Adapted by Stefano Fusi from a preliminary version of:

Appendix: Theoretical Approaches to Neuroscience

L.F. Abbott, Stefano Fusi, Kenneth D. Miller

Networks of Firing-Rate Neurons

We now consider the effects of connecting neurons together into networks. To do this, we make use of a description of neural responses that dispenses with individual action potentials and instead describes the inputs and outputs of neurons solely in terms of firing rates. A neuron firing at a rate r will produce an action potential during a short interval of duration Δt with probability $r\Delta t$. Because neuronal firing is often erratic and variable, there may be cases for which this probability provides the most detailed prediction of a neuronal response that we can give. More importantly, for our purposes, describing neuronal responses in terms of rates greatly simplifies the mathematical analysis of neural networks.

A firing rate network is specified by two critical elements. The first is the relationship between the total synaptic current that a neuron receives, which we denote by $I_{\rm tot}$, and its firing rate r. For constant currents, this relationship is given in terms of a firing-rate function, $r = F(I_{\rm tot})$. $F(I_{\rm tot})$ is a function with the general shape of the curve in Figure 3b, although it may include saturation effects at higher rates than are shown in that figure. When the current varies with time, we assume that the firing rate lags behind but approaches this function exponentially with time constant τ, so that

$$\tau \frac{dr}{dt} = -r + F(I_{\rm tot})\,. \tag{1}$$

The time constant τ incorporates the influences of both the membrane time constant and synaptic time constants. For a constant external current, τ determines how rapidly r approaches its steady-state value $r = F(I_{\rm tot})$. Note that this steady-state firing rate is obtained by setting the right side of Equation 1 to zero and solving for r, a procedure that we will follow repeatedly below.

The second element we need is the relationship between $I_{\rm tot}$ and the activity of other neurons in the network. The rule is simple: each presynaptic neuron contributes an amount to $I_{\rm tot}$ given by the product of its firing rate and a weight factor that characterizes the strength and type of the synapse through which it acts. Excitatory synapses have positive weights and inhibitory synaptic weights are negative. The total current for each neuron is the sum of all such terms. If the network we are studying receives input from other areas, this is included as an additional term in $I_{\rm tot}$ that we denote by h.

A network model of N neurons consists of N equations of the form 1, plus equations giving the total current for each neuron in terms of external sources and the firing

rates of other network neurons. Specifying the weights multiplying the different firing rates in these sums and the form of the firing-rate function F completely defines the network model.

Network models are complex and difficult to analyze for two reasons: the firing-rate function is typically nonlinear and the activities of all the neurons interact through the effects of each neuron's activity on the total currents of the other neurons. The second of these complications is intimately linked to the first. If the firing-rate function of a network model is linear, methods of linear algebra involving eigenfunctions and eigenvalues can be used to unscramble the interactions between the network neurons, yielding a fairly simple and understandable description. For this reason, modelers often approximate nonlinear firing-rate functions with linear approximations, valid over a certain range. We will do this below, taking $F(I_{\rm tot}) \approx gI_{\rm tot}$, where g is a constant known as the gain of the neuronal response. For simplicity, we will assume that the total current is normalized so that we can set $g = 1$. Thus, when we make the linear approximation we will simply replace $F(I_{\rm tot})$ by $I_{\rm tot}$.

Although this firing-rate network model, and especially its linear approximation, is extremely simplistic, it can provide deep and important insights into the properties of neuronal circuits, some of which we now discuss. We begin by considering uniform populations of either excitatory or inhibitory neurons. These populations can be characterized by a single firing rate, either $r_{\rm E}$ for the excitatory population or $r_{\rm I}$ for the inhibitory population. We then expand to consider non-uniform networks with different firing rates for each neuron, introducing the eigenvector and eigenvalue techniques that allow such networks to be analyzed. These methods show that, although our initial approach, in which all members of a neural population of a given type fire at the same rate, may seem overly simplistic, it nevertheless illustrates basic features found in more complex models. However, when recurrent inhibition largely cancels the effects of recurrent excitation, the results of eigenvector analysis can be insufficient, as we illustrate in a simple case in which a uniform excitatory and a uniform inhibitory population interact. Such networks can have a hidden feedforward structure within what appears to be a fully connected recurrent network, and also show quite nonintuitive effects when excitation is added to the inhibitory population. These effects, again, can be generalized to non-uniform networks. To illustrate networks that model decision making, we finally consider models that include two uniform excitatory populations as well as a uniform inhibitory population.

Many of the ideas we discuss are basic insights from linear algebra applied in a neural context. They serve as examples of the intuitions that can arise from formulating and analyzing even the simplest of models. These intuitions should be a part of every systems neuroscientist's toolkit much as understanding cable and channel properties are essential tools for a cellular neurophysiologist.

Purely Excitatory or Inhibitory Networks

We begin by considering a population of excitatory neurons with recurrent connections between them. To keep the network as simple as possible, we connect every pair of neurons in the network with the same synaptic weight and provide them all with the same external input. Because all the neurons are the same, we can characterize the entire network with a single firing rate $r_{\rm E}$. The total current for each neuron can then be written as the product of a synaptic weight factor $w_{\rm EE}$ (for excitatory-excitatory) and the rate $r_{\rm E}$. We also include an externally generated term labeled $h_{\rm E}$ in the total current. If we use the linear approximation $F(I_{\rm tot}) \approx I_{\rm tot}$ the equation governing this network is

$$\tau \frac{dr_{\rm E}}{dt} = -r_{\rm E} + I_{\rm tot} = -r_{\rm E} + w_{\rm EE} r_{\rm E} + h_{\rm E} = -(1 - w_{\rm EE}) r_{\rm E} + h_{\rm E} \tag{2}$$

If $w_{\rm EE} > 1$, the factor multiplying $r_{\rm E}$ on the right side of this equation is positive, which causes $r_{\rm E}$ to grow exponentially away from zero. This means that the linear system is unstable. In the full nonlinear network model, bounds on the neuronal firing rate will eventually halt this growth, but at that point the firing rate is likely to be quite far from zero. If we want a network model with low firing rates and for which the linear approximation is valid, we must for the moment restrict ourselves to the case $w_{\rm EE} < 1$, which keeps the linear system stable.

When $w_{\rm EE} < 1$, it is convenient to rewrite the above equation as

$$\left(\frac{\tau}{1 - w_{\rm EE}}\right) \frac{dr_{\rm E}}{dt} = -r_{\rm E} + \frac{h_{\rm E}}{1 - w_{\rm EE}} \, . \tag{3}$$

This leads us to our first lesson in network dynamics: excitatory feedback has two effects. First, it amplifies the input, as can be seen because $h_{\rm E}$ is multiplied by the factor $1/(1 - w_{\rm EE})$, which is greater than 1. Second, it slows the dynamics because $r_{\rm E}$ approaches $h_{\rm E}/(1 - w_{\rm EE})$ with a time constant $\tau/(1 - w_{\rm EE})$. When recurrent excitation is used to amplify input signals in a network, the price is a slowing down of the response dynamics (Figure 4a). Responses in primary visual cortex typically last considerably longer than the responses of the inputs from lateral geniculate nucleus (LGN). Recurrent excitation provides one possible explanation for this slowing.

One may worry that these results depend on the very simple model used, but instead the simple model yields an insight that is much more general. Without recurrent connections, the steady-state response to a constant input $h_{\rm E}$ would be $r_{\rm E} = h_{\rm E}$. If we now add in the effect that this non-recurrent response has when it is fed back monosynaptically, we find $r_{\rm E} = h_{\rm E} + w_{\rm EE} h_{\rm E}$. Continuing to add in terms that are disynaptic, trisynaptic and so on, we obtain $r_{\rm E} = h_{\rm E} + w_{\rm EE} h_{\rm E} + w_{\rm EE}^2 h_{\rm E} + w_{\rm EE}^3 h_{\rm E} + \ldots$, a series that sums to $h_{\rm E}/(1 - w_{\rm EE})$. Furthermore each term in this series takes a

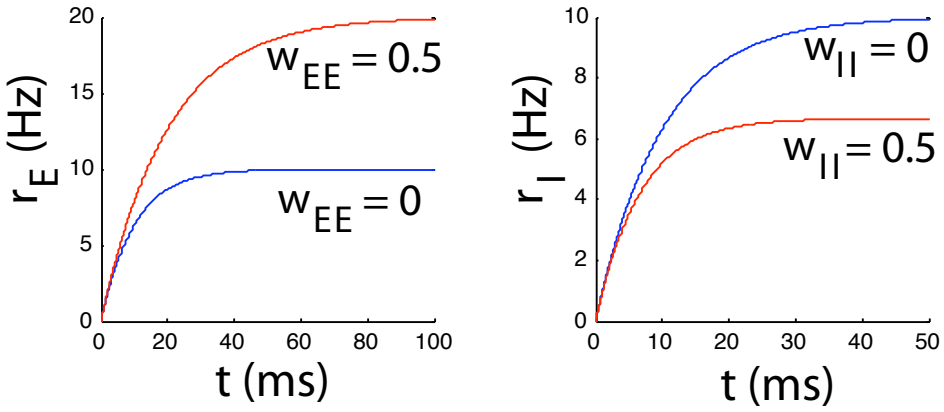

Figure 1: Uniform excitatory and inhibitory networks. a) Response of a recurrent excitatory network to a constant input $h_E = 10$ Hz, activated at time $t = 0$. Increasing the strength of the recurrent excitation from 0 to 0.5 increases the response but makes it rise more slowly. b) Response of a recurrent inhibitory network to a constant input $h_I = 10$ Hz, activated at time $t = 0$. Increasing the strength of the recurrent inhibition from 0 to 0.5 decreases the response but makes it rise more rapidly.

time proportional to the size of the change to be generated by the network, so if the initial change $r_E \rightarrow h_E$ takes time τ, the total change will take $\tau(1 + w_{EE} + w_{EE}^2 + \ldots) = \tau/(1 - w_{EE})$. Thus, we can think of Equation 3 as having summed the polysynaptic series. For a more realistic model, the mathematical expressions may change, but the basic ideas that an input is augmented and integration time slowed by the reverberating circuit does not.

In the limit $w_{EE} \rightarrow 1$, the effective time constant for this network, $\tau/(1 - w_{EE})$, becomes arbitrarily large. Indeed for $w_{EE} = 1$, Equation 2 becomes an equation for a perfect integrator, $\tau dr_E/dt = h_E$. In other words, r_E is simply the time integral of h_E. This means that the neural population adds up or integrates its input without any forgetting or decay. Networks with w_{EE} near 1 are often used as models for neuronal circuits that integrate and remember signals.

A uniform inhibitory population with average rate r_I and average synaptic strength $-w_{II}$ (with $w_{II} > 0$) receiving additional input h_I, is described by an equation similar to 2,

$$\tau \frac{dr_I}{dt} = -r_I - w_{II} r_I + h_I = -(1 + w_{II}) r_I + h_I \tag{4}$$

or

$$\left(\frac{\tau}{1 + w_{II}}\right) \frac{dr_I}{dt} = -r_I + \frac{h_I}{1 + w_{II}}\,.$$

Because of the sign change in the recurrent weight, recurrent inhibition has effects opposite to those of recurrent excitation: it diminishes and speeds up responses (figure 4b). Neural circuits often have extensive mutual inhibition between inhibitory

neurons, something that at first might seem paradoxical. We now see that such recurrent inhibition speeds up the inhibitory response. This can be important for stabilizing a network because inhibition that arrives more quickly gives excessively strong recurrent excitation less time to generate runaway activity.

Networks Stabilized by Inhibition

In our analysis of the single integrate-and-fire neuron, we noted that the excitatory input alone, without inhibitory cancellation, would drive the neuron at extremely high rates. Similarly, experiments with inhibitory blockers suggest that many neuronal circuits are unstable if the effects of inhibition are decreased. In terms of the models we have been discussing, this suggests that the excitatory recurrent strength $w_{\rm EE} > 1$ which, as shown previously, causes a linearized purely excitatory network to be unstable. We now show how inhibition can tame this instability and illustrate an interesting phenomenon related to strong excitatory feedback first revealed by Tsodyks and colleagues.

In this and the following sections, we make an assumption about inhibition that simplifies the analysis considerably without changing the basic phenomena being studied. Inhibitory responses are, in general, quite rapid and they can be accelerated by recurrent inhibition, as discussed previously. We therefore assume that the inhibitory response can be approximated as instantaneous. In the linear approximation, this means that we can set $\frac{dr_{\rm I}}{dt}$ to zero, and thus write

$$r_{\rm I} = \alpha r_{\rm E} + h_{\rm I} \tag{5}$$

where $\alpha = w_{\rm IE}/(1 + w_{\rm II})$ and we now use $h_{\rm I}$ for what would previously have been $h_{\rm I}/(1 + w_{\rm II})$. If we substitute this expression for $r_{\rm I}$ into the linearized equation for the excitatory firing rate,

$$\tau \frac{dr_{\rm E}}{dt} = -r_{\rm E} + w_{\rm EE} r_{\rm E} - w_{\rm EI} r_{\rm I} + h_{\rm E} = (w_{\rm EE} - 1) r_{\rm E} - w_{\rm EI} r_{\rm I} + h_{\rm E}\,, \tag{6}$$

we obtain

$$\tau \frac{dr_{\rm E}}{dt} = (w_{\rm EE} - 1 - \alpha w_{\rm EI})\, r_{\rm E} + h_{\rm E} - w_{\rm EI} h_{\rm I}\,. \tag{7}$$

We have assumed that excitation by itself is unstable, which means, as stated above, that $w_{\rm EE} > 1$. Overall stability requires that the term multiplying $r_{\rm E}$ in Equation 7 be negative: $w_{\rm EE} < 1 + \alpha w_{\rm EI}$. Thus, sufficiently strong inhibition can stabilize the network despite excitation alone being unstable.

For reasons discussed below, we consider the response of this network to an external input purely to the inhibitory neurons, that is, $h_{\rm E} = 0$ and $h_{\rm I} =$ a non-zero constant.

The steady-state firing rate in response to a constant external input h is obtained by setting the right side of Equation 7 to zero and solving for r_{E}. The result is

$$r_{\mathrm{E}} = \frac{-w_{\mathrm{EI}} h_{\mathrm{I}}}{1 - w_{\mathrm{EE}} + \alpha w_{\mathrm{EI}}} .$$

Recalling that $1 - w_{\mathrm{EE}} + \alpha w_{\mathrm{EI}} > 0$, this shows, not surprisingly, that the excitatory rate varies oppositely from the inhibitory input: an input that would appear to suppress inhibition ($h_{\mathrm{I}} < 0$) causes an increase in excitatory firing ($r_{\mathrm{E}} > 0$). The surprise comes when we substitute this expression for r_{E} into Equation 5 for the inhibitory firing rate, which gives the result

$$r_{\mathrm{I}} = \frac{-(w_{\mathrm{EE}} - 1) h_{\mathrm{I}}}{1 - w_{\mathrm{EE}} + \alpha w_{\mathrm{EI}}} .$$

Because of the assumption of instability without inhibition, $w_{\mathrm{EE}} - 1 > 0$, the inhibitory firing rate also varies oppositely to h_{I}. In other words, attempting to drive up the inhibitory firing rate from an external source causes both it and the excitatory firing rate to go down rather than up. Conversely, attempting to suppress inhibitory firing causes both rates to increase.

This effect becomes slightly more intuitive when the temporal sequence involved is considered (Fig. 2). If external excitation is added to the inhibitory population (Fig. 2a), inhibitory firing rates initially increase (Fig. 2b). This drives down excitatory firing rates, resulting in a withdrawal of recurrent excitation onto the inhibitory cells (Fig. 2c). This withdrawal of excitation is *greater* than the increase in excitation to the inhibitory cells that started the process, provided that the network is stable but has $w_{\mathrm{EE}} > 1$. As a result, in the final condition, the inhibitory cells receive less excitation than they did initially, and accordingly their firing rate is decreased (Fig. 2d).

Given the strong recurrent excitation received by cortical excitatory neurons, as well as the instability of cortex when inhibition is reduced, its seems likely that cortical circuits operate in a regime where they are stabilized by inhibition. Thus, we would expect to see the effect just described quite commonly. Though the model is simple, it reveals a more general intuition: if the excitatory subnetwork alone is unstable, it means that an increase in excitatory firing rates causes an increase in recurrent excitation sufficient to drive excitatory rates still higher, and similarly a decrease in excitatory firing rates causes a withdrawal of recurrent excitation sufficient to drive excitatory rates still lower (*e.g.*, when $w_{\mathrm{EE}} > 1$, the change in $\frac{dr_{\mathrm{E}}}{dt}$ induced by a decrease in r_{E} is negative, Equation 6). Thus, if a change in steady state involves a decrease in overall excitatory firing rates, the inhibition received by the excitatory cells must also decrease to compensate for their large withdrawal of recurrent excitation. Similarly, a steady state with increased excitatory firing rates requires an increase in inhibition. Thus, ultimately, both excitation and inhibition must change in the same direction because changes in opposite directions cannot yield a new steady state.

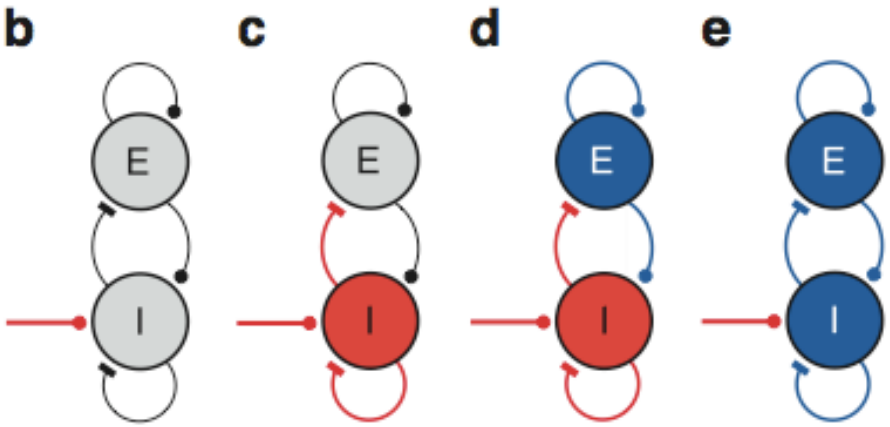

Figure 2: (NOTE: need to redo this fig to change labels b-e to a-d.) Illustration of the sequence of events following addition of excitatory external input to the inhibitory population, signified by red input line, in a network stabilized by inhibition. Gray indicates activity levels before addition of the external input, red indicates increased activity levels, and blue indicates decreased activity levels. See text for details. Figure modified with permission from Ozeki et al., 2007.

In many neurons in the primary visual cortex, an appropriate visual stimulus within the "center" region of the receptive field yields an optimal response, but increasing the size of the stimulus so that it covers a surrounding area (the "surround") reduces the response, a phenomenon known as surround suppression. A stimulus covering only the surround, and not the center, yields no response. It is believed that the center stimulus provides external excitatory input to both excitatory and inhibitory populations, whereas a surround stimulus provides external excitation predominantly onto the inhibitory population. Results from David Ferster's lab indicate that both the inhibition and the excitation that the neuron in primary visual cortex receives are reduced by surround suppression. This may provide an example of the effect we have been discussing.

Nonlinear Analysis of Circuits With Excitatory and Inhibitory Populations

We now consider the effects of nonlinearities on our simple model of one excitatory and one inhibitory population. We use Equations 2 and 4, linking the two populations through weights $w_{\rm EI}$ and $w_{\rm IE}$ as in the linear model. The resulting equations are

$$\begin{aligned} \frac{dr_{\rm E}}{dt} &= -r_{\rm E} + F_{\rm E}[w_{\rm EE}r_{\rm E} - w_{\rm EI}r_{\rm I} + h_{\rm E}] \\ \frac{dr_{\rm I}}{dt} &= -r_{\rm I} + F_{\rm I}[w_{\rm IE}r_{\rm E} - w_{\rm II}r_{\rm I} + h_{\rm I}]\,. \end{aligned} \tag{8}$$

Here we allow for different firing-rate functions for the excitatory and inhibitory

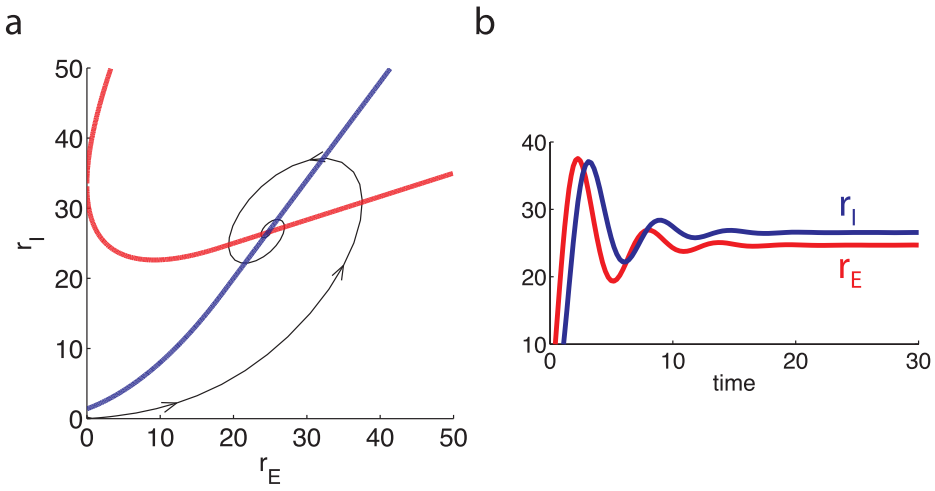

Figure 3: Dynamics of an E-I network. a) Phase plane (r_I versus r_E) for an excitatory-inhibitory network. The red and blue curves are the excitatory and inhibitory nullclines, respectively. The point where these two curves cross determines the equilibrium steady-state values of r_E and r_I. The curve with arrows shows the trajectory of the excitatory and inhibitory rates over time, starting at $r_\mathrm{E} = r_\mathrm{I} = 0$ and ending at the equilibrium point. b) The excitatory and inhibitory rates corresponding to the trajectory in a, plotted as a function of time.

neurons. These equations support a variety of behaviors for different parameters, including steady-state and oscillatory activity, as first explored by Wilson and Cowan.

Suppose that h_E and h_I are constant inputs. The first step in analyzing Equations 8 is to determine whether they produce constant excitatory and inhibitory firing rates in this case and, if so, to determine the steady-state values of r_E and r_I. By definition, steady-state values do not change in time, so their derivatives must be zero. Therefore, constant steady-state firing rates produced by this system must satisfy the conditions

$$r_\mathrm{E} = F_\mathrm{E}(w_\mathrm{EE} r_\mathrm{E} - w_\mathrm{EI} r_\mathrm{I} + h_\mathrm{E}) \quad \text{and} \quad r_\mathrm{I} = F_\mathrm{I}(w_\mathrm{IE} r_\mathrm{E} - w_\mathrm{II} r_\mathrm{I} + h_\mathrm{I})$$

that make the right sides of Equations 8 zero. These two equations define two different relationships between r_E and r_I. The curves along which these relationships hold are called nullclines. In Figure 3, the red curve in the plane defined by the variables r_E and r_I (known as the phase plane) is the set of points for which $r_\mathrm{E} = F_\mathrm{E}(w_\mathrm{EE} r_\mathrm{E} - w_\mathrm{EI} r_\mathrm{I} + h_\mathrm{E})$, and the blue curve is where $r_\mathrm{I} = F_\mathrm{I}(w_\mathrm{IE} r_\mathrm{E} - w_\mathrm{II} r_\mathrm{I} + h_\mathrm{I})$. For points along the red curve, $dr_\mathrm{E}/dt = 0$, so r_E cannot change, and similarly for points along the blue curve, $dr_\mathrm{I}/dt = 0$. For a stable network, Equations 8 sets up flows in the r_E-r_I plane that move toward these nullclines. The point where the two nullclines cross is the equilibrium point of the system where neither r_E nor r_I changes, and the values of r_E and r_I at the crossing point determine the steady-state excitatory and inhibitory firing rates.

The evolution of $r_{\rm E}$ and $r_{\rm I}$ from initial values $r_{\rm E} = r_{\rm I} = 0$ to their steady-state values is indicated by the curve with arrows on it in Figure 3a and, as a function of time in Figure 3b. For the parameters we have used, the trajectory follows a spiral in the phase plane or, equivalently, the firing rates oscillate over time as they approach their steady-state values. Note that when the trajectory crosses a nullcline, it is always either vertical (when crossing the excitatory nullcline) or horizontal (when crossing the inhibitory nullcline). This is required by the fact that the derivative of the corresponding firing rate must vanish along the nullcline.

If we are only interested in small deviations of the firing rates about their equilibrium values, we can approximate the nullclines near their crossing point by straight lines. This is equivalent to the linear approximation we have discussed. The slopes of these straight lines determine the response gains, equivalent to the factor g we introduce previously. An important feature of nonlinear systems is that their gains can change depending where on the curvilinear nullclines the equilibrium point lies. Thus, a nonlinear system can be approximated by a series of linear systems with different gains, something called a piecewise linear approximation.

In terms of the phase-plane analysis, the assumption we made in the previous section that inhibitory responses are instantaneous means that we assume that the inhibitory firing rate always stays on its nullcline (the blue curve in Figure 3a). Our analysis of networks stabilized by inhibition can be generalized: the condition that the excitatory network alone be unstable, which means $w_{\rm EE} > 1$ for the linear network, corresponds more generally to the condition that the excitatory nullcline (the red curve in Figure 3a) has a positive slope at the equilibrium point (where the red and blue curves cross)). Indeed we can expand F_E around the fixed point r_E^0, r_I^0:

$$F_E(I) = F_E(I^0) + \frac{\partial F_E}{\partial I}\frac{\partial I}{\partial r_E}\Delta r_E + \frac{\partial F_E}{\partial I}\frac{\partial I}{\partial r_I}\Delta r_I + ...$$

where $I = w_{\rm EE} r_{\rm E} - w_{\rm EI} r_{\rm I} + h_{\rm E}$. Given that $-r_E^0 + F_E(I^0) = 0$ (because r_E^0, r_I^0 is a fixed point and, there, both derivatives are zero), we get that the nullcline is given by the equation:

$$\Delta r_I = \frac{1}{F_E' w_{\rm EI}}(F_E' w_{\rm EE} - 1)\Delta r_E$$

where $F_E' = \frac{\partial F_E}{\partial I}$.

Adding external excitation to the inhibitory population corresponds to shifting the inhibitory nullcline to the left (because of the additional excitation, a smaller value of $r_{\rm E}$ yields $\frac{dr_{\rm I}}{dt} = 0$ for a fixed $r_{\rm I}$). If the excitatory nullcline has positive slope, as in Figure 3a, then moving the inhibitory nullcline leftward causes the equilibrium values of both $r_{\rm E}$ and $r_{\rm I}$ to decrease. That is, adding external input to the inhibitory population causes $r_{\rm E}$ and $r_{\rm I}$ to move in the same direction.

Circuits for Decision Making

We now apply these mathematical tools to a network model of decision making. Following X-J. Wang, we consider two populations of excitatory neurons, each corresponding to one decision outcome. For example, suppose a person is driving and needs to decide whether to turn right or left. We assume that there are two populations of excitatory neurons, one active when the decision is a right turn, the other when it is a left turn. We do not model neurons that then transform this decision pattern of activity into a motor act, such as turning the steering wheel. Under some circumstances no decision needs to made made, so neither population should be active. When a decision is required, it can either be biased or unbiased by sensory input. If, for example, if the sensory stimulus is a road sign that says "turn left', the sensory input should favor a decision and bias it toward a left turn. If the sensory stimulus is an obstacle in the middle of a three lane highway, there is a need to turn but the direction is irrelevant. In this case, the sensory input should favor a decision without biasing the decision. Between these extremes, input may provide a range of biasing effects.

In order to simplify the analysis, we will consider specifically the example of two possible decisions, which would correspond to the simplest circuit. The network will be made of two excitatory and one inhibitory populations. We would like to know whether it is possible to build a pattern of synaptic connections between the different populations such that: 1) In the absence of relevant sensory stimuli there is only one stable pattern of spontaneous activity which corresponds to a state of "no decision". 2) The sensory stimulus that triggers the decision should select one of the two stable patterns of activities corresponding to the decisions. In particular, in each case the excitatory population corresponding to the intended motor response should be active, and the activity of the excitatory population corresponding to the alternative decision should be suppressed. One possible mechanism to implement such a system would be based on inhibition-mediated competition between the two excitatory populations: when the activity of one population grows, inhibition increases and it suppresses the activity of the other population. The input to these populations generated by the sensory stimulus should trigger the decision and it should generate the bias for one decision or another.

We let r_1 and r_2 be the firing rates of the two excitatory populations, and r_I be the firing rate of the inhibitory population. We begin by assuming no excitatory coupling between the excitatory populations, but that each receives identical input from the inhibitory population:

$$\tau \frac{dr_1}{dt} = -r_1 + F(w_{\mathrm{EE}} r_1 - w_{\mathrm{EI}} r_{\mathrm{I}} + h_1)$$

$$\tau \frac{dr_2}{dt} = -r_2 + F(w_{EE} r_2 - w_{EI} r_{\mathrm{I}} + h_2)$$

As in the previous section, we assume that the inhibition responds instantaneously. We do not include any external input for the inhibition, but we assume that it responds equally to the firing of both excitatory populations. Thus, we write $r_{\mathrm{I}} = \alpha(r_1 + r_2)$. Substituting this into the above equations gives

$$\tau \frac{dr_1}{dt} = -r_1 + F((w_{\mathrm{E}} - w_{\mathrm{I}}) r_1 - w_{\mathrm{I}} r_2 + h_1)$$
$$\tau \frac{dr_2}{dt} = -r_2 + F((w_{\mathrm{E}} - w_{\mathrm{I}}) r_2 - w_{\mathrm{I}} r_1 + h_2)$$

where we have defined $w_{\mathrm{E}} = w_{\mathrm{EE}}$ and $w_{\mathrm{I}} = \alpha_{\mathrm{I}} w_{\mathrm{EI}}$ to simplify the notation. Note that, if we had included excitatory coupling between r_1 and r_2 of strength $w_{12} = w_{21}$, the equations would have the identical form with $w_{\mathrm{I}})$ and $w_{\mathrm{E}})$ replaced by $w_{\mathrm{I}}) - w_{12}$ and $w_{\mathrm{E}}) - w_{12}$ respectively.

In order to study the system we compute the nullclines. We first assume that the sensory inputs h_1 and h_2 are equal ($h_1 = h_2 = h$), which would correspond to the unbiased case in which both decisions are equivalent. In all the interesting cases the nullclines cross at least at one symmetric point, where $r_1 = r_2$. Such a point corresponds to a pattern of activity which does not express any preference for one choice or another, and hence it is a "no decision" configuration. The stability of the symmetric point depends on the slope of the nullclines, which in turn, depends on both the synaptic weights, and on the shape of the neuronal response function, as explained in the case of one excitatory and one inhibitory population. Depending on the external input h, the system can operate in different regimes and the symmetric point can be either stable or not. Indeed h determines the target firing rates of the neurons and hence their sensitivity to modifications of the input. In particular the neurons are maximally reactive when they operate in the linear regime, where the slope of the neuronal response function is maximal. The sensitivity decreases smoothly if the neuronal firing rate goes to zero or if it saturates.

Let us now analyze in detail the symmetric case. The fact that the two inputs are equal, implies symmetry, that in turn, has as a consequence that for any fixed point r_1, r_2, there is a corresponding point r_2, r_1. These two points collapse into a single fixed point if the nullclines cross at a point where $r_1 = r_2$. In all the interesting cases, such a symmetric fixed point exists and it is useful to start the analysis by focusing on the stability properties of such a point. Indeed all patterns of activity for which $r_1 = r_2$ do not express any preference for one choice or another, and hence correspond

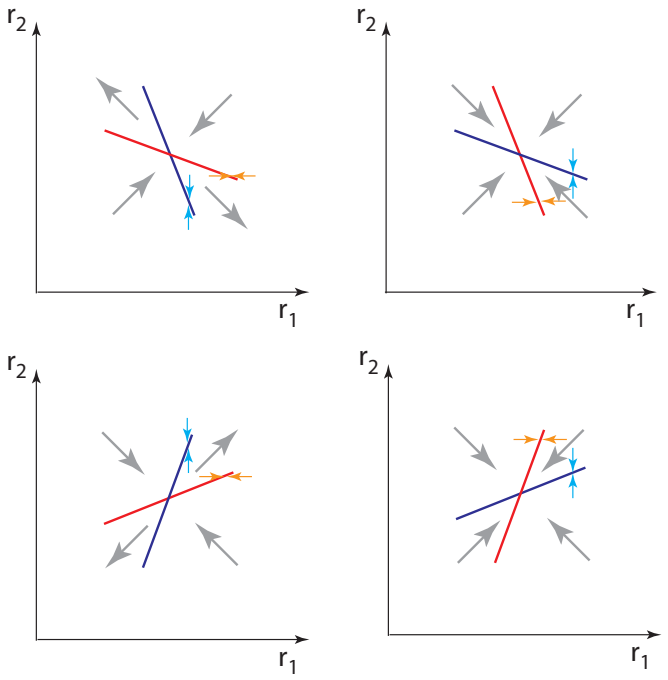

Figure 4: The stability of the symmetric "no decision" point: the two nullclines (red for r_1 and blue for r_2) cross in one point. We focus on what happens in a neighborhood of such a point. Here we illustrate 4 cases corresponding to 4 different slopes of the nullclines. The grey arrows indicate the final direction of the movement. Case A is the desired configuration: attractive along the diagonal of the r_1, r_2 plane, and repulsive in the orthogonal direction.

to a "no decision" configuration. If we want to force the system to decide, we then need to make the symmetric point repulsive. The best way would be to make it what is called a saddle point, something resembling a mountain pass connecting two valleys corresponding to the two decisions (i.e. the two desired attractors in which r_1 is high and r_2 is low and vice versa). The system would naturally flow into either one or the other valley, but it would not be able to climb up from the pass in the direction of the surrounding peaks. These peaks lie along the diagonal $r_1 = r_2$ and soar on the "no decision" path, which is to be avoided. Summarizing, any fixed point along the line over which $r_1 = r_2$ should be attractive in the direction of this line, because we do not want configurations in which both r_1 and r_2 go to infinity or to zero, but along the orthogonal direction it should be repulsive to force the system to either activate r_1 and suppress r_2 or vice versa. A simple analysis of all possibilities shows that there is only one way to obtain such a behavior: the parameters should tuned in such a way that: 1) both nullclines have a negative slope, and 2) the nullcline of r_2 is steeper than the nullcline of r_1 (see Figure 4).

How do we determine quantitatively the slopes of the nullclines given a particular set of parameters? If we approximate the neuronal transfer function with its tangent at the fixed point (linearization), we find after some simple algebra that the two slopes are

$$\frac{\tilde{w}_{\mathrm{E}} - \tilde{w}_{\mathrm{I}} - 1}{\tilde{w}_{\mathrm{I}}} \quad \text{and} \quad \frac{\tilde{w}_{\mathrm{I}}}{\tilde{w}_{\mathrm{E}} - \tilde{w}_{\mathrm{I}} - 1}$$

for r_1 and r_2 respectively, where $\tilde{w}_{\mathrm{E}} = F' w_{\mathrm{E}}$ and $\tilde{w}_{\mathrm{I}} = F' w_{\mathrm{I}}$ and F' is the derivative of the neuronal response function F with respect to the total current evaluated at the fixed point. F' depends on the shape of neuronal response function in our specific case, but it might depend on several other sources of nonlinearity as adaptation or synaptic depression and facilitation.

In the case of small input h, the neurons operate in low firing rate, fluctuation dominated regime, in which F' is small. The nullclines cross in such a way that the symmetric, no decision point, is stable (Figure 5a). In particular we have that both nullclines have a negative slope ($\tilde{w}_{\mathrm{E}} < \tilde{w}_{\mathrm{I}} + 1$) and the r_2 nullcline (blue) is less steep than the r_1 nullcline (red). The second condition corresponds to $\tilde{w}_{\mathrm{E}} > 1$ (indeed it is corresponds to imposing that the slope of the r_1 nullcline is negative, but not too steep $(\tilde{w}_E - \tilde{w}_I - 1)/\tilde{w}_I > -1$).

If we increase the external input h, as in presence of a cue which should trigger a decision, the system switches to a different regime in which it is forced to decide because the symmetric point becomes repulsive. Indeed, as h increases, the symmetric point becomes what is called a saddle point, something resembling a mountain pass connecting two valleys corresponding to the two decisions (i.e. the two desired attractors in which r_1 is high and r_2 is low and vice versa). The system would naturally flow into either one or the other valley, but it would not be able to climb up from the pass in the direction of the surrounding peaks. These peaks lie along the diagonal $r_1 = r_2$ and soar on the "no decision" path, which is to be avoided. Summarizing, the symmetric fixed point along the line over which $r_1 = r_2$ would be attractive in the direction of this line, avoiding configurations in which both r_1 and r_2 go to infinity or to zero, but along the orthogonal direction it would be repulsive to force the system to either activate r_1 and suppress r_2 or vice versa.

As h increases, two things happen: 1) the symmetric fixed point moves to higher firing rates, and 2) F' also increases, making the r_2 nullcline (blue) steeper, and the r_1 nullcline (red) less steep to the extent that both nullclines have a negative slope, and the nullcline of r_2 becomes steeper than the nullcline of r_1 (see Figure 5b).

In this case $\tilde{w}_{\mathrm{E}} < \tilde{w}_{\mathrm{I}} + 1$, which essentially means that inhibition is large enough to dominate over excitation. This is not surprising given that we need to build a sufficiently strong competition between the two excitatory populations. Moreover the r_2 nullcline is steeper than the r_1 nullcline, which happens when $\tilde{w}_{\mathrm{E}} > 1$. The second condition says that each excitatory population would be unstable, like the

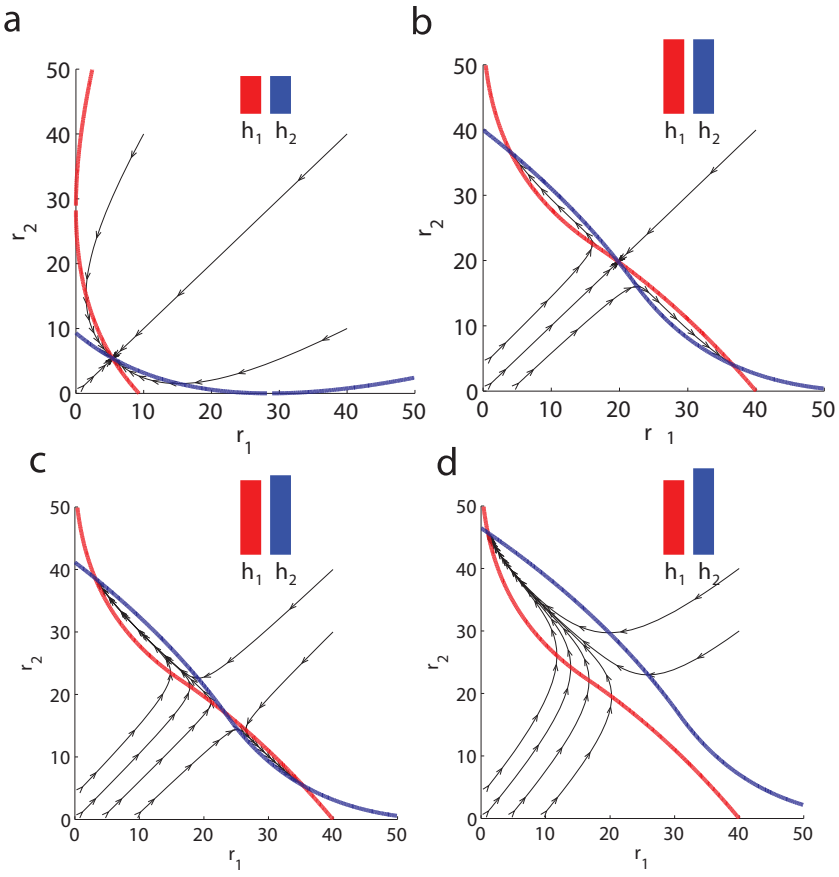

Figure 5: Nullclines for different regimes. a) No decision ($h_1 = h_2 = 20$). b) Unbiased decision ($h_1 = h_2 = 40$). c) Weakly biased decision ($h_1 = 40, h_2 = 43$). d) Strongly biased decision ($h_1 = 40, h_2 = 46$). In all cases $w_{\mathrm{E}} = 1.4, w_{\mathrm{I}} = 0.7$

excitatory population in an ISN, and its activity would explode were it not for the regulatory effects of inhibition. Such a condition is needed to amplify any deviation in the direction orthogonal to the diagonal $r_1 = r_2$. One of the two populations is then guaranteed to run away rather then returning back to the diagonal.

What happens when the network moves away from the symmetric fixed point? If r_1 is larger than r_2, then r_1 keeps increasing at the expense of r_2. As r_2 cannot become negative (i.e. the response function F is zero for negative currents), the process would stop when the activity of the winning population r_1 brings r_2 to zero. However, as r_1 increases, and r_2 decreases, both populations tend to leave the regime in which the neuronal response function is linear. In particular, the saturation due to the refractory period slows down the growth of r_1, and r_2 decays smoothly to zero because when the average input goes below threshold, the neurons in population 2 can still fire, driven by the subthreshold fluctuations. This translates into a bending of the nullclines which leads to the formation of other two fixed points (see Figure

5b). One corresponds to a pattern of activity in which r_1 is elevated and r_2 is low, and the other to one in which r_2 wins over r_1. The two points lie symmetrically around the diagonal $r_1 = r_2$. These are the two stable fixed points which correspond to the two possible decisions.

Notice that if h increases even further F' again decreases to zero, and the slopes of the nullclines change in such a way that the symmetric point becomes stable again, as in case a of Figure 5. The bending of the nullclines is due to saturation, in the case of large sensory input. It is surprising that for too large inputs the two excitatory populations would stop competing.

Every time the sensory input is modified, the nullclines shift. In the case in which the input is biased, for example when h_2 increases and h_1 remains the same, the r_2 nullcline shifts up (see Figure 5c and d) and the intersection point with the r_1 nullcline moves right-down. So now any trajectory that starts from a symmetric point $r_1 = r_2$ falls into the basin of attraction of the stable fixed point corresponding to the decision 2. In other words, as h_2 increases, the bias to be attracted toward a point at which r_2 suppresses r_1 becomes progressively larger and eventually it makes the decision 1 attractor disappear (see Figure 5d).

References

Destexhe A, Par, D (1999) Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J Neurophysiol 81:1531-1547.

Fusi S, Wael A, Miller EK, Wang X-J (2007) A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales, Neuron, in press

Hô N, Destexhe A (2000) Synaptic background activity enhances the responsiveness of neocortical pyramidal neurons. J Neurophysiol 84:1488-1496.

Holt GR, Softky WR, Koch C, Douglas RJ (1996) Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. J Neurophysiol 75:1806-1814.

Kenet T, Bibitchkov D, Tsodyks M, Grinvald, A, Arieli, A (2003) Spontaneously emerging cortical representations of visual attributes. Nature 425:954-956.

Ozeki H, Schaffer ES, Miller KD, Ferster D (2007) Surround suppression in cat visual cortex: Evidence that V1 operates as an inhibition-stabilized network. (submitted).

Rapp M, Yarom Y, Segev I (1992) The impact of parallel fiber background activity on the cable properties of cerebellar Purkinje cells. Neural Comp. 4: 518-532.

Ricciardi LM (1977) Diffusion processes and related topics in biology. Berlin: Springer-Verlag.

Rinzel J, Ermentrout GB (1998) Analysis of neural excitability and oscillations. In C Koch, I Segev, eds. Methods in Neuronal Modeling. MIT Press: Cambridge MA. pp. 251-291.

Salinas E, Sejnowski TJ (2002) Integrate-and-fire neurons driven by correlated stochastic input. Neural Comp 14:2111-2155.

Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. Curr. Opin Neurobiol 4:569-579.

Shelley M, McLaughlin D, Shapley R, Wielaard J (2002) States of high conductance in a large-scale model of the visual cortex. J Comput Neurosci 13:93-109.

Shriki O, Hansel D, Sompolinsky H (2003) Rate models for conductance-based cortical neuronal networks. Neural Comput 15:1809-1841.

Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13:334-350.

Troyer TW, Miller KD (1997) Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. Neural Comp 9:971-983.

Tsodyks MV, Skaggs WE , Sejnowski, TJ, McNaughton BL (1997) Paradoxical effects of external modulation of inhibitory interneurons. J Neurosci 17:4382-4388.

Wilson, HR, Cowan, JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12:1-24.

Wang X-J (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron, 36, 955-968

Wong K-F and Wang X-J (2006) A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. J. Neurosci., 26:1314-1328