Energy approach to rivalry dynamics, soliton stability, and pattern formation in neuronal networks

P. N. Loxley[1,2] and P. A. Robinson[1,2,3]
[1]*School of Physics, The University of Sydney, Sydney, New South Wales 2006, Australia*
[2]*Brain Dynamics Center, Westmead Millennium Institute, Westmead Hospital and The University of Sydney, Westmead, New South Wales 2145, Australia*
[3]*Faculty of Medicine, The University of Sydney, Sydney, New South Wales 2006, Australia*
(Received 8 June 2007; revised manuscript received 25 July 2007; published 31 October 2007)

Hopfield's Lyapunov function is used to view the stability and topology of equilibria in neuronal networks for visual rivalry and pattern formation. For two neural populations with reciprocal inhibition and slow adaptation, the dynamics of neural activity is found to include a pair of limit cycles: one for oscillations between states where one population has high activity and the other has low activity, as in rivalry, and one for oscillations between states where both populations have the same activity. Hopfield's Lyapunov function is used to find the dynamical mechanism for oscillations and the basin of attraction of each limit cycle. For a spatially continuous population with lateral inhibition, stable equilibria are found for local regions of high activity (solitons) and for bound states of two or more solitons. Bound states become stable when moving two solitons together minimizes the Lyapunov function, a result of decreasing activity in regions between peaks of high activity when the firing rate is described by a sigmoid function. Lowering the barrier to soliton formation leads to a pattern-forming instability, and a nonlinear solution to the dynamical equations is found to be given by a soliton lattice, which is completely characterized by the soliton width and the spacing between neighboring solitons. Fluctuations due to noise create lattice vacancies analogous to point defects in crystals, leading to activity which is spatially inhomogeneous.

DOI: 10.1103/PhysRevE.76.046224 PACS number(s): 89.75.Kd, 87.18.Sn, 05.45.Yv, 87.18.Hf

I. INTRODUCTION

Dynamical models of interacting neural populations proposed originally to describe key aspects of brain activity [1,2] and cortical map development [3,4] have been used more recently to successfully uncover important features of binocular rivalry [5–7], visual attention [8], visual hallucinations [9–11], and development of orientation preference in the primary visual cortex [4,12]. The main assumptions in these models include details of the connections between different neural populations. In particular, networks with recurrent connections are known to lead to a range of multistable phenomena, including hysteresis, winner-take-all dynamics, and the formation of spatiotemporal patterns (for a review, see Ref. [13]).

Although model networks represent a vast simplification of reality, they can still be difficult to gain insight from. One reason is that networks which are computationally useful, or which display a range of dynamical behavior, often also contain a significant number of equilibria. This is seen in artificial neural networks for associative memory, where retrieving a memory from incomplete information is equivalent to the dynamics of moving towards a single attractor in a space of states containing many attractors, some of which are analogous to spin glass states [14,15]. In addition, biologically motivated neural networks, often called *neuronal networks* [13], have equilibria that change in stability over time due to processes such as synaptic plasticity [16], spike-rate adaptation [8,17,18], and time-varying input stimuli [7]. Such dynamical features further complicate analysis.

The purpose of this work is to present an energy approach to understanding the dynamics of activity in neuronal networks. The approach allows the stability and topology of equilibria to be viewed—including the basins of attraction, basin boundaries, etc.—in terms of the maxima, minima, and saddle points of an appropriate energy surface. This provides a particularly intuitive way to understand the dynamics and uncovers new results which would be more difficult to show using alternative methods. In order to illustrate this, we consider models for visual rivalry and neuronal pattern formation. The energy surface is constructed using Hopfield's Lyapunov function [19], which is analogous to the potential energy function of a nonconservative system and is minimized by all stable equilibria of the dynamical equations. Although Hopfield's Lyapunov function is only valid for networks with symmetric connections, we show how it can be used to understand behavior in networks with asymmetric connections when a clear separation of dynamical time scales exists.

An energy approach to understanding behavior in neuronal networks has several advantages over the usual phase-plane methods for nonlinear differential equations. These typically involve constructing nullclines—that is, curves along which the time rate of change of a particular variable is zero—and finding where they intersect to give an equilibrium, then determining local stability from consideration of small deviations from each equilibrium. Equilibria and equilibrium stability can be seen directly from an energy surface. In addition, the energy surface shows the basins of attraction, allowing global dynamics to be visualized in a manner similar to predicting the motion of a particle on a potential energy surface in the limit of large friction. This is used in Sec. III to find the dynamical mechanism for limit-cycle oscillations in the activity of two interacting neural populations and

generalizes the energy approach of Ref. [18] for a single population. It is also used to predict when such oscillations may take place.

Another advantage of an energy approach is that it enables the use of variational techniques for finding equilibria. This is illustrated using *collective coordinates*, developed in condensed matter physics [20] and quantum field theory [21] for treating the particlelike nature of solitons. In Sec. IV, stable equilibria are found for spatially localized regions of high neural activity, also called bumps [17,22,23] or solitary peaks [13]. We use the term soliton here, as it is more widely recognized, and the stable equilibria are spatially localized solutions of a nonlinear field equation [21]. In systems containing solitons only a few degrees of freedom may be important. For example, during interactions between solitons there may be little change in the soliton shape, but a large change in its position, the separation of two or more solitons, or the position of the "center of mass" of a collection of solitons [20]. It is often then possible to derive a variational energy depending only on these coordinates, such that its maxima, minima, and saddle points give equilibria of the system.

A final advantage of an energy approach is that the effects of fluctuations due to random noise can be investigated using analogies with statistical mechanics (cf. Ref. [15]). In Sec. IV C, we show how noise fluctuations can decrease long-range order in patterned neural activity, generating activity which is spatially inhomogeneous. The general model and corresponding Lyapunov function are described in Sec. II. In Sec. III, we include reciprocal inhibition and slow adaptation to yield a simple model of visual rivalry. Two alternative mechanisms for oscillations are found in Sec. III A, and limit-cycle behavior is described in Sec. III B. We consider a spatially continuous network with lateral inhibition in Sec. IV and investigate pattern formation by using collective coordinates and constructing a variational energy. Equilibria are found for a single soliton in Sec. IV A and for soliton interactions in Sec. IV B. A soliton lattice is treated in Sec. IV C. The main results of this approach, as well as its advantages and limitations, are summarized in Sec. V. Alternative energy approaches are also discussed.

II. MODEL AND HOPFIELD ENERGY

The model we consider describes activity aggregated over many neurons. The activity u_i of a neural population i (as given by the mean cell-body potential) depends on the activity of other neural populations, j, through a weight W_{ij} representing the mean synaptic connectivity, and on the level of input stimuli h_i. Connections are termed *excitatory* if $W_{ij}>0$ and *inhibitory* if $W_{ij}<0$. The mean-field dynamical equation is given by

$$\frac{1}{\alpha_i}\frac{du_i}{dt} = -u_i + \sum_j W_{ij}S(u_j) + h_i, \tag{1}$$

where $S(u_j)$ is the spiking rate of population j, α_i^{-1} is the time scale over which action potential spikes are smoothed by the synapses and dendrites of population i, and the sum is over all neural populations [1,2,13,14,24]. Assuming a distribution of firing thresholds as in Ref. [24], with mean θ_i, and standard deviation $\sigma_i\pi/\sqrt{3}$, the firing rate function is chosen to have the form

$$S(u_i) = \frac{1}{1+\exp[-(u_i-\theta_i)/\sigma_i]}, \tag{2}$$

which increases from 0 to 1 as u_i crosses the mean firing threshold in the step function limit $\sigma\to 0$; otherwise, the transition is more gradual. Equilibria of Eq. (1) satisfy $du_i/dt=0$.

A Lyapunov function $E(u_i)$ exists for Eq. (1) when connections are symmetric: $W_{ij}=W_{ji}$, as originally shown by Hopfield [19]. This function is a generalization of the potential energy function of a nonconservative system and shares the important property that $dE/dt\leqslant 0$ for every solution to Eq. (1) with $W_{ij}=W_{ji}$. That is, all trajectories go "downhill" on the energy surface, although not necessarily taking the path of steepest descent, until reaching a minimum of E. This minimum corresponds to a stable equilibrium of Eq. (1) and is often termed a fixed-point attractor. The Lyapunov function corresponding to Eq. (1) with $W_{ij}=W_{ji}$ is given by

$$E = -\frac{1}{2}\sum_{i,j} S(u_i)W_{ij}S(u_j) + \sum_i \int_0^{S(u_i)} S^{-1}(V)dV - \sum_i h_i S(u_i). \tag{3}$$

For the choice of firing rate function in Eq. (2), we find

$$\int_0^{S(u_i)} S^{-1}(V)dV = \theta_i S(u_i) + \sigma_i[1-S(u_i)]\ln[1-S(u_i)] + \sigma_i S(u_i)\ln S(u_i). \tag{4}$$

The equilibria of Eq. (1) are the critical points of E—that is, its maxima, minima, and saddle points. For N populations there will be a maximum of 3^N critical points when the firing rate has the form of a sigmoid function as in Eq. (2). The minima of E are stable equilibria, as mentioned, while the maxima and saddle points are unstable equilibria. The advantage of Eq. (3) is that the stability of equilibria and the topology of the dynamics can be directly seen from the curvature of E. At a critical point of E the curvature is given by the matrix of second derivatives:

$$\frac{\partial^2 E}{\partial u_i \partial u_j} = \rho_j(\delta_{ij} - \rho_i W_{ij}), \tag{5}$$

where $\rho_i=S'(u_i)$ is evaluated at the critical point. Considering small deviations δu_i from a critical point allows Eq. (3) to be approximated as

$$E \simeq E_0 + \frac{1}{2}\sum_i \lambda_i \delta u_i^2, \tag{6}$$

where E_0 is the value of E at the critical point and λ_i are eigenvalues of the matrix given by Eq. (5). It is clear from Eq. (6) that the signs of these eigenvalues determine the type of critical point of E: When $\lambda_i>0$ for all i, the critical point is a minimum; when $\lambda_i<0$ for all i, it is a maximum; and

when $\lambda_i > 0$ for some i, and $\lambda_i < 0$ for others, the critical point is a saddle point.

III. VISUAL RIVALRY

During experiments on binocular rivalry, a separate image is presented to each eye. When these images are conflicting the visual system is thrown into oscillations, so that only one of the images is perceived at any time, and this perception continually alternates between the two images [5–7]. Here, we present a simple model of rivalry that contains the essential elements from previous models [5–8,25]. Our purpose is to show how an energy approach can be used to understand rivalry. In particular, we use it to find the dynamical mechanism responsible for rivalry oscillations.

Consider two neural populations (for example, left-eye and right-eye monocular neurons in the primary visual cortex) with activities u_1 and u_2, respectively. A simple neuronal network which includes adaptation is given by the set of equations

$$\frac{1}{\alpha}\frac{du_1}{dt} = -u_1 + W_{11}S(u_1) + W_{12}S(u_2) + h_1, \tag{7}$$

$$\frac{1}{\alpha}\frac{du_2}{dt} = -u_2 + W_{21}S(u_1) + W_{22}S(u_2) + h_2, \tag{8}$$

$$\frac{1}{\eta}\frac{d\theta_1}{dt} = -\theta_1 + \Lambda u_1, \tag{9}$$

$$\frac{1}{\eta}\frac{d\theta_2}{dt} = -\theta_2 + \Lambda u_2, \tag{10}$$

where h_1 and h_2 are inputs from the left and right eyes, for example; η^{-1} is the time scale of adaptation; Λ gives the strength of negative feedback due to adaptation; and the $S(u_i)$ depend on the θ_i through Eq. (2). The network following from Eqs. (7)–(10) is asymmetric. Connections in Eqs. (7) and (8) are symmetric when $W_{ij} = W_{ji}$; however, in Eqs. (7)–(10) the coupling between u_i and θ_i breaks this symmetry. Lyapunov functions usually do not exist for asymmetric networks [26], and limit cycles become possible when asymptotic dynamics is not restricted to a set of stable equilibria.

A. Oscillation mechanisms

For slow adaptation $\eta \ll \alpha$, and a clear separation of time scales exists in Eqs. (7)–(10). Over the fast time scale α^{-1}, the variables θ_1 and θ_2 are approximately constant, and the dynamics reduces to Eqs. (7) and (8). The fast dynamics can then be characterized using Hopfield's Lyapunov function when $W_{12} = W_{21}$, and solution trajectories go downhill on an energy surface that varies slowly in time, depending on which trajectory is taken. A qualitative dynamical description is then possible by making use of both the topology of the energy surface and the dynamical behavior of Eqs. (9) and (10). Similar approaches can also be applied to any neuronal network with short-time-scale dynamics that matches a known Lyapunov function.

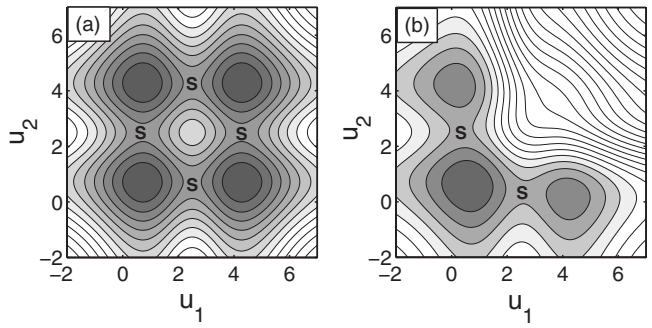

FIG. 1. Energy from Eq. (3) versus u_1 and u_2 for $W_{11} = W_{22} = 5$, $\theta_1 = \theta_2 = 2.5$, $\sigma_1 = \sigma_2 = 1$, and $h_1 = h_2 = 0$. In (a), there is a central maximum (light circle), four surrounding minima (dark circles), and four saddles points (S) when $W_{12} = W_{21} = 0$. In (b), there are three minima and two saddles points when $W_{12} = W_{21} = -0.3$.

With no interaction between two neural populations (but allowing for recurrent excitation within each population), $W_{12} = W_{21} = 0$, and the energy surface following from Eqs. (3) and (4) is shown in Fig. 1(a). There are nine critical points of E in this figure: a central maximum, four surrounding minima, and four saddle points, one separating each pair of minima. The four minima are stable equilibria, where either both populations have the same activity (top-right and bottom-left minimums) or one population is at high activity and the other is at low activity (top-left and bottom-right minima). The maximum and saddles points are unstable equilibria. It is possible to keep track of the number, location, and stability of equilibria by seeing how the energy surface changes as the parameters are varied continuously. Solution trajectories of Eqs. (7) and (8) can also be visualized by considering the basins of attraction and basin boundaries formed by the critical points of E.

Reciprocal inhibition is a key element in rivalry and implies $W_{12} = W_{21} < 0$ when connections are symmetric. The resulting energy surface is shown in Fig. 1(b), where it is seen that the minimum corresponding to the stable equilibrium with both populations at high activity has merged with the maximum and two neighboring saddles points, following a saddle-node bifurcation. Now only one population can be in a stable high-activity equilibrium at any time—a key feature of rivalry. Spike-rate adaptation causes a neural population to switch from high activity to low activity after a characteristic period of time, generally tens of milliseconds to several seconds in the visual system [6–8]. This is included by allowing θ_1 and θ_2 to vary over the slow time scale η^{-1} according to Eqs. (9) and (10). The energy surface then evolves slowly in time, and two different dynamical mechanisms for oscillations become possible, depending on the initial state.

1. Rivalry oscillations

Starting at the top-left minimum of Fig. 1, corresponding to population 1 at low activity and population 2 at high activity, leads to an increase in θ_1 via Eq. (9) and a decrease in θ_2 via Eq. (10). After a time η^{-1}, the asymmetry in θ_1 and θ_2

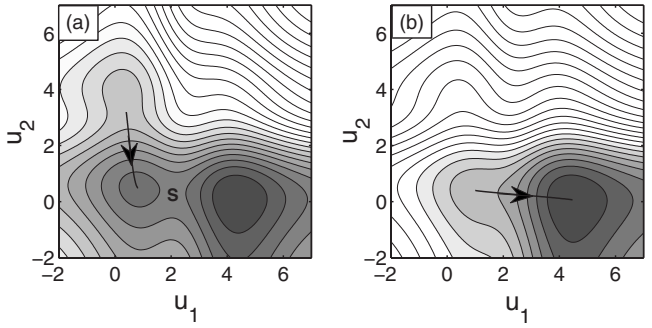

FIG. 2. Energy from Eq. (3) versus u_1 and u_2, and a trajectory solving Eqs. (7) and (8) (arrows), for different values of θ_i. In (a), the top-left minimum vanishes when $\theta_1=2.4$ and $\theta_2=2.6$. In (b), the bottom-left minimum vanishes when $\theta_1=2.3$ and $\theta_2=2.7$.

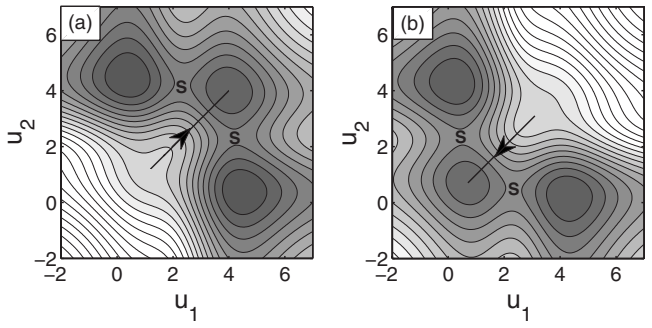

FIG. 3. Energy from Eq. (3) versus u_1 and u_2, and a trajectory solving Eqs. (7) and (8) (arrows), for different values of θ_i. In (a), the bottom-left minimum vanishes when $\theta_1=\theta_2=2.25$. In (b), the top-right minimum vanishes when $\theta_1=\theta_2=2.45$.

results in the disappearance of the occupied top-left minimum through a saddle-node bifurcation, as shown in Fig. 2(a), followed by downhill motion towards the bottom-left minimum where both populations are at low activity. As the asymmetry continues to increase, the bottom-left minimum also vanishes, as shown in Fig. 2(b), followed by downhill motion towards the remaining minimum at the bottom-right, where population 1 is at high activity and population 2 is at low activity. The motion is downhill in both cases because the height of an energy minimum can increase during a saddle-node bifurcation. A similar sequence of steps then returns the system to the top-left minimum. This mechanism gives rivalry oscillations, as the activities oscillate between states where one population is at high activity and the other is at low activity (out-of-phase oscillations). In Fig. 2, it is seen that going via the bottom-left minimum can lead to an L-shaped trajectory for rivalry oscillations.

2. In-phase oscillations

Starting at the bottom-left minimum of Fig. 1, where both populations are at low activity, and requiring that $\theta_1=\theta_2$ initially causes both θ_1 and θ_2 to decrease at exactly the same rate. Now there is no asymmetry between θ_1 and θ_2 or between u_1 and u_2. After a time η^{-1}, the bottom-left minimum vanishes through a saddle-node bifurcation with the maximum and two neighboring saddles, as shown in Fig. 3(a), followed by downhill motion towards the top-right minimum where both populations are at high activity. Following this, θ_1 and θ_2 both begin to increase and the top-right minimum vanishes after another η^{-1} period, as in Fig. 3(b). Motion is then downhill towards the bottom-left minimum where both populations start at low activity again. This mechanism gives oscillations which do not correspond to rivalry, as the activity oscillates between states where both populations have the same activity (in-phase oscillations). The lack of asymmetry means this trajectory does not go via an intermediate-energy minimum as is possible in the rivalry case, but always follows a straight line between the top-right and bottom-left minima.

B. Limit-cycle behavior

The energy given by Eq. (3) can also be used to predict when oscillations take place. To simplify analysis we consider a network with a high degree of symmetry: $W_{11}=W_{22}$, $W_{12}=W_{21}$, and $h_1=h_2$. Equilibria of Eqs. (7)–(10) are then given by $u_1=u_2=u_0$ and $\theta_1=\theta_2=\theta_0$, where u_0 and θ_0 are intersections of the nullclines of Eqs. (7)–(10),

$$\theta = u + \sigma \ln\left(\frac{W_{11}+W_{12}}{u-h}-1\right), \tag{11}$$

$$\theta = \Lambda u, \tag{12}$$

and where $h=h_1=h_2$ has been defined. This is shown in Fig. 4(a) for $\Lambda=2.5$. The line given by Eq. (12) (dashed line) intersects the curve given by Eq. (11) (solid line) exactly once at $(u,\theta)=(u_0,\theta_0)$, corresponding to a single equilibrium of Eqs. (7)–(10). When $\eta \ll \alpha$, the variables θ_1 and θ_2 remain approximately constant over the characteristic time α^{-1} and the fast dynamics depends only on Eqs. (7) and (8). Upon using $\theta_1=\theta_2=\theta_0$ in Eqs. (3) and (4), the equilibrium stability can be determined from the curvature of the E surface at $u_1=u_2=u_0$.

The energy surface given by Eqs. (3) and (4) with $\theta_1=\theta_2=\theta_0$ is plotted in Fig. 4(b), and the position of the equilibrium at $u_1=u_2=u_0$ implies that the equilibrium is a minimum of E and is therefore stable. This is confirmed from the

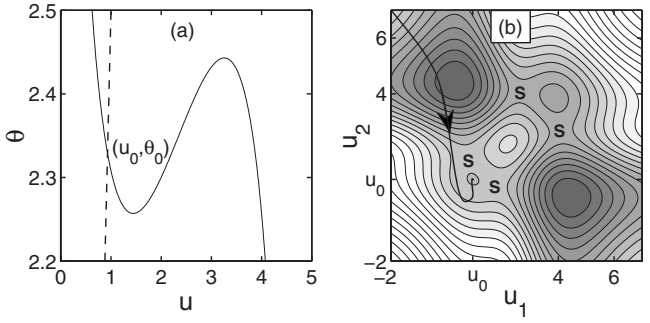

FIG. 4. Dynamics of Eqs. (7)–(10) with $W_{11}=W_{22}=5$, $W_{12}=W_{21}=-0.3$, $h_1=h_2=0$, $\sigma_1=\sigma_2=1$, $\Lambda=2.5$, $\alpha^{-1}=20$ ms, and $\eta^{-1}=900$ ms. In (a), the intersection of Eq. (11) (solid line) and Eq. (12) (dashed line) results in a single equilibrium at (u_0,θ_0). In (b), a plot of Eq. (3) with $\theta_1=\theta_2=\theta_0$ yields a minimum at $u_1=u_2=u_0$. A trajectory solving Eqs. (7)–(10) is also shown (arrow).

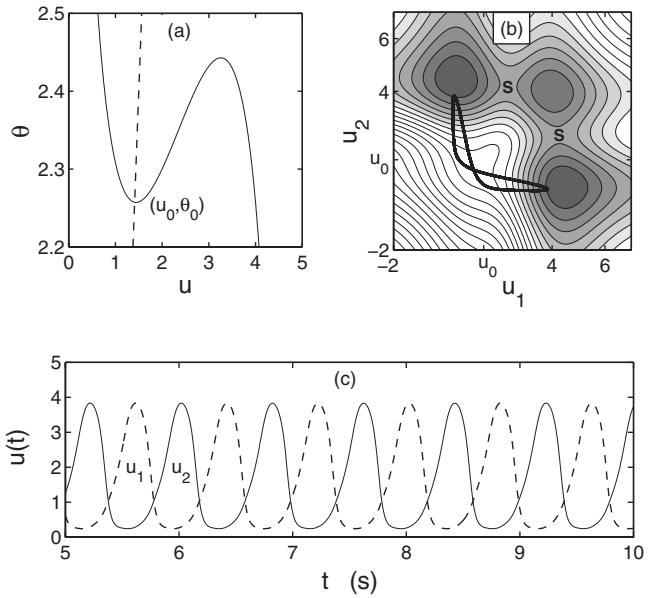

FIG. 5. Dynamics of Eqs. (7)–(10) for $\Lambda=1.5$. In (a), the intersection of Eq. (11) (solid line) and Eq. (12) (dashed line) is shown. In (b), a plot of Eq. (3) with $\theta_1=\theta_2=\theta_0$ yields a saddle point at $u_1=u_2=u_0$. The dynamics given by $u_1(t)$ and $u_2(t)$ after neglecting transient behavior is also shown (thick line). In (c), $u_1(t)$ (dashed line) and $u_2(t)$ (solid line) versus t is shown.

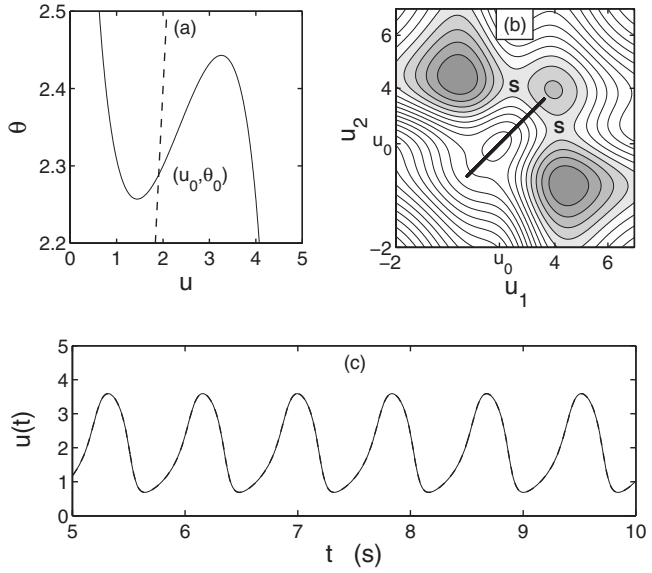

FIG. 6. Dynamics of Eqs. (7)–(10) for $\Lambda=1.2$ and initial condition $u_1=u_2$ and $\theta_1=\theta_2$. In (a), the intersection of Eq. (11) (solid line) and Eq. (12) (dashed line) is shown. In (b), a plot of Eq. (3) with $\theta_1=\theta_2=\theta_0$ yields a maximum at $u_1=u_2=u_0$. The dynamics given by $u_1(t)$ and $u_2(t)$ after neglecting transient behavior is also shown (thick line). In (c), $u_1(t)$ (dashed line) and $u_2(t)$ (solid line) versus t is shown.

trajectory of a point evolving under Eqs. (7)–(10), which is also shown in Fig. 4(b).

Changing Λ changes the slope of the dashed line in Fig. 4(a) and yields new values for u_0 and θ_0. This is shown in Fig. 5(a) for $\Lambda=1.5$. It is seen from the energy surface in Fig. 5(b) that an increase in u_0 and a decrease in θ_0 result in u_0 no longer being a minimum. To confirm this, we consider the matrix of second derivatives at a critical point of E. From Eq. (5), this can be written as

$$\begin{pmatrix} \rho(1-G_a) & \rho G_b \\ \rho G_b & \rho(1-G_a) \end{pmatrix},$$

where we have defined $\rho=\rho_1=\rho_2$, $G_a=\rho W_{11}=\rho W_{22}$, and $G_b=\rho|W_{12}|=\rho|W_{21}|$, and we have assumed $W_{11}=W_{22}>0$ and $W_{12}=W_{21}<0$ due to recurrent excitation and reciprocal inhibition, respectively. The eigenvalues of this matrix are

$$\lambda_\pm=\rho(1-G_a\pm G_b). \tag{13}$$

When $G_a+G_b<1$, both eigenvalues are positive and the critical point is a minimum. When $G_a-G_b>1$, both eigenvalues are negative and the critical point is a maximum. Finally, if $G_a+G_b>1$ and $G_a-G_b<1$, one eigenvalue is positive while the other is negative, so the critical point is a saddle point. The critical point at $u_1=u_2=u_0$ in Fig. 5(b) is found to be a saddle point, with the direction of decreasing energy given by the $u_1=-u_2$ line. This path of decreasing energy means the equilibrium has exactly one unstable direction on the energy surface. The instability leads to a limit cycle in Eqs. (7)–(10), which is shown as a projection of phase space onto u_1 and u_2 in Fig. 5(b) and as $u_1(t)$ and $u_2(t)$ versus t in Fig. 5(c). It is seen that u_1 oscillates out of phase with u_2, so when the activity of population 1 is high, population 2 has low activity and vice versa. This limit cycle corresponds to the rivalry mechanism discussed in Sec. III A.

Decreasing Λ to $\Lambda=1.2$ in Fig. 6(a) increases both u_0 and θ_0 as the intersection point passes a local minimum in the nullcline. The relevant critical point in Fig. 6(b) changes from a saddle point to a maximum, with the second direction of decreasing energy given by the $u_1=u_2$ line. The equilibrium now has two unstable directions on the energy surface. Choosing any initial condition along this second unstable direction, as long as it is not the equilibrium, leads to a second limit cycle, as shown in Figs. 6(b) and 6(c). It is seen that u_1 and u_2 now oscillate in phase, corresponding to the mechanism for in-phase oscillations discussed in Sec. III A. The key point, however, is that the basin of attraction for this second limit cycle consists only of those points satisfying $u_1=u_2$ and $\theta_1=\theta_2$; all other points are attracted to the first limit cycle. This is shown in Fig. 7, where we choose similar conditions to Fig. 6, but now with the initial condition just off the $u_1=u_2$ line. After initially oscillating in phase, u_1 and

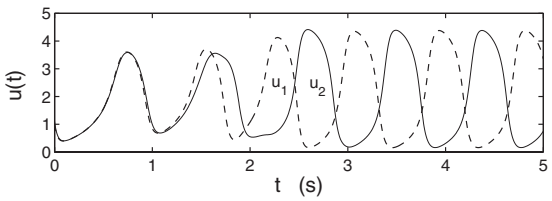

FIG. 7. Dynamics of Eqs. (7)–(10) for the same parameter values as in Fig. 6, but choosing an initial condition with $u_1\neq u_2$.

u_2 eventually settle into out-of-phase oscillations, as in Fig. 5. The limit cycle for in-phase oscillations is therefore unstable, while that for out-of-phase oscillations is stable. The reason is that any initial asymmetry in u_1 or u_2 leads to asymmetry in θ_1 or θ_2 via Eqs. (9) and (10), and subsequently triggers the rivalry mechanism described in Sec. III A.

What is the role of this second limit cycle? When $G_a - G_b < 1$, oscillations are dominated by a saddle point on the energy surface. The stable direction is given by the $u_1 = u_2$ line, and choosing any initial condition satisfying $u_1 = u_2$ and $\theta_1 = \theta_2$ leads to the equilibrium at (u_0, θ_0). Choosing any other initial condition, on the other hand, leads to the limit cycle for out-of-phase oscillations shown in Fig. 5. So phase space is divided into two basins of attraction: one given by all points satisfying $u_1 = u_2$ and $\theta_1 = \theta_2$ and one given by all other points. When $G_a - G_b > 1$, oscillations are dominated by a maximum on the energy surface. The first basin of attraction now leads either to the limit cycle for in-phase oscillations shown in Fig. 6 or to the equilibrium if $u_1 = u_2 = u_0$. So, although the mechanism for in-phase oscillations exists even when reciprocal inhibition is present, the relevant basin of attraction is very small. This means in-phase oscillations are unlikely ever to be observed in time series data, found either numerically or experimentally.

Experimental measurements of binocular rivalry show there is little correlation between successive high-activity durations [5,25]. It is expected that including random noise in Eqs. (7)–(10) will lead to this behavior, as it does in similar deterministic models of rivalry [5]. Within the energy approach, it is clear that switching between different high-activity states would then depend on the relative time scales of adaptation versus activation over an energy barrier.

IV. NEURAL PATTERN FORMATION

Spatially nonuniform neural activity has been widely investigated in various neuronal models (for a recent review, see Ref. [13]) and is thought to be important for phenomena such as short-term memory [1,22], visual hallucinations [9–11], cortical map formation [4,12], and rate coding in binocular rivalry [25]. In such models, it is generally assumed that the neural connectivity has a characteristic spatial range. Our purpose is to use an energy approach to investigate patterns in a neuronal network with spatially distributed connections. In this case we consider the continuum limit of Eqs. (1)–(3).

For a one-dimensional (1D) network and a translation-invariant synaptic density $W(x-x')$, the continuum limit of Eq. (1) leads to the replacements $u_i \to u(x)$, $u_j \to u(x')$, $h_i \to h(x)$, and $\Sigma_j W_{ij} \to \int_{-\infty}^{\infty} W(x-x')dx'$, yielding

$$\frac{1}{\alpha}\frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_{-\infty}^{\infty} W(x-x')S[u(x',t)]dx' + h(x), \tag{14}$$

where

$$S[u(x,t)] = \frac{1}{1+\exp\{-[u(x,t)-\theta]/\sigma\}}. \tag{15}$$

These equations describe a single neural population with spatially distributed connections. The corresponding Lyapunov function follows from the continuum limit of Eq. (3) and is given by

$$E = -\frac{1}{2}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx' S[u(x)]W(x-x')S[u(x')] + \int_{-\infty}^{\infty} dx \int_{0}^{S[u(x)]} S^{-1}(Q)dQ - \int_{-\infty}^{\infty} dx h(x)S[u(x)]. \tag{16}$$

The replacements $S_i \to S[u(x)]$, $\theta_i \to \theta$, and $\sigma_i \to \sigma$ are also used in Eq. (4). In the following, we will mainly be concerned with the properties of this Lyapunov function.

The structure of connections in many pattern-forming neural models is assumed to include *lateral inhibition*, meaning the connections are excitatory over short range and inhibitory over a longer range. The synaptic weight is often then chosen to have the Mexican hat form

$$W(R) = \frac{1}{B\sqrt{\pi}}\left[\exp\left(-\frac{R^2}{r_1^2}\right) - A\exp\left(-\frac{R^2}{r_2^2}\right)\right], \tag{17}$$

where $r_1 < r_2$, $A < r_1/r_2$, and $B = r_1 - Ar_2$. This function describes connections with $W(R) > 0$ for small R (short-range excitatory coupling) and $W(R) < 0$ for larger R (long-range inhibitory coupling). We now investigate the stability and interactions of spatially localized regions of high neural activity in a continuum network with $W(R)$ given by Eq. (17).

A. Single soliton

Spatially localized regions of high neural activity, termed *solitons* here, and interactions between solitons can be investigated using a variational energy. This is constructed by choosing a variational form for $u(x)$ in terms of one or more collective coordinates, substituting it into Eq. (16) and performing the integrals over x and x' to obtain an expression that depends only on the collective coordinates. Minimizing the variational energy with respect to the collective coordinates yields stable equilibria.

The variational form we use for a single soliton is given by an exact solution of Eqs. (14) and (15) in the limit as $\sigma \to 0$. This is constructed by integrating $W(x-x')$ over a high-activity region of size m:

$$u(x) = u_0 + \int_{-m/2}^{m/2} W(x-x')dx', \tag{18}$$

leading to a soliton of width m, centered at the origin, as shown in Fig. 8(a). Here, $u = u_0$ is chosen to extremize E at $m = 0$. It was originally shown by Amari [22] that Eq. (18) gives a stable equilibrium of Eqs. (14) and (15) in the $\sigma \to 0$ limit for a specific choice of m. More generally, we find stable equilibria for $\sigma > 0$ by minimizing a variational energy with respect to m.

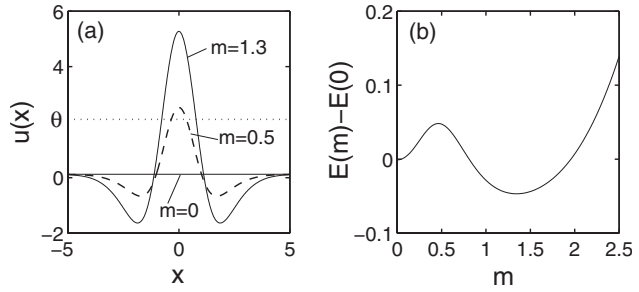

FIG. 8. Energy of configurations given by Eq. (18) for different values of m. In (a), plots of $u(x)$ are shown for the uniform state ($m=0$), unstable soliton ($m\simeq 0.5$), and stable soliton ($m\simeq 1.3$) using $r_1=1$, $r_2=2$, and $A=0.47$. In (b), the variational energy calculated using Eqs. (16)–(18) with $\theta=2.1$, $\sigma=1$, and $I=0$ is shown. $E(0)$ is the energy at $m=0$.

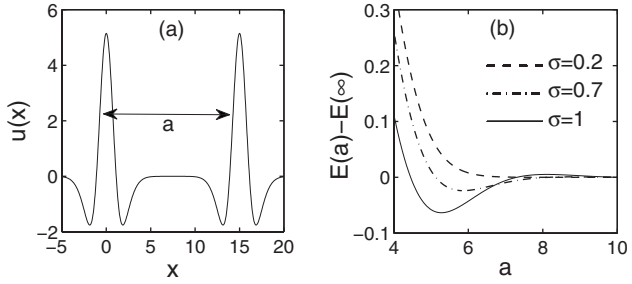

FIG. 9. Energy of configurations given by Eq. (19) for different values of a and σ. In (a), a plot of $u(x)$ for two solitons of width m_0 is shown using parameter values from Fig. 8. In (b), the variational energy calculated using Eqs. (16), (17), and (19) for $\sigma=0.2$ 0.7, and 1 is shown.

The variational energy following from Eqs. (16)–(18) is shown in Fig. 8(b) and has three critical points: two local minima and one local maximum. The minima are stable equilibria at the center of two basins of attraction: one for the uniform state $m=0$ and one for a stable soliton given by Eq. (18) with m equal to $m_0\simeq 1.3$. The maximum located at m equal to $m_c\simeq 0.5$ is an unstable equilibrium and forms the basin boundary. The unstable equilibrium is also a *nucleus of critical size*: a stable soliton will form when $m>m_c$.

A stable soliton given by Eq. (18) represents a balance between nonlinearity and dissipation. From Fig. 8(a), it is seen that $u(x)<\theta$ when $m\ll m_c$. In this case the first term in Eq. (14) dominates and dissipation will cause the system to relax towards the uniform state at $m=0$. However, when $m\simeq m_c$, we find $u(x)\simeq\theta$ over a region of size m and the second term in Eq. (14) becomes significant. Short-range excitatory coupling will stabilize a local region if $m>m_c$. Long-range inhibitory coupling increases the energy when $m>m_0$, so the activity remains localized.

B. Soliton interactions

Translational invariance of the synaptic density means that many widely spaced solitons are also equilibria of Eqs. (14) and (16). However, interactions begin to play a role as the soliton separation decreases. In Ref. [23], it was found that the Mexican hat form of lateral inhibition is not sufficient to stabilize two-soliton solutions of Eq. (14) when a Heaviside step function is used for S. Here, we show the more general result that two-soliton solutions can be stable when S is given by a sigmoid function.

Interactions between two solitons of equal width can be investigated by defining a collective coordinate a as the distance between two solitons when $a\gg m$, and assuming

$$u(x)=u_0+\int_{-m/2}^{m/2}[W(x-x')+W(x-a-x')]dx', \quad (19)$$

as shown in Fig. 9(a) for fixed m and a. The corresponding variational energy is shown in Fig. 9(b) for fixed m and different values of σ. Above a certain critical value of σ, a local energy minimum appears at a equal to $a_0\simeq 5.3$. This leads to a force of attraction between solitons, and bound states consisting of two or more solitons spaced a_0 apart become stable. For small σ the minimum at a_0 vanishes and bound states are no longer stable without the inclusion of additional forces. An increase in the energy at small values of a leads to a repulsive force that keeps solitons separated from one another.

Interactions between solitons depend on the coupling term in Eq. (16), which is given by

$$E_{\text{int}}=-S[u(x)]W(x-x')S[u(x')] \quad (20)$$

and can be understood qualitatively in the following manner. It is seen from Figs. 8 and 9 that an individual soliton consists of a peak of high activity and two regions of low (or minimum) activity. Let the high-activity region be described by $u(x)$ and the low-activity regions by $u(x')$. Interactions between $u(x)$ and $u(x')$ have $W(R)<0$ due to the characteristic range of excitation and inhibition, implying that $E_{\text{int}}\geqslant 0$ for interactions between these regions. Since $S[u(x)]$ is large and positive, one possibility for minimizing E_{int} is by decreasing $S[u(x')]$. This can be achieved by moving two solitons together so their low-activity regions overlap, decreasing $u(x')$ and therefore decreasing $S[u(x')]$. In the step function limit $\sigma\to 0$, $S[u(x')]=0$ cannot be decreased any further, so the energy is not minimized by moving solitons together and no force of attraction is found. When solitons become close enough such that $W(R)<0$ for interactions between neighboring high-activity regions, the solitons begin to repel each other, regardless of σ.

It has previously been proposed that localized regions of high activity may play a role in working memory [1,22]. In this case, an elevated firing rate within a local region of the cortex would be associated with the memory of a particular event. The existence of bound states now demonstrates that interactions between high-activity regions may need to be taken into account if any of these local regions are close enough together.

C. Soliton lattice

Lowering the firing threshold lowers the barrier to soliton formation. This is shown in Fig. 10, where the energy surface

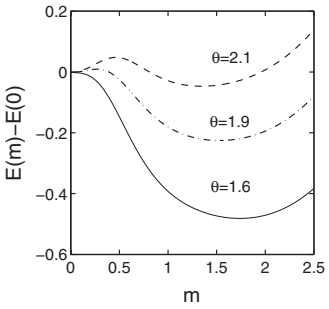

FIG. 10. Energy of configurations given by Eq. (18) for different values of θ. The variational energy calculated using Eqs. (16)–(18) with $\sigma=1$ and $I=0$ is shown for $\theta=2.1$, 1.9, and 1.6.

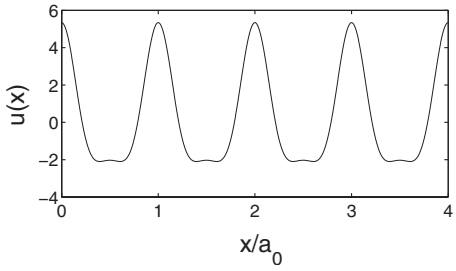

FIG. 11. A soliton lattice with lattice constant a_0.

from Fig. 8 is plotted for different values of θ. As θ is lowered, the local maximum moves towards the minimum at $m=0$, until they merge in a saddle-node bifurcation at $\theta \simeq 1.6$. The equilibrium at $m=0$ then becomes unstable to soliton formation, and solitons are generated spontaneously—the nucleus of critical size mentioned previously now has zero width. More generally, a pattern-forming instability (also known as a Turing instability) takes place at $\theta=\theta_c$ when the critical point at $m=0$ goes from being a minimum in E, to a saddle point in E, with the direction of decreasing energy corresponding to a deviation of nonzero wave number.

1. Pattern-forming instability

The curvature of E at the critical point $m=0$ can be found from the continuum limit of Eq. (5) evaluated at $u=u_0$ [which is also the second variational derivative [21] of Eq. (16)], yielding

$$\left.\frac{\delta^2 E}{\delta u(x)\,\delta u(x')}\right|_{u=u_0} = \rho[\delta(x-x') - \rho W(x-x')], \quad (21)$$

where $\rho=S'(u_0)$. The eigenvalues λ_k and eigenfunctions χ_k describing the curvature of E then satisfy the following eigenvalue equation:

$$\int_{-\infty}^{\infty} dx' \rho[\delta(x-x') - \rho W(x-x')]\chi_k(x') = \lambda_k \chi_k(x). \quad (22)$$

Assuming an eigenfunction of the form $\chi_k \propto \exp(-ikx)$ in Eq. (22) yields the eigenvalue

$$\lambda_k = \rho[1 - \rho \tilde{W}(k)], \quad (23)$$

where $\tilde{W}(k)=\int_{-\infty}^{\infty} W(R)e^{-ikR}dR$ is the Fourier transform of $W(R)$. For the Mexican hat form in Eq. (17), $\tilde{W}(k)$ has a global maximum at $k=k_c \neq 0$. Lowering θ causes ρ in Eq. (23) to increase, until at $\theta=\theta_c$ it is found that $\lambda_k=0$ for $k=k_c$, while $\lambda_k>0$ for all other k.

Turing instabilities lead to spontaneous pattern formation in a diverse range of systems and are responsible for patterns ranging from animal coat markings to those of fluid convection cells [27,28]. Here, we show that one type of stationary pattern resulting from a Turing instability is given by a *soliton lattice*, a pattern also found in several condensed matter systems (e.g., Refs. [30,31]). The energy approach is now used to give the precise form of this soliton lattice.

Consider a multi soliton configuration of the form of Eq. (19). The energy from Eq. (16) can then be written as

$$E = -N_s \Delta E, \quad (24)$$

where $N_s=L/a$ is the total number of solitons, a is the spacing between each soliton, and L is the length of the region under consideration, and where $\Delta E>0$ is the energy of each soliton, given by

$$\Delta E = E_{sol} + E_{bind}, \quad (25)$$

where $E_{sol}=E(m=0)-E(m)$ and $E_{bind}=E(a\to\infty)-E(a)$ have been defined. The first expression gives the energy of a single unbound soliton and is maximized at the equilibrium soliton width $m=m_0$. The second expression is the binding energy of a single soliton at a distance a from each of its neighbors. The key point is that E_{bind} will not necessarily be maximized at a Turing instability; instead, the value of a must be chosen to maximize both N_s and E_{bind} according to Eq. (24). In particular, it is seen in Fig. 9 that choosing a below a certain critical value makes $E_{bind}<0$ due to the short-range repulsion between neighboring solitons. However, the resulting increase in N_s may still lead to a minimum of Eq. (24)—and therefore to a stable bound state, even when $\sigma=0$.

If the global maximum of $\tilde{W}(k)$ is at $k\simeq 2\pi/a_0$, where a_0 is the value of a which minimizes Eq. (24), then the number of solitons will increase rapidly at $\theta=\theta_c$ due to a Turing instability. This will eventually saturate at $N_s=L/a_0$, and the result is a regularly spaced lattice of solitons, as shown in Fig. 11. According to Eq. (19), the soliton lattice is completely specified by the connectivity function $W(R)$, the equilibrium soliton width m_0, and the equilibrium spacing between neighboring solitons a_0. These last two quantities can be found from minimizing Eq. (24).

Spontaneous pattern formation has also recently been investigated in a continuum neuronal network consisting of two populations [29]. In that case two different Turing instabilities were identified, one leading to stationary patterns such as those considered here, and the other to oscillating patterns, but only when the time constant for inhibition takes certain values [29]. Numerical simulations were used in Ref. [29] to follow the nonlinear development of these patterns. The single-population model analyzed here does not contain

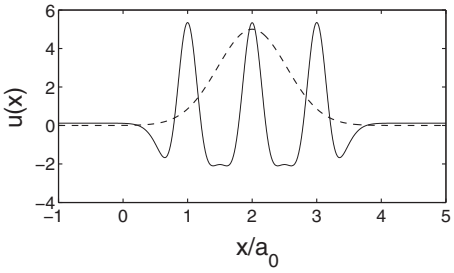

FIG. 12. Solitons (solid line) at a site of maximum stimulus (dashed line).

an oscillating Turing instability, since two or more populations are necessary for this. However, the model given by Eqs. (14) and (15) is adequate for illustrating the usefulness of the energy approach and has allowed us to find a parametrized *nonlinear* solution, yielding insights into the nonlinear properties of patterns.

2. Effect of noise fluctuations

Crystal lattices often contain vacant sites due to thermal fluctuations [32]. These vacancies allow atoms to diffuse from one part of a crystal to another [32]. In our continuum network, fluctuations due to noise in the neural activity play an analogous role to thermal fluctuations in crystals—namely, creating soliton vacancies by removing solitons from the soliton lattice. The energy required to remove a single soliton from the lattice is given by ΔE from Eq. (25). When noise fluctuations are present, this energy can be offset by the number of extra states which are created—that is, ΔE is balanced by $T\Delta S$ due to the entropy ΔS of mixing of solitons and soliton vacancies. Using arguments from point defects in crystals [32], the equilibrium number of lattice vacancies N_v is then

$$N_v = N_s \exp(-\Delta E/k_B T), \tag{26}$$

where $k_B T$ is the strength of noise fluctuations. Similar arguments are used in Ref. [12] to find the "temperature" at which structural singularities in cortical maps begin to annihilate each other.

Soliton lattice vacancies allow solitons to be redistributed on the lattice via processes analogous to "consecutive slipping" in crystals. This can lead to high neural activity becoming concentrated in particular regions of the network, such as at sites of maximum stimulus. According to Eq. (16), the energy due to a nonuniform stimulus $h(x)$ is given by

$$E_{\text{stim}} = -\int_{-\infty}^{\infty} dx h(x) S[u(x)] \tag{27}$$

and is minimized when the overlap between the firing rate $S[u(x)]$ and the stimulus $h(x)$ is maximized. An example involving solitons is shown in Fig. 12.

When noise fluctuations satisfy $k_B T \simeq \Delta E$, the number of lattice vacancies becomes large and long-range order in the neural activity decreases. In this case a soliton liquid or gas may provide a more appropriate description of the neural activity.

V. SUMMARY AND DISCUSSION

In this paper, we have presented an energy approach to understanding the dynamics of activity in neuronal networks. The purpose was to find an intuitive way of understanding the neural dynamics and to uncover new results which would be more difficult to show using standard methods. In order to illustrate this approach we considered a discrete network of two neural populations with reciprocal inhibition and slow adaptation and a spatially continuous network consisting of a single population with lateral inhibition. Important nonlinear phenomena were investigated, including limit cycles, solitons, and pattern formation.

The main results from our approach are as follows: (i) In a two-population network with reciprocal inhibition and slow adaptation, two limit cycles were discovered—one where the activities of both populations oscillate out of phase, as in rivalry, and one where they both oscillate in phase. The energy approach was used to find the dynamical mechanisms for oscillations, as well as the point of bifurcation and basin of attraction of each limit cycle. This extends the work of Ref. [18] for adaptation in a single neural population to adaptation in two interacting populations. (ii) In a continuum network with lateral inhibition, stable equilibria were found for a single soliton and for bound states of two or more solitons when the firing rate was described by a sigmoid function. Bound states become stable when moving two solitons together minimizes the Lyapunov function, a result of decreasing activity in regions between peaks of high activity. This finding extends the results of Ref. [23] for a step-function firing rate, where no stable bound states were found using Mexican hat—type lateral inhibition. (iii) At the pattern-forming instability, a nonlinear solution of the dynamical equations was found to be given by a soliton lattice, which was completely characterized by the equilibrium soliton width and the equilibrium spacing between neighboring solitons. It was shown how to calculate these quantities in terms of an energy minimization. An analogy with the thermal creation of lattice vacancies in crystals was also made, leading to a mechanism for liberating solitons from the lattice—decreasing long-range order and generating spatially inhomogeneous activity within the network.

The advantages of an energy approach for understanding dynamics in neural networks include visualization of the stability and topology of equilibria—and therefore, prediction of global dynamics in terms of trajectories which tend towards minimums of the energy; the use of variational techniques for finding equilibria; and the construction of analogies with other physical systems—including the treatment of fluctuations due to noise using results from statistical mechanics. Limitations include the treatment of asymmetric networks with short timescale dynamics which does not match any known Lyapunov function; visualization of energy surfaces in high dimensions—an appropriate choice of collective coordinates may not always be possible; and evaluation of the energy generally requires a double sum or integral to be performed.

Finally, it is worth noting that many of the results discussed here can be qualitatively reproduced using alternative energy expressions to Eq. (3). One example is the energy given by

$$E = -\sum_{i,j} (u_i - \theta_i) W_{ij} S(u_j) + \frac{1}{2}\sum_i u_i^2 - \sum_i h_i u_i. \quad (28)$$

Each population with $W_{ii} > 0$—i.e., each population with recurrent excitatory connections—can be bistable [18], with a stable equilibrium at high and low activities. The energy given by Eq. (28) is then equivalent to a set of double-well potentials. Such potentials have been thoroughly investigated in physics and are capable of describing a diverse range of phenomena. A simple interpretation of this type has not been found for the more complicated Lyapunov function due to Hopfield.

ACKNOWLEDGMENT

This work was supported by the Australian Research Council.

[1] H. R. Wilson and J. D. Cowan, Kybernetik **13**, 55 (1973).
[2] S. Amari, IEEE Trans. Syst. Man Cybern. **2**, 643 (1972).
[3] C. von der Malsburg, Kybernetik **14**, 85 (1973).
[4] N. V. Swindale, Network Comput. Neural Syst. **7**, 161 (1996).
[5] S. R. Lehky, Perception **17**, 215 (1988).
[6] H. R. Wilson, R. Blake, and S. H. Lee, Nature (London) **412**, 907 (2001).
[7] H. R. Wilson, Proc. Natl. Acad. Sci. U.S.A. **100**, 14499 (2003).
[8] H. R. Wilson, B. Krupa, and F. Wilkinson, Nat. Neurosci. **3**, 170 (2000).
[9] G. B. Ermentrout and J. D. Cowan, Biol. Cybern. **34**, 137 (1979).
[10] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas, and M. C. Wiener, Philos. Trans. R. Soc. London, Ser. B **356**, 299 (2001).
[11] P. C. Bressloff, Phys. Rev. Lett. **89**, 088101 (2002).
[12] M. W. Cho and S. Kim, Phys. Rev. Lett. **92**, 018101 (2004).
[13] B. Ermentrout, Rep. Prog. Phys. **61**, 353 (1998).
[14] J. J. Hopfield, Phys. Today **47**, 40 (1994).
[15] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA 1991).
[16] C. Koch, *Biophysics of Computation* (Oxford University Press, Oxford, 1999).
[17] S. Coombes and M. R. Owen, Phys. Rev. Lett. **94**, 148102 (2005).
[18] P. N. Loxley and P. A. Robinson, Biol. Cybern. **97**, 113 (2007).
[19] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **81**, 3088 (1984).
[20] A. R. Bishop, in *Solitons in Action*, edited by K. Lonngren and A. C. Scott (Academic Press, New York, 1978).
[21] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).
[22] S. Amari, Biol. Cybern. **27**, 77 (1977).
[23] C. R. Laing and W. C. Troy, Physica D **178**, 190 (2003).
[24] H. R. Wilson and J. D. Cowan, Biophys. J. **12**, 1 (1972).
[25] C. R. Laing and C. C. Chow, J. Comput. Neurosci. **12**, 39 (2002).
[26] Z. Li and P. Dayan, Network Comput. Neural Syst. **10**, 59 (1999).
[27] J. D. Murray, *Mathematical Biology* (Springer-Verlag, Berlin, 1989).
[28] M. van Hecke, P. C. Hoenberg, and W. van Saarloos, in *Fundamental Problems in Statistical Mechanics VIII*, edited by H. van Beijeren and M. H. Ernst (North-Holland, Amsterdam, 1994).
[29] J. Wyller, P. Blomquist, and G. T. Einevoll, Physica D **225**, 75 (2007).
[30] Y. R. Lin-Liu and K. Maki, Phys. Rev. B **22**, 5754 (1980).
[31] C. B. Hanna, A. H. MacDonald, and S. M. Girvin, Phys. Rev. B **63**, 125305 (2001).
[32] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1967).