, J. Phys. A **20**, L1107 (1987).
[38] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1994), Chap. 4.