Neural networks as spatio-temporal pattern-forming systems

1998 Rep. Prog. Phys. 61 353

(http://iopscience.iop.org/0034-4885/61/4/002)

Rep. Prog. Phys. **61** (1998) 353–430. Printed in the UK PII: S0034-4885(98)15500-8

Neural networks as spatio-temporal pattern-forming systems

Bard Ermentrout
Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received 13 October 1997, in final form 2 February 1998

Abstract

Models of neural networks are developed from a biological point of view. Small networks are analysed using techniques from dynamical systems. The behaviour of spatially and temporally organized neural fields is then discussed from the point of view of pattern formation. Bifurcation methods, analytic solutions and perturbation methods are applied to these models.

0034-4885/98/040353+78$59.50

Contents

1. Introduction

The advent of multi-electrode recordings and imaging techniques has enabled neuroscientists to obtain high spatial and temporal resolution of the electrical activity for populations of neurons in the brain. This has led to a resurgence of interest in the biological community in large-scale models of neural activity. This in turn has engaged the interest of physicists and mathematicians in the behaviour of these systems. There are many levels in which one can model neural activity from detailed models of single-ion channels and synapses (Destexhe *et al* 1994) up to 'black box' models used to understand psychological phenomena (McClelland and Rumelhart 1988). An 'intermediate' approach to modelling is the use of 'neural networks' where each neuron is represented by a low-dimensional dynamical system. In this review, such models will be considered from the perspective of nonlinear dynamics and general theories of spatio-temporal pattern formation. Most of the familiar work on neural networks treats their behaviour using the methods of statistical mechanics. In these models, the connectivity between elements has no topological or spatial structure. The types of theories and the models that will be explored in this paper are structured models in that the interactions between elements have some topological structure be it a line, circle, or two-dimensional array. Furthermore, I will usually be interested in the steady-state behaviour of these systems where inputs are either static in time or periodically varying. These types of models were popular two decades ago, but there has been a renewed interest in their behaviour now that experimental methods are able to provide the kinds of qualitative and quantitative information required to make these models useful.

Most of the articles and reviews about neural networks have been concerned with their computational properties in a formal sense. There has been little attempt to associate the networks with any specific biological behaviour. In particular, the connections between neural networks and physiology have not been made. In this review, our primary applications are taken from the biological literature. Thus, for us, the main motivation for the study of neural networks is to enable us to explain some particular biological experiment. Some of the physics of neural networks will be very briefly described but the emphasis in this review is on the classical methods of applied mathematics such as bifurcation theory and singular perturbation and how the neural network equations can be used to understand biological systems.

The first part of the review provides a brief introduction to computational neuroscience. A number of related neural network models are then derived. Small networks of one, two, three, and four neurons are analysed next. A brief overview of the physics of neural networks is provided with attention paid to networks with no topological structure. Finally, the bulk of this review describes the analysis of spatially structured neural networks.

2. An introduction to computational neuroscience

The brains of animals contain millions of interconnected neurons, specialized cells that are responsible for both processing incoming information from the environment and producing motor output. Here, we briefly review the biological details required for modelling interconnected neurons. A comprehensive introduction to neuroscience can be found in Kandel and Schwartz (1985) or Shepherd (1990). A single nerve cell consists of three main

parts: the soma or cell body, the axon, and the dendrites. In the typical scenario (although this has been overturned by recent experiments) information comes into the cells through the dendrites where it is 'integrated' at the cell body and this transformed signal is output through the axon. Thus, in the models that we will look at and, in most computational models, the interactions between one cell and another occur via the axon of the first cell to the dendrite of the second cell. The level of detail for each of these three parts of the cell as well as the mechanism of 'communication' of the axon of one to the dendrite of the other depends on the kinds of questions one is interested in and the sizes of the networks that one wishes to simulate. Since our approach will be mainly from the point of view of analysis, one of our aims is to simplify or reduce the single cell as much as possible and yet still retain some of the important properties. Typically, if one wants a detailed model of a single cell, the cell is broken into many smaller compartments each of which consists of a simple RC circuit. Each of these compartments represents a single differential equation. The resistances (called conductances in many cases) can be simple passive resistors or themselves nonlinear dynamic variables. Indeed it is the latter property that gives rise to the nonlinear and threshold behaviour of real and model neurons. The active conductances are a consequence of ionic channels in the cell membrane whose channel states can depend on many aspects of the cellular milieu, notably the transmembrane potential. Calcium and other ionic species can also be responsible for the rates of opening and closing of channels. The current through one of these resistors is usually modelled as ohmic and satisfies:

$$I = \bar{g} m^p h^q (V - V_{\mathrm{eq}})$$

where, $\bar{g}$ is the maximal conductance, m, h are fractions of open channels, p, q are non-negative integers, and V_{eq} is the equilibrium potential of the channel. The gating variables themselves satisfy differential equations of the form:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \alpha(1 - m) - \beta m \tag{2.1}$$

which should be recognized as the master equation for a two-state Markov process (see Destexhe *et al* 1994). The rates, α, β are often functions of the voltage or other quantities. If p, q are both zero, then the channel is said to be passive and is then just like a resistor. In many modelling studies of single neurons, all the active conductances are on the soma and the dendrites are passive.

Biophysically based (that is based on voltage clamp studies of individually recorded nerve cells) computational models that ignore the dendritic architecture have the following form:

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = -\sum_k \bar{g}_k m_k^{p_k} h_k^{q_k} (V - E_k) + I \tag{2.2}$$

where, m_k, h_k are gating variables that satisfy differential equations like (2.1), E_k is the reversal potential of the kth channel, and I is an applied current.

A typical mammalian neuron has a resting potential of -65 mV. There are two basic types of active currents: inward currents which when active tend to depolarize the nerve cell and outward currents which hyperpolarize the cell. The main inward currents are due to sodium and calcium while outward currents are due to potassium and chloride. There are dozens of different currents and the interested reader is urged to consult Johnston and Wu (1995). When the potential of a resting cell is raised sufficiently, inward currents are activated and this leads to a further rise in the potential. This positive feedback results in the production of an action-potential or spike of voltage which is recorded by the experimentalist. The voltage usually comes back down due to two effects: (i) turning

on of outward currents and (ii) inactivation of the inward currents. Thus, application of a constant current causes the neuron to spike, repolarize, and spike again. The steady rate of firing that occurs in the presence of a constant current results in the so-called FI curve which will play a major role in the derivation and behaviour of neural networks. Many cortical cells have the additional property that they initially fire quite rapidly but as the stimulus persists, the firing rate goes down to a lower level. This is called spike adaptation and it will also play an important role in the behaviour of models in networks. Huguenard and McCormick (1996) have written a small manual on voltage gated models which gives explicit formulae for many different channels and gates that are found in thalamic and cortical neurons.

2.1. Firing properties of neurons

Since it will be important later in our discussion, we describe a number of different firing functions and how they are related to the onset of rhythmic activity in biophysical models. As the current, I is increased, most cortical neurons switch from a resting constant potential to an active mode. In the active mode, either trains of spikes are generated or burst of spikes. The firing pattern of bursting cells consists of a brief episode of spikes followed by a period of quiescence, etc. We will concern ourselves with the cells that fire repetitive spike trains, so-called 'regular' and 'fast' spike units which comprise the majority of cells in cortical networks (White 1989). Bursting cells are important and have been the subject of much recent research (Gray 1994). There are basically three ways in which a model can switch from silence to repetitive firing: (a) Hopf bifurcation, (b) saddle–node limit cycle, and (c) homoclinic. In the Hopf case, the oscillations appear with a finite frequency that is bounded away from 0. Typically, the Hopf bifurcation is subcritical which means that in a limited range of currents, there is bistability between a periodic and resting state. This behaviour is known to occur in the squid axon, both by modelling and by experiments (see Guttman *et al* 1980). The appearance of repetitive firing at a non-zero frequency is called class or type II firing and is rare in models of cortex (see Rinzel and Ermentrout 1989). More typical of cortical models is that the firing rate appears at a zero frequency via one of the two mechanisms, (b) or (c). Figure 1 illustrates these bifurcations. Firing models of the type illustrated in (*b*) become active via a saddle–node bifurcation on an invariant circle. The firing rate as a function of the current is

$$F = A\sqrt{I - I_c} \tag{2.3}$$

near the critical value of the current. In the third scenario illustrated in (*c*), the periodic behaviour arises via a homoclinic bifurcation. In this case the firing rate obeys

$$F = -A\frac{1}{\ln(I - I_c)} \tag{2.4}$$

near the critical current. These latter firing patterns are difficult to distinguish experimentally. Guckenheimer *et al* (1997) recently used the behaviour of bursting cells to experimentally distinguish the saddle–node and homoclinic bifurcations. Note that the homoclinic bifurcation model also has a current regime where the dynamics is bistable with a stable fixed point and a stable limit cycle.

2.1.1. Integrate-and-fire models and their relatives. No discussion of spiking models would be complete without a description of the 'integrate-and-fire' model:

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = -g(V - E) + I$$

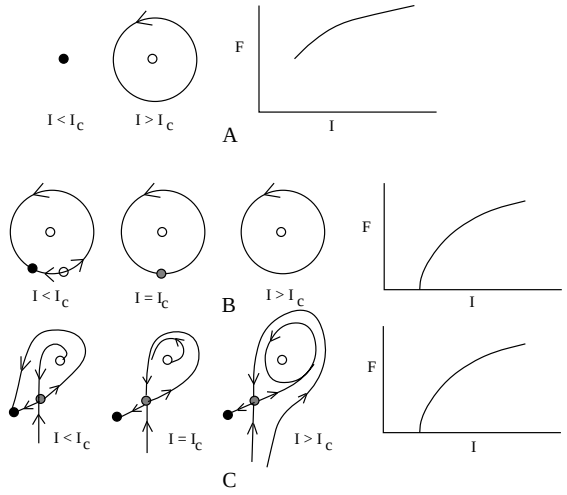

Figure 1. Firing rate functions of neurons and the onset of rhythmicity. (*a*) Subcritical Hopf bifurcation. (*b*) Saddle–node limit cycle bifurcation. (*c*) Regular homoclinic bifurcation. (Full circles are stable, open circles are unstable nodes, and grey circles are saddles.)

which is a simple linear integrator. The idea of this model is that the voltage rises until it reaches a threshold potential, $V_{\rm thr}$ at which point it 'fires a spike' and is reset to some new potential, $V_{\rm reset}$. This model has been the subject of many theoretical papers because of its simplicity. The firing rate is just:

$$F = -\frac{g}{C \ln\left(\frac{E+I/g-V_{\rm thr}}{E+I/g-V_{\rm reset}}\right)}. \tag{2.5}$$

Note that this model is a hybrid of the two cases, (b) and (c) above in that it has the monostability of the saddle–node model but has the logarithmic firing rate of the homoclinic bifurcation.

Recently, Gerstner (1995) introduced a generalization of the integrate-and-fire model that incorporates many more aspects of real neurons, such as refractory periods, adaptation, and synaptic time constants. By integrating the integrate and fire model, we can rewrite it as:

$$V(t) = V_0 + (V_{\rm reset} - V_0)\mathrm{e}^{-t/\tau}$$

where, $V_0 = (I + gE)/g$ and $\tau = C/g$. Gerstner and his collaborators, generalized this concept to incorporate more complex integration and recovery of the neuron into their 'spike response model' which has the following form. Let $v(t, s)$ be the potential of a neuron at time t that last fired a spike at $t - s$, so that s is the time since the last spike. Then:

$$v(t, s) = V_{\rm ext}(t) + V_{\rm syn}(t, s) + \eta^{\rm refr}(s)$$

where the first two terms are the external inputs and the synaptic inputs respectively and $\eta^{\rm refr}(s)$ is the refractoriness of the neuron. For the integrate and fire, this is just

$$\eta^{\rm refr}(s) = E + (V_{\rm reset} - E)\mathrm{e}^{-s/\tau}.$$

The spike-response model generalizes the functional form of the refractoriness.

2.2. Synaptic connections

Now that we have described the firing properties of isolated nerve cells (and only in a very cursory fashion—an entire book could easily be written on the dynamics of a single nerve cell), we want to describe the mechanism by which they are coupled. When the soma of a neuron fires, the potential invades a region called the axon hillock. This causes a train of impulses to travel down the axon which terminates in a series of regions called synaptic terminals. The potential invades these terminals and causes the release of a chemical called a neurotransmitter. This transmitter binds to the dendritic or somatic membranes of other cells and causes channels to open up. This in turn causes current to pass through the dendrites. The current can either be depolarizing (excitatory), hyperpolarizing (inhibitory), or shunting. Most neural network modellers are concerned with the first two types of interactions. Abbott (1992b, 1994) derived a neural network model (which we discuss below) that takes shunting interactions into account. Some of Grossberg's (1990) models also take this into account. Chemical synapses are modelled at the cellular level in the same way as the channels; the current is ohmic:

$$I_{\text{syn}} = \bar{g}s(t)(V - V_{\text{syn}}).$$

The function $s(t)$ is either of a predefined form or itself satisfies a differential equation like the gating variables m, h described above. In the case of predefined forms, when the presynaptic cell fires a spike, then $s(t)$ is reset to a function of the form:

$$s(t) = \frac{\mathrm{e}^{-at} - \mathrm{e}^{-bt}}{1/a - 1/b}.$$

When $a = b$, this simplifies to $a^2 t\mathrm{e}^{-at}$ and is often called an 'alpha' function. When $b \to \infty$ this is just the exponential function. The parameter b characterizes the rise time of the synapse and a the decay time. A synapse is said to be excitatory if the reversal potential, V_{syn}, of the synapse is positive relative to the resting potential of the cell. The primary neurotransmitter responsible for excitatory synapses is called glutamate. The reversal potential of these synapses is usually around 0 mV, so that they provide quite a large inward current for a cell that is resting at -65 mV. An inhibitory synapse is one for which the synaptic reversal potential is negative relative to the resting potential. The main neurotransmitter for inhibitory synapses is γ-aminobutyric acid or GABA. Reversal potentials vary between -75 and -90 mV. Excitatory synapses are faster acting than inhibitory synapses but inhibition often occurs close to the soma and so it can have a larger effect on the neuron.

The response of a passive membrane to a synapse has a characteristic shape and is called the post synaptic potential (PSP) or post synaptic current (PSC). If the decay of the synapse is very fast, then the decay of the PSP is dominated by the passive decay of the cell membrane. On the other hand, if the synapse decays slowly, then the decay of the PSP is dominated by the synaptic time course. This difference will play a role in our derivation of neural networks in the next section.

We have treated dendrites as discrete compartments in this discussion. However, they are more truly continuous membranes and better modelled by partial differential equations (PDEs) of the form:

$$\tau \frac{\partial V}{\partial t} = -(V - V_{\text{rest}}) + \lambda^2 \frac{\partial^2 V}{\partial x^2} + I(x, t). \tag{2.6}$$

The parameter λ is called the space constant of the dendrite and is proportional to the square root of the diameter of the dendrite. The constant τ is called the time constant and

is independent of the geometry of the dendrite. A typical value for τ is 10 ms. In a highly branched neuron, each branch would be modelled by this cable equation and the boundary conditions determined by continuity of voltage and conservation of current. Abbott (1992a) provided a lovely diagrammatic method for solving large sets of such PDEs iteratively. We will return to this dendritic model later when we describe more complex models for neural networks. For a good review of models that use the full dendritic structure see Bressloff and Coombes (1997). Note that the solution to the steady-state cable equation consists of a sum of exponentials, $e^{\pm x/\lambda}$. Thus, steady stimuli are degraded as the distance from the dendrites increases and the smaller λ is, the faster this decay. λ is proportional to the square root of the transmembrane resistance so that the attenuation of current due to a shunt with conductance g is

$$A_{\text{shunt}} = e^{-a\sqrt{g}x} \tag{2.7}$$

where a is a constant depending on the membrane properties and x is the distance from the current source. Equation (2.7) is important, as a shunting synapse decreases the resistance (increases g) and thus dynamically 'lengthens' the dendrite.

3. What constitutes a neural network?

A neural network can mean different things depending on who is describing it. Our perspective is to try to keep some connection to biology. Thus, such constructions as asynchronously updated symmetric neural networks, while of great interest theoretically, will not be considered in much detail in this review. There are many books available that describe this perspective (Anderson and Rosenfeld 1988, Amit 1989, Hertz *et al* 1991, Muller *et al* 1995, Arbib 1995).

Perhaps, we might better say 'neuronal networks' rather than 'neural' networks where the difference lies in the use of the word neuronal instead of neural. Indeed, neural networks have come to mean something quite formal that is related to algorithms and computation and not to questions about how brains might work. Our emphasis will be on dynamics and spatio-temporal behaviour. This type of organized activity is not often addressed in the neural network literature. However recent experimental methods now allow brain scientists to examine spacetime activity in real neural systems and this has led to some new modelling efforts and the need for more structured types of networks.

Below we derive sets of discrete or continuous spacetime models and will examine these from the point of view of stability, bifurcation, and pattern formation. In some ways, this review will have the flavour of the article by Cross and Hohenberg (1993) but will be narrowly focused on the particular class of models that arise in neural modelling. The types of questions addressed are biological and mathematical.

We now turn to a derivation of general models for neural networks. There are many approaches that can be used and there are almost as many variations on the derivation as there are models. We will look at several derivations and examine the types of equations that arise from these derivations.

The most straightforward derivation occurs when one considers the basic idea illustrated in figure 2. We will cast this in a very general form. The general idea of this derivation goes back to Freeman (1972) although his models were usually linear. A similar derivation (although less general) is given in Ermentrout and Cowan (1980a). Let i and j be the indices of two neurons which are connected in the sense that the axon of neuron j terminates on the dendrite of neuron i. Let $V_i(t)$ be the potential at the ith neuron at time t. Suppose that such a potential is converted by the axon hillock to a firing rate $u_i(t) = S(V_i(t))$. This firing

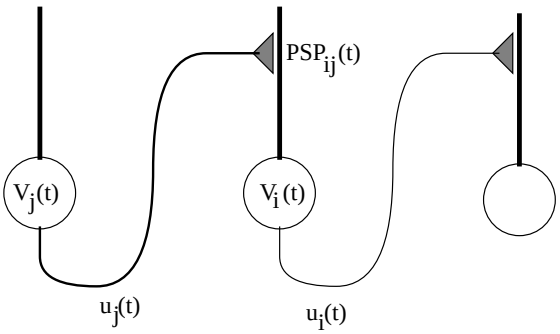

Figure 2. Synaptically coupled neurons.

rate is often what is measured by biologists. There can be some delay due to the distance travelled down the axon before the firing releases a neuro-transmitter. This delay could depend on distance as well as the type of neurons. Call it δ_{ij}. A single-action potential on a dendrite is seen as a post-synaptic potential, $PSP_{ij}(t-s)$ where s is the time of the spike hitting the terminal and t is the time after the spike. Suppose that the spikes come at times $t_1, t_2, \ldots$. Then the total potential is given by

$$\Phi_{ij}(t) = G_{ij}(t_1, t_2, \ldots)$$

where G_{ij} depends on many aspects of the membrane. For example, in more complicated neural models, the effect of successive firings can actually depend on how frequently they come. The result of two closely timed pulses can be more or less than the sum of the two impulses depending on things like synaptic facilitation, fatigue, or random misfires. These aspects are usually ignored in neural network literature; we will also ignore them for now. However, we want to point out that it is at this point in the modelling that one would incorporate such processes. Often spike adaptation or facilitation is modelled as an amplitude factor.

Returning to our derivation, we assume that the spikes sum linearly so that the network potential is

$$\Phi_{ij}(t) = \sum_k PSP_{ij}(t - t_k).$$

To close the equations, recall that $u_j(t)$ is the instantaneous firing rate. The number of impulses arriving between t and $t + \mathrm{d}t$ is $u_j(t)\,\mathrm{d}t$. Thus, the total potential contributed by the cell firing at rate $u_j(t)$ to the cell i is

$$\Phi_{ij}(t) = \int_{t_0}^{t} PSP_{ij}(t-s)u_j(s-\delta_{ij})\,\mathrm{d}s.$$

This closes the system and we are left with a set of nonlinear Volterra integral equations:

$$V_j(t) = \sum_j \int_{t_0}^{t} PSP_{ij}(t-s)S_j(V_j(s-\delta_{ij}))\,\mathrm{d}s. \tag{3.1}$$

The PSP_{ij} itself may depend on other variables in order to account for such phenomena as adaptation, learning, facilitation and fatigue. The model as described above is not well defined as we see at $t = t_0$ we have set the potentials all to be 0. Since one often takes $t_0 = -\infty$ this is not a big problem. At this point our neural model bears little resemblance

to the types of models that one usually associates with neural network literature. We will now manipulate it into the form that is most well known.

First, we note that changing the 'starting point' of the model we can also write:

$$u_i(t) = S_i\left(\sum_j \int_{t_0}^{t} PSP_{ij}(t-s)u_j(s-\delta_{ij})\,\mathrm{d}s\right) \tag{3.2}$$

which is called the firing rate formulation.

3.0.1. The firing rate function. The nonlinear form of the firing rate function is very important; there are many choices that various authors have used. The simplest is the step function, where the neuron fires maximally or not at all depending on whether the potential is above or below threshold. If one uses a statistical-mechanical approach to derive 'mean-field' equations (Cowan 1968, Amari 1972) then this sharp step function is smoothed out to the well known squashing functions:

$$S(V) = \frac{S_{\max}}{1+\mathrm{e}^{-(V-V_T)/V_s}} \qquad \text{logistic}$$

$$S(V) = \frac{S_{\max}}{2}(1+\mathrm{erf}(V-V_T)/V_s) \qquad \text{Gaussian.}$$

The logistic is usually easier for use in computational models. Both of these may be regarded, not as firing rates, but as probabilities of a cell firing in which case, $S_{\max} = 1$. For $V_s \to 0$ both go to the step function, often called the 'zero-temperature' model.

Another firing rate function that is convenient is the piecewise linear model,

$$S(V) = \begin{cases} 0 & \text{for } V < V_T \\ a(V-V_T) & \text{for } V_T \leqslant V \leqslant S_{\max}/a + V_T \\ S_{\max} & \text{for } S_{\max}/a + V_T < V. \end{cases}$$

If $S_{\max} = \infty$ this is the linear rectifier model and if $a = \infty$, this is the step function model. Equations (2.3) and (2.4) provide another set of firing rate models that are quantitatively based on biophysical properties of neurons.

3.1. Simplification

There are two simplifying assumptions that lead to two different and not quite equivalent models.

3.1.1. The voltage-based model. This is the model that Hopfield (1984) used and is well known to physicists. Suppose that the post-synaptic potential on neuron i always has the same shape no matter which presynaptic cell j caused it. The sign and the amplitude can be different, but the shape of the potential is the same. That is

$$PSP_{ij}(t) = w_{ij}PSP_i(t).$$

As we mentioned in section 2.2, this is the case if the time course of the post-synaptic potential is dominated by the membrane properties of the post-synaptic cell and not the kinetics of the synapse. Next, suppose that the PSP is a sum of exponentials (either real or complex) and powers, $PSP(t) = E(t)$. Then it is well known that the inverse of the linear integral operator:

$$Ku \equiv \int_{t_0}^{t} E(t-s)u(s)\,\mathrm{d}s$$

is just a linear differential operator L with constant coefficients. Thus, (3.1) becomes

$$L_i V_i = \sum_j w_{ij} S_j(V_j(t - \delta_{ij}))$$

which is just a delay-differential equation. Finally, if L_i is first order and there are no delays, we obtain

$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} + V_i = \sum_j w_{ij} S_j(V_j)$$

which is the Hopfield (1984) model. Recapitulating, to obtain the Hopfield's model, the shape of the PSP depends only on the *postsynaptic cell*. Since the decay is governed by the membrane properties of the post-synaptic cell, τ_i is legitimately called the membrane time constant of the model.

3.1.2. The activity-based model. Suppose that the shape of a PSP depends only on the nature of the presynaptic cell except for possibly the amplitude. (Although here, the sign will usually be the same from cell to cell since the sign of a synaptically induced voltage change is determined by the transmitter released from the presynaptic cell.) That is

$$PSP_{ij}(t) = w_{ij} PSP_j(t).$$

As above suppose that the PSP is also a function of exponentials and powers. We define

$$U_j(t) = \int_{t_0}^{t} PSP_j(t - s) u_j(s)\, \mathrm{d}s$$

to be the time-averaged firing rate. Then once again, we see that

$$L_j U_j(t) = u_j(t)$$

where L_j is a linear differential operator. Substituting this into (3.2) we see that

$$L_i U_i(t) = S_i\left(\sum_j w_{ij} U_j(t - \delta_{ij}) \right).$$

which is now a delay-differential equation. Finally if L_i is a first-order differential equation and there is no delay we find that

$$\tau_i \frac{\mathrm{d}U_i}{\mathrm{d}t} + U_i = S_i\left(\sum_j w_{ij} U_j \right)$$

which is another standard neural network model. Since the τ_i depends on the *presynaptic* neuron, this time constant is related to the synaptic decay and not the membrane time constant. Furthermore, it is incorrect to call U_i the 'firing rate' of the cell. The true firing rate is in fact the right-hand side. Pinto *et al* (1996) referred to U_i as the 'synaptic drive'.

3.1.3. Which model to use. It is useful to think about which of these models is the more reasonable. If the synapses are very short lasting, then the dominant time constant is the membrane time constant and the assumptions underlying the voltage-based model are more reasonable. On the other hand, if the membrane time constant is small and the synaptic time course is longer, the activity or firing rate model is the reasonable choice. These may seem like minor points, but if one wants to say something about biology, then these considerations are necessary (see for example, Pinto *et al* 1996).

More usually, can we use the same inversion tricks if the PSPs are sums of exponentials and powers but do not have time courses that are dependent *only* on the pre- or post-synaptic cell? The answer is yes, although the solution is hardly satisfying. Let

$$Q_{ij}(t) = \int_{t_0}^{t} PSP_{ij}(t-s)u_j(s)\,\mathrm{d}s.$$

Then $L_{ij}Q_{ij}(t) = u_j(t)$ so (3.2) becomes

$$L_{jk}Q_{jk}(t) = S_j\left(\sum_i Q_{ji}(t-\delta_{ji})\right) \qquad j,k = 1,\ldots,N.$$

Note that the right-hand sides of these equations are the same for all k for any particular j. However, if the linear differential operators are all different, then this is not a redundantly determined system. As far as the author knows, differences between this type of model and the simple firing rate model have not been explored. The simplest situation would involve two mutually coupled neurons with independent exponential time courses resulting in a four-dimensional dynamical system. For example, a pair of mutually coupled excitatory/inhibitory neurons would have the form:

$$\begin{aligned}
\tau_{ee}Q'_{ee} + Q_{ee} &= S_e(w_{ee}Q_{ee} - w_{ie}Q_{ie})\\
\tau_{ei}Q'_{ei} + Q_{ei} &= S_e(w_{ee}Q_{ee} - w_{ie}Q_{ie})\\
\tau_{ie}Q'_{ie} + Q_{ie} &= S_i(w_{ei}Q_{ei} - w_{ii}Q_{ii})\\
\tau_{ii}Q'_{ii} + Q_{ii} &= S_i(w_{ei}Q_{ei} - w_{ii}Q_{ii}).
\end{aligned}$$

Under which conditions is there a difference in this model and the simpler case where there are only two time constants remains an open question. For example, let $S_j(u) = 1/(1+\mathrm{e}^{-(u-\theta_j)})$ and set $w_{ee} = w_{ei} = w_{ii} = w_{ie} = 8$, $\theta_e = -1$, $\theta_i = -1.5$. The excitatory time constants are both 1 and the inhibitory time constants are one of two values, 1.2 or 0.1. If $\tau_{ie} = \tau_{ii}$ then the fixed point is stable but if $\tau_{ii} < \tau_{ie}$ then the fixed point is unstable and there is an oscillatory solution. This makes intuitive sense since the inhibitory cell inhibits itself much faster than the excitatory cell which essentially disinhibits the network. This example makes it clear that the behaviour of this more complicated system is different from that of the activity- or voltage-based analogues.

3.2. Generalizations and other 'neural network' models

Abbott (1992a, 1994) generalized the ideas of the previous section to derive a model for a neural network in which there are two types of inhibition: shunting and hyperpolarizing. The hyperpolarizing inhibition is slower than the shunting inhibition but is due to the same population of cells. Both the excitation and the hyperpolarizing inhibition occur at the terminal end of the dendrite and the shunting inhibition occurs on the body of the dendrite. Since this shunting inhibition is assumed to uniformly cover the dendrite, its effect on any inputs at the terminal end of the dendrite is to attenuate the network input by an amount:

$$A_{\text{shunt}} = \mathrm{e}^{-\beta\sqrt{G}}$$

where β is a constant and G is the conductance of the shunting synapse, proportional to the activity of the inhibitory cell (see equation (2.7)). For example, in a small network consisting of an inhibitory and excitatory cell with 'exponential' synapses, the model equations are:

$$\tau_e\frac{\mathrm{d}U_e}{\mathrm{d}t} = -U_e + F_e(\mathrm{e}^{-w_{ge}\sqrt{U_g}}(w_{ee}U_e - w_{ie}U_i))$$

$$\tau_i \frac{\mathrm{d}U_i}{\mathrm{d}t} = -U_i + F_i(\mathrm{e}^{-w_{gi}\sqrt{U_g}}(w_{ei}U_e - w_{ii}U_i))$$
$$\tau_g \frac{\mathrm{d}U_g}{\mathrm{d}t} = -U_g + F_i(\mathrm{e}^{-w_{gi}\sqrt{U_g}}(w_{ei}U_e - w_{ii}U_i)).$$

The analysis of this simple network has not yet been done in any comprehensive fashion.

Milton *et al* (1993, 1995) described a Markov model for a neural network which arises from an attempt to produce a dynamic model from a cellular automaton (CA) formulation. In the CA, a cell has a rest state, 0, an excited state, E and k refractory states. When $k = 1$, this is the Greenberg–Hastings model and was first used as a model for neurons by Beurle (1956). This model has been popular as the simplest example of an excitable medium. For simplicity, suppose that $k = 1$ so the cells have only one refractory state. Let u denote the fraction of cells that are at rest. Let f denote the fraction that are firing and let r be the fraction that are refractory. The total population of cells in each of these three states is 1 so that $u + f + r = 1$. The rate at which cells leave the resting state and enter the firing state is some function of all cells that are firing and the number at rest. The number of cells that re-enter the resting state is proportional to the rate at which they leave the refractory state. Thus,

$$u' = -K_{uf}u + K_{ru}(1 - u - f)$$

and

$$f' = K_{uf}u - K_{fr}f.$$

In general, only K_{uf} will be dependent on other cells and coupling between them. The generalization to many populations and multiple degrees of refractoriness is obvious. As with the previous model, little analysis has been done on this particular class of models.

The original Wilson and Cowan (1972) model incorporated refractoriness by premultiplying the firing rates by

$$1 - \int_{t-r}^{t} u(s)\,\mathrm{d}s$$

to compensate for all cells that had fired within a period, r, of t. Wilson and Cowan approximated this integral by:

$$1 - ru.$$

Thus, their firing rate model has the form:

$$\tau_i \frac{\mathrm{d}U_i}{\mathrm{d}t} = -U_i + (1 - r_i U_i)F_i \tag{3.3}$$

where F_i is the usual firing rate term. In their small network model, there were two populations of neurons, excitatory and inhibitory.

Grossberg (1990) considered models of the form:

$$C\frac{\mathrm{d}V_i}{\mathrm{d}t} = -g_L(V_i - E_L) - (V_i - E_{ex})\sum_j A_{ij}(V_j) - (V_i - E_{\mathrm{in}})\sum_j C_{ij}(V_j) + I_i \tag{3.4}$$

which is like a passive membrane coupled with instantaneous nonlinear synapses.

There have been numerous other derivations of neural models which include a variety of anatomical and physiological aspects. Mallot and Giannakopoulos (1996) considered continuum models which have very general spatial and temporal properties including axonal spread and delays that depend on the axonal distance. Wright *et al* (1987) and Wright and Liley (1995) considered continuum models where there are distributions of delays, axonal

diameters, and varying density of cells. Koch and Leisman (1996) also developed a general theory for neuronal networks with an emphasis on how delays can destabilize certain spatial modes.

3.2.1. Averaging models. We mention several methods that have been successfully employed to obtain simplified neural models from biophysical models such as (2.2). These techniques have an advantage in that every step of the way, we can exactly quantify the parameters. The idea is that there is a small parameter that allows us, through changes of variables, to apply the method of averaging and thus obtain a simplified model. We only describe one method in detail and then discuss other methods that lead to a variety of different greatly simplified models. In all cases, we start with biophysically based models. We want to emphasize the point that in networks with many neurons and in systems where the neurons have complex firing patterns, such as bursting, then these reduction methods are not likely to be a good representation of the actual firing patterns of individual cells. The averaging methods, however, do suggest certain reasonable forms for firing rate models that go beyond an *ad hoc* choice for a nonlinear 'squashing function'.

Suppose that the time course of the synapses is 'slow' compared with the firing patterns of the neuron. That is,

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \epsilon[\alpha(V)(1-s) - \beta s] \tag{3.5}$$

where $\alpha(V)$ is a sigmoidal nonlinearity and ϵ is a small positive parameter. We assume that s is the gating variable for the synapse so that if s is held constant, the voltage dynamics go to a steady state; either repetitive firing or a fixed point. For simplicity in the derivation, we assume only a single self-coupled cell with this slow synapse:

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = -I_{\mathrm{ion}} - gs(V - V_{\mathrm{rev}})$$

where I_{ion} contains all the gating variables for the active channels and V_{rev} is the reversal potential of the synapse. We now treat $G \equiv gs$ as a bifurcation parameter. For the moment assume that the synapse is 'excitatory' so that increasing G results in the system switching from rest to periodic behaviour via a saddle–node limit cycle bifurcation at $G = G_c$ (see figure 1(*b*)). We have:

$$V(t) = v(t; G)$$

where v is a function that is constant in t for $G < G_c$ and periodic with period $T(G - C_c)$ for $G > G_c$. Since ϵ is small, we can substitute this value of voltage into equation (3.5) to obtain:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \epsilon[\alpha(v(t; gs))(1-s) - \beta s]. \tag{3.6}$$

For $G < G_c$ v is a constant. For $G > G_c$ the system oscillates and we can average over the period T. Thus, we obtain the closed scalar system:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \epsilon[S(gs)(1-s) - \beta s] \tag{3.7}$$

where

$$S(G) = \frac{1}{T}\int_0^T \alpha(v(t; G))\,\mathrm{d}t.$$

(In the case of a fixed point take $T = \infty$.) Now if we assume that when the cell is not firing, the potential is below the threshold for the synapse, $S(G) = 0$ for $G < G_c$. If the spikes are very narrow, then the average $S(G)$ is approximately

$$S(G) = \frac{\kappa}{T(G)} = \kappa F(G)$$

where $F(G)$ is the firing rate as a function of the synaptic conductance. Thus, we recover essentially a firing rate model. For simple additive synaptic currents, the general model has the form:

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = \epsilon\left(S_i\left[\sum_i g_{ij}s_j + I_i\right](1 - s_i) - \beta_i s_i\right)$$

which is quite similar to the original Wilson–Cowan equation (3.3), with $r = 1$. The main point of this derivation is that the function S can be numerically evaluated and then fitted with known functions to yield a quantitatively correct model of the full membrane dynamics (Chen *et al* 1997a, Ermentrout 1994). Adaptation and other slow processes are easily included within this formulation.

In the same manner as was done for general ionic models, a reduction of integrate-and-fire models can also be made. Here, I will take the model of Somers *et al* (1995) and reduce it to a simple neural network. Somers *et al* considered a spatially interconnected network of excitatory and inhibitory cells. We ignore the spatial aspect. The model cells have the form:

$$C\frac{\mathrm{d}V_j}{\mathrm{d}t} = -g_{jl}(V_j - V_L) - g_{ja}z_j(V - V_a) - g_{ej}s_e(V_j - V_{ex}) - g_{ij}s_i(V_j - V_{\mathrm{in}}) + I_j$$

where z_j, s_e, s_i are time-dependent gating variables, and V_j is the potential of the cell. The index, $j = \{e, i\}$. The gating variables, z_j represent adaptation; in the simplest case, there is no adaptation for the inhibitory cells. If V_j crosses V_{thr} the threshold, then, the synaptic gating variable, s_j is 'turned on' and decays with some characteristic time course. If we fix s_e, s_i, z_j to be constant, then the voltage equation is readily solved for any initial condition. From this, we can determine the firing time of the next spike. This gives us essentially a firing rate model. This can then be substituted into the dynamics for the gating variables to obtain a closed system involving no actual potentials. If we assume that the potential is reset to V_{reset} after a spike, and synapses are simple exponentials, then the formally reduced version of Somers *et al* is just:

$$\begin{aligned}
&\tau_e s_e' = -s_e + G_e f_e/C\\
&\tau_i s_i' = -s_i + G_i f_i/C\\
&\tau_a s_a' = -s_a + G_e f_e/C\\
&G_e = g_{el} + g_{ee}s_e + g_{ie}s_i + g_a s_a\\
&G_i = g_{il} + g_{ei}s_e + g_{ii}s_i\\
&v_e = (I_e + g_{ee}s_e V_{ex} + g_{ie}s_i V_{\mathrm{in}} + g_{el}V_l + g_a s_a V_a)/G_e\\
&v_i = (I_i + g_{ei}s_e V_{ex} + g_{ii}s_i V_{\mathrm{in}} + g_{il}V_l)/G_e\\
&f(v) = -1/\ln[(v - V_{\mathrm{thr}})/(v - V_{\mathrm{reset}})]\\
&f_e = \min(f(v_e), f_{e,\max})\\
&f_i = \min(f(v_i), f_{i,\max})
\end{aligned}$$

where we have clipped the firing rates to be less than some prescribed maximum (since there is no refractory period) and define the firing rate function, $f(v)$ to be 0 if $v < V_{\mathrm{thr}}$.

While this looks complicated a plot of the nullclines of the (s_e, s_i)-system reveals a nearly piecewise linear system.

3.2.2. Weakly connected neural networks. Hoppensteadt and Izhikevich (1997) developed a general theory of weakly connected neural models. The idea is to suppose that for each neuron, the dynamics are near a bifurcation and then allow these neurons to interact weakly with other similar neurons. The result is that each high-dimensional 'neuron' is represented by a much lower-dimensional system that is the normal form for this bifurcation, and coupling is linear. For example, Hoppensteadt and Izhikevich showed that near a Hopf bifurcation, each neuron is represented by a complex scalar, z_j and the equations have the form:

$$\frac{\mathrm{d}z_j}{\mathrm{d}t} = z_j(a_j - b_j z_j \bar{z_j}) + \sum_k w_{jk} z_k$$

where all constants are complex numbers. The analysis of models of this form remains an open question. Even in the case of just two cells, the behaviour can be remarkably complex (Aronson *et al* 1990).

Another reduction presumes that each cell is firing periodically at roughly the same rate. Each neuron can be represented as a phase, θ_i and the reduction (Kuramoto 1984) is:

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = \omega_i + \sum_j H_{ij}(\theta_j - \theta_j).$$

These models have been the object of a great deal of recent research along two lines: (i) general behaviour and (ii) the form of the interaction functions H_{ij}. In section 4.5.1 we discuss these models in detail.

A final example which we will return to in section 7.2.3 is the behaviour of a network with weak coupling near a saddle–node bifurcation on a circle. That is, there is an invariant circle with two fixed points and at the bifurcation, these merge and then disappear leaving a limit cycle oscillation (cf figure 1(*b*)). Hoppensteadt and Izhikevich showed that in this case, each neuron can be represented by a single-phase variable, θ_i which satisfies the differential equation:

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = 1 - \cos\theta_i + r_i(1 + \cos\theta_i) + \sum_j g_{ij}(\theta_i)\delta(\theta_j - \pi) \tag{3.8}$$

$$g_{ij}(\theta) = 2\arctan\left(\tan\frac{\theta}{2} + w_{ij}\right) - \theta \tag{3.9}$$

where w_{ij} is proportional to the coupling strength, $\delta(\theta)$ is the Dirac delta or unit impulse function, and r_i is a bias term which is positive if the system oscillates and negative if the system is excitable. (Note that the formula in proposition 8.12 in their book contains a misprint and (3.8) is the correct form.)

These three examples of weakly connected networks arise in three different regimes. The first when there are *small amplitude* limit cycles weakly connected, the second when there are large amplitude limit cycles weakly connected, and the last when there are excitable systems near threshold weakly connected. The latter two are more relevant to neurobiology as most neural oscillations occur far from the small amplitude bifurcation.

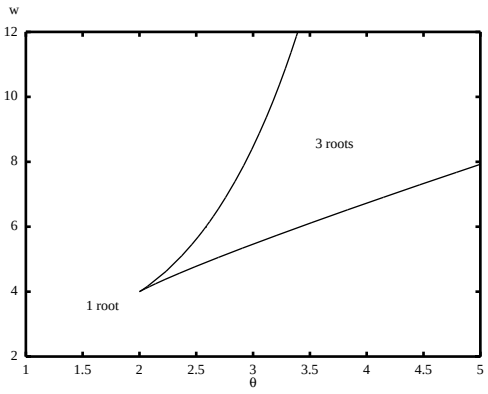

Figure 3. Two-parameter bifurcation diagram of the scalar neural network. Inside the cusp, there are three fixed points and outside there is one. The curves outlining the cusp are saddle–node or fold points.

4. Small networks

We now address the behaviour of small neural networks. We are interested in the dynamics of these systems. We will not attempt to completely classify their behaviour, although this is probably possible for small systems with first-order or exponential synapses. Beer (1995) made the effort to classify two-dimensional models with a variety of connections. Here we will build on this work adding a few cases that he missed. We will work in the 'activity' coordinate system, so that the models that we are interested in have the form:

$$\tau_j u_j' = -u_j + F_j\left(\sum_k w_{jk} u_k\right).$$

4.1. Scalar first-order networks

The simplest neural network is the first-order scalar one:

$$u' = -u + f(wu - \theta) \tag{4.1}$$

where f is the usual squashing function, for example, $f(u) = 1/(1 + \mathrm{e}^{-u})$. There are two parameters in this simple network, the synaptic weight, w and the threshold, θ. Inputs and time constants can all be scaled out of this model. Since the model is a scalar one, all solutions tend to fixed points and these lie between 0 and 1. Thus, all that remains to determine is the number of fixed points and their stability. It is easy to plot a two-parameter bifurcation diagram which we summarize in figure 3. Along the cusp are fold points where an unstable and stable fixed point coalesce. Inside the cusp are two stable fixed points and one unstable fixed point. Outside the cusp, there is a single stable fixed point. While this is a very simple model, it is canonical and it is the ability of the neural network through recurrent excitation to produce bistability that allows these networks to generate such a vast array of spatio-temporal patterns.

4.2. More general scalar networks

The simplest scalar network uses a first-order temporal operator. Can we expect anything new to happen if we use a higher-order PSP? The steady states remain the same in both

models. Thus, one possible difference could be the appearance of limit cycles. There are two ways to increase the order of the synapse. One is to use a sum of more than one exponential for the PSP function. Since a synaptic time course is well fitted by a pair of exponentials and the membrane properties of a passive dendrite are fit by a single exponential, the maximum number of exponentials that we will consider is three. A second way to increase the order is to incorporate fixed time delays (an der Heiden 1979).

4.2.1. Multiple exponentials. Suppose the PSP is the sum of two exponentials, $e^{-\beta_1 t}$, $e^{-\beta_2 t}$. Then the simple nonlinear feedback model becomes:

$$u_t = v$$
$$v_t = -(\beta_1 + \beta_2)v + \beta_1\beta_2(-u + f(wu - \theta)).$$

Bendixson's negative criterion (Edelstein-Keshet 1988) implies that there are no oscillations for this model. After transients, there is little difference between the first- and second-order PSP models.

The same cannot be said of third-order models. Suppose that the PSP is the sum of three real exponentials, $e^{-\beta_1 t}$, $e^{-\beta_2 t}$, $e^{-\beta_3 t}$. The equations are:

$$u_t = v$$
$$v_t = z$$
$$z_t = -(\beta_1 + \beta_2 + \beta_3)z + (\beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3)v\beta_1\beta_2\beta_3(-u + f(wu - \theta)).$$

Oscillations can occur in this network if the weight is negative! It is simple to show that if we let $wf'(w\bar{u} - \theta) \equiv -\kappa$ be a parameter, then as κ increases past

$$\kappa_c = -1 + \frac{(\beta_1 + \beta_2 + \beta_3)(\beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3)}{\beta_1\beta_2\beta_3} \tag{4.2}$$

there is a Hopf bifurcation and oscillations can occur. (This is a consequence of the Routh–Hurwitz criterion. See Edelstein-Keshet (1988).)

4.2.2. Delays. Consider the simplest delayed-feedback neural model:

$$u_t(t) = -u(t) + f(wu(t - \delta) - \theta).$$

This is the essence of a model that has been used by Milton in his analysis of the pupillary reflex (Mackey and Milton 1988). The parameter, δ is the delay. Following the prescription of MacDonald (1989) it is easy to determine whether or not the delay can destabilize a fixed point. Letting $\kappa = wf'(w\bar{u} - \theta)$ where $\bar{u}$ is the fixed point, the fixed point is stable if and only if all roots, s to

$$s + 1 = \kappa e^{-s\delta}$$

have negative real parts. It is easy to show that if $\kappa > 0$, no delay-induced stability can occur, but if $\kappa < 0$, then for delays that are large enough, oscillations occur. Similar delayed negative feedback systems are analysed in Glass and Mackey (1988) and Murray (1989). Large networks with delays have been considered by Marcus and Westervelt (1989).

4.3. Two cells

Now we consider two coupled cells:

$$\tau_1 u_1' = -u_1 + f(w_{11}u_1 + w_{12}u_2 - \theta_1)$$
$$\tau_2 u_2' = -u_2 + f(w_{21}u_1 + w_{22}u_2 - \theta_2)$$

where f is a monotonic function like the squashing function. The behaviour of this model system can be analysed by using phase-plane methods. Since this is a planar system, the only attractors are limit cycles and fixed points. Limit cycles are impossible in the one-dimensional network with single and double exponential synapses so that they represent a new class of behaviour in this simple network. Many theorems on neural networks apply when the weights are symmetric and these invariably show that all initial conditions go to fixed points (see section 5.4). We can eliminate most of the 'uninteresting' cases by using a result from Ermentrout (1995): if $w_{12}w_{21} \geqslant 0$ then there are no limit cycles. This is a generalization of the symmetry condition on the weights; certainly if $w_{12} = w_{21}$ then the non-negativity condition holds. Furthermore, we can apply the Bendixson's criterion to conclude that if both w_{11} and w_{22} are negative, then there can also be no limit cycles. In the cases: (i) $w_{12}w_{21} \geqslant 0$ *or* (ii) $w_{11}, w_{22} < 0$; the only possible long-term behaviour is the approach to fixed points. The use of nullclines serves to completely classify the number of the fixed points. Recall that nullclines are the curves defined in the plane where $u_j' = 0$. Thus, their intersections yield fixed points. For the logistic squashing function, f, the nullclines can be monotone or 'cubic' like. This implies that there are at most nine fixed points or as few as one fixed point. If there is a unique fixed point, it is globally asymptotically stable. Generically there will be an odd number of fixed points. Cases where there are an even number of fixed points, say, for example four, will perturb to three or five fixed points with a small change in some parameter. Since there are always an odd number of fixed points, the number of stable and unstable fixed points is not equal. There will usually be stable nodes, saddle points, and unstable nodes. The saddle points have one-dimensional stable manifolds which play the role of separating the domains of stability of the stable fixed points. Figure 4 shows several examples of this type of network. There are several observations which are useful. The nullclines have outer and inner branches if they are cubic in shape. If the nullcline is not kinked then there is one branch and it is defined to be an outer one. If the nullclines intersect so that the resulting fixed point is on an outer branch from each nullcline, then that fixed point is a stable node. If the fixed point is on an inner branch of one nullcline and the outer branch of the other, then it is a saddle point. Finally, if the fixed point lies on inner branches of both nullclines, it is an unstable node. Thus, one needs only to draw the nullclines and look at the intersections to determine both the fixed point and stability. A similar situation occurs in higher dimensions (Beer 1995) although it is not as simple to visualize.

In Beer's analysis, he found 11 different phase planes in his two-neuron network. We will show that there are several others that he missed in his analysis that involve multiple limit cycles. We do not claim that this appended list is exhaustive, although it is tempting to conjecture that there can be no more than two stable limit cycles in these systems. He found the following systems ordered by the number of fixed points (1) 1 stable FP, (1lc) 1 unstable FP and 1 stable LC, (3a) 2 stable FP, 1 saddle; (3b) 1 stable FP, 1 saddle, 1 unstable FP; (3lc) same as (3b) with 1 stable LC; (5a) 3 stable FP and 2 saddles; (5b) 2 stable FP, 2 saddles, 1 unstable FP; (5lc) same as (5b) with a stable limit cycle; (7) 3 stable FP, 1 unstable FP, 3 saddles; (9) 4 stable fixed points, 4 saddles, 1 unstable FP. (Here

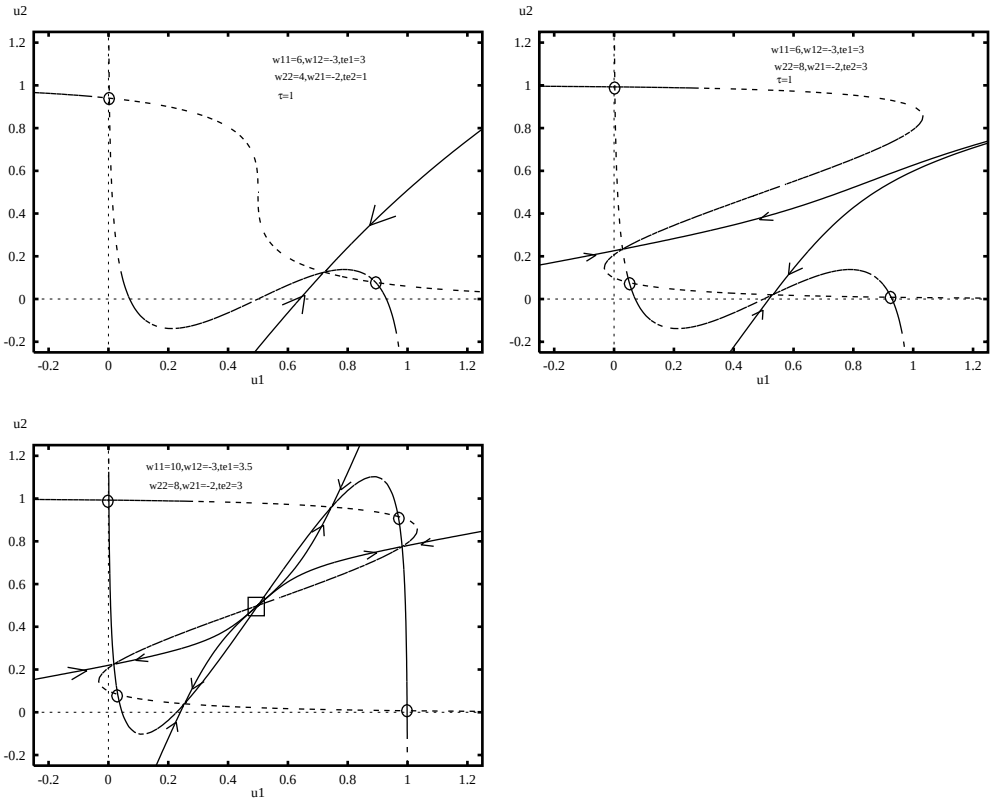

Figure 4. Phase-plane of two-neuron models which have only fixed points. Nullclines and stable manifolds of the saddle points are shown. Circles are stable nodes and squares are unstable nodes. (*a*) Three fixed points: one saddle and two attractors. (*b*) Five fixed points: two saddles and three attractors. (*c*) Nine fixed points: four attractors, four saddles, and one unstable node.

when we say 'unstable FP', we mean either a node or a spiral but not a saddle which is topologically distinct.)

The most interesting dynamics occur when there is a negative feedback loop between the two neurons and there is some self-excitation ($w_{jj} > 0$). In this case, we cannot eliminate the possibility of limit cycles. We restrict our attention to networks that obey 'Dale's principle' which for our purposes means that a cell that exerts an inhibitory (excitatory) influence on some other cell will inhibit (excite) any other cell that it is connected to. Since we have shown that limit cycle dynamics occur only in systems with negative feedback and some self-excitation, we briefly look at two neuron models where w_{11}, w_{21} are non-negative and w_{22}, w_{12} are non-positive. These can have all of the behaviours in Beer's list except for seven fixed points and nine fixed points. There are several interesting cases that Beer missed. These are shown in figure 5 and include (1lc2) stable FP, stable LC, unstable LC and (1lc3) unstable FP, 2 stable LC, 1 unstable LC. Case (1lc2) has bistability between a fixed point and a limit cycle but an unstable limit cycle separates the basins of attraction in contrast to Beer's (3lc1) where the saddle point separates the basin. Case (1lc3) has bistability between two limit cycles.

The point of this analysis is that the complete behaviour of two cell neural networks

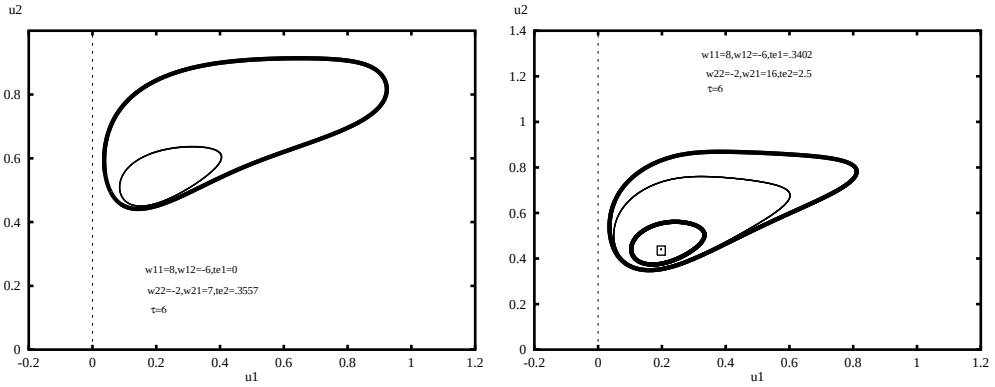

Figure 5. Multistability with limit cycles in two-neuron networks. (*a*) Stable fixed point and stable limit cycle coexisting ($w_{11} = 8$, $w_{12} = -6$, $w_{21} = 7$, $w_{22} = -2$, $\theta_1 = 0$, $\theta_2 = 0.355$, $\tau_1 = 1$, $\tau_2 = 6$, $f(u) = 1/(1 + e^{-u})$). (*b*) Two stable coexistent limit cycles (parameters as in (*a*), except $\theta_1 = 0.34$, $\theta_2 = 2.5$, $w_{21} = 16$).

can be classified and determined by looking at the nullclines and with the use of a decent bifurcation package.

Borisyuk and Kirillov (1992) studied (3.3) in which only two parameters are allowed to vary; the input to the excitatory neurons and the connection from the excitatory cell to the inhibitory cell. They numerically compute a complete two-parameter bifurcation for this system. They find a variety of special points such as the coalescence of limit cycles, Takens–Bogdanov points, and several other codimension 2 bifurcations. We refer the reader to Borisyuk and Kirillov or Beer to see examples beyond those that we have described here.

4.4. Chaos in forced two- or three-dimensional networks

Chaotic behaviour in neural networks and in brain theory has been an active, if controversial, area of research (Baird 1986, Basar 1990, Freeman 1987). We make no claim here as to its relevance to biology but mention it in the interest of completeness.

Topological considerations eliminate the possibility of chaos in two-cell neural networks. Periodically driving these systems can result in chaos. Beer (1995) found such behaviour by driving the system (3lc1) which is near a homoclinic bifurcation. Driving systems near homoclinic bifurcations is known to produce chaotic behaviour (Guckenheimer and Holmes 1983).

It is easy to create an autonomous neural network that exhibits chaotic behaviour. Beer did this for a three-neuron network. We can start with case (3lc1) in which the two-dimensional network has a stable fixed point and stable limit cycle separated by a saddle point. We then simply add a third variable that slowly passes through the homoclinic point. Figure 6 shows an example of the attractor for this. A numerically computed Poincaré map reveals an essentially one-dimensional structure that is in the form of a cap map.

Another approach to creating chaotic neural networks is to find networks with fixed points that have high co-dimension singularities. Ermentrout (1984a) constructed three-variable neural networks with chaotic behaviour by choosing parameters so that there is a triple zero eigenvalue. Below, we discuss another class of chaotic neural networks found by coupling two neural oscillators. van Vreeswijk and Sompolinsky (1996) have recently shown that large networks of coupled neurons where the excitation and inhibition balance (in

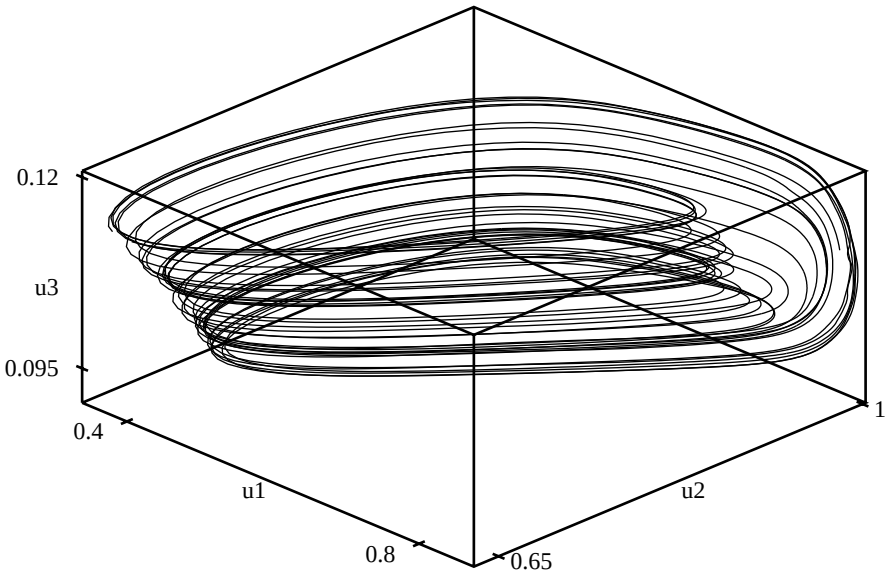

Figure 6. Chaotic solution to a three-neuron network, $\tau_i u_i' = -u_i + f(\sum_j w_{ij} u_j - \theta_i)$ for $i = 1, 2, 3$. $f(u) = 1/(1 + e^{-u})$, $\theta_1 = -2$, $\theta_2 = 4.5$, $\tau_1 = 1$, $\tau_2 = 1.35$, $w_{11} = 8$, $w_{12} = -6$, $w_{21} = 16$, $w_{22} = -2$, $\theta_3 = 8$, $w_{13} = -10$, $w_{31} = 8$, $\tau_3 = 60$, $w_{23} = w_{32} = w_{33} = 0$.

a mean-field sense) exhibit chaotic behaviour that suggests a reason for the high variability of cortical spikes.

4.5. Coupled oscillatory nets

We review the behaviour of a pair of excitatory–inhibitory neural networks that have autonomous oscillations in the absence of coupling. There has been a great deal of recent interest in the mechanisms of synchronization between neurons in cortex. This theoretical and experimental interest has been due to the experiments of Gray and Singer (1989) in which synchronized 40 Hz oscillations are found in the cat visual cortex during the presentation of certain stimuli. This experimental result has spawned a virtual industry of models about synchrony in networks of neurons. The basic questions are what mechanisms of coupling between neurons guarantee synchronization over large distances in the presence of noise. Grannan *et al* (1993) suggested that the tight coherence between cells that are some distance apart cannot be achieved with localized coupling and requires long-distance coupling between the distant units. The reason for this assertion is that the noise will overwhelm any coherence that could be induced through strictly local connections. Given that two cells or local cortical circuits are connected and that they both oscillate with roughly the same frequency, the question becomes: what are the stable allowable phase relationships between the cells? This question has led to a large amount of theoretical work with far fewer experimental results (Grannan *et al* 1993, Hansel and Sompolinsky 1996, Hansel *et al* 1995, Usher *et al* 1993, van Vreeswijk *et al* 1994, Wang and Rinzel 1992, Borisyuk *et al* 1995, Coombes and Lord 1997, Ermentrout and Kopell 1991). There are some general principles that seem to have emerged mainly in the 'weak coupling regime'. For integrate-and-fire models, the behaviour of pairs of mutually coupled cells and globally coupled arrays has been thoroughly analysed, (Usher *et al* 1993, Abbott and van Vreeswijk 1993, van Vreeswijk *et al* 1994).

4.5.1. Weak coupling. Suppose that two nearly identical networks are uncoupled from each other and have a unique stable autonomous limit cycle with period, T. Since this periodic orbit is hyperbolic (that is, the linearized system has only one unit multiplier and all the others are bounded away from the unit circle) this means that the two limit cycles will persist when the systems are coupled if the coupling is sufficiently weak. There are generically at least two things that can happen when two identical oscillators are coupled weakly: (i) they can synchronize and oscillate with no phase shift; or, (ii) they can oscillate a half cycle out of phase. Other behaviours can occur, but the synchronous and out-of-phase oscillations are always solutions to the weakly coupled system.

By using the method of averaging, it is possible to compute the interaction between the oscillators in the limit when the interactions are vanishingly small. Uncoupled, each oscillator can be described by a single variable, the phase, θ_j where $j = 1, 2$ corresponds to the index of the oscillator. Kuramoto (1984) was among the first to describe how to compute the interaction between oscillators in the limit of infinitesimally weak coupling. Suppose, X_1, X_2 are two vectors corresponding to identical weakly coupled limit cycle oscillators with period T satisfying:

$$\frac{\mathrm{d}X_j}{\mathrm{d}t} = F(X_j) + \epsilon G_j(X_j, X_k) \qquad j, k = 1, 2 \qquad k \neq j.$$

Let $X_0(t)$ satisfy

$$\frac{\mathrm{d}X_0}{\mathrm{d}t} = F(X_0)$$

and let, $X^*(t)$ satisfy:

$$\frac{\mathrm{d}X^*(t)}{\mathrm{d}t} = -D_X F(X_0(t))^T X^*(t)$$

with normalization

$$\frac{1}{T}\int_0^T X^*(t) \cdot X_0'(t)\,\mathrm{d}t = 1.$$

Then, for ϵ sufficiently small, the solutions to the weakly coupled system have the form, $X_j(t) = X_0(t + \theta_j) + \mathrm{O}(\epsilon)$ and the phases, θ_j satisfy:

$$\frac{\mathrm{d}\theta_1}{\mathrm{d}t} = 1 + \epsilon H_1(\theta_2 - \theta_1) \tag{4.3}$$

$$\frac{\mathrm{d}\theta_2}{\mathrm{d}t} = 1 + \epsilon H_2(\theta_1 - \theta_2) \tag{4.4}$$

where H_j is a T-periodic function of its argument. The functions, H_j are given by the averages:

$$H_j(\phi) = \frac{1}{T}\int_0^T X^*(t) \cdot G_j(X_0(t), X_0(t + \phi))\,\mathrm{d}t.$$

Letting $\phi = \theta_2 - \theta_1$ denote the phase difference between the two cells, we see that

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \epsilon(H_2(-\phi) - H_1(\phi)) \equiv -\epsilon g(\phi).$$

Thus, phase-locked solutions (solutions which maintain a constant phase difference) are just fixed points of the function g and if $g' > 0$ then they are stable fixed points. If the oscillators are symmetrically coupled and identical, then $G_1(X, Y) = G_2(X, Y)$ and g is just twice the odd part of H. Because H is periodic, g will have two roots, 0 and $T/2$,

the synchronous and antiphase solutions. (Unless, H is purely even, which occurs in some physical systems such as Josephson junctions, but never in any biological examples.)

Obviously g can have other roots besides 0 and $T/2$ but the main solutions that are of interest are the synchronous and antiphase solutions. Thus, a great deal of recent research has been aimed at finding what aspects of the coupling and oscillator properties determine the stability of the synchronous state (see, e.g. van Vreeswijk *et al* 1994, Wang and Rinzel 1992, Hansel *et al* 1995, Coombes and Lord 1997).

Haken (1996) and Kelso (1995) as well as Hansel *et al* (1995) studied a model where $g(\phi) = a\sin\phi - b\sin 2\phi$. If $a > 0$ and b is small enough, then the only roots are 0 and $\pi/2$. As b increases in magnitude two new roots appear, and depending on parameters can be stable or unstable. This type of behaviour is found in 'realistic' models of neurons (see for example, van Vreeswijk *et al* 1994, Hansel *et al* 1995). The key point in the weakly coupled analysis is to find the functions H and g which in turn requires finding the function, $X^*(t)$. The latter is a simple numerical exercise (see Williams and Bowtell, 1997) for a given model and a given type of connection. This local analysis informs us of all possible phase-locked solutions in the case of weak coupling.

Hoppensteadt and Izhikevich (1997) proved a useful result for coupled networks of the form:

$$u_j' = -u_j + f(a_{ee}u_j - a_{ie}v_j + b_{ee}u_k - b_{ie}v_k + p_e) \tag{4.5}$$

$$\tau v_j' = -v_j + f(a_{ei}u_j - a_{ii}v_j + b_{ei}u_k - b_{ii}v_k + p_i) \tag{4.6}$$

where $k \neq j$. Here b_{jk} are small parameters. When they are zero, each of the uncoupled systems, (u_j, v_j) have a limit cycle. There are four ways to couple the two networks together by letting each of the four b_{jk}'s be non-zero. Let $g_{jk}(\phi)$ be the odd part of the interaction function for each of the four possibilities. Hoppensteadt and Izhikevich proved that

$$g_{ie}'(0) = g_{ei}'(0).$$

That is, the stability of synchrony is independent of whether the coupling is from the excitatory cell of one cell to the inhibitory cell of the other or vice versa.

Example. Suppose that $a_{ee} = 12$, $a_{ie} = 8$, $a_{ei} = 16$, $a_{ii} = 2$, $p_e = -2$, $p_i = -6$, $\tau = 1$ and $f(u) = 1/(1 + e^{-u})$. Then a computation of the interaction functions $H(\phi)$ for each of the four possibilities in equations (4.5) yields the following approximations:

$$H_{ee}(\phi) = -1.225\cos\phi - 0.29\sin\phi$$

$$H_{ie}(\phi) = 0.4124\cos\phi + 0.71\sin\phi$$

$$H_{ei}(\phi) = 0.245 + 0.145\cos\phi + 0.71\sin\phi$$

$$H_{ii}(\phi) = -0.257 + 0.546\cos\phi - 0.77\sin\phi.$$

Thus, for cross connections, synchrony is stable but for connections between the same cell types, synchrony is unstable since in the former, the odd part of H is positive and in the latter it is negative.

4.5.2. Beyond weak coupling. Weak coupling is a rather stringent requirement for the analysis. In many numerical cases, it has 'predicted' far better than would be expected for an asymptotic theory. Hoppensteadt and Izhikevich suggested that the normal coupling between pyramidal cells is weak although it is not clear in what sense any interactions are 'weak'. Ermentrout and Kopell (1991) showed that 'weak' coupling is not necessary in order

to apply averaging and if the 'unit' producing oscillations is able to produce outputs with a wide distribution of phases, then one can still use averaging even with strong coupling.

It is difficult to imagine any general theory of moderate and strong coupling of neural oscillators; the systems are just too complicated. (However, Terman and Wang (1995) have developed a general theory for moderately coupled relaxation oscillators.) Most results to date are numerical and thus very specific to the particulars of the model. As we noted above, the behaviour of integrate-and-fire neurons for arbitrary coupling strengths is fairly well characterized because closed form solutions can be found.

The most systematic investigation of the interactions between neural network oscillators was done by Borisyuk *et al* (1995). They consider the Wilson–Cowan firing rate model (3.3) in which each oscillator consists of a pair of excitatory and inhibitory cells. In the same manner as (4.5) they look at all four different types of coupling between the excitatory–inhibitory network. They use the weak coupling analysis as a starting point of their work in which they follow the behaviour as the coupling strength increases and the weakly coupled assumption no longer holds. They use a continuation package to follow the branches of the phase-locked solutions as the the coupling parameter varies. In addition to the various types of phase-locked solutions expected from the weak coupling hypothesis, they find bifurcation to tori, chaos, period-doubling, and oscillator death as the strength of the coupling increases.

5. General considerations about large networks with no inherent spatial structure

In this section, we review some of the literature on recurrent neural networks with no imposed spatial structure. These types of networks have been attractive to physicists due to their isomorphism to spin glass systems in some cases. From a biological point of view, they can be regarded as models of local regions of cortical tissue where there is not a global spatial or other topological structure. The literature on these types of networks is large and there have been many reviews and books on the subject. The basic models of interest have the form:

$$\tau_j \frac{\mathrm{d}u_j}{\mathrm{d}t} + u_j = F_j\left(\sum_k w_{jk} u_k\right) \tag{5.1}$$

with either $w_{jk} = w_{kj}$ or in some cases, no structure at all on the weights. (However, often time is discrete and additionally, in some 'models', the units are updated asynchronously.) The 'voltage' formulation in which the nonlinearity is inside the summation will be considered as well. A very nice history of the development of these networks as well as applications can be found in Cowan and Sharp (1988).

5.1. McCulloch–Pitts neural networks

Among the earliest attempts at modelling the dynamics of neural assemblies is the classic paper of McCulloch and Pitts (1943). In this work, the authors considered a network of binary neurons of the form:

$$N_i(t+1) = H\left(\sum_j w_{ij} N_j(t) + \xi_i\right)$$

where H is the unit step function and w_{ij}, ξ_i are real numbers. At each timestep, all 'neurons' are updated according to the dynamic rule above. McCulloch and Pitts showed that these networks can carry out any logical calculation and thus could simulate any digital calculation. The main problem faced by the user of such a network is how to choose the

weights, w_{ij} for the desired calculation. An alternative to the McCulloch–Pitts model is to update the network asynchronously, that is randomly pick an index and update that particular cell. All of these models avoid the issue as to whether and under what circumstances, the network will settle into some type of regular behaviour such as a fixed point or limit cycle. Hopfield, in his pivotal work, answered this question for a class of neural networks with symmetric weights, $w_{ij} = w_{ji}$.

5.2. Feedforward networks

Models such as (5.1), in which the outputs of some units become feedback to the input units, are called recurrent networks. The simplest neural networks are feedforward models in which connections only go in one direction. Typically there are several 'layers' of cells with the first layer called the 'input' layer and the last layer called the 'output' layer. Each layer receives weighted inputs from the layer below it and sends output to the layer above it.

The earliest feedforward networks were perceptrons (Rosenblatt 1958) in which there are only two layers, an input and an output layer and one set of weights from the input to the output layer. Each neuron is a McCulloch–Pitts unit (that is 0 or 1 output) but the weights between units are allowed to vary. The idea is to choose the weights so that the desired output pattern is produced for a given input pattern. Rosenblatt introduced a rule to adjust the weights: if the unit's response was correct, make no adjustment. If it is incorrect then increase or decrease all weights which arise from active input units depending on whether the unit is supposed to be active. A related model is the adaline invented by Widrow and Hoff (1960) which uses a gradient descent algorithm to vary the weights by minimizing the mean square error. We discuss this rule more generally in section 5.2.1. Two-layer feedforward networks are very restrictive and cannot solve some very simple classification problems. The classic example is the 'exclusive or' problem. For binary inputs, s_1, s_2 produce an output that is zero if $s_1 = s_2$ and 1 otherwise. Minsky and Papert, who advocated an artificial intelligence view of the brain, proved that perceptrons could not distinguish between the letter T and the letter C (see Cowan and Sharp 1988).

This obvious weakness of perceptrons can be solved by adding extra layers. Layers that are not the input or output layer are called 'hidden layers'. The dynamics of feedforward networks is simple and ultimately represents a static nonlinear transformation of the input. Nevertheless, these networks are extremely useful in a variety of applications such as character and speech recognition as well as many other engineering examples. See Hertz *et al* (1991) or Muller *et al* (1995) for many applications.

5.2.1. Back propagation. The main problem in all of these feedforward models is how to choose the weights. This requires 'training' the network—that is, giving input/output pairs and adjusting weights so that the error is minimized. The most common method for doing this is called 'back propagation'. The error in the output layer (layer M) is used to adjust the weights from layer $M-1$ to M and this error is propagated back to adjust weights from layer $M-2$ to layer $M-1$, etc. This technique is so important that we summarize here the general algorithm as described by Hertz *et al*. Let V_k^m be the output of layer m:

$$V_k^m = S(h_k^m) \equiv S\left(\sum_j w_{kj}^m V_j^{m-1}\right)$$

where S is the sigmoid function. (Note that this is the steady state of the network after all temporal transients have died away.) Let $V_k^0 \equiv \xi_k$ be input and Y_k be the desired output.

Define

$$\delta_k^M = S'(h_k^M)(Y_k - V_k^M)$$

where

$$h_k^m = \sum_j w_{jk}^m V_j^{m-1}.$$

Then define

$$\delta_k^{m-1} = S'(h_k^{m-1}) \sum_j w_{kj}^m \delta_j^m.$$

Now adjust the weights by an amount

$$\Delta w_{kj}^m = \eta \delta_k^m V_j^{m-1}$$

where η is an adjustable parameter for the learning rate. Repeat this many times for all patterns until some convergence is obtained. The error in the top layer is propagated back to lower layers, hence the name of the rule. Hertz *et al* (1991) showed that this is equivalent to using gradient descent on the mean-squared error. In a recent paper, Fitzsimonds *et al* (1997) showed that in cultures of hippocampal neurons, long-term depression (a type of synaptic weight modification) was 'back propagated' the synapses of the presynaptic neurons thus providing a possible biological analogue of back propagation.

5.3. Hopfield networks as spin glasses

The theory of neural networks was reborn (at least in the eyes of physicists) with the publication of two papers by Hopfield (1982, 1984) in which he pointed out that neural networks with certain symmetries are analogous to spin glasses. We describe these results here, following the development by Hertz *et al*. We first consider a network of binary neurons which are updated asynchronously. (That is, an index is randomly chosen and the corresponding unit is updated.) Then we look at the global stability of continuous systems of the form (5.1).

Consider the system:

$$x_i(t+1) = H\left(\sum_j w_{ij} x_j - \mu_i\right)$$

where H is the Heaviside step function. In order to make the association with spin systems clearer, we introduce the spin variables S_i where $S_i = 2x_i - 1$ take on values of $+1$ (firing) or -1 (not firing). We then obtain:

$$S_i(t+1) = \mathrm{sgn}\left(\sum_j w_{ij} S_j - \theta_i\right)$$

where

$$\theta_i = 2\mu_i - \sum_j w_{ij}$$

and $\mathrm{sgn}(x)$ is the signum function; $+1$ for non-negative arguments and -1 for negative ones. If we are using asynchronous updating for the network, then it is clear that there will be trouble if w_{ii} is not zero. For simplicity, we set $\theta_i = 0$. Hopfield's most important contribution was to notice that this system has an energy function:

$$H = -\tfrac{1}{2} \sum_{ij} w_{ij} S_i S_j.$$

To see that this is indeed an energy function, let S_i' be the new state of one of the cells, i. If $S_i' = S_i$ then there is no change in the energy. Otherwise, $S_i' = -S_i$ so that the change in energy is

$$\Delta H = -\tfrac{1}{2}(S_i' - S_i)\sum_j w_{ij}S_j.$$

Since

$$S_i' = \mathrm{sgn}\left(\sum_j w_{ij}S_j\right)$$

both S_i' and the sum have the same sign. Thus, ΔH is negative and the energy decreases. This shows that eventually the system will settle into one of several possible minima. These states are the so-called memories of the system. This is the simplest model for an associative memory; all initial conditions settle into some particular attractor. That is, an initial input is 'associated' with a particular final state. This network and related ones such as (5.1) are called 'attractor' neural networks since all initial data eventually settle into one of the stable fixed points of the system.

We now want to determine how to 'set' the memories to a desired pattern. Suppose that we want the system to always settle into a single pattern (as opposed to one of several), say ξ_j. We want to choose w_{ij} so that

$$\mathrm{sgn}\sum_j w_{ij}\xi_j = \xi_i.$$

Consider

$$w_{ij} = \frac{1}{N}\xi_i\xi_j$$

where N is the number of units. Then it is clear that such a choice of weights will always have $\boldsymbol{\xi}$ as a fixed point. In fact, if fewer than half the bits of the starting pattern are wrong, then the output will still be the desired pattern. The network is an attractor network and nearby initial data will converge to the desired pattern. Note that another attractor of this network is $-\xi_j$ the state with every bit reversed. If more than half the initial bits of the starting state S_j are the opposite of ξ_j then the network will converge to the attractor, $-\xi_j$. Globally, the binary phase space is partitioned into two subspaces separated by the 'hyperplane' in which there are exactly half the bits reversed. All initial data in one half are attracted to ξ_j and those in the other half to $-\xi_j$.

If we want to store M attractors, $\{\xi_j^1, \ldots, \xi_j^M\}$ then we choose

$$w_{ij} = \sum_{m=1}^{M}\xi_i^m\xi_j^m.$$

If these are all mutually orthogonal, then each pattern is a fixed point of the map. In general, however, we cannot expect this to be true. We want to know how 'close' the output:

$$h_i^n \equiv \sum_j w_{ij}\xi_j^n$$

is to the desired pattern, $\mathrm{sgn}h_i^n = \xi_i^n$. By breaking the sum up we see that

$$h_i^n = \xi_i^n + \frac{1}{N}\sum_j\sum_{m\neq n}\xi_i^m\xi_j^m\xi_j^n \equiv \xi_i^n + E_i^n.$$

As we noted above, if the patterns are mutually orthogonal, then the error E_i^n is identically zero. If the absolute error is less than one for each i then the output will have the desired

sign and will be a local attractor. An elegant calculation by Hertz *et al* shows that if the choice of patterns is random in that there are no correlations between the patterns, then the capacity (number of stable patterns) of the network can be computed for any given error tolerance. They introduce the cross-talk term

$$C_i^n = -\xi_i^n E_i^n.$$

If C_i^n is negative, then the error is in the right direction and will not affect the sign; if it is greater than 1, then the output sign will be in error. They show that

$$\text{Prob}(C_i^n > 1) = \tfrac{1}{2}\left(1 - \text{erf}\sqrt{N/2m}\right).$$

For example, if the tolerance for error is 0.001, then the maximum capacity is about 10% of the network size.

In addition to the attractors that are desired in these networks, there are spurious states that fall into three classes. We have already seen that the reversed states are also attractors. Mixed states correspond to linear combinations of an odd number of the desired patterns. For example, the pattern $\xi_j^1 + \xi_j^2 + \xi_j^3$ when plugged into the network yields:

$$\tfrac{1}{2}(\xi_j^1 + \xi_j^2 + \xi_j^3) + \text{cross terms}.$$

Using the symmetry, it is easy to show that this is a stable pattern. A mixture consisting of an even number of states will have some zero states which means that it is not a valid pattern. Finally, there are states which are not a combination of any finite set of the original patters; the so-called spin-glass states.

5.4. Statistical mechanics and the capacity of attractor networks

Amit *et al* (1987) and Sompolinsky (1987) studied the behaviour of networks in which $\alpha \equiv p/N$, the ratio of the number of stored patterns to the number of neurons approached a limiting value as $N \to \infty$. They showed that in the limit of 'zero temperature' (that is the units are binary) that the capacity of the network is roughly, $\alpha = 0.144$. That is, if one tries to store more than this critical number of patterns, then the network will be overwhelmed by the spurious states. The books by Muller *et al*, Amit (1989) and Hertz *et al* give many more details on these calculations.

5.4.1. Energy and Liapunov functions for general continuous recurrent networks. Hopfield (1984) showed that just as in the asynchronous binary units, there is an energy equation for (5.1) if the weights are symmetric. The function is

$$E = -\tfrac{1}{2}\sum_{ij} u_i u_j + \sum_i \int_0^{u_i} g^{-1}(z)\,\mathrm{d}z$$

where $u_i = g(V_i)$. We show that $\mathrm{d}E/\mathrm{d}t < 0$. Differentiation yields

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} &= -\frac{1}{2}\sum_{ij} w_{ij}\frac{\mathrm{d}u_i}{\mathrm{d}t}u_j - \frac{1}{2}\sum_{ij} w_{ij}u_i\frac{\mathrm{d}u_j}{\mathrm{d}t} + \sum_i g^{-1}(u_i)\frac{\mathrm{d}u_i}{\mathrm{d}t} \\
&= -\sum_i \frac{\mathrm{d}u_i}{\mathrm{d}t}\left(\sum_j w_{ij}u_j - V_i\right) \\
&= -\sum_i \frac{\mathrm{d}u_i}{\mathrm{d}t}\tau_i\frac{\mathrm{d}V_i}{\mathrm{d}t} \\
&= -\sum_i \tau_i g'(V_i)\left(\frac{\mathrm{d}V_i}{\mathrm{d}t}\right)^2 \leqslant 0.
\end{aligned}$$

The key part of this argument is arriving at the second line from the first one; this is valid because $w_{ij} = w_{ji}$. That is, the network is symmetric. Furthermore, g is a sigmoid function and thus monotonic. If g is strictly monotonic, then $\mathrm{d}E/\mathrm{d}t$ vanishes only when $\mathrm{d}V_i/\mathrm{d}t = 0$ for all i, that is at a fixed point. These symmetric networks have no non-trivial steady states (that is, steady states which are not fixed points) such as limit cycles and chaotic behaviour. Note that this does not imply that every fixed point is stable. *Hopfield's result implies that all initial data relax to some fixed point, not that all fixed points are attractors.*

5.4.2. Cohen–Grossberg theorem. We close this section with another important result in symmetric attractor neural networks, the Cohen–Grossberg (1983) theorem. We will first state the theorem and then show how it applies to a number of standard neural models. The general model takes the form:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = a_i(x_i)\left[b_i(x_i) - \sum_{j=1}^{N} c_{ij}d_j(x_j)\right]. \tag{5.2}$$

This model is a generalization of the famous Lotka–Volterra model for species competing for a common resource. The assumptions are quite simple,

$$c_{ij} = c_{ji} \qquad \text{symmetry} \tag{5.3}$$

$$a_i(x_i) \geqslant 0 \qquad \text{positivity} \tag{5.4}$$

$$d_i'(x_i) \geqslant 0 \qquad \text{monotonicity.} \tag{5.5}$$

We have seen assumption (5.3) in all of the examples so far. Without this symmetry assumption, the convergence of these networks to fixed points would only occur in very special circumstances. The other two assumptions are rather mild. The latter two assumptions imply that there is a global Liapunov function:

$$V = -\sum_{i=1}^{N}\int^{x_i} b_i(y)d_i'(y)\,\mathrm{d}y + \tfrac{1}{2}\sum_{j,k=1}^{N} c_{jk}d_j(x_j)d_k(x_k).$$

Differentiating V with respect to t we find

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\sum_{i=1}^{N} a_i(x_i)d_i'(x_i)\left(b_i(x_i) - \sum_{j=1}^{N} c_{ij}d_j(x_j)\right)^2$$

which is negative if assumptions (5.4), (5.5) hold. (Note that we really only need $a_i d_i' \geqslant 0$.) The existence of a function which decreases along trajectories implies that all solutions tend to fixed points. If inequalities (5.4), (5.5) are strict, then the only fixed points satisfy:

$$b_i(x_i) = \sum_{j=1}^{N} c_{ij}d_j(x_j).$$

This result includes the Hopfield result as a special case. To see this consider a slight generalization:

$$\tau_i\frac{\mathrm{d}V_i}{\mathrm{d}t} + V_i = \sum_j w_{ij}g(V_j) + I_i$$

where we have included inputs I_i. Make the following identifications:

$$a_i(V_i) = \frac{1}{\tau_i}$$

$$b_i(V_i) = I_i - V_i$$

$$d_i(V_i) = g(V_i)$$
$$c_{ij} = -w_{ij}.$$

It is clear that the assumptions of the Cohen–Grossberg theorem are met.

A subcase of the Cohen–Grossberg class of models (used by Destexhe 1994) is the shunting network:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -A_i x_i + (B_i - x_i)[I_i + f_i(x_i)] - (x_i + C_i)\left[J_i + \sum_{j=1}^{N} D_{ij} g_j(x_j)\right] \tag{5.6}$$

where $D_{ij} = D_{ji} \geqslant 0$ and $g_j'(x_j) \geqslant 0$. The constants A_i, B_i, C_i are all non-negative. This model is somewhat like a membrane model where there is self-excitation if $f_i > 0$ and lateral inhibition. We will analyse models like this in a later part of this paper in the limit as $N \to \infty$ and there is spatial structure in the coefficients D_{ij}. Clearly if A_i, B_i, C_i are all non-negative and the inputs (I_j, J_j, g_j, f_j) are positive, then if $x_i(0) \in (-C_i, B_i)$ then it will remain in this interval for all time. To make the association with the Cohen–Grossberg equation (5.2) we let $y_i = x_i + C_i$. Then make the identifications:

$$a_i(y_i) = y_i$$
$$b_i(y_i) = \frac{1}{y_i}\{A_i C_i - (A_i + J_i)y_i + (B_i + C_i - y_i)[I_i + f_i(y_i - C_i)]\}$$
$$c_{ij} = D_{ij}$$
$$d_i(y_i) = g_i(y_i - C_i).$$

All of the assumptions obviously hold so that this shunting network is again a special case of the Cohen–Grossberg theory.

Ye *et al* (1995) and Herz (1995) studied the dynamics of Cohen–Grossberg networks in the presence of multiple delays. Instead of terms of the form

$$c_{ij} d_j(x_j(t))$$

in (5.2), they incorporate multiple delays:

$$c_{ij}^k d_j(x_j(t - \tau_k)).$$

They find conditions on the delays, τ_k such that there is still global stability. They basically show that if the delays are small enough, then all the results of the Cohen–Grossberg theorem still hold. This result generalizes that of Marcus and Westervelt (1989) who analysed the Hopfield model with delays and showed that with small delays global stability still holds but with larger delays oscillations occur. These effects are analogous to the delay-induced instability discussed in section 4.2.2.

Koch and Leisman (1996) studied the effect of delays in a spatially continuous neural network and how these delays increase the number of unstable modes of behaviour. Their analysis is for a strictly linear network, so it is likely that many of the modes that they find will be unstable in a full nonlinear analysis. (See section 8 for examples of this type of pattern selection.)

5.5. Line attractors

Line attractors are a type of attractor that has recently been considered as a model for occulomotor integrator which transforms velocity-coded visual and vestibular signals into

eye position. Experimentally, what is found is that the firing rates of individual neurons is a linear function of the eye position, E. That is

$$u_i = h_i + \xi_i E \tag{5.7}$$

where h_i and ξ_i are constants that depend on the individual neuron. The eye position is a dynamic variable that the integrator determines by the summed activity of the neurons in the network. The eye position can be held over many seconds, long after the stimulus has disappeared, thus, it has been suggested that the firing rates are 'attractor' states of some sort of neural network. Because the firing rates of each of the neurons are parametrized by the constant E, the network has been called a line attractor. Seung (1996) proposed the simple model:

$$u_i' = -u_i + h_i + \sum_j w_{ij} f(u_j).$$

The idea is to choose the weights and inputs so that the firing rates are as close to (5.7) as possible. Seung, motivated by Hopfield's results, chose

$$W_{ij} = \xi_i \eta_j.$$

The steady states are then:

$$u_i = h_i + C\xi_i$$

where C must satisfy

$$C = \sum_j \eta_j f(C\xi_j + h_j). \tag{5.8}$$

The idea is to choose ξ_i, η_i, so that there are as many values of C as possible that satisfy (5.8). If f is linear, then choose $\eta_j = \xi_j$ to be normalized and orthogonal to h_j. For the nonlinear problem, this choice works fairly well and can lead to many roots of this equation.

5.6. Winner-take-all networks

An important class of networks that has been studied by a large number of authors are the so-called winner-take-all (WTA) networks. In these networks, input may be given or not, but the end result is that only one unit is turned on and the other units are suppressed. These units are important in the formation of self-organized maps (Kohonen 1989) since the assumption in the map models is that one unit or small group of local units is excited. Chialvo and Bak (1997) showed that the WTA behaviour is very good for building systems that are able to rapidly adapt to changing environments. The architecture of WTA networks is usually of the form:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = F\left(u_i, I_i, \sum_j u_j\right) \tag{5.9}$$

where F is an increasing function of its second argument and a decreasing function of the third argument. If the global inhibition of the network is strong enough, then all but the system receiving the strongest input will be suppressed. Yuille and Gryczwyz (1989) showed that for $F(u, I, \Sigma) = -u + If(u - \Sigma)$ with f a sigmoid nonlinearity, then WTA behaviour can occur. In the case of equal inputs, it is also possible for WTA to occur, but in this case, the 'winner' is the unit that initially starts highest. Models such as the Yuille and Gryczwyz are easy to understand by looking at $N = 2$ neurons and $f(u) = 1/(1 + \mathrm{e}^{-u})$. There will be a unique winner no matter what the initial conditions are if both inputs I_i are

small enough so that there is a unique fixed point. As both inputs increase in strength, a bifurcation to a pair of new fixed points occurs. This leads to bistability of the network and so even though the network has asymmetric inputs, some initial conditions will go to the weaker one. In most cases that arise in neural models, the Cohen–Grossberg theorem can be applied to prove that all solutions tend to fixed points.

Fukai and Tanaka (1997) recently derived a simple Lotka–Volterra model from a general neural network. The equations have the form:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = u_i\left(1 + I_i + (\beta - 1)u_i - \beta \sum_j u_j\right) + \epsilon$$

where ϵ is a small positive parameter. For $\epsilon = 0$ they prove that when $\beta = 1$, the network has WTA behaviour in that the largest input suppresses all the other inputs. When $\beta_- < \beta < 1$ the network can have more winners, say n, as long as

$$1 + I_{\min} > \frac{\beta n}{1 - \beta}(\bar{I} - I_{\min})$$

where $I_{\min}$ is the minimum input among the winners and $\bar{I}$ is the mean input among winners. Finally, they show that if $\beta > 1$ the system is a variable WTA in which one unit wins and the others are 0, but the winner is not necessarily the unit with the largest input. Unit j can be a winner as long as

$$\beta(1 + I_j) > (1 + I_i) \qquad \text{for all } i \neq j.$$

Ermentrout (1992a) showed that if the global inhibition is not instantaneous, then the WTA behaviour can be disrupted. For example, consider the WTA model:

$$u_j' = -u_j + f(14u_j - 15v - 1)$$
$$\tau v' = -v + f(15(u - 1 + u_2 + u_3) - 8)$$

where $f(u) = 1/(1 + \mathrm{e}^{-u})$. Then for τ smaller than $\tau_H \approx 0.4$, the network is a WTA. At $\tau = \tau_H$ the system remains a WTA network, but the activity of all the units oscillates with the 'winner' having a much bigger amplitude than the other two units. As τ continues to increase, the difference in magnitude of the three units decreases until at $\tau = \tau_L \approx 4.24$, the stable asymmetric state disappears and the only stable solution is a synchronous oscillation. Thus, the WTA phenomena depends on the ability of the inhibitory influence to act fast.

5.7. Oscillatory networks revisited

Gerstner *et al* (1996) recently considered generalizations of the results of Abbott and van Vreeswijk (1993) van Vreeswijk *et al* (1994) by considering their version of the integrate-and-fire model. They considered a network of neurons of the form:

$$V_i(t) = \sum_f \eta(t - t_i^f) + \sum_j w_{ij} \sum_f \alpha(t - t_j^f - \Delta).$$

Here, w_{ij} are the weights, α is the post-synaptic potential, Δ is the synaptic delay, and $\eta(t)$ is the refractory function (see section 2.1.1). $V_i(t)$ is the potential of the cell, i and t_i^f is the set of firing times of cell, i. They analyse stability of the homogeneous oscillatory solution for networks in which the row sums of w_{ij} are identical, say, w, and $w_{ij} > 0$. They give conditions on the postsynaptic function and the refractory function which guarantee stability of the synchronous state. The conditions for stability are easily illustrated (see figure 7). If the refractory curve, $\theta - \eta(t)$ (where θ is the threshold to fire) intersects the synaptic potential curve $w\alpha(t)$ where the latter has a positive slope, then the synchronous state is

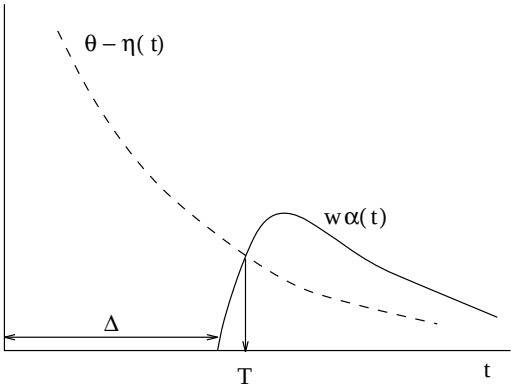

Figure 7. Criterion for stability of the synchronous solution in the 'spike-response' model. The curve of refractoriness, $\theta - \eta(t)$ must intersect the synaptic response, $w\alpha(t)$, where it is increasing.

stable; otherwise it is unstable. Clearly, if the delay is too short, then excitation is unstable as was shown for the integrate-and-fire model by van Vreeswijk *et al* (1994).

A similar result for weakly coupled oscillator networks can also be proven (Ermentrout 1992b). Consider the analogue of (4.3) in a network of N coupled oscillators:

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = \omega_i + H_i(\theta_1 - \theta_i, \ldots, \theta_N - \theta_i). \tag{5.10}$$

Each argument of the functions H_i is periodic with the same period. Suppose that there is a phase-locked solution:

$$\theta_i(t) = \Omega t + \bar{\theta}_i + C$$

where Ω is the ensemble frequency, C is an arbitrary constant, and $\bar{\theta}_i$ is independent of time and represents the relative phases of each of the oscillators. Let

$$w_{ij} = \partial H_i / \partial \phi_j$$

be the partial derivative of H_i with respect to each of its N arguments evaluated at this phase-locked solution. If $w_{ij} \geqslant 0$ and the matrix w_{ij} is irreducible, then the phase-locked solution is asymptotically stable.

6. Models with spatial structure

The models that we have explored so far essentially model a single layer of neurons. That is the implicit assumption made when there are symmetric connections. Furthermore, the structure of the connections is distributed and associative; there is no clear topographic organization of the connections. There clearly exist spatially distributed connections in cortical networks and much of the early theoretical work on neural networks took this viewpoint. Recent experimental advances have enabled neuroscientists to look at spatially distributed activity in cortex and other brain regions. Section 4, on small networks, assumes a layered organization with a clear distinction between excitatory and inhibitory cells. This contrasts with the general work that we described in section 5 on attractor neural networks where there were no restrictions on the values of the weights.

We now turn to the dynamics of neural networks which have spatially structured connections in the sense that there are either layers or topological organization of connections

between units. We will first introduce continuum networks. Then we will describe general results on single-layer models for convolution-type coupling. We will also look at the dynamics of propagation. We then turn to multilayer models in which there is both excitation and inhibition. We apply a variety of mathematical techniques that can be used to attack these problems such as bifurcation and perturbation methods. Finally, we look at more complicated networks in which additional structure is introduced. In all cases, we will try to make analogies with the behaviour of the networks and real biological systems.

6.1. Continuum models

The number of neurons and synapses in even a small piece of cortex is immense so that a natural approach is to take a continuum limit and study neural networks in which space is continuous. We will look at general considerations before turning to simple convolution-type interactions. Consider M populations of neurons with activities, $u_\ell(x,t)$ where x is a spatial variable. It could be in one, two, or perhaps three dimensions. Then the general neural network has the form:

$$L_\ell(t)u_\ell(x,t) = F_\ell\left(\sum_{\ell'}\int_\Omega w_{\ell\ell'}(x,y)u_{\ell'}(y,t)\,\mathrm{d}y\right)$$

where L_ℓ is a temporal integrator; in most of the examples we consider, it will be the operator, $Lu = \tau\,\mathrm{d}u/\mathrm{d}t + u$. Here Ω is the spatial extent of the network. The connectivity functions, $w(x,y)$ determine the total strength of connections between cells at y to cells at x. The most typical form for $w(x,y)$ is a form that is dependent only on the difference between x and y,

$$w(x,y) = W(x-y) \qquad \text{homogeneous.} \tag{6.1}$$

Such interaction functions are called homogeneous since they are translation invariant. An additional simplification is that the interactions are isotropic so that, $W(z)$ is only a function of $|z|$ the absolute value of the vector z. Most people who use continuum models use this approximation. Another issue is the problem of what to do at the boundaries. For example in a one-dimensional domain, what happens at the edges. There are several different approaches that can be taken. To illustrate these, we fix the model to lie on the interval $[0,1]$, so that the interaction is just:

$$\int_0^1 w(x,y)u(y,t)\,\mathrm{d}y.$$

The simplest approach is to just leave the interaction function as $w(x,y) = W(|x-y|)$. This implies that cells near the edges receive fewer connections than cells in the middle. This means that spatially uniform activities are not transformed to spatially uniform outputs. On a large domain or with sufficiently localized connections, this is not a problem for transient phenomena. However, for recurrent activity, the local boundary effects can have global consequences. Thus, this is not particularly suitable when one wants to study long-term spontaneous activity.

The second approach is to let the domain be periodic. This is reasonable for some types of networks where x is not space but orientation (see sections 8.1.2 or 7.3.4 for examples). However, in general this does not make much biological sense. Related to this simplification is to let the domain be infinite. This itself presents some mathematical difficulties. The advantage of periodic and infinite domains will become evident when we apply bifurcation methods to the equations.

The third way to deal with the issue of the edges is to 'homogenize' the medium. That is, let

$$r(x) = \int_0^1 W(|x - y|)\,\mathrm{d}y$$

and define:

$$w(x, y) = W(|x - y|)/r(x).$$

This weight function will now transform spatially homogeneous input to spatially homogeneous output. If the domain is periodic, all that this does is normalize the integral of w. We will call interaction functions of this type 'homogenized'.

6.1.1. What is the form of the interaction function? If the density of cells in a given layer is uniform, and the interactions between cells depend only on distance, then the interaction function from layer ℓ to layer ℓ' is:

$$w(x - y) = \int_\Omega D_{\ell'}(x - y)A_\ell(y)\,\mathrm{d}y$$

where $D_{\ell'}(z)$ is the arborization function for the dendrites of the postsynaptic cell population and $A_\ell(z)$ is the axonal arborization of the presynaptic population. Thus, the contribution from cells at point x to cells at point y is the convolution of the arborization function. In section 8.3.3, we take a more ambitious approach due to Bressloff (1995) and include the fact that synapses may occur at different distances from the cell body along the postsynaptic dendrite.

Given a distance-dependent interaction function, what should the shape be? The classic model has W peaked at the origin and decaying with distance. For example, W may be a Gaussian, $\exp(-(|x|/\sigma)^2)$, exponential, $\exp(-|x|/\sigma)$, or with finite support. One issue is whether a cell can connect to itself. This issue is not yet resolved anatomically although there is strong evidence that at least for inhibitory cortical cells, self-connections are quite prevalent (Tamas *et al* 1997). There is some scant evidence for self-synapses (called autapses) in excitatory pyramidal cells (Lubke *et al* 1996). If self-connections are prohibited by introducing a 'gap' in connectivity, it is possible that new phenomena could occur; this is a subject of current research. For example, Rinzel *et al* (1998) showed that in a network of inhibitorily coupled neurons that have post-inhibitory rebound, smooth propagation occurs only if there is a gap. There are recent examples in the prefrontal cortex of anatomical connections which branch and tuft, branch and tuft, etc in a periodic fashion. That is, the connectivity of cells decreases with distance but is highly modulated by a periodic function. Again, the functional implications of this type of interaction are still an open question.

If the density of cells is non-uniform, then there will be an obvious inhomogeneity in the connections:

$$w_{\ell\ell'}(x, y) = W(|x - y|)\rho_\ell(y)\rho'_\ell(x).$$

The implications of this remain a subject of current research.

In general, spatial interactions are symmetric and similar to one of the four types illustrated in figure 8 which are called respectively: (a) excitatory; (b) inhibitory; (c) lateral-inhibitory; and (d) lateral-excitatory. All of the examples that we analyse in the remainder of this paper will be of one of these four types. All four can be written in the following form:

$$w(x) = a_1 \mathrm{e}^{-(x/\sigma_1)^2} - a_2 \mathrm{e}^{-(x/\sigma_2)^2}$$

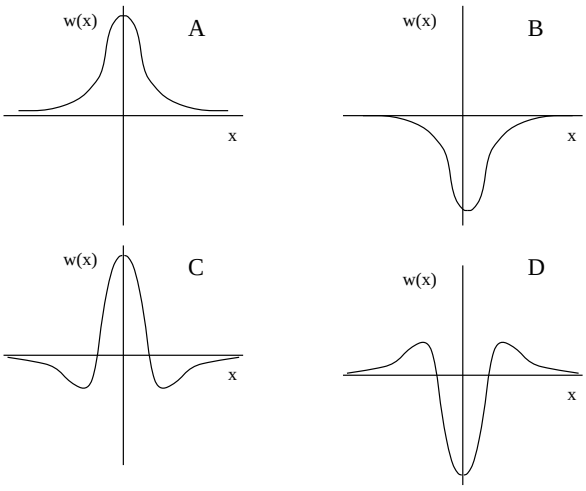

Figure 8. The most typical spatial interaction functions for spatially continuous neural networks. (*a*) Purely excitatory interactions; (*b*) purely inhibitory interactions; (*c*) 'lateral-inhibition' or 'Mexican hat' interactions in which there is local excitation and distant inhibition; (*d*) 'lateral excitation', in which there is local inhibition and distant excitation.

where all parameters are non-negative. Purely excitatory connections have $a_2 = 0$, inhibitory have $a_1 = 0$, lateral inhibitory have $\sigma_2 > \sigma_1, a_1 > a_2$ and lateral excitatory have $\sigma_1 > \sigma_2, a_2 > a_1$.

6.1.2. 'Hebbian' or product connectivity. A simple model for connectivity is the analogue of Hopfield's 'Hebbian' interactions:

$$w(x, y) = \sum_{k=1}^{m} \xi_k(x)\eta_k(y)$$

where $\xi_k(x)$ are continuous functions. We will look at an example of this type of interaction in a model for orientation response proposed by Ben Yishai *et al* (1995, 1997). To see how such a model could be justified, suppose that the domain is periodic. Then

$$w(x, y) = W(|x - y|)$$

is periodic and we can expand W in a cosine series (since it is symmetric):

$$W(z) = J_0 + J_1 \cos ax + J_2 \cos 2ax + \cdots.$$

If we truncate this after finitely many terms, then the interaction will be a sum of terms of the form:

$$\int \cos[na(x - y)]u(y)\,\mathrm{d}y = \int (\cos nay \cos nax + \sin nay \sin nax)u(y)\,\mathrm{d}y$$

which is just a sum of 'Hebbian' interactions.

7. Pattern formation in active media

Pattern formation in nonlinear spatially distributed systems has a long history in physics and biology. The mechanisms responsible for pattern formation are well known (see Fife 1979, Murray 1989, or Cross and Hohenberg 1994). In neuronal networks, we have all

the necessary interactions to lead to a variety of spatial and temporal behaviour. The response of a neural network to inputs is obviously very important, but the analysis of such systems remains a difficult question both analytically and numerically. Thus, we must turn to approximations and perturbative methods. These methods have proven to be very useful in elucidating the behaviour of neural networks. We are particularly interested in the spontaneous excitation of neural networks as well as how the presence of weak inputs can alter this behaviour. Strong inputs will presumably dominate the behaviour of the networks (see Amari 1977, for a justification of this). In a recent series of papers, Hansel and Sompolinsky (1996, 1997) have also assumed that the inputs are weak. Roque Da Silva Filho (1992) studied the equilibrium properties of a two-layer continuous neural network with localized inputs. He shows that the sizes of response of the inhibitory layer is inversely related to the size of the input fields to the excitatory and inhibitory layers. A similar 'paradoxical' behaviour was found in a simple variable excitatory–inhibitory model (Tsodyks *et al* 1997).

In one space dimension, there are several types of behaviour that were found numerically by Wilson and Cowan (1973) and Ellias and Grossberg (1975) and it is these behaviours that we will consider here from an analytic point of view. We will review the following points.

(1) Waves of various types:

(a) wavefronts joining an excited state to a resting state;

(b) solitary waves or travelling pulses in which the medium is briefly excited and then returns to rest;

(c) periodic wave trains.

(2) Stationary pulses or local excitations.

(3) Oscillatory behaviour.

(4) Spatially periodic stationary patterns.

In two dimensions, the same behaviours can be found, however, there are many other interesting patterns that can occur such as spiral waves, target waves, and doubly periodic patterns. We will describe the existence of these patterns in multilayer spatially distributed neural networks. The main techniques are singular perturbation, bifurcation, and linear approximation methods.

Before beginning with the analysis, we briefly discuss the relevance of these patterns to biology. A more detailed discussion can be found in the specific analysis sections. Waves such as wavefronts and pulses have now been observed in slice preparations (Chagnac-Amitai and Connors 1989, Traub *et al* 1993, Golomb and Amitai 1997) using voltage-sensitive dyes and multiple electrodes. Periodic wave trains are common in many systems such as the Limax olfactory bulb (Kleinfeld *et al* 1994) and the locomotor generators of some fish (Grillner, 1974). Standing solitary spatial pulses are believed to be the analogue of short-term memory. There are many examples of delayed-response tasks in which an initiating stimulus is removed, but localized neural activity remains (Fuster and Alexander 1971, Funahashi *et al* 1989). Standing pulses are believed to be the analogue of this phenomenon. Oscillations in neural networks have been of great interest ever since Gray and Singer (1986) observed coherent oscillations in response to a stimulus in the cat visual cortex. Finally, a number of authors have suggested that spatially periodic activity profiles may be the analogue of simple visual hallucinations (Ermentrout and Cowan 1979b, Tass 1995).

In this section, we focus on *active media*. That is, we will describe behaviour of systems that are highly nonlinear and far away from any spatially and temporally uniform states. In the next section, we will look at bifurcation of spatio-temporal patterns from rest. Active

media can be characterized by their local dynamics. We point out three different types of locally active media: (i) bistable, (ii) excitable, and (iii) oscillatory. Locally bistable behaviour occurs when there are two stable states for the network, resting and active. A locally excitable medium has a unique stable fixed point and a threshold-like behaviour. Small perturbations return to rest, but larger ones are amplified before returning to rest. The behaviour of these systems when coupled in a spatially extended network can be quite complex. Vasiliev *et al* (1987) provided a comprehensive review of the behaviour of *diffusively* coupled active systems using singular perturbation methods. In this section, we will use similar techniques to examine the highly nonlinear behaviour of *synaptically* coupled neural networks which are locally active.

The local dynamics combined with different spatial connectivities leads to a variety of large amplitude patterns. Bistable networks with local excitation lead to travelling wavefronts. If there is lateral inhibition, then the same bistable network will lead to localized regions of excitation. In the excitable system, local excitation leads to travelling wave pulses and if there is lateral inhibition, stationary pulses occur. Locally oscillatory networks lead to travelling waves.

7.1. Wavefronts

We will start with some results about one-dimensional models on the real line as these models provide the starting point for singular perturbation methods.

Consider the following network:

$$V(x,t) = \int_0^\infty h(s)\,\mathrm{d}s \int_{-\infty}^{\infty} w(x-y)S(V(y,t-s))\,\mathrm{d}y \tag{7.1}$$

which represents a one-dimensional network of neurons distributed on the real line. We normalize the temporal operator, h and the spatial weight function so that they both integrate to 1 over their respective domains. We assume that w is non-negative and symmetric (like figure 8(a)), h is positive and monotone decreasing, and S is monotone. Finally, we assume that the equation:

$$f(V) \equiv -V + S(V) = 0$$

has precisely three roots, $V_1 < V_2 < V_3$ such that $f'(V_j)$ is negative for $j = 1, 3$ and positive for $j = 2$. This last assumption implies that the spatially homogeneous equation is bistable in that there are two stable fixed points, a low potential, V_1 and a high potential, V_3. The middle potential, V_2 separates the two fixed points.

In the reaction–diffusion literature, the following bistable model is often studied:

$$u_t = u_{xx} + f(u)$$

where $f(u)$ has three roots $u_1 < u_2 < u_3$ with $f'(u_j)$ negative for $j = 1, 3$ and positive for $j = 2$. Fife and McLeod (1977) showed that this model equation has as a solution a unique travelling wavefront, $u(x,t) = U(\xi)$ where $\xi = x - ct$. The front satisfies $U(-\infty) = u_3$ and $U(+\infty) = u_1$ and the velocity, c, is unique.

Based on this reaction–diffusion equation, Ermentrout and McLeod (1993) were able to prove that there is a travelling front solution to (7.1), $V(x,t) = v(\xi)$. The wavefront joins the two stable solutions, V_1 and V_3. They are able to show that it is unique as well as stable. One property that will be useful later is that the sign of the velocity c is the same as the sign of the integral:

$$I = \int_{V_1}^{V_3} f(V)\,\mathrm{d}V.$$

The function $f(V)$ is 'cubic-like', If I is positive, then the positive area exceeds the negative area. This occurs if the threshold of the sigmoid, S is low. The wavespeed is positive so that if the system is perturbed enough from the lower state, it will 'ignite' and switch the medium from low potential (firing rate) to high potential (firing rate). On the other hand, if the threshold is high, then I will be negative and the medium will switch from the high state to the low state. In particular, there is a unique value of the threshold for which the wave is 'frozen' and does not propagate. At this value, $I = 0$.

Idiart and Abbott (1993) considered (7.1) under the conditions that $h(t) = \mathrm{e}^{-t}$ and constructed an approximate solution by linearizing S around the middle fixed point. Ermentrout and McLeod (1993) solved the piecewise constant case, $S(u) = H(u - \theta)$ where H is the step function. Pinto (1997) also constructed a solution for general h and axonal delay. We will review these constructions in the next section. Chen *et al* (1997a) generalized the Ermentrout and McLeod results to include models of the form:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = F\left(\int_{-\infty}^{\infty} w(x-y)g[u(y,t)]\right)P(u) - C(u).$$

Choosing $P(u) = 1-u$, $g(u) = C(u) = u$, we obtain the continuum analogue of the spike-averaged system (3.7). This result was used to prove the existence of travelling spindle waves in a biophysically based model for the reticular nucleus of the thalamus (Chen *et al* 1997b).

7.1.1. Piecewise linear models. We now explicitly construct wavefronts for a piecewise constant model:

$$u(x,t) = \int_{-\infty}^{t} h(t-s)\,\mathrm{d}s \int_{-\infty}^{\infty} w(x-y)H(u(y,s)-\theta)\,\mathrm{d}y \tag{7.2}$$

where H is the step function and θ is the threshold. We look for a solution $u(x,t) = U(z)$ where $z = x - ct$ is the travelling coordinate. Since the wave is translation invariant, we fix the origin by requiring that $U(0) = \theta$. We want wavefronts which are monotone and satisfy $U(-\infty) = 1, U(\infty) = 0$. Using the fact that the step function is 0 for $z > 0$ by our choice of the origin, the travelling solution satisfies:

$$U(z) = \int_{z}^{\infty} h\left(s - \frac{z}{c}\right)\int_{sc}^{\infty} w(y)\,\mathrm{d}y\,\mathrm{d}s.$$

The condition that $U(0) = \theta$ determines the velocity, c:

$$\theta = \int_{0}^{\infty} h(s)\int_{cs}^{\infty} w(y)\,\mathrm{d}y\,\mathrm{d}s. \tag{7.3}$$

For a given PSP function h and weight function, w, we need only evaluate these integrals. Since the weight function and the PSP function have been normalized, it is clear that when $\theta = \frac{1}{2}$, that $c = 0$. Furthermore, the velocity increases as the threshold decreases with an infinite positive velocity as $\theta \to 0$ and an infinite negative velocity as $\theta \to 1$. Finally, uniqueness of the velocity is also obvious from the monotonicity of this function of c. Thus for each threshold between 0 and 1, there is a unique velocity, c. Pinto (1997) extended this theory to the case in which there are axonal delays. That is, (7.2) is replaced by:

$$u(x,t) = \int_{-\infty}^{t} h(t-s)\,\mathrm{d}s \int_{-\infty}^{\infty} w(x-y)H(u(y,s-|x-y|/v)-\theta)\,\mathrm{d}y. \tag{7.4}$$

Here v is the axonal velocity and $|x-y|/v$ is the conduction delay. As $v \to \infty$ we recover the original model. Since the axonal conduction velocity is two orders of magnitude faster

than typical cortical waves, one is justified in ignoring the conduction delays. However, for waves arising in low-threshold models, this velocity does have an effect. Pinto found that if $|c| < v$, then

$$\theta = \int_0^\infty h(s) \int_{\frac{c}{|v|-c}}^\infty w(y)\,\mathrm{d}y\,\mathrm{d}s.$$

Note that as c approaches v from below, the integral vanishes, and $\theta = 0$. Similarly, as $c \to -v^+$, $\theta \to 1$. Thus, the velocity is pinned between $-v$ and v.

In another calculation, Pinto and also Chen *et al* (1997b) considered the equation:

$$\frac{\partial u}{\partial t}(x,t) = AH\left(\int_{-\infty}^\infty w(x-y)u(y,t)\,\mathrm{d}y - \theta\right)(1-u(x,t)) - Bu(x,t) \tag{7.5}$$

where A, B are positive parameters. Recall that this model arises when averaging is done in a full ionic channel model. The velocity is found by solving:

$$\theta = \frac{A}{A+B}\left(\frac{1}{2} + \int_0^\infty w(z)\mathrm{e}^{-\frac{A+B}{c}z}\,\mathrm{d}z\right).$$

Idiart and Abbott (1993) adapted a similar approach and obtained the velocity for an equation of the form:

$$\tau u_t = -u + \int_{-\infty}^\infty w(x-y)S(u(y,t-|x-y|/v))\,\mathrm{d}y.$$

They showed that if $S'(\theta) = g$ and the slope of the front at threshold is K, then for infinite conduction velocity ($v \to \infty$)

$$c = \left(\frac{gKw_2}{2\theta\tau}\right)^{\frac{1}{2}}$$

where

$$m_2 = \int_{-\infty}^0 w(y)y^2\,\mathrm{d}y.$$

They did not provide a formula for the slope, K, which must still be determined, but their observation is useful; the slope of the front and the velocity of the wave are intimately related.

7.2. Travelling pulses

Travelling fronts are not particularly interesting in biological applications, since cells or layers of cells do not stay at the excited state forever. Rather, additional *slow* processes gradually bring the neurons back to rest. There are two possible sources for the repolarization of the network: (i) synaptic inhibition, (ii) slow ionic processes within the cell such as adaptation. Normal synaptic inhibition is fast and strong and thus will usually disrupt any type of wave propagation. In cortical slice preparations (Chagnac-Amitai and Connors 1989, Golomb and Amitai 1997) the inhibition has to be blocked or dramatically reduced pharmacologically before any propagation can occur. Adaptation is the likely mechanism for the termination of propagating waves. We will now look at several examples of the analysis of travelling pulses. There are two different ways to approach the problem. Amari (1977) wrote down a piecewise constant model and then constructed a solution. Pinto (1997) constructed solutions in general by assuming a slow local negative feedback that can be viewed as adaptation.

7.2.1. Amari's analysis. Amari (1977) considered a network in which there are excitatory–excitatory connections and inhibitory–excitatory connections, but the excitatory–inhibitory connections are local. Letting u_e and u_i denote the respective potentials of the excitatory and inhibitory populations, Amari considered:

$$\frac{\partial u_e}{\partial t} = -u_e + w_{ee}(x) * H(u_e) - w_{ie}(x) * H(u_i) + P_e \tag{7.6}$$

$$\frac{\partial u_i}{\partial t} = -u_i + w_{ei} H(u_e) + P_i \tag{7.7}$$

where $w(x) * f$ is the convolution of w with f. He set the thresholds to the step functions to be 0 and then added inputs to the model to adjust the bias. One could easily have set the inputs to zero and have the threshold as a parameter.

Amari sought travelling wave solutions with velocity c that represent propagating regions of excitation where the excitatory population is positive for a length, a. He showed that the length of the excited region and the velocity of the waves are found by solving:

$$0 = \int_0^\infty \mathrm{e}^{s/c} K(s)\,\mathrm{d}s + cP_e$$

$$0 = \int_0^\infty \mathrm{e}^{-s/c} K(s+a)\,\mathrm{d}s + cP_e$$

where

$$K(s) = \int_0^a w_{ee}(s-s')\,\mathrm{d}s' - \int_{s_1}^{s_2} w_{ie}(s-s')\,\mathrm{d}s'$$

and

$$s_1 = -c \ln \frac{w_{ei}(\mathrm{e}^{-a} - 1)}{P_i}$$

$$s_2 = a + c \ln \left(1 + \frac{P_i}{w_{ei}}\right).$$

Remarks.

(1) The interval (s_1, s_2) is where the inhibitory layer is excited.

(2) If the inhibitory population is positively stimulated, $P_i > 0$, then there are no propagating waves since s_1 is undefined.

(3) There is no difference in the time constants between inhibition and excitation in this model. For continuous models, it is very difficult to get travelling wave pulses unless there are big differences in the time constants. Rinzel and Keller (1973) found travelling pulses for the piecewise constant Fitzhugh–Nagumo model and found that they have no need for a big difference in time constants, but in continuous versions of the same model, a big difference in time constants is required. Thus, the piecewise constant model seems to be a special case.

Pinto obtained a result like Amari's in a simpler system in which the inhibition is purely linear and the effect of the excitation on the inhibition is also linear. Pinto used a shooting argument after converting the equations to a boundary value problem. He showed that there are *two* velocities which solve the equations; one narrow and slow and one wide and fast. This is typical of the results found for the reaction–diffusion models; in these models the slow wave is unstable and the fast wave is stable. Based on numerical calculations, Pinto arrived at the same conclusion.

As a final example of the usefulness of the piecewise constant models, we can also consider the models that arise from averaging of ionic equations:

$$\frac{\partial s}{\partial t}(x,t) = AH(w * s - bz - \theta)(1-s) - s \tag{7.8}$$

$$\tau \frac{\partial z}{\partial t}(x,t) = BH(w * s - bz - \theta)(1-z) - z \tag{7.9}$$

where z is the adaptation and s is the synaptic activity. Note that in a sense, this is the simplest model of all since the same nonlinearity modulates the rates of the synaptic drive and the adaptation. It is easy to verify that if there is a travelling pulse solution to this with velocity c and width, a, it must satisfy the two equations:

$$\theta = \int_{-a}^{0} w(y)s_1(y)\,\mathrm{d}y + \int_{-\infty}^{-a} w(y)s_2(y)\,\mathrm{d}y$$

$$\theta = \int_{-a}^{0} w(y+a)s_1(y)\,\mathrm{d}y + \int_{-\infty}^{-a} w(y+a)s_2(y)\,\mathrm{d}y - \frac{bB}{B+1}(1 - \mathrm{e}^{-\frac{B+1}{c\tau}a}).$$

The functions s_1, s_2 are defined by:

$$s_1(y) = \frac{A}{A+1}(1 - \mathrm{e}^{\frac{A+1}{c}y})$$

$$s_2(y) = \frac{A}{A+1}(1 - \mathrm{e}^{-\frac{A+1}{c}a})\mathrm{e}^{\frac{y+a}{c}}.$$

7.2.2. Singular perturbation construction of solutions. We now return to the general model:

$$u_t = -u + w * f_1(u) - f_2(v) \tag{7.10}$$

$$v_t = \epsilon(-av + f_3(u)) \tag{7.11}$$

which can be regarded as an excitatory neural network with slow adaptation. Alternatively, we could also consider

$$u_t = AF(w * u - bv)(1-u) - u \tag{7.12}$$

$$v_t = \epsilon[BF(w * u - bv)(1-v) - v]. \tag{7.13}$$

The idea is to construct travelling waves from these equations exploiting the slowness of the second variable. Casten *et al* (1975) applied a technique similar to this to a version of the Hodgkin–Huxley equations. If $\epsilon = 0$ then v is just a constant and the travelling wavefront results that we discussed above allow us to create a front joining the unexcited and excited states. The velocity is parametrized by the constant value of the recovery variable, v which acts like a threshold.

In figure 9, the phase plane of (7.10), (7.11) is shown when there are no spatial interactions. There is a unique fixed point and if the initial conditions are sufficiently large, there is a large amplification before decay to rest. This is the hallmark of an excitable medium and thus if excitation is allowed to spread, we expect travelling pulses to exist. Note that the nullclines have the classic shape for excitable systems; the excitatory nullcline is kinked and has three branches, the two outer branches and an inner branch.

We follow the method used in Pinto (1997) with the obvious generalizations. (Pinto has $f_3(u) = u, a = 0$, and $f_3(v) = bv$.) Introduce the travelling coordinate, $\xi = x - ct$ so that (7.10) becomes

$$-cu_\xi = -u + w * f_1(u) - f_2(V) \tag{7.14}$$

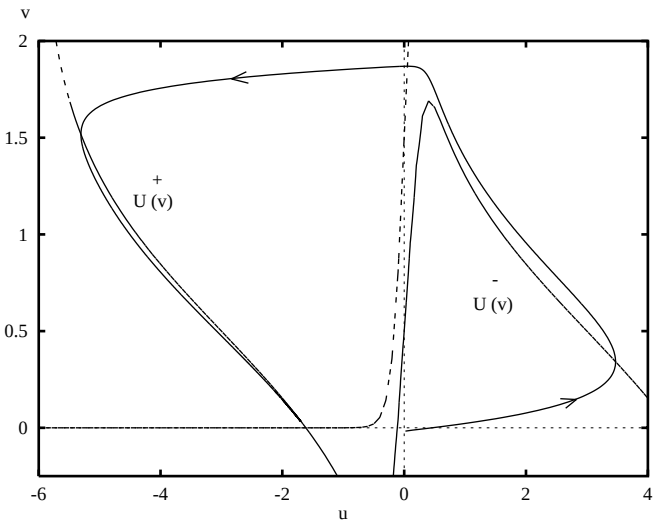

Figure 9. Phase-plane of an excitable system showing representative trajectories. The u nullcline is cubic-like with two stable branches, $U^{\pm}(v)$.

where V is a constant. This is the 'inner' equation. Let $\eta = \epsilon\xi$ so that we obtain in the limit of $\epsilon \to 0$:

$$-c\partial v_\eta = (-av + f_3(U^{\pm}(v))) \equiv G^{\pm}(v) \tag{7.15}$$

where $U^{\pm}(v)$ are the two outer branches of the solution to

$$0 = -u + f_1(u) - f_2(v)$$

(see figure 9).

Figure 10 shows a travelling wave solution in the moving coordinates. Initially, v is at its resting value, $v = V_{\text{rest}}$. From the theoretical results of section 7.1, we know that there is a unique travelling wavefront with velocity, c to (7.14) that switches the excitatory network from rest, $U_{\text{rest}} = U^-(V_{\text{rest}})$ to the excited state, $U^1_{\text{exc}} = U^+(V_{\text{rest}})$. We turn to the outer equation. From the phase-plane picture, $G^+(V)$ is strictly positive so that along this curve, as η decreases, $v(\eta)$ increases. From the nullclines, we see that u is also decreasing in this regime. We must determine when u makes the jump to the left branch. Suppose that this happens at the value $v = V_{\text{jump}}$. Then the inner equation will be used to make the jump from the new excited state, $U^2_{\text{exc}} = U^+(V_{\text{jump}})$ to the inhibited state, $U_{\text{in}} = U^-(V_{\text{jump}})$. The velocity of this backward front must match the velocity of the forward front. Under certain technical assumptions, there is a unique value of V_{jump} for which the velocities match. Thus, the velocity of onset of excitation determines where the jump occurs. We finish the construction by considering the outer equation with $G^-(v)$ and initial condition, $v = V_{\text{jump}}$. Since $G^-(v)$ vanishes at $v = V_{\text{rest}}$ and is negative for $v > V_{\text{rest}}$, the solution to this first-order equation with $v = V_{\text{jump}}$ as the initial data will decay to V_{rest} as $\eta \to -\infty$ as required. This completes the construction of the singular travelling pulse. The velocity is determined solely from the up-jump and all of the rest of the dynamics follows.

The same approach as that used here could also be applied to model (7.12), (7.13) that arise from averaging. We could then compare the results with the simulations of Golomb and Amitai (1997) in which a full biophysical model is derived.

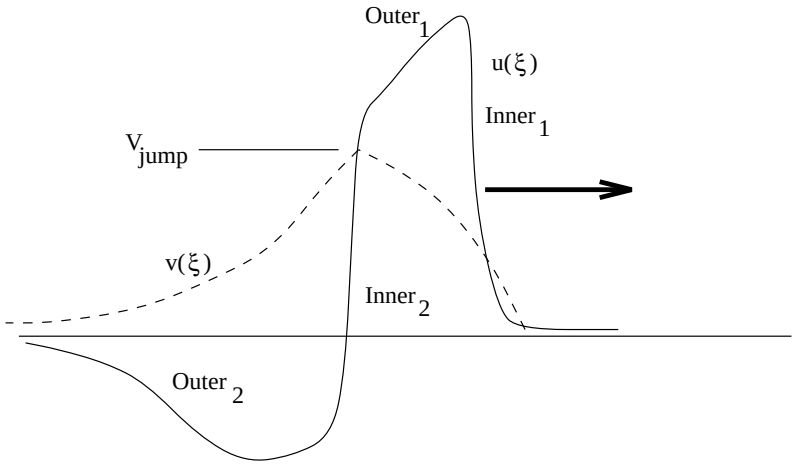

Figure 10. The singular travelling wave solution. Full curves are excitatory activity and the broken curve is the recovery variable. Inner and outer parts of the solution are labelled.

7.2.3. A different singular mechanism. Sakaguchi (1988) studied the discrete firing rate analogue of (7.6), (7.7):

$$\frac{\mathrm{d}u_j}{\mathrm{d}t} = -u_j + f_e\left(\sum_k w_{jk}u_k - bv_j\right)$$
$$\frac{\mathrm{d}v_j}{\mathrm{d}t} = -v_j + f_i(cu_j - dv_j)$$

when the mechanism for excitability is a saddle–node loop. That is, rather than a single fixed point and widely disparate timescales, Sakaguchi studied the network when there are three fixed points and the local phase-plane is as shown in figure 1(*b*). He considered local (nearest-neighbour) coupling and found that there are travelling pulse waves. We can use a different singular perturbation scheme to compute the velocity of these waves. Recall from section 3.2.2 that if we assume weak coupling with each local system near threshold, we can reduce this network to a scalar network with just a phase variable. Using (3.8) and assuming nearest-neighbour coupling, the model equations become:

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = 1 - \cos\theta_i - I(1+\cos\theta_i) + g(\theta_i)[\delta(\theta_{i-1}-\pi) + \delta(\theta_{i+1}-\pi)]$$

where

$$g(\theta) = 2\arctan\left(\tan\frac{\theta}{2} + w\right) - \theta$$

and w is the coupling strength. The velocity of an excitation wave can be easily computed. Suppose that cell i is at rest and its neighbour has just fired. Then the new phase is

$$\theta_{\mathrm{new}} = g(\theta_{\mathrm{rest}}) + \theta_{\mathrm{rest}}.$$

If this is past threshold, $\arccos\frac{1-I}{1+I}$, then the phase will increase past π at which point it will cause the next cell in the chain to fire. Thus, the velocity is the reciprocal of this integration time:

$$c^{-1} = \int_{\hat{\theta}}^{\pi} \frac{\mathrm{d}\theta}{1 - I - (1+I)\cos\theta}$$

$$\hat{\theta} = 2\arctan\left(\tan\frac{\theta_{\text{rest}}}{2} + w\right)$$

$$\theta_{\text{rest}} = -\arccos\frac{1-I}{1+I}.$$

As the coupling strength gets larger, the velocity tends to infinity and if the coupling strength is too weak, the cell will never fire. These ideas can be used to look at a continuum network of neurons, however, the explicit calculation of velocity cannot usually be done.

7.3. Solitary standing pulses

There has been a great deal of recent interest in the modelling of results from delayed-response tasks. Animals are required to maintain a memory of the position of a stimulus for a short time after the stimulus has disappeared. What makes these experiments interesting from a modelling point of view is that physiological recordings in a region of the brain called the prefrontal cortex showed that spatially localized groups of neurons fired during the recall task and then stopped firing once the task was finished (Fuster and Alexander 1971, Funahashi *et al* 1989). This has led to a theory that recurrent excitation coupled with lateral inhibition keeps localized groups of neurons active. There have been a number of neural network models for this phenomena (Wilson and Cowan 1973, Amari 1977, Camperi and Wang 1997). We will describe these models and their solution by again either constructing explicit solutions or using perturbation methods.

Wilson and Cowan used a two-layer network model with excitatory and inhibitory neurons interacting in a spatially distributer manner. When the local dynamics is like figure 9 (that is excitable) but the inhibition has a larger spread than the excitation, a local area of excitation develops but does not spread. The strong self-excitation keeps the centre excited but the lateral inhibition prevents spread. Ellias and Grossberg (1975) also used numerical solutions to show the same type of behaviour. This was interpreted to be the analogue of short-term memory. More recently, Salinas and Abbott (1996) interpreted this local activity as a representation of the retinal position in higher visual cortex.

A one-layer model in which the spatial interactions are not of one sign (such as the 'Mexican-hat' in figure 8(*c*)) can be regarded as an approximation of the multilayer system. We will consider:

$$\frac{\partial u}{\partial t}(x,t) = -u(x,t) + \int_\Omega w(x-y)S(u(y,t))\,\mathrm{d}y + I(x,t) \tag{7.16}$$

where, $w(x)$ has the form shown in figure 8(*c*) or 11.

Before we turn to the analysis of this model, we briefly comment on the implicit assumptions of such an interaction function. Consider the system (7.10), (7.11) with spatial interactions between all layers:

$$u_t = -u + w_1 * f_1(u) - w_2 * f_2(v) \tag{7.17}$$

$$v_t = \epsilon(-v + w_3 * f_3(u)). \tag{7.18}$$

Suppose that $f_3(u) = f_1(u)$, $f_2(v) = v$, and inhibition is fast, that is, $\epsilon \gg 1$. Then from (7.18),

$$v = w_3 * f_3(u).$$

Substitution of this into (7.17) yields

$$u_t = -u + w * f_1(u)$$

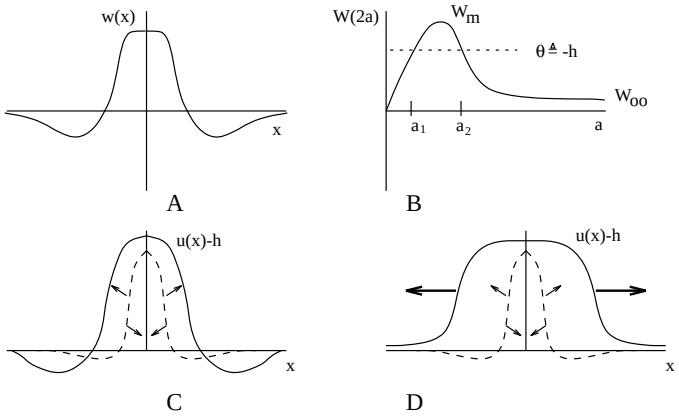

Figure 11. Construction of the solitary pulse in the piecewise linear model. (*a*) The lateral-inhibitory kernel, $w(x)$. (*b*) Integral of $w(x)$. Horizontal line shows the threshold or stimulus. Intersections correspond to pulse widths. (*c*) Unstable pulse (broken) acts as a separatrix between the stable pulse and the rest state. (*d*) Unstable pulse acts as a separatrix between a wavefront and the rest state.

where

$$w(x) = w_1(x) - w_2(x) * w_3(x).$$

So the one-layer model can be regarded as a two-layer model in which the inhibition is linear and very fast.

7.3.1. Amari's model. If we choose f_1 to be the step function in equations (7.17) and (7.18) then we obtain the model that Amari analysed. We now turn to that analysis. Consider:

$$\frac{\partial u}{\partial t}(x,t) = -u(x,t) + \int_\Omega w(x-y)H(u(y,t))\,\mathrm{d}y + h + s(x,t) \tag{7.19}$$

where $w(x)$ is a symmetric integrable function, $H(u)$ is the step function, and h is a homogeneous and $s(x,t)$ is a spatially patterned input pattern. The domain Ω is either the whole real line or a circular domain of length L. Since the circular domain is equivalent to a periodic pattern on the whole line, we will consider only the infinite domain. Amari restricted the function $w(x)$ to be positive for x near 0 and negative for x sufficiently large and have exactly one positive 0 (see figure 11(*a*)).

We suppose that there is no spatial input, $s(x,t) = 0$ and look for solutions to this model. As we saw earlier, there can be travelling wave solutions that represent fronts of excitation. (Indeed, simply let $v = u - h$ be a new variable and we regain the model (7.2) where $\theta = -h$.) Amari found four different types of solutions: (i) 0-solutions in which $u < 0$ for all x, (ii) ∞-solutions in which the entire medium is excited, $u > 0$, (iii) a-solutions in which the medium is excited over an interval, $(-a, a)$; (iv) (a, b)-solutions which are b-periodic with an excited region of length a. Solitary peak solutions (a-solutions) can be arbitrarily translated to be centred at the origin. Let

$$W(x) = \int_0^x w(y)\,\mathrm{d}y$$

let W_m be the maximum of $W(x)$, and let $W_\infty = W(\infty)$. The the behaviour of the system depends only on the relative sizes of W_m, W_∞, and h. For the periodic solutions with period,

b, let

$$w_b(x) = \sum_{n=-\infty}^{\infty} w(x + nb)$$

and let

$$W_b(x) = \int_0^x w_b(y)\,\mathrm{d}y.$$

Theorem 1 (Amari 1977). There are three cases with the indicated solutions occurring in the specified intervals of h. For the a-solutions, $a > 0$ satisfies $W(2a) + h = 0$ (see figure 11(b)). There may be 0, 1, or 2 solutions for a given value of h. An a-solution is stable if and only if $w(2a) < 0$. The infinite family of periodic solutions, (a, b) must satisfy $W_b(2a) + h = 0$ and are stable if and only if $w_b(2a) < 0$.

(1) $2W_\infty > W_m$

$$h < -2W_\infty \to \{0\}$$
$$-2W_\infty < h < -W_m \to \{0, \infty\}$$
$$-W_m < h < -W_\infty \to \{0, a_1, a_2, \infty\}$$
$$-W_\infty < h < 0 \to \{0, a, \infty\}$$
$$0 < h \to \{\infty\}.$$

(2) $W_m > 2W_\infty > 0$

$$h < -W_m \to \{0\}$$
$$-W_m < h < -2W_\infty \to \{0, a_1, a_2\}$$
$$-2W_\infty < h < -W_\infty \to \{0, a_1, a_2, \infty\}$$
$$-W_m < h < 0 \to \{0, a, \infty\}$$
$$0 < h \to \{\infty\}.$$

(3) $W_\infty < 0$

$$h < -W_m \to \{0\}$$
$$-W_m < h < 0 \to \{0, a_1, a_2\}$$
$$0 < h < -2W_\infty \to \{(a, b)\}$$
$$-2W_\infty < h \to \{\infty\}.$$

Remarks.

(1) The constant solutions are stable when they exist.

(2) The a-solutions occur either individually or in pairs. When individually, they are unstable. When there are two of them, the wider one is stable and the thinner unstable.

(3) In the cases in which there is an unstable a-solution along with a 0 and ∞ solution, the standing pulse acts as a separatrix between the 0 solution and the ∞ solution. Any initial data which lies above this will develop into a travelling wave that eventually excites the whole medium (see figure 11(d)). Similarly, in the case where there are two roots to $W(x) + h = 0$, the smaller root acts as a separatrix; any initial data that lies above this will develop into the stable solitary pulse (see figure 11(c)).

(4) If $W(x)$ has finite support, that is, it vanishes for large enough x, say, for $|x| > C$ then there can be many independent solitary peaks which do not interact as long as their distance is at least C. Similarly, if W has 'wiggles' in it, there may be more than two solitary peaks.

(5) If $W_\infty > 0$, then for small enough b, there are only unstable periodic solutions since $w_b(x)$ is always positive.

Since the ideas of the proof of the theorem are quite easy to sketch and since the same methods can be used to generalize the types of weight functions, we briefly describe the solutions. Suppose that $u(x) > 0$ in the interval $(-a, a)$. Then

$$u(x) = h + \int_{-a}^{a} w(x - y)\,\mathrm{d}y.$$

Since $u(a) = 0$ by definition, we obtain

$$0 = h + \int_{-a}^{a} w(a - y)\,\mathrm{d}y = h + \int_{0}^{2a} w(y)\,\mathrm{d}y.$$

The rest of the theorem follows from the analysis of the roots of this equation.

Amari proved stability by looking at the movement of the boundaries of the pulses. A standard stability analysis using linearization and exploiting the fact that the formal derivative of the step function is the Dirac delta function leads to the same result.

Amari then assumed that there is weakly inhomogeneous input of the form, $s(x, t) = \epsilon S(x)$. He showed that the excited region moves to the peak of the inhomogeneity. We will revisit this point below using perturbation methods.

7.3.2. Generalization of the Amari model. Kishimoto and Amari (1979) proved the existence of solitary pulse solutions for (7.16) using a fixed point theorem. The idea of the proof is the find two values, $\theta_1 < \theta_2$ where $H(u - \theta_1) \leqslant S(u) \leqslant H(u - \theta_2)$ such that the stable pulse solution exists for the piecewise constant functions. Using these 'upper' and 'lower' solutions they prove existence. Kopecz and Schoener (1995) used a variant Amari's equations to model saccadic motor planning by integrating visual inputs with the localized spatial attractors in the Amari system.

Pinto (1997) reconsidered the single-layer Amari model as a two-layer model and looked at the stability of the stationary peaks of this model. In particular, he considered:

$$\frac{\partial u_1}{\partial t} = -u_1 + w_1(x) * H(u_1) - w_2(x) * u_2 + h_1 \tag{7.20}$$

$$\tau \frac{\partial u_2}{\partial t} = -u_2 + w_3(x) * H(u_1) + h_2. \tag{7.21}$$

Note that this is equivalent to (7.17), (7.18) with linear inhibition and with $\epsilon \equiv \tau^{-1}$. Time-independent solutions satisfy:

$$u(x) = w(x) * H(u(x)) + h$$

where

$$w(x) = w_1(x) - w_2(x) * w_3(x)$$

and

$$h = h_1 - \bar{w}_2 h_2.$$

Here $\bar{w}_2$ is the total integral of $w_2(x)$. Thus, the analysis of the stationary states for this model is the same as Amari's analysis. The time constant of inhibition does not come into play for the existence. Pinto showed that the a-solutions are stable solutions to this extended model if $w(2a) < 0$ (Amari's original condition) and

$$\frac{w_1(0) + w_1(2a)}{|w(0) - w(2a)|} < 1 + 1/\tau.$$

If this latter condition is violated, then stability is lost via imaginary eigenvalues and we expect to see a transition to either travelling pulses or perhaps breathing solutions. Instability of localized pulses as the time course of inhibition is slowed was studied extensively for reaction–diffusion equations by Nishiura and Mimura (1989). They found breathers, local pulses whose widths oscillate.

We can formally analyse the breathing solutions by following the width of the pulse as a function of time. We integrate the inhibitory equation and arrive at the scalar integro-differential equation:

$$u_t(x,t) = -u(x,t) + w_1(x) * H(u(x,t)-\theta) - L\{w_4(x) * H(u(x,t)-\theta)\}$$

where $w_4(x) = w_2(x) * w_3(x)$ and L is the integral operator:

$$Lf(t) = \frac{1}{\tau}\int_{-\infty}^{t} \mathrm{e}^{-(t-s)/\tau} f(s)\,\mathrm{d}s.$$

We suppose that the pulse remains symmetric about the origin and has width, $a(t)$. By definition, $u(a(t),t) = \theta$ the threshold. Differentiating this, we see that

$$u_x(a,t)\frac{\mathrm{d}a}{\mathrm{d}t} + u_t(a,t) = 0.$$

We approximate u_x by its steady state value and evaluate everything at $x = a$. This leads to the following evolution equation for $a(t)$:

$$\kappa(a)\frac{\mathrm{d}a}{\mathrm{d}t} = -\theta + W_1(2a) - L\{W_4(2a)\}$$

where

$$W_1(x) = \int_0^x w_1(y)\,\mathrm{d}y$$
$$W_4(x) = \int_0^x w_4(y)\,\mathrm{d}y$$
$$\kappa(a) = w_1(0) - w_1(2a) - (w_4(0) - w_4(2a)).$$

The function, $\kappa(a)$ is just $-u_x(a(t),t)$ evaluated at the steady state. Since L is just an exponential operator, we can convert this problem into a two-dimensional differential equation:

$$\kappa(a)\frac{\mathrm{d}a}{\mathrm{d}t} = -\theta + W_1(2a) - z \tag{7.22}$$

$$\tau\frac{\mathrm{d}z}{\mathrm{d}t} = -z + W_4(2a). \tag{7.23}$$

For any choice of parameters, this system can be analysed by standard phase-plane methods. In particular, if the threshold is chosen appropriately, then there are two fixed points. One of these is an unstable saddle point and the other is stable if τ is small enough. As τ increases, a Hopf bifurcation occurs. Unfortunately for all the examples that we have considered, the bifurcation is subcritical and only unstable periodic orbits bifurcate. Once past the critical time constant, we numerically find the appearance of travelling pulses.

Pelinovsky and Yakhno (1996a, b) performed a similar stability analysis on an analogous model:

$$\frac{\partial u_1}{\partial t} = -u_1 + F(w_1 * u_1 - \alpha u_2) \tag{7.24}$$

$$\tau\frac{\partial u_2}{\partial t} = -u_2 + w_2 * u_1. \tag{7.25}$$

They are interested in the stability of the stationary peaked solution as a function of the parameter τ. They obtained results similar to Pinto. They also showed some numerical simulations which exhibit the aforementioned breathing. They performed an analysis of the pulse-width equations similar to the one given above. They also considered systems in which the interaction functions have more than one 'bump'. They found m-pulse solutions where $m \geqslant 1$ which can stably occur when the weight functions have several sign changes. They studied the stability of the m-pulse solutions and found stable oscillatory structures for these as well. Their technique was somewhat different; they differentiated the first equation with respect to t and then used the second equation to reduce the system to a second-order differential equation for $\Phi(x) = w_1(x) * u_1(x) - \alpha u_2$. They obtained:

$$\tau \frac{\partial^2 \Phi}{\partial t^2} + (\tau + 1)\frac{\partial \Phi}{\partial t} + \Phi = w_3 * F(\Phi) + \tau w_1 * \left[F'(\Phi)\frac{\partial \Phi}{\partial t} \right] \tag{7.26}$$

where $w_3(x) = w_1(x) - \alpha w_2(x)$. They showed that the number of standing pulse solutions is related to the number of sign changes of the effective interaction function, $w_3(x)$.

7.3.3. Singular perturbation construction of solitary pulse solutions. The key to obtaining solitary pulses is the existence of long-range inhibition coupled with the 'excitable' local dynamics. This suggests that perhaps one could perform a singular perturbation calculation using the ratio if the excitatory spread to inhibitory spread as a parameter. This idea was successfully applied to reaction diffusion equations (Nishiura and Mimura 1989, Ermentrout *et al* 1984). Pinto (1997) recently used a singular perturbation technique applied to the time-independent equations:

$$u_1(x) = \int_{-\infty}^{\infty} w_1(x-y) f_1(u_1(y))\,\mathrm{d}y - \epsilon \int_{-\infty}^{\infty} w_2(\epsilon(x-y)) f_2(u_2(y))\,\mathrm{d}y \tag{7.27}$$

$$u_2(x) = \int_{-\infty}^{\infty} w_3(x-y) f_1(u_1(y))\,\mathrm{d}y. \tag{7.28}$$

Let

$$\bar{w_j} = \int_{-\infty}^{\infty} w_j(y)\,\mathrm{d}y$$

and define $U^{\pm}(z)$ to be the left and right branches of the solutions to

$$-u + \bar{w_1} f_1(u) - z = 0.$$

The ideas are similar to those used in the construction of the travelling wave. We follow Pinto's analysis here. Since the medium is symmetric, we need only solve the equations on the positive x-axis. We want the pulse to look like figure 12. There is one discontinuous outer solution and one inner solution. The inner equation is found by setting $\epsilon = 0$. Thus u_1 satisfies:

$$u_1(x) = \int_{-\infty}^{\infty} w_1(x-y) f_1(u_1(y))\,\mathrm{d}y - C \tag{7.29}$$

where C is an unknown constant. From the results of Ermentrout and McLeod (see section 7.1), it follows that there is a unique value of C for which there is a transition front *with 0 velocity* with u_1 jumping from $U^+(C)$ to $U^-(C)$. This value of C satisfies:

$$\int_{U^-(C)}^{U^+(C)} \bar{w_1} f_1(u) - u - C\,\mathrm{d}u = 0$$

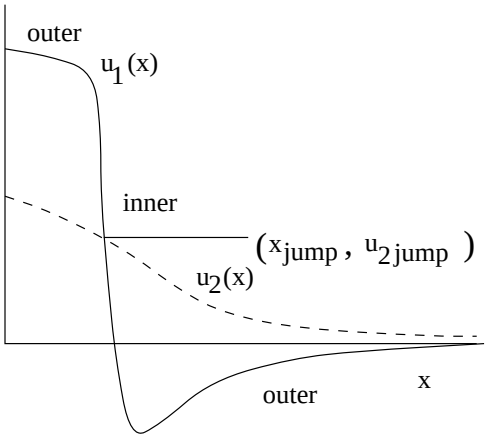

Figure 12. Construction of the singular standing pulse.

that is, the area of the positive and negative parts of the 'cubic', $-u + \bar{w_1} f_1(u) - C$, exactly cancel.

The outer equation is obtained by rescaling space by $1/\epsilon$

$$u_1(X) = \bar{w_1} f_1(u_1(X)) - w_2(X) * f_2(u_2(X)) \tag{7.30}$$

$$u_2(X) = \bar{w_3} f_1(u_1(X)). \tag{7.31}$$

The pulse width is determined by matching the inner and outer regions. Pinto showed that the value of X at which the jump occurs, x_{jump} must satisfy:

$$C = w_2(X) * f_2(u_2(X))(x_{\text{jump}})$$

where

$$u_2(X) = \bar{w_3} f_1(U^{\pm}(X)).$$

The choice of $U^{\pm}$ is determined by whether or not $|X| < x_{\text{jump}}$. Pinto showed that under fairly general circumstances such a solution can be found. He finally showed that the outer solution goes to the rest state as $X \to \pm\infty$. Thus, a pulse can be constructed in the singular limit of either local excitation or 'global' inhibition. Note that in the singular construction of the travelling wave (in section 7.2.2), the outer equation was an ordinary differential equation and not an integral equation as it is here. Thus, we were able to break the outer equation into separate pieces. With the integral equation, the outer equation must be solved globally.

7.3.4. A model for head-direction. Zhang (1996) is interested in modelling the head-direction cells. These cells have a remarkable property: they signal the position of the head independent of the location of the animal; the system is perfectly world centred and thus acts like a gyroscope, even in total darkness. His model lies on a ring and he wants to design a network which can follow a stimulus, that is, it can dynamically shift a localized peak of excitation. His starting point is a model like the one-layer lateral inhibition network (7.16). He wishes to find weight distributions which can dynamically shift the peak of the firing rate without disrupting its shape. One way to do this is to add a slow inhibitory process which will result in a travelling wave around the circle. However, this approach makes it difficult to control both the direction and velocity of the shift. Zhang's approach

is to dynamically modify the weight function in order to achieve the translation invariance. Let $w(x)$ be a fixed symmetric weight distribution and suppose that:

$$U(x) = \int_0^{2\pi} w(x-y)S(U(y))\,\mathrm{d}y$$

is a stable stationary peaked distribution of the firing rates. (For example, it could arise in the manner of Amari's model or as a symmetry-breaking bifurcation of a uniform state as in section 8.1.2; in Zhang, the latter mechanism holds.) Consider the dynamic weight function:

$$W(x;t) = w(x) + \gamma(t)w'(x)$$

where $w'(x)$ is the derivative of the weight function w and $\gamma(t)$ is the input to the system which causes the head-direction shift. The scalar network with this dynamic weight function is:

$$\tau\frac{\partial u}{\partial t}(x,t) = -u(x,t) + \int_0^{2\pi} W(x-y;t)f(u(y,t))\,\mathrm{d}y.$$

With this choice of weights, it is immediately clear that:

$$u(x,t) = U\left(x + \frac{1}{\tau}\int_0^t \gamma(s)\,\mathrm{d}s\right)$$

solves this equation, once one observes that

$$U'(x) = \int_0^{2\pi} w'(x-y)f(U(y))\,\mathrm{d}y.$$

Thus, the profile remains invariant with an instantaneous angular velocity $\omega = -\gamma(t)/\tau$. If $\gamma(t)$ is on briefly, then the peak distribution shifts to the new position and stops. Redish *et al* (1996) proposed a similar two-layer model for the head-direction system.

7.3.5. Response of a pulse system to weak, slowly varying input. We return to (7.16) to examine the response of the model to slowly changing small inputs. Recall that the solitary peak is translation invariant. Thus, small perturbations can be expected to move it by a small amount. Amari (1977) considered the dynamics of this in the case of the step function nonlinearity. He found that with small stationary inputs, the peak moves towards the maximum. He also showed that local maxima of inputs will not block the movement as long as they are not too large. This is intuitively appealing from a biological point of view since if this is a model for short-term memory, then the spatial excitation should occur where the largest stimulus appears. The role of external inputs in the form of noise has also been of interest to Camperi and Wang (1997) who are concerned with the drift in the peak induced because of random perturbations. They point out that if the drift is too great, then mechanism will not provide a very robust short-term memory.

With this motivation consider:

$$\frac{\partial u}{\partial t}(x,t) = -u(x,t) + \int_\Omega w(x-y)S(u(y,t))\,\mathrm{d}y + \epsilon I(x,\epsilon t).$$

The input is weak and slow. When $\epsilon = 0$, we assume the system has a peaked solution, $U(x)$ such that

$$U(x) = \int_\Omega w(x-y)S(U(y))\,\mathrm{d}y.$$

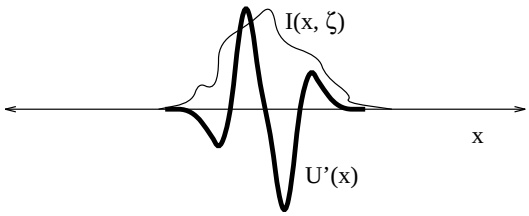

Figure 13. Effect of slowly varying input on the peak of the pulse. The pulse moves so that the integral of the product of its derivative (smooth curve) and the input, $I(x,\zeta)$, is zero.

Thus, we look for solutions of the form:

$$u(x,t) = U(x+\phi(\zeta)) + \epsilon r(x,\zeta) + \cdots$$

where $\zeta = \epsilon t$ is a slow timescale, r is orthogonal to $U'(x)$ and ϕ is a slowly varying spatial shift. Plugging this into the full equation, we obtain at the next order:

$$r(x,\zeta) - \int_\Omega w(x-y)S'(U(x+\phi))r(y,\zeta)\,\mathrm{d}y = -U'(x+\phi(\zeta))\frac{\mathrm{d}\phi}{\mathrm{d}\zeta} + I(x,\zeta).$$

The linear operator on the left has a one-dimensional nullspace spanned by $U'(x+\phi)$ and since U and w are symmetric, the operator is self-adjoint. Thus, there is a bounded solution for r is and only if the right-hand side is orthogonal to $U'(x+\phi)$ yielding an evolution equation for ϕ:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\zeta} = \frac{\int_\Omega U'(y+\phi)I(y,\zeta)\,\mathrm{d}y}{\int_\Omega U'(y)^2\,\mathrm{d}y} \equiv G(\phi,\zeta).$$

Thus, ϕ will move so as to make G vanish (see figure 13). In particular, suppose the input is stationary and symmetric. Then because of the antisymmetry of $U'(x)$, the pulse will move so that it is centred on stimulus. If the stimulus slowly varies, the pulse will move with it. Note that if the weight functions have compact support, then so will the pulse, and thus, a stimulus that is too far away will never interact with the pulse and it will not move. In section 8.1.2 we look at the response of a pulse to a rotating stimulus.

7.3.6. Two-dimensional pulses. The construction of pulses in two dimensions is somewhat more complicated, but can be done fairly simply by looking for radially symmetric solutions. These will be solutions if the weight functions have the same symmetry, a reasonable assumption for an isotropic network. For example, the construction of a pulse in the Amari model just means solving:

$$\theta = \int_0^a r\,w(r)\,\mathrm{d}r$$

for the pulse radius, a.

7.4. Continuous oscillatory networks

Consider the following continuous two-layer network which is the analogue of (4.5):

$$\frac{\partial u}{\partial t}(x,t) = -u(x,t) + f_e\left(\frac{a_{ee}}{\sigma_{ee}}\int_{-\infty}^{\infty} w((x-y)/\sigma_{ee})u(y,t)\,\mathrm{d}y - \frac{a_{ie}}{\sigma_{ie}}\int_{-\infty}^{\infty} w((x-y)/\sigma_{ie})v(y,t)\,\mathrm{d}y\right) \tag{7.32}$$

$$\tau \frac{\partial v(x,t)}{\partial t} = -v(x,t) + f_i\left(\frac{a_{ei}}{\sigma_{ei}} \int_{-\infty}^{\infty} w((x-y)/\sigma_{ei}) u(y,t)\,\mathrm{d}y \right.$$
$$\left. -\frac{a_{ii}}{\sigma_{ii}} \int_{-\infty}^{\infty} w((x-y)/\sigma_{ii}) v(y,t)\,\mathrm{d}y \right) \tag{7.33}$$

where σ_{jk} are the space constants of the interactions between layers. Suppose that the spatially homogeneous network has a stable limit cycle solution. By assuming that the space constants are small, it is possible to develop a perturbation for the spacetime behaviour of the phases of the oscillators in much the same way as Kuramoto did for reaction–diffusion equations. Ermentrout (1982) performed this calculation and found that the solution to (7.32) has the form, $(u(x,t), v(x,t)) = (U_0(t+\theta(x,t)), V_0(t+\theta(x,t)))$ where the phase evolves according to the equations:

$$\theta_t = 1 + A\theta_x^2 + B\theta_{xx} \tag{7.34}$$

where A, B depend on the space constants in the following manner:

$$2A = a_{ee}\sigma_{ee}^2 c_{ee} + a_{ie}\sigma_{ie}^2 c_{ie} + a_{ei}\sigma_{ei}^2 c_{ei} + a_{ii}\sigma_{ii}^2 c_{ii} \tag{7.35}$$
$$2B = a_{ee}\sigma_{ee}^2 d_{ee} + a_{ie}\sigma_{ie}^2 d_{ie} + a_{ei}\sigma_{ei}^2 d_{ei} + a_{ii}\sigma_{ii}^2 d_{ii}. \tag{7.36}$$

The coefficients c_{jk} and d_{jk} are related to the derivatives of the interaction functions, H computed for a pair of coupled oscillators in section 4.5.1:

$$c_{jk} = \langle w^2 \rangle H''_{jk}(0)$$
$$d_{jk} = \langle w^2 \rangle H'_{jk}(0)$$

where

$$\langle w^2 \rangle = \int_{-\infty}^{\infty} x^2 w(x)\,\mathrm{d}x.$$

Stability of any spatially diffusing patterns requires that the coefficient B is positive. For the example we looked at in section 4.5.1, the cross connections had better dominate. The spatial phase model (7.34) has been the subject of a great deal of study. In particular, if the coefficient B is small, then one has to go to higher order in the expansion. The result of this calculation is the Kuramoto–Shivashinsky equation:

$$\theta_t = 1 + A\theta_x^2 + B\theta_{xx} + C\theta_{xxxx}.$$

This equation admits spatio-temporal chaos. Thus, there is no reason why we should not expect to see the same sort of behaviour in neural networks.

7.5. Two-dimensional active media

In two dimensions, many of the analogues of one-dimensional structures are found when considered in radially symmetric coordinates. We saw this with the standing pulse. A pulse wave becomes an expanding pulse ring and the wavefront becomes a radially symmetric front. However, there are a number of patterns that are not trivial extensions of the one-dimensional patterns. The best known of these patterns are spiral waves. These are known to occur in reaction–diffusion equations (Keener and Tyson 1986, Tyson and Keener 1988, Winfree 1980, Greenberg and Hastings 1978). Keener and collaborators used singular perturbation methods (see Tyson and Keener 1988) to construct spiral waves when there is slow recovery. Such methods have not been applied to neural networks although there is probably no reason why they would not work, at least in the long wave approximation where the integral operators resemble diffusion. There have been a few numerical simulations

of spiral waves in locally active media. Milton *et al* (1993) found spiral waves in an integrate-and-fire model with localized coupling. In a similar model, Fohlmeister *et al* (1995) found the same behaviour in a two-dimensional network of simulated neurons using the 'spike-response' model. They asserted that these patterns are related to hallucinations (see section 8.4). Destexhe (1994) studied a two-layer network of the form:

$$C\frac{\mathrm{d}X_i}{\mathrm{d}t} = -g_L(X_i - V_L) - \sum_k g_{ki}^{EE} F_E(X_k(t-\delta))(X_i - V_E) - \sum_l g_{li}^{IE} F_I(Y_l(t-\delta))(X_i - V_I)$$

$$C\frac{\mathrm{d}Y_i}{\mathrm{d}t} = -g_L(Y_i - V_L) - \sum_k g_{ki}^{EI} F_E(X_k(t-\delta))(Y_i - V_E) - \sum_l g_{li}^{II} F_I(Y_l(t-\delta))(Y_i - V_I)$$

where the connections are locally made and X, Y represent the excitatory and inhibitory populations respectively. He found spiral waves as well as spatio-temporal chaos in the networks with nearest-neighbour and second-nearest-neighbour connections.

One obvious question is whether spiral wave activity has any relevance to biology. Petsche *et al* (1974) observed rotating waves with a frequency of about 8 Hz in rabbit cortex that was treated with topical penicillin (which causes seizure-like activity.) Rotating and expanding two-dimensional waves have been observed in the developing retina (Meister *et al* 1991). Spreading depression in retina (Leao) has the form of a spiral wave although this is thought to be due mainly to diffusion-driven extracellular potassium (Tuckwell and Miura 1978). In a recent experiment, Prectl *et al* (1997) observed rotating waves in the turtle cortex that are stimulus driven and not due to any epileptogenic agent.

The standard paradigm for the generation of spiral and rotating waves is to assume a locally excitable dynamics with short-range coupling. However, it is possible to couple locally oscillatory networks together to generate spiral waves. For example, Sakaguchi *et al* (1987) found spiral-like behaviour in a network of locally coupled phase models of the form:

$$\theta'_{i,j} = \omega + \sum_{\{i',j'\}\in \mathrm{nnbs}} H(\theta_{i',j'} - \theta_{i,j})$$

where $H(\theta) = \sin(\theta + \xi)$. Here the sum is over nearest neighbours. Paullet and Ermentrout (1994) proved that these solutions exist and are stable for H, an odd periodic function.

8. Bifurcation methods and neural networks

Perhaps the most general techniques that can be applied to the analysis of spatially distributed neural networks are those of bifurcation from a uniform state. These methods are useful more in getting an idea about how to choose parameters to obtain a particular type of behaviour than they are for global analysis of the network. The types of behaviour that can bifurcate from a uniform state are usually of four varieties: (i) new uniform states, (ii) spatially uniform periodic states, (iii) spatially periodic states, and (iv) spatially and temporally periodic states. For example, it is not possible to get a solitary peak or travelling solitary wave to bifurcate from a uniform state on the infinite line. Bifurcation methods were first applied to neural networks by Ermentrout and Cowan (1979a, b, 1980a, b) in a rather general fashion. There is a renewed interest in the onset of pattern formation in

network via loss of stability of homogeneous states. Ben-Yishai *et al* (1997) and Hansel and Sompolinsky (1997) recently examined a model for orientation tuning and used methods related to bifurcation theory to obtain waves and peaked solutions on a ring model. In a pair of papers, Tass (1995, 1997) re-examined the model of Ermentrout and Cowan (1979b) in much greater detail and computed the normal forms for the bifurcations. Bressloff and collaborators recently analysed the role that dendrites can take in modulating pattern formation. Finally, the greatest explosion of work in spontaneous pattern formation has been in the area of models for the development of neural maps (von der Malsburg 1973, Oja 1982, Kammen and Yuille 1988, Linsker 1988, MacKay and Miller 1990, Miller 1994, Krekelberg and Taylor 1996).

In this section, we will review the methods and results concerning spontaneous symmetry breaking in neural models. In particular, we will focus on some simplified models such as that proposed by Hansel and Sompolinsky. We start with a one-layer model which illustrates almost everything that can happen in these models. After some examples, we look at two-spatial dimensions and their added complexity.

8.1. Single-layer models

Almost everything that can occur in multilayer models can occur in single-layer models that have sufficiently complicated connectivity and post-synaptic potential functions. We will assume a one-spatial dimension continuum network which is homogeneous on the interval (a, b).

$$V(x,t) = \int_0^\infty h(s) \int_a^b w(x,y) f(V(y, t-s)) \, \mathrm{d}y \, \mathrm{d}s + I(x,t) \tag{8.1}$$

where h is the synaptic post-synaptic potential function and w is the spatial weighting function. (This is the same form as (7.1) but with possible infinite extent.) As usual, we normalize the post-synaptic potential function, h so that it integrates to 1. We make assumptions about w:

$$\text{symmetry } w(x,y) = w(y,x)$$

$$\text{homogeneity } \int_a^b w(x,y) \, \mathrm{d}y = C$$

$$\text{boundedness } \int_a^b \int_a^b w(x,y)^2 \, \mathrm{d}y \, \mathrm{d}x < \infty.$$

Homogeneity is necessary in order for the network to have a spatially constant solution. This model was the subject of some analysis by Coleman and Renninger (1974, 1975) when $(a, b) = (-\infty, \infty)$, w is a convolution weight, and f is piecewise linear. Hansel and Sompolinsky (1997) also analysed this network when the domain is 2π-periodic, $h(t) = \mathrm{e}^{-t}$, $w(x, y) = J_0 + J_2 \cos 2(x - y)$ and f is piecewise linear. Ermentrout *et al* (1986) analysed the discrete time version of this equation as a model for sea-shell coloration patterns (see also Murray 1989, for a description of this model). an der Heiden (1979) studied (8.1) when the weights are exponential so that he could convert the system into a reaction–diffusion system.

The general idea of bifurcation and pattern formation is to look at the stability of the constant solution as some parameter varies. Using normal forms and symmetry, the infinite-dimensional system is reduced to a low-dimensional system of usually simple amplitude equations that correspond to different spatial modes of the system. Often, many researchers stop after linear stability analysis to get an idea of the possible patterns that can form. This is

fine for situations in which there is only a single unstable mode, but the nonlinear analysis is necessary in order to correctly *select* which of the modes will stably appear. Haken (1996) formally used these techniques to derive *ad hoc* amplitude equations. Our approach is to use the more traditional approach of Liapunov–Schmidt and normal form analysis. However, it is useful to point out that Haken's methods are not restricted to equations but can be used to understand complex spatio-temporal data as well.

Let $I(x,t)$ be constant. Since the network is homogeneous, there is a spatially constant solution, $V(x,t) = \bar{V}$ satisfying

$$\bar{V} = f(\bar{V}) + I.$$

In order to use bifurcation methods, there must be some parameter which is varying. We will use the parameter, $\nu = f'(\bar{V})$ as a parameter that characterizes the general excitability of the network. If $|\nu|$ is small, then the network is doing very little to the input. The stability of the homogeneous state is found by looking at the linearized equation:

$$v(x,t) = \nu \int_0^\infty h(s) \int_a^b w(x,y) v(y, t-s)\, \mathrm{d}y\, \mathrm{d}s. \tag{8.2}$$

Consider the linear integral operator:

$$(Ku)(x) = \int_a^b w(x,y) u(y)\, \mathrm{d}y. \tag{8.3}$$

Since w is symmetric, K is self-adjoint and thus all eigenvalues of K are real and the eigenfunctions are mutually orthogonal. Let $\{\mu_j, \varphi_j(x)\}$ be the eigenvalue–eigenfunctions ordered by the size of the number of zeros of the eigenfunction. The solution to the linear problem is thus

$$v(x,t) = \mathrm{e}^{\lambda t}\varphi(x)$$

where

$$1 = \nu \mu_j \hat{h}(\lambda) \tag{8.4}$$

and $\hat{h}$ is the Laplace transform of h. Stability is assured if $\mathrm{Re}(\lambda) < 0$ for all λ and μ_j. We can now distinguish the four different ways in which stability can be lost. Since the weight function, $w(x,y)$ is homogeneous, the constant function, **1** is an eigenfunction with eigenvalue C. Suppose that for $|\nu|$ small, all values of λ have negative real parts. (This just means that the poles of $\hat{h}$ have negative real parts since the eigenvalues are bounded in magnitude for this compact operator, so that small enough values of ν will perturb the poles of $\hat{h}$ by only a small amount.) As ν increases in magnitude at least one of the roots λ may cross the imaginary axis for $j = j_c$. The four different cases can now be described.

- New constant solution: $\lambda = 0$ and $\varphi_{j_c}(x) = \mathbf{1}$.
- Homogeneous periodic solution: $\lambda = \pm \mathrm{i}\omega$ and $\varphi_{j_c}(x) = \mathbf{1}$.
- Spatial pattern: $\lambda = 0$ and $\varphi_{j_c}(x) \neq \mathbf{1}$.
- Spacetime pattern: $\lambda = \pm \mathrm{i}\omega$ and $\varphi_{j_c}(x) \neq \mathbf{1}$.

Depending on the details of the interaction functions and the PSP function, any of these behaviours can occur.

We show that a necessary condition for spatial pattern formation is that the interactions not be strictly positive. This follows from the properties of strictly positive weight functions, $w(x,y)$. From the analogue of the Frobenious–Perron theorem, operator (8.3) has a positive maximal eigenvalue and the eigenfunction corresponding to it is also positive. Since all eigenfunctions are mutually orthogonal and the operator is homogeneous, this eigenfunction is **1**. Thus, the eigenfunction corresponding to the maximal eigenvalue of the operator K

is the spatially constant one. Suppose that there is a zero eigenvalue. Then, because of normalization, $\hat{h}(0) = 1$ so that (8.4) implies:

$$\nu = 1/\mu_j.$$

Since the maximal eigenvalue is the one corresponding to the homogeneous solution, stability will be lost first at the constant mode. In most bifurcation problems, the first bifurcating branch is the only one which has a chance to be stable and successively branching solutions are unstable. Thus, in order to get pattern formation, we need some inhibition in the weight function. We now turn to some examples.

8.1.1. Exponential PSPs. We consider some specific temporal operators, namely the exponentials up to order 3 mentioned in section 4.2. As above, let $\{\mu_j, \varphi_j(x)\}$ be eigenvalue–eigenfunction pairs for the operator K in equation (8.3). Then the uniform state is stable if and only if the roots, λ to equation (8.4) have negative real parts for all μ_j. Let β_k be decay rates, so that, for example, in the second-order PSP function, the response if a sum of $e^{-\beta_1 t}$ and $e^{-\beta_2 t}$. Then for up to order 3, (8.4) becomes:

$$\lambda + \beta_1 = \beta_1 \nu \mu_j \tag{8.5}$$

$$(\lambda + \beta_1)(\lambda + \beta_2) = \beta_1 \beta_2 \nu \mu_j \tag{8.6}$$

$$(\lambda + \beta_1)(\lambda + \beta_2)(\lambda + \beta_3) = \beta_1 \beta_2 \beta_3 \nu \mu_j. \tag{8.7}$$

As shown in section 4.2.1, the only way stability can be lost for order 1 and 2 exponential synapses is through a zero eigenvalue which leads to stationary solutions. This will only occur for μ_j positive. Since the parameter is ν, then as soon as ν increases past

$$\nu_c = \min_j \frac{1}{\mu_j}$$

the uniform state will be unstable. A new solution will bifurcate that has the form:

$$V(x) \approx (\nu - \nu_c)^p \varphi_{j_c}(x)$$

where j_c is the index corresponding to the maximum eigenvalue. The power, p is generically 1 but if there is translational symmetry as would be the case on a ring or on an infinite line, then the power p is $\frac{1}{2}$. We will consider the ring case in more detail below.

We turn to the triple exponential case. There are two ways that the rest state can become unstable. One occurs when there is a zero eigenvalue and this happens exactly as in the single- and double-exponential cases and the same formula for $V(x)$ obtains. However, if the eigenvalue with maximum magnitude is *negative*, then it is possible to lose stability at a Hopf bifurcation. Using (4.2) the critical values of ν are:

$$\nu_j^H = -\frac{1}{\mu_j}\left(-1 + \frac{(\beta_1 + \beta_2 + \beta_3)(\beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3)}{\beta_1\beta_2\beta_3}\right). \tag{8.8}$$

Since all the β_k are positive, the term in the parentheses is always positive so that instability occurs only for values of μ_j which are negative.

Let

$$\nu_j^0 = \frac{1}{\mu_j}.$$

Then the first destabilization of the uniform state occurs when ν exceeds

$$\nu_c = \min_j(\nu_j^0, \nu_j^H).$$

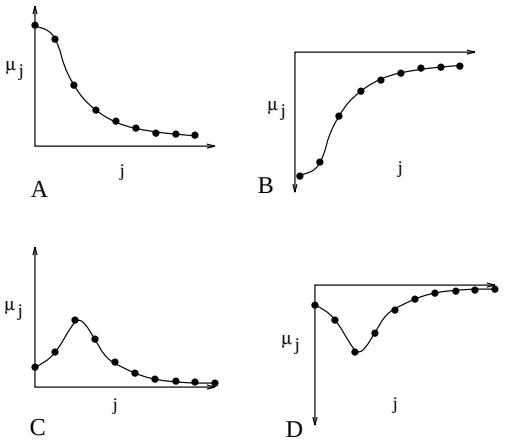

Figure 14. The approximate eigenvalues for the integral operator, K, when the interaction functions are of the four types in figure 8. Index j is number of zeros of the eigenfunction. (a) purely excitatory interactions; (b) purely inhibitory interactions; (c) lateral inhibition and local excitation; (d) lateral excitation and local inhibition.

If the critical value occurs at one of the ν_j^0 then stationary patterns bifurcate. If the critical value occurs at one of the ν_j^H, then time-periodic patterns bifurcate that have the form:

$$V(x,t) = \sqrt{\nu - \nu_c}\varphi_{j_c}(x)\cos\omega t \tag{8.9}$$

where

$$\omega = \sqrt{\beta_1\beta_2 + \beta_2\beta_3 + \beta_1\beta_3}.$$

If the eigenfunction is non-constant, then, spatio-temporal patterns bifurcate. If there is translational symmetry in the problem, then the behaviour is more complicated. We discuss this case in detail in section 8.1.2.

We close this discussion with an association of the connection kernels in figure 8 with the bifurcations that occur. The eigenfunctions for the connection kernels are similar to cosines and sines (and are precisely the trigonometric functions on the ring and infinite line). Thus, the eigenvalues lie close to the curves of the Fourier transforms of the connection kernels. These transforms are sketched in figure 14 for the four different connectivities along with the discrete eigenvalues. We can thus obtain all four types of bifurcation with the three-exponential PSP according to the nature of the connectivity.

- Pure excitation leads to spatially uniform constant solutions.
- Pure inhibition leads to spatially uniform time-periodic solutions.
- Lateral inhibition leads to spatially patterned solutions.
- Lateral excitation leads to spatially patterned time-periodic solutions.

All forms of pattern formation that occur in more complex models appear in this simple scalar model with higher-order PSPs.

8.1.2. Ring model Suppose that the network lies on a periodic domain ($a = 0, b = 2\pi$), the synapses are first order, and the weight function is a function of the distance only:

$$\frac{\partial V}{\partial t}(x,t) = -V(x,t) + \int_0^{2\pi} w(x-y)f(V(y,t))\,\mathrm{d}y.$$

Clearly, due to the periodicity and symmetry of $w(x)$

$$w(x, y) = w(x - y) = \sum_{k=0}^{\infty} J_k \cos k(x - y).$$

The eigenfunctions are $\{\sin(2\pi kx), \cos(2\pi kx)\}$ with eigenvalues, J_k each of multiplicity 2 except $k = 0$. The most 'unstable' mode is that one for which J_k is maximal. Hansel and Sompolinsky (1997) recently studied this network as a mechanism for orientation tuning in cortex. In this case, x does not mean physical space but is instead the preferred orientation of a cortical cell. They considered the situation where only the first two modes are non-zero and $J_0 < J_1$. In this case, we expect the first mode that bifurcates to be the one corresponding to the eigenfunctions $\cos(x), \sin(x)$. These have a single peak in the interval $(0, \pi)$ and thus, the network generates a solitary peak solution. You may recall that we stated that peak solutions could not bifurcate from rest. This is actually a periodic solution if one 'unravels' the ring. Note that this mechanism for peak formation is very different from the mechanisms discussed in section 7.3. In active media, standing pulses occur concurrently with stable uniform states; here the uniform state is unstable. In order to determine if the bifurcating solution is stable, we must perform the local nonlinear analysis.

We expand f about the fixed point

$$f(V) = \nu(V - \bar{V}) + c_2(V - \bar{V})^2 + c_3(V - \bar{V})^3 + \cdots.$$

We assume that J_n is the maximal coefficient of the Fourier series for w and that this occurs for $n > 0$ in order to get a spatial pattern.

Remark on global inhibition. In systems which combine *global inhibition* with a decaying Gaussian-type excitation, the most unstable mode will be $n = 1$ corresponding to a single peak in the ring. This is why single peaks are seen in both simple model of Hansel and Sompolinsky as well as the more complicated model of Somers *et al*. To see why this is, the Fourier coefficients of the positive Gaussian are monotonically decreasing with the maximal one being the constant ($n = 0$) mode (see figure 14). All Fourier components of global inhibition vanish except for the zeroth one. If this is large, then the maximal component of the sum of the local excitation and the global inhibition will be $n = 1$.

Without loss of generality, we assume $J_n = 1$ so that $J_j < 1$ for all other values, $j \neq n$. The translational symmetry of the model is why the nullspace is two-dimensional; the solution can be rotated around the ring arbitrarily. The instability occurs when the 'excitability' parameter, $\nu = \nu_c = 1$. For $\nu > \nu_c$ the uniform state is unstable. It is convenient to represent the solution in terms of a complex variable:

$$V(x, t) - \bar{V} = (z\mathrm{e}^{\mathrm{i}nx} + \bar{z}\mathrm{e}^{-\mathrm{i}nx}) + \cdots.$$

A standard calculation shows that z evolves like:

$$z' = z(\nu - \nu_c + Az^2\bar{z}) \tag{8.10}$$

$$A = 2c_2^2\left(\frac{1}{1 - J_{2n}} + \frac{2}{1 - J_0}\right) + 3c_3. \tag{8.11}$$

Since all coefficients are real, there is an arbitrary phase for the complex number z and only the amplitude is determined:

$$|z| = \sqrt{-\frac{\nu - \nu_c}{A}}.$$

Stable solutions will bifurcate if and only if $A < 0$. Since the maximal coefficient is $J_n = 1$, all of the fractions are positive, so that for stability, the third derivative of f had better be negative. For example, if the nonlinearity is a standard sigmoid, then fixed points that lie near the inflection point will lead to stable spatial patterns. In particular, if $n = 1$, we obtain the analogue of Hansel and Sompolinsky. They go beyond the bifurcation calculations here by using the piecewise linear function, $f(u) = \max(u, 0)$. However, the final results are essentially the same.

The presence of small inhomogeneities changes the normal form slightly. Suppose that the inputs are moving slowly around the circle, $I(x - \omega t)$. Then the normal form becomes

$$z' = z(\nu - \nu_c + Az\bar{z}) + C(t)$$

where

$$C(t) = \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}nx} I(x - \omega t)\,\mathrm{d}x \equiv p\mathrm{e}^{\mathrm{i}\psi - \mathrm{i}n\omega t}.$$

Taking $n = 1$ without loss in generality (we can always shorten the ring) and rewriting the equations in terms of $z = r\mathrm{e}^{\mathrm{i}\theta}$, we obtain

$$r_t = r(\nu - \nu_c + Ar^2) + p\cos(\psi - \omega t - \theta)$$
$$\theta_t = \frac{p}{r}\sin(\psi - \omega t - \theta).$$

In the rotating frame, $\phi = \psi - \omega t - \theta$ the equations are:

$$r_t = r(\nu - \nu_c + Ar^2) + p\cos(\phi)$$
$$\phi_t = \omega - \frac{p}{r}\sin(\phi).$$

This is exactly the same equation as that obtained by Kath (1981) in his analysis of a periodically driven Hopf bifurcation near resonance. If $\omega = 0$ then the system will lock onto the peak of the stimulus. As long as the stimulus does not move too fast, the network will follow it as it moves around the ring; it is *phase-locked* to the driving stimulus. Thus, the inhomogeneities have the effect of uniquely determining the phase of the spatial pattern. For example, in the Hansel and Sompolinsky model, weakly tuned inputs uniquely turn on the network at the appropriate orientation. In the absence of inputs, Hansel and Sompolinsky call this peak a 'marginal' solution due to its translation invariance. Like the active peaks in the previous section, the presence of inhomogeneities selects a spatial position.

If the ring model is combined with a three-exponential PSP, then the form of the solution is more complicated than in the case of a non-ring network, equation (8.9). Because of the additional translational symmetry, the solutions are of the form:

$$V(x, t) - \bar{V} = z\mathrm{e}^{\mathrm{i}(\omega t - nx)} + \bar{z}\mathrm{e}^{-\mathrm{i}(\omega t - nx)} + w\mathrm{e}^{\mathrm{i}(\omega t + nx)} + \bar{w}\mathrm{e}^{-\mathrm{i}(\omega t + nx)} + \cdots.$$

This is the sum of two wavetrains, one going left and one going right. In Ermentrout (1979) I derived the normal forms for a two-layer model. In generic cases, the two complex variables, w, z satisfy

$$z' = z(c_1(\nu - \nu_c) + c_3 z\bar{z} + d_3 w\bar{w})$$
$$w' = w(c_1(\nu - \nu_c) + c_3 w\bar{w} + d_3 z\bar{z})$$

where d_3 is an additional complex coefficient. I showed that there are only two types of solutions: either one of the two z, w is zero or both are equal and non-zero. In the former case, the solutions to (8.1) are travelling waves while in the latter case they are standing waves. The two different possibilities cannot be mutually stable. In fact, travelling waves require, $\mathrm{Re}\,d_3 < \mathrm{Re}\,c_3 < 0$ and standing waves require that $\mathrm{Re}\,c_3 < \mathrm{Re}\,d_3 < 0$. If the

nonlinearity is strictly odd, then it is simple to show that $d_3 = 2c_3$ so that travelling waves are the only solutions that can be stable. This is an example of nonlinear pattern selection; in the absence of the normal form, one has no *a priori* reason to suspect that there will be travelling waves, standing waves, or any mixture of the two.

8.2. Multiple bifurcations

In the previous section, we assumed that only one spatial mode became unstable at the critical amount of excitation. However, it is possible that by introducing another parameter, two modes could become unstable simultaneously. In this case, a variety of phenomena can occur such as secondary bifurcation from the principal branches (Bauer *et al* 1975, Cohen 1977, Keener 1976). Ermentrout and Cowan (1980b) analysed a two-layer neural network when there are multiple modes bifurcation at a complex eigenvalue. They found a variety of travelling and standing waves.

8.3. Multilayer models

The single-layer model that we discussed in the previous sections incorporates most of the behaviour that can be expected from bifurcating solutions in one spatial dimension. The same ideas can be applied to multilayer models with similar results. In all of these models, the symmetry of the interactions makes it easy to derive the form of the nonlinear mode equations. This observation was first noted by Sattinger (1977) and greatly expanded upon by Golubitsky and Schaefer (1985). The main point is that even if the network has many layers, the form of the bifurcation and the symmetry of the interactions leads to the same type of behaviour. Details matter primarily in the interpretation in terms of the biology and the behaviour of the reduced bifurcation equations. (The latter is true since, although the form of the bifurcation equations is the same from model to model, the coefficients can be quite different.) The other important detail is the mechanism by which the uniform state becomes unstable. We will now discuss a number of examples.

8.3.1. Standard two-layer models. In Ermentrout and Cowan (1980a) the following model was considered:

$$\tau_e \frac{\partial u_e}{\partial t} = -u_e + S_e(w_{ee} * u_e - w_{ie} * u_i)$$
$$\tau_i \frac{\partial u_i}{\partial t} = -u_i + S_e(w_{ei} * u_e - w_{ii} * u_i).$$

The domain of the model was the infinite line, but a ring of tissue could be likewise analysed. The crucial question is as some parameter varies: How does a uniform state lose stability? Assuming for simplicity that $S_e(0) = S_i(0) = 0$ we can linearize about $(u_e = u_i = 0)$. Since the domain is periodic (or infinite), the eigenfunctions for the integral operators are of the form:

$$\mathrm{e}^{\mathrm{i}kx}$$

where k is real in the case of the infinite domain and $k = nL/2\pi$ where n is an integer and L is the length of the ring of tissue for the circular domain. Let

$$\hat{w_{jl}}(k) \equiv \int_\Omega \mathrm{e}^{-\mathrm{i}kx} w_{jl}(x)\,\mathrm{d}x$$

be the Fourier modes of the weight functions. Without loss of generality, let $\tau_e = 1$ and $S_e'(0) = S_i'(0) = 1$. (This can be done since the derivatives can be absorbed into the weight functions.) Then we show that the uniform state is stable if and only if both

$$D(k) \equiv \hat{w_{ie}}(k)\hat{w_{ei}}(k) + (1 + \hat{w_{ii}}(k))(1 - \hat{w_{ee}}(k)) > 0 \tag{8.12}$$

$$T(k) \equiv -\frac{1 + \hat{w_{ii}}(k)}{\tau_i} - 1 + \hat{w_{ee}}(k) < 0 \tag{8.13}$$

for all k. The four simple ways to lose stability as some parameter varies are that one of these conditions is violated at either a non-zero or zero value of k. If (8.12) is violated then time-independent solutions will bifurcate with spatial patterns if the critical value of k at which instability occurs is non-zero. If (8.13) is violated, then a Hopf bifurcation to temporally oscillatory behaviour occurs and will be spatially uniform if the critical k is zero and spatially periodic otherwise. In Ermentrout and Cowan (1979a) the relative interactions between the excitatory and inhibitory cells that lead to spatially uniform stable periodic solutions was classified. We can give a general feel for the interactions necessary if we make a few simplifying assumptions. We suppose that

$$w_{jl}(x) = \frac{\alpha_{jl}}{\sigma_{jl}} W\left(\frac{x}{\sigma_{jl}}\right)$$

where W is a Gaussian-like non-negative weight function, symmetric and peaked at the origin. Then,

$$\hat{w_{jl}}(k) = \alpha_{jl}\hat{W}(\sigma_{jl}k).$$

The function $\hat{W}$ is positive, symmetric, and decreasing. The numbers, σ_{jl} can be thought of as space-constants, larger ones imply a greater connection spread. We can now see that a necessary condition for instability leading to spatially periodic patterns that are constant in time is that α_{ee} must be large enough and the cross interaction terms must have a greater spread than the self-interaction terms. In other words, one requires 'lateral inhibition'. Furthermore, we want τ_i to be small enough so that (8.13) is not violated. Fast, spread-out inhibition will lead to spatially patterned activity. The analogue to this was found as a requirement in the two-layer model for orientation tuning by Ben-Yishai *et al* (1995, 1997) as well as the more complicated spiking model proposed by Somers *et al* (1995). The latter paper was a simulation study, but by using the reduction methods described in section 3.2.1 we could presumably reduce their spatial model to a network similar to the present calculation. Ben-Yishai *et al* (1997) performed a more detailed analysis than we give here and pieced together the behaviour by letting the nonlinearity be piecewise linear. The calculations in this paper border on the heroic as the algebra quickly becomes overwhelming!

The second condition (8.13) will be first violated at a non-zero wavenumber, k if there is slow inhibition (τ_i large enough) sufficient recurrent excitation and self-inhibition (α_{ee}, α_{ii} large enough), there is strong feedback coupling (($\alpha_{ei}\alpha_{ie}$ large enough), and the inhibitory–inhibitory interactions have greater spread than the excitatory–excitatory ($\sigma_{ii} > \sigma_{ee}$). Only the latter condition is biologically questionable. If this last condition is violated, then, bifurcation to spatially homogeneous oscillations will occur.

8.3.2. Orientation selection. Ben Yishai *et al* (1997) proposed a model for a cortical hypercolumn that consists of two layers of spatially coupled neurons in a topological ring.

Hansel and Sompolinsky (1997) considered a simplified model:

$$u_t = -u + F\left[\int_0^\pi W(\theta - \theta')u(\theta', t)\,\mathrm{d}\theta' + I(\theta, t) - z(\theta, t)\right]$$
$$\tau z_t = \beta u - z$$

where $F(u) = A\max(0, u)$, $W(\theta) = J_0 + 2J_2\cos 2\theta$, and I represents inputs to the network. In addition, if $\beta \neq 0$ then there is adaptation via the variable z.

This model is meant to represent the response of a local cortical column to oriented input. (This is why the angle only traverses 0 to π as a bar which is oriented at 0 and one which is oriented at π are indistinguishable.) $u(\theta, t)$ is the firing rate of the neuron that represents the orientation, θ. We have already discussed this type of network in the absence of adaptation in section 8.1.2. Thus, we will look at the effects of adaptation. Before doing this, we summarize Hansel's and Sompolinsky's results on the adaptation-free model. They write down closed form solutions as a function of the parameters, J_0, J_1. They found two curves in the (J_0, J_2)-parameter plane. If $J_0 < 1$ and $J_2 < 1$, then the homogeneous solution is the only stable solution. For $J_2 > 1$, $J_0 < 1$ the peak solution is the only stable solution. Finally if J_2, J_0 are too large all solutions tend to infinity. The reason for this is that their model never saturates. If saturation is added, then a new homogeneous steady state where all cells fires obtains. Essentially, the behaviour is like that of the linearized system. They then showed that if the input is slightly tuned to favour a particular orientation, the firing rate will reflect this inhomogeneity. Furthermore if the input moves, the firing rate will 'follow' it as long as the rotation velocity is not too great. In other words, their model behaves just like the normal form analysis in section 8.1.2.

Hansel and Sompolinsky gave a thorough analysis of the adaptation model. They found three different types of behaviour in the absence of inputs: (i) homogeneous, (ii) standing pulse, (iii) travelling pulse. For a given value of the adaptation time constant, they created a phase diagram showing the regimes of stability for the three types of behaviours as a function of the two parameters, J_2 and β. Because they considered the piecewise linear model, they are able to construct all the transitions from the standing to the travelling pulse. They found that as J_2 increases, the homogeneous solution becomes unstable to either a stationary peak or if the adaptation is large enough, to a moving pulse. They never find standing wave oscillations to their model although there is no reason why they cannot occur.

It is easy to apply bifurcation methods to their model by linearizing about the homogeneous state. We omit the details noting that the uniform state, in the absence of inputs, is stable if and only if:

$$-1 + 1/\tau + J_l < 0$$
$$\beta - (J_l - 1) > 0$$

for $l = 0, 2$. Since J_0 is small or negative to produce lateral inhibition, the conditions mainly constrain J_2:

$$J_2 < 1 + 1/\tau$$
$$J_2 < 1 + \beta.$$

If the first condition is violated, then travelling waves bifurcate and if the second is violated then standing pulses bifurcate. They find the same critical behaviour as a function of the combined parameter, $\beta\tau$ according to whether it is greater than 1 (travelling waves) or less than 1 (standing pulse). Thus, the amplitude and the speed of adaptation determine the behaviour of the network as the tuning parameter, J_2 increases.

8.3.3. Dendritic effects. Bressloff and colleagues (1995–1997) described a new mechanism for the production of spatial and spatio-temporal patterns in neural networks. Their clever idea is the add dendritic effects the standard model. Each cell at position x has associated with it, a dendrite parametrized by its distance from the soma, ξ. The network lies on the infinite line and the dendrites are semi-infinite, originating at the soma. Using Green's functions for the passive dendrites, they derived the following equation for the somatic potential:

$$\frac{\partial U}{\partial t}(x,t) = -U(x,t) + \nu \int_{-\infty}^{t} \int_{-\infty}^{\infty} G[\alpha(x-x'), t-t']W(x-x')f(U(x',t'))\,\mathrm{d}x'\,\mathrm{d}t' \tag{8.14}$$

$$G(\xi,t) = \frac{1}{\sqrt{\pi Dt}} \exp\left(-\frac{\xi^2}{4Dt} - t/\tau\right) \tag{8.15}$$

where $\alpha(x) = \xi_0 + C|x|$. The function G is the spacetime Green's function of a dendritic cable with diffusion length D and time constant τ. The function $\alpha(x)$ models the correlation between dendritic position of a synapse and distance of the presynaptic cell. If $C = 0$ there is no correlation and all synapses land on the same point on the dendrite, ξ_0. If $C > 0$, then synapses from farther away land more distantly from the soma (see e.g. Shepherd 1990). The function $W(x)$ is the usual spatial decay of the synaptic weights, and f is the usual nonlinear firing rate. What is novel about Bressloff's mechanism is that he *does not* require a lateral-inhibition type of interaction function to obtain spatial pattern formation. This point is important as most of the long-distance interactions in cortex are excitatory and not inhibitory. They found a number of interesting phenomena. First, they showed that if $W(x)$ is purely excitatory and decreasing with distance, then including dendritic effects does not change the fact that the model still cannot generate stable spatially periodic temporally constant structures. However, if the weights *increase* with distance and the dendritic is correlated with presynaptic distance, then they showed that spatial patterns can bifurcate. Their most interesting result is that the presence of dendrites leads to an effective spatio-temporal delay and a bifurcation to spatially periodic and temporally periodic solutions is possible. They assume that there is *local inhibition and long-range excitation*, the opposite of the usual models that have 'lateral inhibition'. They make the dendrites uncorrelated with distance for simplicity and find that if the distance of the synapses from soma is large enough, the instability occurs. The mechanism for this is that the distant synapses provide enough of a delay for this negative feedback system to oscillate (see sections 4.2.1 and 8.1.1). In a related mechanism, Crook *et al* (1997) showed how distance-dependent delays due to finite axonal conduction can lead to a spatial instability in an array of weakly coupled oscillators.

8.3.4. A three-layer example. Ermentrout and Lewis (1997) showed that a neural network with only excitatory connections can result in bifurcation of spatially and temporally periodic behaviour. They considered a model of the form:

$$\begin{aligned}
\frac{\partial u_1}{\partial t} &= f(a_{11} w * u_1 - a_{21}u_2 + a_{31}u_3) \\
\frac{\partial u_2}{\partial t} &= -a_{22}u_2 + g(a_{12}u_1) \\
\frac{\partial u_3}{\partial t} &= -a_{33}u_3 + h(a_{13}u_1)
\end{aligned}$$

where all parameters are positive and $w(x)$ is similar to a Gaussian. The model can be considered as an excitatory population with two locally acting populations; an excitatory one and an inhibitory one. The mechanism for this resides in the linearization which leads to a third-order characteristic polynomial. General conditions for this and related three-component models are derived that guarantee loss of stability to a spatio-temporal pattern. As in Bressloff's work, there cannot be static spatially periodic patterns unless the weights have lateral inhibitory structure. This example and the Bressloff dendritic model are related to the single-layer model with lateral-excitation and local inhibition (section 8.1.1) and a three-exponential PSP.

8.4. Two spatial dimensions

The cortex is actually better modelled as a two-dimensional layered sheet than as a one-dimensional system. We have already discussed some of the behaviour that can arise in two spatial dimensions when the medium is active. The bifurcation methods described in the previous section are easily generalized to two dimensions, however, the additional degree of freedom confers additional symmetry and this in turn leads a much higher multiplicity in critical eigenvalues. This leads to a more complicated set of normal forms and a much greater variety in the types of patterns that can bifurcate. Before we discuss the mathematics and the recent work on this problem, we will first give some biological motivation for the analysis. The analogue of the periodic patterns that bifurcate from rest in one dimension are doubly periodic patterns in two dimensions. These take the form of stripes and chessboard-like patterns. Thus, our task is to interpret spatially structured activity patterns in a two-dimensional cortical network. Ermentrout and Cowan first (1979b) suggested that these patterns may be related to simple geometric structures that are perceived during drug-induced hallucinations. Other authors have made related claims (Tass 1995, 1997, Fohlmeister *et al* 1995, Ermentrout 1984, Reggia and Montgomery 1996). To see the connection, suppose that the activity pattern on the cortex is one of these simple geometric patterns. Then, how is this perceived? That is, what sort of retinal pattern would produce an analogous pattern of stimulus related activity on the cortex? Schwartz (1977) showed that there is a regular topographic mapping of 'retinal space' to 'cortical space'. Away from the fovea it is essentially the complex logarithm. In other words, (r, θ) in retinal coordinates goes to $(\ln r, \theta)$ in cortical coordinates. Inverting this transformation, it is easy to see that stripes are transformed into bulls-eyes, spirals, and exploding rays while chessboards are transformed into cobweb and mosaic-like patterns. (See figure 15 for these transformed patterns.) In his classic treatise on mescaline, Klüver (1967) classified the patterns seen during the early stages of hallucination into a small number of 'form constants':

- funnels and tunnels;
- spirals;
- mosaics and cobwebs.

Siegel (1977) also described a number of other altered states in which simple geometric patterns occur. Thus, Ermentrout and Cowan suggested that the patterns that arise when the cortical network is destabilized due to the effects of the hallucinogen are the neural correlates of simple visual hallucinations. Ermentrout (1984b) suggested that premigrainous auras could also arise in a similar manner, but did not provide a mechanism. More recently, Reggia and Montgomery (1996) suggested a way that the excitation could be increased. Associated with migraines is a well known phenomena called a scotoma which consists of a slowly moving blind region in the visual field. It has been suggested (Richards 1971) that this may be associated with cortical spreading depression (CSD) which is

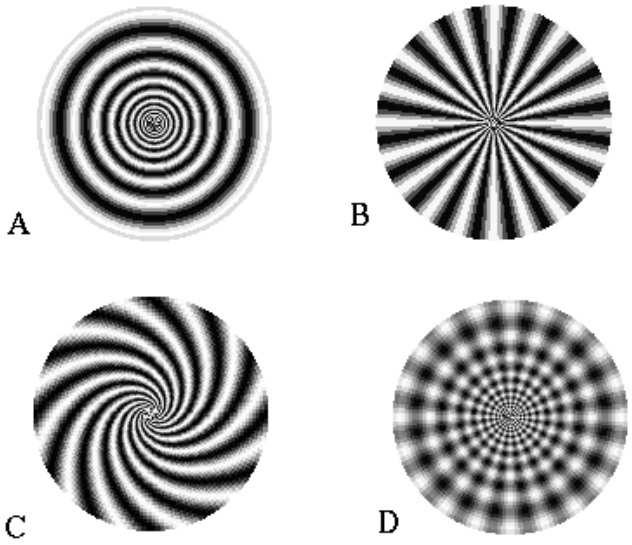

Figure 15. Visual patterns seen in the case of a mixture of one or two modes. Critical wavenumbers are $(\pm 12, \pm 16)$, $(\pm 16, \pm 12)$, $(0, \pm 20)$ and $(\pm 20, 0)$. (*a*)–(*c*) Single stable mode; (*d*) mixture of two modes.

caused by the diffusion of large amounts of extracellular potassium (Leao, Bures *et al* 1974). (However, this point is controversial (Kranda and Kulikowski 1984). Reggia and Montgomery created a computer model for CSD and coupled it to a one-layer lateral inhibitory network. Since small increases in extracellular potassium causes depolarization of neurons, this provided them with the mechanism for the enhanced cortical excitability. Tass (1995, 1997) proposed a model based on the equations analysed by Ermentrout and Cowan (1979b) but assumed that the excitation is epileptogenic in origin. Finally, Fohlmeister *et al* (1995) proposed that intrinsic spiral waves produced in an excitable medium are responsible for the patterns of hallucination. That is, they assert that it is not an instability which causes the formation of the patterns but instead a property of reduced inhibition in an excitable system.

We now turn to the analysis of two-dimensional networks when a bifurcation at a non-zero spatial wavenumber occurs. To avoid technical difficulties, we will follow Tass and assume a periodic domain rather than the infinite domain considered by Ermentrout and Cowan. Without loss of generality, we assume the domain is $[0, 2\pi) \times [0, 2\pi)$. There are two cases that are of interest; bifurcation to stationary spatially periodic patterns and bifurcation to spatially and temporally periodic patterns. We have already described mechanism for producing these instabilities in one-dimensional neural networks. The same mechanisms work well in two dimensions, however, the additional spatial dimension increases the dimension of the nullspace. As the mechanisms are all different but produce essentially the same result, we

consider the simplest one-layer model which is the two-dimensional analogue of (8.1):

$$V(x, y, t) = \int_0^\infty h(s) \int_\Omega w((x - x')^2 + (y - y')^2) S(V(x', y', t - s)) \, dx' \, dy' \, ds.$$

As in section 8.1, we will assume that there is some parameter which causes instability of the uniform state. Let

$$\hat{w}(n^2 + m^2) = \int_0^{2\pi} \int_0^{2\pi} w(x'^2 + y'^2) e^{-inx'} e^{-imy'} \, dx' \, dy'.$$

We assume a 'lateral inhibitory' or 'lateral excitatory' weight function so that $\hat{w}(R^2)$ has a maximum (minimum) value at $|R_0| > 0$. Depending on the PSP function and the choice of weight functions, the bifurcation will occur at a zero eigenvalue leading to stationary patterns or at an imaginary eigenvalue leading to spatially and temporally periodic patterns. Thus, our task is two-fold; first determine the dimension of the nullspace and then determine the form of the bifurcation equations so that we can decide which patterns are stable.

8.4.1. Bifurcation to stationary patterns. Suppose that instability is lost at a critical absolute wavenumber, R_0 and a zero eigenvalue. Then the nullspace is spanned by terms of the form

$$\varphi(x, y) = e^{inx} e^{imy} \tag{8.16}$$

where

$$n^2 + m^2 = R_0^2.$$

Tass stated that there are two cases:

(1) $n = \pm m$ or $nm = 0$ (four-dimensional nullspace);

(2) $n \neq \pm m$ & $nm \neq 0$ (eight-dimensional nullspace).

In fact, there are infinitely many cases. For example, suppose, $R_0 = 5$. Then the pairs, $(\pm 5, 0)$, $(0, \pm 5)$, $(\pm 4, \pm 3)$, $(\pm 3, \pm 4)$ yield a 16-dimensional nullspace. In any case, near the bifurcation, the solution to the full problem will be of the form:

$$\begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix} = \begin{pmatrix} U_0 \\ V_0 \end{pmatrix} \sum_k z_k \varphi_k(x, y)$$

where $\varphi_k(x, y)$ are the eigenfunctions (8.16). Since we are interested in real solutions, half of these will be complex conjugates of the other half, so that there will be only half as many bifurcation equations. The following theorem characterizes the bifurcation equations in the two cases Tass considered.

Theorem 2.

(i) For the case of a four-dimensional nullspace, the generic bifurcation equations are:

$$z_1' = z_1(\nu - a z_1 \bar{z_1} - b z_2 \bar{z_2})$$
$$z_2' = z_2(\nu - b z_1 \bar{z_1} - a z_2 \bar{z_2})$$

and the corresponding complex conjugate equations.

(a) The pure solutions, $(z_1, z_2) = (r, 0)$, $(z_1, z_2) = (0, r)$ where $r = \sqrt{\nu/a}$ are linearly stable if and only if $a < b$.

(b) The mixed solution, $(z_1, z_2) = (r, r)$ where $r = \sqrt{\nu/(a + b)}$ are linearly stable if and only if $a > b$.

(ii) For the case of an eight-dimensional nullspace, the generic bifurcation equations are:

$$\begin{aligned}
z_1' &= z_1(\nu - az_1\bar{z_1} - bz_2\bar{z_2} - cz_3\bar{z_3} - dz_4\bar{z_4}) \\
z_2' &= z_2(\nu - bz_1\bar{z_1} - az_2\bar{z_2} - dz_3\bar{z_3} - cz_4\bar{z_4}) \\
z_3' &= z_3(\nu - cz_1\bar{z_1} - dz_2\bar{z_2} - az_3\bar{z_3} - bz_4\bar{z_4}) \\
z_4' &= z_4(\nu - dz_1\bar{z_1} - cz_2\bar{z_2} - bz_3\bar{z_3} - az_4\bar{z_4}).
\end{aligned}$$

(a) The pure solutions, $z_j = r$, $z_k = 0$ for $k \neq j$ and $r = \sqrt{\nu/a}$ are linearly stable if and only if $a < b, a < c, a < d$.

(b) There are six different pairwise solutions, $z_j = z_k = r$, $z_l = 0$ for $l \neq j, k$. Since there is symmetry, consider only those where $z_1 \neq 0$:

$$\begin{aligned}
z_1 = z_2 &= \sqrt{\frac{\nu}{a+b}} \\
z_1 = z_3 &= \sqrt{\frac{\nu}{a+c}} \\
z_1 = z_4 &= \sqrt{\frac{\nu}{a+d}}.
\end{aligned}$$

These respective solutions are linearly stable if and only if

$$\begin{aligned}
a > b \quad &\text{and} \quad a < d + c - b \\
a > c \quad &\text{and} \quad a < d + b - c \\
a > d \quad &\text{and} \quad a < b + c - d
\end{aligned}$$

respectively.

(c) The triplet solutions where $z_j = z_k = z_l = r$ and $z_i = 0$ for $i \neq j, k, l$ are all unstable.

(d) The fully non-zero solution, $z_j = \sqrt{1/(a+b+c+d)}$ is linearly stable if and only if

$$\begin{aligned}
a &> d + c - b \\
a &> d + b - c \\
a &> b + c - d.
\end{aligned}$$

Remarks.

(1) It is assumed that all coefficients, a, b, c, d are also positive.

(2) In both cases, all the different possible solutions are mutually exclusive. That is, in the four-dimensional nullspace, the pure and mixed solutions cannot be simultaneously stable. Similarly, in the eight-dimensional case, there are three different mutually stable solutions: the pure mode solutions, the pairwise-mixed solutions, and the solutions comprised of all modes. Clearly the one- and two-mode solutions cannot be simultaneously stable nor can the two- and four-mode solutions. To see that the one- and four-mode solutions cannot be simultaneously stable note that the stability of the one-mode solution implies that $3a < b+c+d$ while the stability of the four mode implies that $3a > b+c+d$.

(3) Pure mode solutions correspond to stripes; pairwise solutions are chessboard-like; and the four-mode solutions are also mosaic-like.

(4) In the analysis by Tass (1995), he derived these equations for the Wilson–Cowan model using 'synergetics'. However, he appeared to have assumed that the nonlinearities

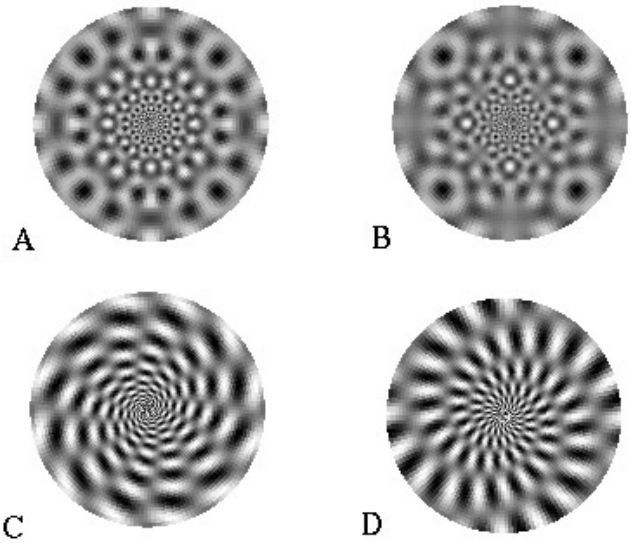

Figure 16. Visual patterns seen in the case of a mixture of four or more modes. Same wavenumbers as in figure 15. (*a*), (*c*), (*d*) Mixture of four modes (*d*) mixture of eight modes.

have a zero second derivative at the homogeneous state. In the absence of these quadratic terms, the coefficients for the bifurcation equations are degenerate in the following sense: $b = c = d = 2a$. With this symmetry assumption, in the four-dimensional case, only single modes are stable which implies that only stripes can be generated. Similarly, in the eight-dimension case, only stripes are stable. By neglecting or ignoring the role of the quadratic terms, whole classes of solutions are unstable.

Tass has some very nice figures which illustrate the patterns when viewed in visual-space coordinates, transformed from the rectilinear cortical coordinates. Some of his patterns are remarkably close to the patterns reported in Siegel (1977) that are drawn by subjects who have taken mescaline.

Figure 15 illustrates some of these patterns in visual coordinates for the four- and eight-dimensional case. Figure 16 shows complex mixtures in the case of a 16-dimensional nullspace.

8.4.2. Bifurcation at an imaginary eigenvalue. In this case, the dimension of the nullspaces in each of the two cases above is doubled since the eigenfunctions are multiplied by $e^{\pm i\omega t}$. Tass (1997) derived the bifurcation equations when there are eight and 16 eigenfunctions. As in the case of bifurcation to stationary patterns, the solutions to the neural network have the form:

$$\begin{pmatrix} u(x, y, t) \\ v(x, y, t) \end{pmatrix} = \begin{pmatrix} U_0 \\ V_0 \end{pmatrix} \sum_k z_k \varphi_k(x, y) e^{\pm i\omega t}.$$

As with his analysis of the stationary solutions, Tass assumed that the nonlinearity has no quadratic terms at the critical value of the bifurcation. This simplifies the calculation of the coefficients of the bifurcation equations. His results are summarized in the (less general) analogue of the above theorem.

Theorem 3.

(i) For the case of an eight-dimensional nullspace, the bifurcation equations in the case of no quadratic terms in the nonlinearity are:

$$z_1' = \nu z_1 - (a+b\mathrm{i})\left(2z_2\sum_{k\neq 1} z_k\bar{z}_k + z_1^2\bar{z}_1 + \bar{z}_3 z_2 z_4\right)$$

$$z_2' = \nu z_2 - (a+b\mathrm{i})\left(2z_2\sum_{k\neq 2} z_k\bar{z}_k + z_2^2\bar{z}_2 + \bar{z}_4 z_1 z_3\right)$$

$$z_3' = \nu z_3 - (a+b\mathrm{i})\left(2z_3\sum_{k\neq 3} z_k\bar{z}_k + z_3^2\bar{z}_3 + \bar{z}_1 z_2 z_4\right)$$

$$z_4' = \nu z_4 - (a+b\mathrm{i})\left(2z_4\sum_{k\neq 4} z_k\bar{z}_k + z_4^2\bar{z}_4 + \bar{z}_2 z_1 z_3\right).$$

There are two different types of stable solutions:

(a) $z_j = \sqrt{\nu/a}\,\mathrm{e}^{-\mathrm{i}t\nu b/a}$ for some j and $z_k = 0$ for $k \neq j$.

(b) $z_j = r\mathrm{e}^{-\mathrm{i}(\zeta_j+\Omega t)}$ for all j where, $r = \sqrt{\nu/(5a)}$, $\Omega = -b\nu/a$, and $\zeta_2+\zeta_4-\zeta_1-\zeta_3 = \pi$.

(ii) For the case of a 16-dimensional nullspace, the bifurcation equations in the case of no quadratic terms in the nonlinearity are:

$$z_j' = \nu z_j - (a+b\mathrm{i})\left\{2z_j\left(\sum_{k\neq j} z_k\bar{z}_k + z_j\bar{z}_j\right) + 2\bar{z}_j^{-}\sum_{k(j)} z_k z_k^{-}\right\}$$

where we have used the notation z_j^{-} to mean the amplitude of the mode whose spatial wavenumber is the negative of the wavenumber corresponding to mode j; and $j(k)$ is the set of all indices that do not include wavenumber j or its negative. There are two types of stable solutions:

(a) $z_j = \sqrt{\nu/a}\,\mathrm{e}^{-\mathrm{i}t\nu b/a}$ for some j and $z_k = 0$ for $k \neq j$.

(b) Four modes coexist and are equal in amplitude, $r = \sqrt{\nu/(5a)}$ and frequency, $\Omega = -b\nu/a$. There are essentially two subtypes, perpendicular pairs, $(1,3)$ and $(2,4)$, and non-perpendicular pairs, $(1,3)$ and $(5,7)$. These all have the same phase relationship as in case (i)b.

Remarks.

(1) If there *are* quadratic terms in the nonlinearity at the bifurcation point, then the coefficients are more complicated than those in the theorem, however, the terms themselves are the same.

(2) As an example of the 16-dimensional case, the eight spatial wavevectors corresponding to these modes are pairs (n,m) with eigenfunctions, $\mathrm{e}^{\mathrm{i}nx+\mathrm{i}my}$. The eight pairs are

$$\{(1,2),(2,-1),(-1,-2),(-2,1),(2,1),(1,-2),(-2,-1),(-1,2)\}.$$

Thus, we see immediately that modes $1,2$ and $4,3$ are perpendicular.

(3) In each of the two cases (i), (ii) the single mode pattern corresponds to travelling waves.

(4) The second class of patterns is more interesting and corresponds to two different patterns blinking in succession. For example, horizontal and vertical strips would, when transformed to visual space, result in starburst patterns slowly changing into bulls eyes and back again. Diagonal stripe pairs lead to spirals which switch from clockwise to anticlockwise.

As in his paper on stationary patterns, Tass (1997) has a number of striking pictures of these patterns and their temporal evolution in the visual space coordinate system.

8.4.3. More complex models. The patterns described in the previous sections are all related to bifurcations due to the rotational symmetry of the two-dimensional cortical sheet. Wiener (1994) extended the purely spatial model to include local orientation cells in the manner of Hansel and Sompolinsky. In the simplest case, the equations are:

$$V(x, y, \theta, t) = \int_{-\infty}^{t} h(t - t')\mathrm{d}t' \int_{0}^{\pi} \mathrm{d}\theta' \times \int_{\Omega} \mathrm{d}x'\,\mathrm{d}y'\, w(x - x', y - y', \theta - \theta') f(V(x', y', \theta', t').$$

The function w could either be of the form:

$$w(x, y, \theta) = W(x, y)Q(\theta)$$

or of a more complicated type. Wiener suggested that the connections between cells may not be isotropic with circular symmetry, but rather anisotropic with elliptical symmetry and the major axis aligned with the preferred orientation. Thus, cells in the 'layer' with preferred orientation of, for example, $\pi/4$ would have stronger coupling along the diagonal than along the orthogonal direction. In his thesis, Wiener considered the consequences of this and used the results to explain some of the fine structures of visual hallucinations such as the zig-zag patterns and cobwebs.

Mogilner and Edelstein-Keshet (1995) analysed equations with orientation and space together in a different context, but many of the behaviours of their models can probably occur in these models. They find patches with preferred orientations as well completely aligned regions. The interaction of orientation and spatial location is also very important in the development of cortical maps (Kohonen 1989) where patches of orientation patterns are distributed across space in a periodic manner.

The difficulty in using the additional coordinate of orientation, is that the map from cortical column activity to visual perception is not entirely clear. The complete bifurcation analysis of these models as well as their neurological interpretation remains an unsolved problem.

9. Conclusions

We have reviewed the behaviour of neural networks that arise from biological principles in a steady-state environment. Our methods are those of dynamical systems and applied mathematics. We have shown that these techniques are particularly useful when the models have some intrinsic structure whether it is spatial or temporal. It is likely that real nervous systems are not randomly or globally connected networks but rather do have some intrinsic spatial order at least in the lower sensory areas. We have used rather simple models for neurons since a rather complete analysis is then possible. However, real neurons are much more complicated and a single neuron is able to produce an astonishing array of dynamic

behaviours. Furthermore, the possibility of modifiable synapses, active dendrite, etc makes the computational ability of even a single neuron incredibly powerful. The analysis of single neuron models with gated channels and multiple compartments is in its infancy. One can only hope that the milieu in which a single neuron lies in some sense simplifies what is important so that the present highly reduced models can shed light on real biology.

Acknowledgment

The author was supported in part by NSF and NIMH.

Appendix. Proof of theorem 2

The actual form of the bifurcation equations can be determined by symmetry arguments, thus we will concentrate on the stability issues. We will assume that all coefficients are positive, since negative coefficients will lead to instabilities. Since all coefficients are real, we only need to look for real solutions. As long as all the coefficients are different, it is then trivial to show that the only fixed points are those that we have described above. Stability of the solutions to the four-dimensional solutions is routine. The eight-dimensional model is more complicated. I only sketch the idea. First look at the real part of the equations; next multiply, for example the $r_1 \equiv \mathrm{Re}\, z_j$ equation by r_1 and introduce the new variable, $R_1 = r_1^2$. Rescaling time, we are left with the competitive logistic equation:

$$R_1' = R_1(\nu - aR_1 - bR_2 - cR_3 - dR_4)$$

and similar equations for the others. Again with no loss in generality let $\nu = 1$. Then linearize the equations about each of the possible fixed points. The single mode is very simple and leads to an upper triangular matrix whose eigenvalues can be picked off by inspection. For the two model solutions, the linearization leads to a block upper-triangular matrix with two 2×2 blocks. The eigenvalues of these two symmetric blocks are easily found. For the triplet solutions, the resultant linearized equations are block triangular with a 1×1 and 3×3 block. Without loss in generality, suppose that only $z_4 = 0$. Then the eigenvalue of the lower 1×1 block is negative if and only if $a < d$. The upper block is a positive constant times

$$\begin{bmatrix} -a & -b & -c \\ -b & -a & -d \\ -c & -d & -a \end{bmatrix}.$$

The characteristic polynomial for this is

$$(x+a)^3 - (d^2+b^2+c^2)(x+a) + 2bcd = 0.$$

Letting $\nu = x + a$, we must solve:

$$f(\nu) = \nu^3 - (d^2+b^2+c^2)\nu + 2bcd = 0.$$

$f(d) = -d(b-c)^2 < 0$ so that there must be a root, ρ, to $f(\nu)$ which is greater than d. Thus, there is an eigenvalue $x = \rho - a$. However, recall that $a < d$ is necessary for stability and thus, this eigenvalue is positive and the instability is assured. Finally, in the last case, the linearized system is a positive constant times the 4×4 matrix:

$$\begin{bmatrix} -a & -b & -c & -d \\ -b & -a & -d & -c \\ -c & -d & -a & -b \\ -d & -c & -b & -a \end{bmatrix}.$$

The eigenvectors for this matrix are $[1, 1, 1, 1]^T$, $[1, 1, -1, -1]^T$, $[1, -1, -1, 1]^T$, and $[1, -1, 1, -1]^T$ from which the result follows.

References

Abbott L F 1992a *Physica* **185A** 343–56
——1992b *From Biology to High Energy Physics* ed O Benhar *et al* (Pisa: ETS EDitrice) pp 179–96
——1994 *Neural Modeling and Neural Networks* ed F Ventriglia (Oxford: Pergamon) pp 57–78
Abbott L F and van Vreeswijk C 1993 *Phys. Rev.* E **48** 1483–90
Amari S-I 1972 *IEEE Trans. Syst. Man Cybern.* **SMC2** 643–57
——1977 *Biol. Cybern.* **27** 77–87
Amit D 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)
Amit D, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30–67
an der Heiden U 1979 *Analysis of Neural Networks* (New York: Springer)
Anderson J A and Rosenfeldt E (ed) 1988 *Neurocomputing: Foundations of Research* (Cambridge, MA: MIT)
Arbib M A (ed) 1995 *The Handbook of Brain Theory and Neural Networks* (Cambridge, MA: MIT)
Aronson D, Ermentrout B and Kopell N 1990 *Physica* **41D** 403–49
Baird B 1986 *Physica* **22D** 150–75
Basar E (ed) 1990 *Chaos in Brain Function* (Berlin: Springer)
Bauer L, Keller H B and Reiss E L 1975 *SIAM Rev.* **17** 101–22
Beer R D 1995 *Adapt. Behav.* **3** 469–509
Ben-Yishai R, Hansel D and Sompolinsky H 1997 *J. Comput. Neurosci.* **4** 57–77
Ben-Yishai R, Lev Bar-Or R and Sompolinsky H 1995 *PNAS USA* **92** 3844–8
Beurle R L 1956 *Phil. Trans. R. Soc.* B **240** 55–94
Borisyuk G N, Borisyuk R M, Khibnik AI and Roose D 1995 *Bull. Math. Biol.* **57** 809–40
Borisyuk R M and Kirillov A B 1992 *Biol. Cybern.* **66** 319–25
Bressloff P C 1995 *Biol. Cybern.* 281–90
——1996 *Phys. Rev. Lett.* **76** 4644–8
Bressloff P C and Coombes S 1997 *Phys. Rev. Lett.* **78** 4665–8
Bressloff P C and DeSouza B 1997 *Preprint*
Bures J, Buresova D and Krvanek J 1974 *Application of Leao's Spreading Depression of Electroencephalographic Activity* (New York: Academic)
Camperi M and Wang X J 1997 *Proc. CNS*96* ed J M Bower (New York: Plenum)
Casten R G, Cohen H and Lagerstrom P A 1975 *Quart. J. Appl. Math.* **32** 365–402
Chagnac-Amitai Y and Connors B W 1989 *J. Neurophysiol.* **61** 747–58
Chen Z, Ermentrout B and McLeod J B 1997a *Applicable Anal.* **64** 235–53
Chen Z, Ermentrout B and Wang X-J 1997b *J. Comput. Neurosci.* in press
Cohen D S 1977 *J. Math. Anal. Appl.* **57** 505–21
Cohen M A and Grossberg S 1983 *IEEE Trans. Syst. Man Cybern.* **13** 815–26
Coleman B D and Renninger G H 1974 *Math. Biosci.* **20** 155–70
——1975 *J. Theor. Biol.* **51** 243–65
Coombes S and Lord G J 1997 *Phys. Rev.* E **56** 5809–18
Cowan J D 1968 *Neural Networks* ed E R Caiianello (Berlin: Springer) pp 181–8
Cowan J and Sharp D H 1988 *Quart. Rev. Biophys.* **21** 365–427
Crook S, Ermentrout G B, Vanier M C and Bower J M 1997 *J. Comput. Neurosci.* **4** 161–72
Cross M C and Hohenberg P C 1993 *Rev. Mod. Phys.* **65** 851–1112
Destexhe A 1994 *Phys. Rev.* E **50** 1594–606
Destexhe A, Mainen Z F and Sejnowski T J 1994 *J. Comput. Neurosci.* **1** 195–230
Edelstein-Keshet L 1988 *Mathematical Models in Biology* (New York: Random House)
Ellias S A and Grossberg S 1975 *Biol. Cybern.* **20** 69–98
Ermentrout G B 1979 Symmetry breaking in stationary, homogeneous, isotropic neural nets *PhD Thesis* University of Chicago, Department of Biophysics and Theoretical Biology
——1982 *Competition and Cooperation in Neural Nets* ed M A Arbib and S Amari (New York: Springer) (New York: Springer) pp 57–70
——1984a *SIAM J. Appl. Math.* **44** 80–95
——1984b *The Neurobiology of Pain* ed A H Holden and W Winlow (Manchester: Manchester University Press) pp 267–83

——1992a *Neural Netw.* **5** 415–31
——1992b *SIAM J. Appl. Math.* **52** 1665–87
——1994 *Neural Comput.* **6** 679–95
——1995 *Handbook of Brain Theory and Neural Networks* ed M A Arbib (Cambridge, MA: MIT) pp 32–735
Ermentrout G B, Campbell J and Oster G 1986 *Veliger* **28** 369–88
Ermentrout G B, Chen X and Chen Z 1997 *Physica* D **108** 147–67
Ermentrout G B and Cowan J D 1979a *J. Math. Biol.* **7** 265–80
——1979b *Biol. Cybern.* **34** 137–50
——1980a *SIAM J. Appl. Math.* **38** 1–21
——1980b *SIAM J. Appl. Math.* **39** 323–40
Ermentrout G B, Hastings S P and Troy W C 1984 *SIAM J. Appl. Math.* **44** 1133–49
Ermentrout G B and Kopell N 1991 *J. Math. Biol.* **29** 195–217
Ermentrout G B and Lewis M 1997 *Bull. Math. Biol.* **59** 533–49
Ermentrout G B and McLeod J B 1993 *Proc. R. Soc.* A **123** 461–78
Fife P C 1979 *Mathematical Aspects of Reacting and Diffusing Systems (Lecture Notes in Biomathematics 28)* (Berlin: Springer)
Fife P C and McLeod J B 1977 *Arch. Ration. Mech. Anal.* **65** 335–61
Fitzsimonds R M, Song H-J and Poo M-M 1997 *Science* **388** 439–47
Fohlmeister C, Gerstner W, Ritz R and van Hemmen J L 1995 *Neural Comput.* **7** 905–14
Freeman W J 1972 *Mass Action in the Nervous System* (New York: Academic)
——1987 *Biol. Cybern.* **56** 139–50
Fukai T and Tanaka S 1997 *Neural Comput.* **9** 77–98
Funahashi S, Bruce C J and Goldman-Rakic P 1989 *J. Neurophys.* **61** 331–49
Fuster J and Alexander G 1971 *Science* **173** 652–4
Gerstner W 1995 *Phys. Rev.* E **51** 738–59
Gerstner W, van Hemmen J L and Cowan J D 1996 *Neural Comput.* **8** 1653–76
Glass L and Mackey M 1988 *From Clocks to Chaos: The Rhythms of Life* (Princeton, NJ: Princeton University Press)
Golomb D and Amitai Y 1997 *J. Neurophysiol.* **78** 1199–211
Golubitsky M and Schaefer D G 1985 *Singularities and Groups in Bifurcation Theory* (New York: Springer)
Grannan E R, Kleinfeld D and Sompolinsky H 1993 *Neural Comput.* **5** 550–69
Gray C M 1994 *J. Comput. Neurosci.* **1** 11–38
Gray C M and Singer S 1989 *PNAS USA* **86** 1698–702
Greenberg J and Hastings S 1978 *SIAM J. Appl. Math.* **34** 515–23
Grillner S 1974 *Expt. Br. Res.* **20** 459–70
Grossberg S 1990 *Computational Neuroscience* ed E Schwartz (Cambridge, MA: MIT) pp 56–68
Guckenheimer J, Harris-Warrick R, Peck J and Willms A 1997 *J. Comput. Neurosci.* **4** 257–77
Guckenheimer J and Holmes P 1983 *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Berlin: Springer)
Guttman R, Lewis S and Rinzel J 1980 *J. Physiol.* **305** 377–95
Haken H 1996 *Principles of Brain Function: A Synergetic Approach to Brain Activity, Behavior, and Cognition* (Berlin: Springer)
Hansel D, Mato G and Meunier C 1995 *Neural Comput.* **7** 307–37
Hansel D and Sompolinsky H 1996 *J. Comput. Neurosci.* **3** 7–34
Hansel D and Sompolinsky H 1997 *Methods of Neuronal Modeling* 2nd edn, ed C Koch and I Segev (Cambridge, MA: MIT)
Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City, CA: Addison Wesley)
Herz A V M 1995 *Phys. Rev.* E **44** 1415–18
Hopfield J J 1982 *PNAS USA* **79** 2554–8
——1984 *PNAS USA* **81** 3088–92
Hoppensteadt F and Izhikevich E 1997 *Weakly Connected Neural Nets* (Berlin: Springer)
Huguenard J and McCormick D A 1996 *Electrophysiology of the Neuron* (Oxford: Oxford University Press)
Idiart M A P and Abbott L F 1993 *Network* **4** 285–94
Johnston D and Wu S 1995 *Foundations of Cellular Neurophysiology* (Cambridge, MA: MIT)
Kammen D M and Yuille A L 1988 *Biol. Cybern.* **59** 23–31
Kandel E R and Schwartz J H 1985 *Principles of Neural Science* (New York: Elsevier)
Kath W 1981 *Stud. Appl. Math.* **65** 95–112

Keener J 1976 *Stud. Appl. Math.* **55** 187–211
Keener J P and Tyson J J 1986 *Physica* D **21** 307–24
Kelso J S 1995 *Dynamic Patterns: The Self-organization of Brain and Behavior* (Cambridge, MA: MIT Press)
Kishimoto K and Amari A 1979 *J. Math. Biol.* **7** 303–18
Kleinfeld D, Delaney K R, Fee M S, Flores J A, Tank D W and Gelperin A K 1994 *J. Neurophysiol.* **72** 1402–19
Kluver H 1967 *Mescal and the Mechanisms of Hallucination* (Chicago, IL: The University of Chicago Press)
Koch P and Leisman G 1996 *Int. J. Neurosci.* **86** 179–96
Kohonen T 1989 *Self-organization and Associative Memory* (Berlin: Springer)
Kopecz K and Schoener G 1995 *Biol. Cybern.* **73** 49–64
Kranda K and Kulikowski J J 1984 *The Neurobiology of Pain* ed A H Holden and W Winlow (Manchester: Manchester University Press) pp 243–66
Krekelberg B and Taylor J G 1996 *J. Chem. Neuroanatomy* **10** 191–6
Kuramoto Y 1984 *Chemical Oscillations, Waves, and Turbulence* (New York: Springer)
Leao A *J. Neurophys.* **7** 357–90
Linsker R 1988 *PNAS USA* **83** 7508–12, 8390–4, 8779–83
Lubke J, Markram H, Frotscher M and Sakmann B 1996 *J. Neurosci.* **16** 3209–18
MacDonald N 1989 *Biological Delay Systems: Linear Stability Theory* (Cambridge: Cambridge University Press)
MacKay D J C and Miller K D 1990 *Neural Comput.* **2** 173–87
Mackey M C and Milton J G 1988 *Ann. NY Acad. Sci.* **504** 16–32
Mallot H A and Giannakopoulos F 1996 *Biol. Cybern.* **75** 441–52
Marcus C M and Westerveld R M 1989 *Phys. Rev.* A **39** 347–59
McClelland J L and Rumelhart D E 1988 *Explorations in Parallel Distributed Processing* (Cambridge, MA: MIT)
McCulloch W S and Pitts W 1943 *Bull. Math. Biophys.* **5** 115–33
Meister M, Wong R O, Baylor D A and Shatz C J 1991 *Science* **252** 939–43
Miller K D 1994 *Prog. Brain Res.* **102** 303–18
Milton J G, Chu P H and Cowan J D 1993 *Neural Information Processing Systems* vol 5, ed S J Hanson, J D Cowan and C L Giles (San Mateo, CA: Morgan Kaufmann) pp 1001–6
Milton J, Mundel T, an der Heiden U, Spire J-P and Cowan J D 1995 *The Handbook of Brain Theory and Neural Networks* ed M A Arbib (Cambridge, MA: MIT) pp 994–7
Mogilner A and Edelstein-Keshet L 1995 *J. Math. Biol.* **33** 619–60
Muller B, Reinhardt J and Strickland M T 1995 *Neural Networks: An Introduction* (Berlin: Springer)
Murray J D 1989 *Mathematical Biology* (New York: Springer)
Nishiura Y and Mimura M 1989 *SIAM J. Appl. Math.* **49** 481–514
Oja E 1982 *J. Math. Biol.* **15** 267–73
Paullet J and Ermentrout G B 1994 *SIAM J. Appl. Math.* **54** 1720–44
Pelinovsky D E and Yakhno V G 1996a *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **6** 81–7
——1996b *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **6** 89–100
Petsche H, Prohaska O, Rappelsberger P, Vollmer R and Kaiser A 1974 *Epilepsia* **15** 439–63
Pinto D 1997 Computational, experimental, and analytical explorations of neuronal circuits in the cerebral cortex *PhD Thesis* University of Pittsburgh, Department of Mathematics
Pinto D J, Brumberg J C, Simons D J and Ermentrout G B 1996 *J. Comput. Neurosci.* **3** 247–64
Prectl J C, Cohen L B, Mitra P P, Pesaren B and Kleinfeld D 1997 *PNAS USA* **94** 7261–626
Redish A D, Elga A N and Touretzky D S 1996 *Network* **7** 671–85
Reggia J A and Montgomery D 1996 *Comput. Biol. Med.* **26** 133–41
Richards W 1971 *Sci. Am.* **224** 88–96
Rinzel J and Ermentrout G B 1989 *Methods in Neuronal Modeling* ed C Koch and I Segev (Cambridge, MA: MIT) pp 135–70
Rinzel J and Keller J B 1973 *Biophysical J.* **13** 1313–37
Rinzel J, Terman D, Wang X J and Ermentrout G B 1998 *Science* to appear
Roque Da Silva Filho A C 1992 *Network* **3** 303–21
Rosenblatt F 1958 *Psychol. Rev.* **62** 386–408
Sakaguchi H 1988 *Prog. Theor. Phys.* **79** 1061–7
Sakaguchi H, Shinomoto S and Kuramoto Y 1987 *Prog. Theor. Phys.* **77** 1005–10
Salinas E and Abbott L F 1996 *PNAS USA* **93** 11 956–61
Sattinger D H 1977 *J. Funct. Anal.* **28** 58–101
Schwartz E 1977 *Biol. Cybern.* **25** 181–94
Seung S 1996 *PNAS USA* **93** 13 339–44
Shepherd G M 1990 *The Synaptic Organization of the Brain* (Oxford: Oxford University Press)

Siegel R K 1977 *Sci. Am.* **237** 132
Somers D C, Nelson S B and Sur M 1995 *J. Neurosci.* **15** 5448–65
Sompolinsky H 1987 *Heidelberg Colloquium on Glassy Dynamics* ed L van Hemmen and I Morgenstern (Heidelberg: Springer)
Tamas G, Buhl E H and Somogyi P 1997 *J. Neurosci.* **17** 6352–64
Tass P 1995 *J. Biol. Phys.* **21** 177–210
——1997 *J. Biol. Phys.* **23** 21–66
Terman D and Wang D 1995 *Physica* **81D** 148–76
Traub R D, Jefferys J G R and Miles R 1993 *J. Physiol.* **472** 267–87
Tsodyks M V, Skaggs W E, Sejnowski T J and McNaughton B L 1997 *J. Neurosci.* **17** 4382–8
Tuckwell H and Miura R 1978 *Biophys. J.* **23** 257–76
Tyson J J and Keener J P 1988 *Physica* **32D** 327–61
Usher M, Schuster H G and Niebur E 1993 *Neural Comput.* **5** 570–86
van Vreeswijk C, Abbott L F and Ermentrout G B 1994 *J. Comput. Neurosci.* **1** 313–21
van Vreeswijk C and Sompolinsky H 1996 *Science* **274** 1724–6
Vasiliev V A, Romanovskii Y M, Chernavskii D S and Yakhno V G 1987 *Autowave Processes in Kinetic Systems* (Dordrecht: Reidel)
von der Malsburg C 1973 *Kybernetik* **14** 85–100
Wang X J and Rinzel J 1992 *J. Neurosci.* **53** 899–904
White E L 1989 *Cortical Circuits: Synaptic Organization of the Cerebral Cortex. Structure, Function, and Theory* (Boston, MA: Birkhauser)
Widrow B and Hoff M E 1960 *WESCON* convention record IV, pp 96–104
Wiener M 1994 Hallucinations, symmetry, and the structure of primary visual cortex: a bifurcation approach *PhD Thesis* University of Chicago, Department of Mathematics
Williams T and Bowtell G 1997 *J. Comput. Neurosci.* **4** 47–55
Wilson H R and Cowan J D 1972 *Biophys. J.* **12** 1–24
——1973 *Kybernetic* **13** 55–80
Winfree A T 1980 *The Geometry of Biological Time* (New York: Springer)
Wright J J, Kydd R R and Lees G J 1987 *J. Neurosci.* **33** 1–13
Wright J J and Liley D T J 1995 *Biol. Cybern.* **72** 347–57
Yuille A and Gryczwyz N M 1989 *Neural Comput.* **1** 334–47
Zhang K 1996 *J. Neurosci.* **16** 2112–26