

Section: Behavioral/Systems Neuroscience
Section Editor: Dr. Thomas J. Carew

Paradoxical Effects of External Modulation of Inhibitory Interneurons

Misha V. Tsodyks^{1,2,3}, William E. Skaggs², Terrence J. Sejnowski^{3,4}
and Bruce L. McNaughton²

1 Department of Neurobiology, Weizmann Institute
Rehovot 76100, Israel

and

2 Arizona Research Laboratories, Division of Neural Systems, Memory and Aging
University of Arizona, Tucson AZ 85724

and

3 Howard Hughes Medical Institute, Computational Neurobiology Laboratory
The Salk Institute for Biological Studies, La Jolla CA 92037

and

4 Department of Biology, University of California at San Diego
La Jolla, CA 92093

Running Title: *Effects of modulation of inhibitory interneurons*

18 text pages, 6 figures

237 words in the abstract, 463 words in the introduction, 687 words in the discussion

Acknowledgments: This research was sponsored in part by the Howard Hughes Medical Institute and the Wasie Foundation (TJS), PHS grant NS20331 (BLM) HFSP short-term fellowship SF-379/95 (MVT), and support from the McDonnell-Pew Program in Cognitive Neuroscience (WES).

ABSTRACT. The neocortex, hippocampus, and several other brain regions contain populations of excitatory principal cells with recurrent connections and strong interactions with local inhibitory interneurons. To improve our understanding of the interactions between these cell types, we modeled the dynamical behavior of this type of network, including external inputs. A surprising finding was that increasing the direct external inhibitory input to the inhibitory interneurons, without directly affecting any other part of the network, can in some circumstances cause the interneurons to increase their firing rates. The main prerequisite for this paradoxical response to external input is that the recurrent connections among the excitatory cells be strong enough to make the excitatory network unstable when feedback inhibition is removed. Because this requirement is met in the neocortex and several regions of the hippocampus, these observations have important implications for understanding the responses of interneurons to a variety of pharmacological and electrical manipulations. The analysis can be extended to a scenario with periodically varying external input, where it predicts a systematic relationship between the phase shift and depth of modulation for each interneuron. This prediction was tested by recording from interneurons in the CA1 region of the rat hippocampus in vivo, and the results broadly confirmed the model. These findings have further implications for the function of inhibitory and neuromodulatory circuits, which can be tested experimentally.

Several regions of the mammalian brain, including the neocortex and hippocampus, consist largely of intermixed excitatory and inhibitory subpopulations (see (Jones, 1986; Amaral and Witter, 1995)). The excitatory cells are generally more numerous and project extensively to each other as well as to the inhibitory cells. The inhibitory cells, which are primarily GABAergic, project strongly to the excitatory cells; recent evidence indicates that they also project to each other (Sik et al., 1995). In the hippocampus and neocortex, complete blockade of inhibition with GABA antagonists leads to runaway activity in the excitatory cells, culminating in an epileptic seizure (Traub and Miles, 1991; Grinvald et al., 1988). Thus interneurons govern the activity of principal cells in somewhat the same sense that shepherds govern the activity of sheep.

Because this type of neural organization appears so commonly in the brain, it is important to have a good understanding of its dynamical properties. The analysis presented here was originally motivated by an observation while simulating an integrate-and-fire model of the hippocampal theta rhythm (Tsodyks et al., 1996) - a strong, regular oscillation that dominates the hippocampal electroencephalogram (EEG) of some mammals during behavioral states of active movement, rapid eye-movement sleep, or light dissociative anesthesia (Vanderwolf, 1969). Theta oscillations are controlled by inputs from the medial septal area, and vanish from the hippocampus if the medial septum is lesioned or inactivated. A substantial fraction—probably more than half—of the projection from medial septum to hippocampus arises from GABAergic cells, and terminates almost exclusively on interneurons (Freund and Antal, 1988). It is therefore generally believed that the theta-rhythmic activity of hippocampal cells is entrained by rhythmic inhibition of inhibitory interneurons.

The model consisted of two pools of neurons, one excitatory and the other inhibitory; and the hippocampal theta rhythm was modeled as an external, rhythmically varying, *inhibitory*

input to the *inhibitory* neurons. When the model was simulated, we noticed, to our surprise, that the excitatory and inhibitory pools both oscillated in synchrony with the external input and thus in synchrony with each other (Fig. 1). This seemed quite paradoxical: if the only external input was to the inhibitory cells, a decrease in their activity would be expected to provoke an increase in the activity of the excitatory cells, so that the two would oscillate 180° out of phase. In an effort to understand this phenomenon, we constructed a simplified average-firing-rate model of the network whose dynamics could be examined analytically. A straightforward phase plane analysis shows that a “paradoxical” response of inhibitory neurons to external modulation is a very general feature of this type of network, and can be expected to be observable in real brains. The model is abstract, but its essential features are quite robust, and the main conclusions of the present analysis have been verified using more realistic integrate-and-fire models. We therefore used simultaneous recordings from interneurons and pyramidal cells of rat hippocampus to test some of the predictions of the model which relate the phase shift and depth of modulation of interneurons.

Methods

Experimental Procedure

The experimental data considered in this paper were recorded using methods that have previously been described in detail (Skaggs et al., 1996). Briefly, hippocampal unit activity was recorded using tetrodes, which consisted of four $12\text{-}\mu\text{m}$ wires twisted together. The tetrodes were placed in or near the CA1 cell body layer of the dorsal hippocampus. Different cells recorded from a single tetrode were distinguished on the basis of their spike amplitudes on the four tetrode channels. A cell was classified as an interneuron if (1) it had a narrow spike waveform, less than $300\ \mu\text{sec}$ from peak to valley, (2) it did not fire in complex-spike bursts, and (3) it had a mean rate greater than 5 Hz averaged across the whole session. The hippocampal EEG was recorded from a separate electrode positioned near the fissure separating the CA1 region from the dentate gyrus, because this is the best location for recording large, robust theta waves. Theta phases were calculated by digitally filtering the EEG signal with a bandpass of 6–10 Hz and then using the peaks of the resulting waves as reference points. The phase of an interneuron and its depth of modulation were obtained by plotting a histogram of the interneuron’s firing rate for different theta phases of EEG, and comparing this with a similar histogram for the whole population of pyramidal neurons recorded simultaneously during the same session. An example of such a histogram for one of the interneurons is shown in Fig. 2. The depth of modulation of the firing rate obtained from this plot was then normalized by the maximal firing rate.

The Model

Consider a recurrent network model consisting of two populations: N_e excitatory neurons and N_i inhibitory neurons. In a coarse-grained description, detailed specification of activity in each individual neuron can be replaced by the average activity of the corresponding population (the fraction of neurons active within a certain time window around t). This

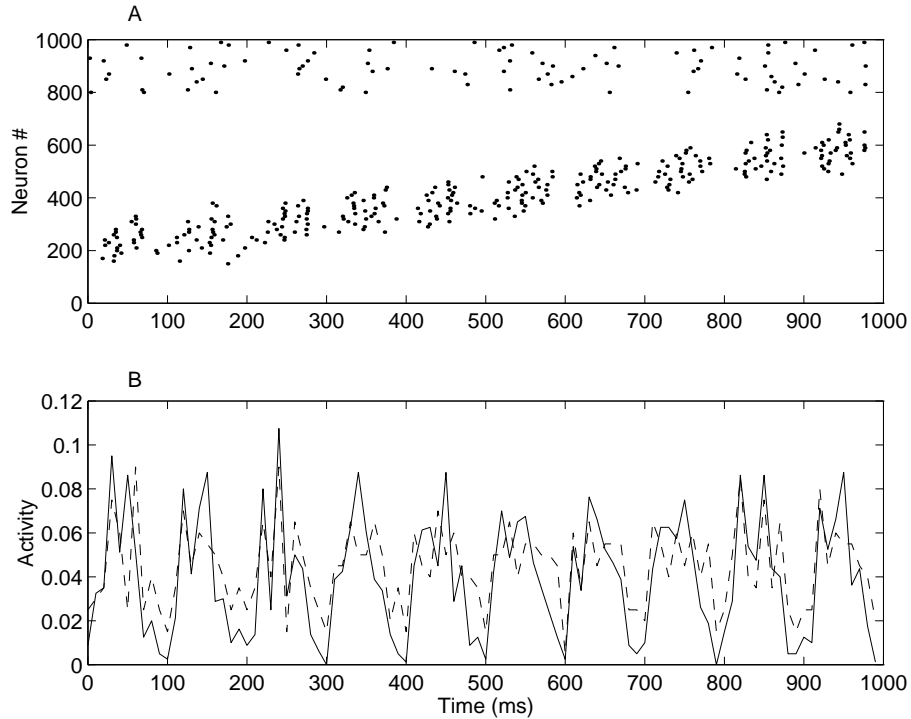


Figure 1: A, Spiking activity in an interconnected network of integrate-and-fire neurons studied as a model of the hippocampus during a simulating a rat running through a linear apparatus (Tsodyks et al., 1996). The network consisted of 800 excitatory and 200 inhibitory neurons. The spike train of each neuron, i , is plotted as ticks along a horizontal line representing the times at which spikes were emitted. Excitatory neurons (1 – 800) were ordered according to the locations of their place fields such that cells with neighboring indices had overlapping fields. The spikes of every tenth neuron are shown. The movement of the rat along the apparatus was simulated by external input to successive groups of pyramidal cells. The theta rhythm was induced in the network by applying oscillatory input (not shown) to the inhibitory neurons. Details of the model are described in (Tsodyks et al, 1996). B, Average activity of excitatory (solid line) and inhibitory (dashed line) populations plotted as a fraction of neurons from each population that fired within time bins of 10 msec. The excitatory and inhibitory populations are in-phase with the driving inhibitory inputs to the inhibitory neurons.

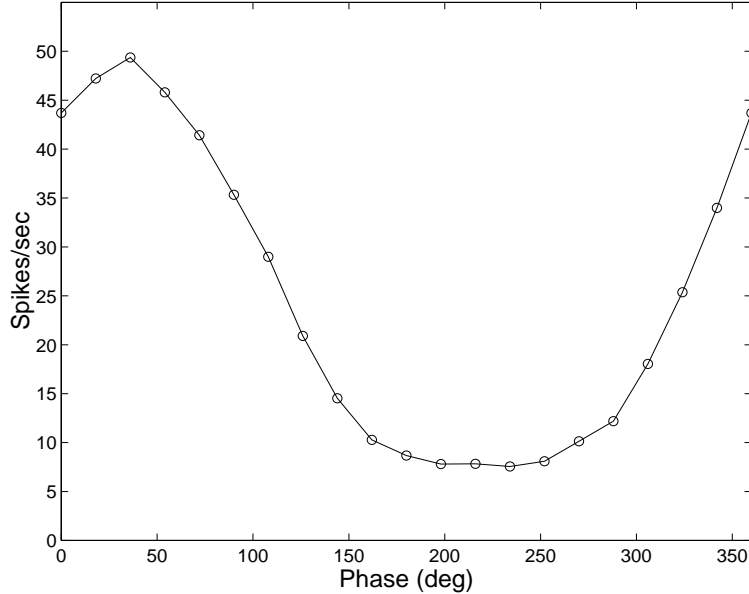


Figure 2: Histogram of firing phases relative to the theta rhythm for one interneuron recorded in the hippocampus. The firing rate in spikes per second plotted a function of phase with bins of 8° . The phases were calibrated so that 0° (360°) corresponds to a maximum activity of the population of pyramidal cells recorded simultaneously with a given interneuron.

level of description can be justified if one is interested in the average behavior of cells with similar response properties (e.g. hippocampal cells with overlapping place fields, or cells with similar receptive fields and preferred orientations in visual cortex). The time evolution of the average excitatory activity $E(t)$ and inhibitory activity $I(t)$ is governed by temporally coarse-grained equations (Wilson and Cowan, 1972):

$$\tau \frac{dE}{dt} = -E + g_e(J_{ee}E - J_{ei}I + e(t)), \quad (1)$$

$$\tau' \frac{dI}{dt} = -I + g_i(J_{ie}E - J_{ii}I + i(t)), \quad (2)$$

where $g_e(x)$ and $g_i(x)$, called the *response functions*, are the proportions of cells firing in the two populations for a given level of input activity x . The strengths of the interactions in the model are controlled by the parameters J . For example, J_{ee} is the product of the average number of recurrent excitatory contacts per cell and the average postsynaptic current due to one presynaptic action potential on the postsynaptic cell. It is assumed that J_{ee} , J_{ei} , J_{ie} , and J_{ii} are all positive. The time constants of the excitatory and inhibitory populations are τ and τ' , respectively; these constants describe the time needed to bring neurons to firing as they receive subthreshold excitation and are comparable to the membrane time constants of these neurons (approximately 10–20 msec (Abeles, 1991)). Finally, $e(t)$ denotes the average external input received by the excitatory population from other brain regions, and $i(t)$ is the external input to the inhibitory population. A schematic diagram of the network is shown in Fig. 3.

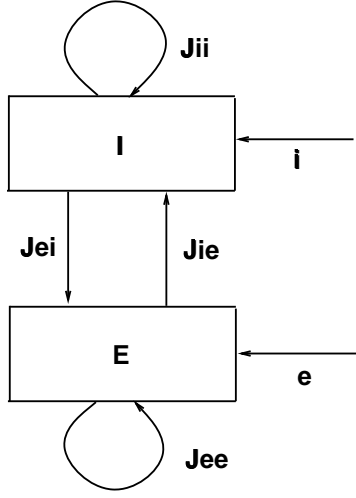


Figure 3: Schematic diagram of the network model. The excitatory population (E) is connected to itself through recurrent excitatory connections (J_{ee}) and the inhibitory population (I) has recurrent inhibitory connections J_{ii} . The strengths of the interactions between these two populations are given by J_{ei} and J_{ie} . The external inputs are e to the excitatory population and i to the inhibitory population.

The response functions $g_e(x)$ and $g_i(x)$ are monotonically increasing functions of x , ranging from 0 to 1, and usually assumed to have a sigmoidal shape. To make analysis of the network simpler, we consider first a threshold-linear function with constant slope and saturation:

$$g(x) = \begin{cases} 0 & \text{if } x < \theta \\ \beta(x - \theta) & \text{if } \theta < x < \theta + 1/\beta \\ 1 & \text{if } x > \theta + 1/\beta \end{cases} \quad (3)$$

The main conclusions of the analysis can be easily extended to the general case in which the slope of the response function β varies with the activity level.

Steady state solutions.

Differential equations 1, 2 define the evolution of the activity state of the network (e.g. relaxation to a steady state), which can be viewed as a line in the *phase plane* of the variables E and I . If the external inputs $e(t)$ and $i(t)$ change slowly compared to the time constants τ and τ' , so that the network has enough time to settle into the steady state solution of Eqns. 1 and 2, then, at each moment, the activities of the two populations are given by:

$$E = \beta(J_{ee}E - J_{ei}I - \theta + e(t)) \quad (4)$$

$$I = \beta(J_{ie}E - J_{ii}I - \theta + i(t)) \quad (5)$$

For fixed values of the parameters and the functions $e(t)$ and $i(t)$, these equations define two curves in the E, I plane, which are called the *nullclines* of the differential equations 1 and

2. If the state of the network is described by a point (E, I) lying on one of the nullclines, say the first one, the corresponding time derivative ($\frac{dE}{dt}$ in this case) is zero and thus the evolution line of the system goes parallel to the I axis. Any point at which the nullclines intersect is a fixed point of the system of equations 1 and 2, where both derivatives are zero and the system is therefore in a steady state.

Note that the solution of Eqs. 4 and 5 is only relevant if it is *locally stable*; that is, if under a small perturbation the network returns to a steady state. There are two conditions under which this is guaranteed to happen: first, if the strength of the recurrent excitation is weak (namely, $\beta J_{ee} < 1$), in which case the excitatory population is stable regardless of the other interactions; second, if the recurrent excitation is strong ($\beta J_{ee} > 1$), but other interactions are also strong enough.

Technically, the stability condition for the fixed point given by eqns. 4 and 5 is that the matrix of the coefficients of eqns. 1 and 2, linearized around the fixed point, must have eigenvalues with negative real part. For our choice of response function, these eigenvalues are given by the expression,

$$\lambda_{\pm} = \frac{1}{2} \left(\frac{\beta J_{ee} - 1}{\tau} - \frac{\beta J_{ii} + 1}{\tau'} \right) \pm \frac{1}{2} \sqrt{\left(\frac{\beta J_{ee} - 1}{\tau} + \frac{\beta J_{ii} + 1}{\tau'} \right)^2 - 4\beta^2 \frac{J_{ei} J_{ie}}{\tau \tau'}}. \quad (6)$$

In order for both eigenvalues to have negative real part, the first term of this expression must be negative. This condition can be solved to yield the requirement,

$$\frac{(\tau'/\tau)(\beta J_{ee} - 1) - 1}{\beta} < J_{ii}. \quad (7)$$

This is sure to hold if $\beta J_{ee} < 1$, because J_{ii} was assumed to be positive; but even if not, the condition will hold if J_{ii} is large enough. It is perhaps surprising that J_{ii} is the necessary factor for stability of the steady state: the reason for this is that when J_{ee} is large and J_{ii} is small, the only possible stable states are $E = 0$ or $E = 1$; the coefficients J_{ie} and J_{ei} can only determine to which of these stable states the system converges. Note also that there is an additional requirement for stability in this case: in order for both eigenvalues given by eqn. 6 to have negative real part, it is also necessary that the product,

$$4\beta^2 \frac{J_{ei} J_{ie}}{\tau \tau'}, \quad (8)$$

be sufficiently large. That is, if the recurrent excitation is strong, then stability exists only if both the recurrent inhibition and the excitation-inhibition interaction are strong enough.

If these conditions are met, then the equilibrium state of the system is given by the solution of equations 4, 5, namely,

$$E = \frac{\beta}{\lambda} ((1 + \beta J_{ii})(e(t) - \theta) - \beta J_{ei}(i(t) - \theta)) \quad (9)$$

$$I = \frac{\beta}{\lambda} (\beta J_{ie}(e(t) - \theta) + (1 - \beta J_{ee})(i(t) - \theta)), \quad (10)$$

where

$$\lambda = \beta^2 J_{ie} J_{ei} + (1 - \beta J_{ee})(1 + \beta J_{ii}). \quad (11)$$

The stability analysis presented above is performed for a simplified choice of the response function with the fixed slope β . As was mentioned, if the response function has a general sigmoid shape, the slope may vary with the activity level. As a result, for a given strength of all the interactions, the steady state may be stable for some level of external inputs and unstable for another level. This fact will be seen in the model of gamma oscillations (see Fig. 5).

Results

Paradoxical Effect of External Input

The most interesting aspect of the model emerges if we examine the consequences of increasing the external drive $i(t)$ onto the inhibitory neurons. Eq. 9 shows that this inevitably leads to a decrease in the equilibrium value of E —assuming that the system remains stable—but eq. 10 shows that the effect on I depends on the sign of the term $(1 - \beta J_{ee})$. If this term is negative, I moves in the opposite direction from $i(t)$; that is, an external input applied directly to the inhibitory cells causes the activity of those cells to move in the opposite direction.

The reasons for this phenomenon can be clarified by a phase plane analysis, which allows the properties of the solution to be analyzed graphically. Eqs. 4 and 5 define the nullclines of the differential equations 1 and 2, respectively. The nullclines in the I, E phase plane are shown in Fig. 4 for the two cases of weak and strong excitatory feedback mentioned above. The arrows near the nullclines indicate the sign of the corresponding derivative (dE/dt or dI/dt) on both sides of the nullclines. In Fig. 4B the arrows are facing away from the E-nullcline (Eq. 4), which means that the excitatory population by itself is unstable for any *fixed* level of inhibition: its activity either dies off or explodes to saturation, depending on initial conditions. Nevertheless, the intersection of the nullclines can still be stable, provided that the inhibitory coefficients J_{ii} , J_{ie} , and J_{ei} are large enough, as described above.

In these plots, an increase in the external input $i(t)$ corresponds to an upward translation of the I -nullcline. Inspection of Fig. 4 shows that this shift of the I -nullcline leads to quite different patterns of movement of the intersection point for the two conditions. In the case of weak excitation, increase in $i(t)$ leads to a downward-rightward shift of the intersection point, which means that the activity of the inhibitory population increases and the activity of the excitatory population decreases. In this scenario, which is intuitively clear, the excitatory cells are modulated out-of-phase with the inhibitory cells; however, in the case of strong excitatory feedback, an increase in the external drive to the inhibitory population leads to a downward-leftward shift of the intersection point, hence a *decrease* in the activity of both populations, as shown in Fig. 4B. Thus, when the recurrent excitation is strong enough, both populations will be modulated in phase with each other, and out-of-phase with the external drive.

It is possible to derive another important result from this phase plane analysis. The relative amplitude of modulation of the two populations, in response to a change in external

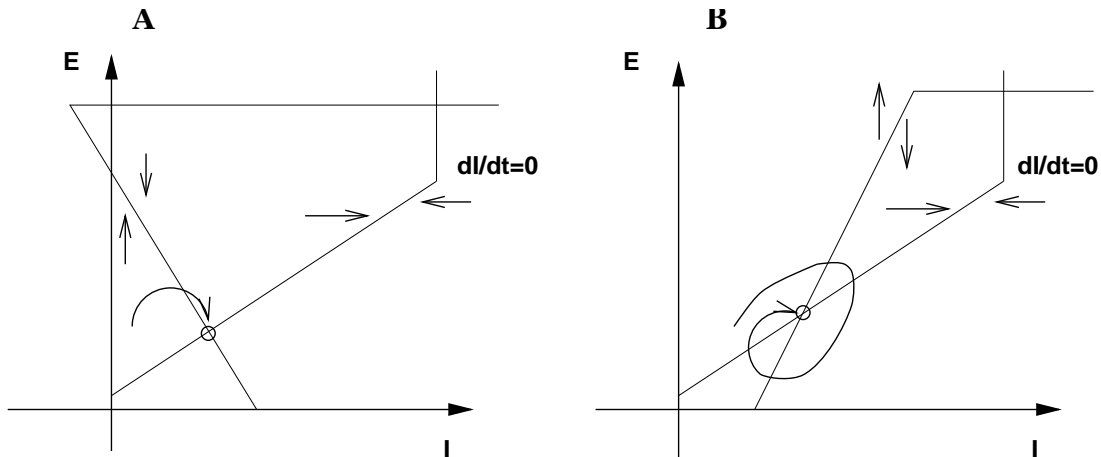


Figure 4: Phase plane analysis of the dynamic population of Equations 1 and 2. The state of the network is given by a point in the E-I plane. The nullclines are given by the steady-state Equations 4 and 5 (the second one is marked by the corresponding derivative, $\frac{dI}{dt}$). Nullclines of the state variables are shown for the cases (A) $J_{ee} < 1$ and (B) $J_{ee} > 1$. The arrows indicate the direction of motion of the state in the vicinity of the corresponding nullcline. A typical trajectory of the variables E and I towards the stable fixed point is shown in both cases.

input to the inhibitory neurons, is given by the slope of the E -nullcline:

$$\frac{\Delta E}{\Delta I} = \frac{\beta J_{ei}}{\beta J_{ee} - 1}. \quad (12)$$

Thus, the relative depth of modulation of the two populations provides an estimate of the relative strength of recurrent inhibition and excitation in the network. Note that, in the critical situation where βJ_{ee} is exactly equal to 1, any external input to the inhibitory population is transmitted directly to the excitatory population without causing any change at all in the firing rates of the inhibitory neurons.

Relation between Phase Shifts and Depth of Modulation of Interneurons

This analysis concerns the fixed points of Eqs. 1 and 2, and hence is strictly valid only in the limit of infinitely slow changes in the external inputs. In practice, these inputs have finite time constants (e.g. around 125 msec, the period of the theta rhythm in the hippocampus), so dynamical effects need to be considered. When there are oscillations in the external inputs, the most prominent effects are phase shifts of the population activities relative to the input. Under some conditions, such as those shown here, these phase shifts can be substantial even for inputs varying slowly compared to the time constants τ and τ' . Suppose the input $i(t)$ in Eq. 2 is given by:

$$i(t) = i_0 + i_1 \cos(2\pi ft), \quad (13)$$

where i_1 is the amplitude and $f \equiv 1/T$ is the period of the input. With our choice of response functions (Eq. 3), standard techniques for analyzing linear differential equations

(e. g., (Braun, 1986)) can be applied to Eqs. 1, 2 and 13. In the case where only the inhibitory inputs are oscillating, the analysis predicts that the phase shifts of the network activity relative to the inhibitory input should have two components:

- The first component, which is common to both excitatory and inhibitory populations, is on the order of $\min(\tau, \tau')/T$. Since this component is the same for both populations, it cannot be detected by recordings exclusively from within the network.
- An additional component, exhibited only by the *inhibitory* population, is given by:

$$\delta\phi = \begin{cases} 180^\circ + \arctan(2\pi f\tau/(1 - \beta J_{ee})) & \text{if } \beta J_{ee} < 1 \\ -\arctan(2\pi f\tau/(\beta J_{ee} - 1)) & \text{if } \beta J_{ee} > 1 \end{cases} \quad (14)$$

Therefore, a phase shift is expected due to the finite period of the oscillations for both strong and weak recurrent excitation; however, this shift is generally small, on the order of τ/T , unless the network operates near the transition point between the two regimes: $\beta J_{ee} \sim 1$. In this transition region, the amplitude of modulation for the inhibitory population becomes small, as is seen from Eq. 12. This is a consequence of the fact that, in this region, the E -nullcline is nearly vertical, so that varying the external input to the *inhibitory* population has only a small effect on its activity, but produces substantial changes the activity of the *excitatory* population.

Note that the analysis presented above was performed for threshold-linear response functions. In the case of general sigmoid functions, the results can still be considered as qualitatively accurate if the modulation of neuronal activity is such that the slope of the sigmoid remains roughly the same.

Limit cycle and fast oscillations

There are a few other features of this model worthy of notice. If the excitatory-inhibitory interaction term given by eq. 8 is sufficiently large, then the eigenvalues given by eq. 6 will be complex-valued, causing any perturbation from the fixed point—or any change in the external input—to result in a series of damped oscillations (as shown in Fig. 4B). This phenomenon may be related to the so-called gamma or “40 Hz” oscillations seen in many parts of the brain (Bragin et al., 1995). There is, however, another closely related mechanism that may also be involved. As mentioned above, the interaction between inhibitory interneurons, J_{ii} , determines the maximal strength of recurrent excitation compatible with the stability of the steady state. If recurrent excitation is even stronger, the stable solution of Eqs. 1,2 is a *limit cycle*, i.e. fast oscillations with the period on the order of τ and τ' . An example of these oscillations is shown in Fig. 5. The sharp peaks would be associated with the timing of spikes in the population. In this condition, the excitatory and inhibitory populations have almost identical phases relative to the theta wave; a slight positive phase shift of the inhibitory population can be seen in Fig. 5. A different model for gamma rhythm generation which is based on synchronization of an interneuron population is proposed in (Whittington et al., 1995; Traub et al., 1995) and also predicts a zero phase lag between pyramidal cell and interneuron firing during 40 Hz.

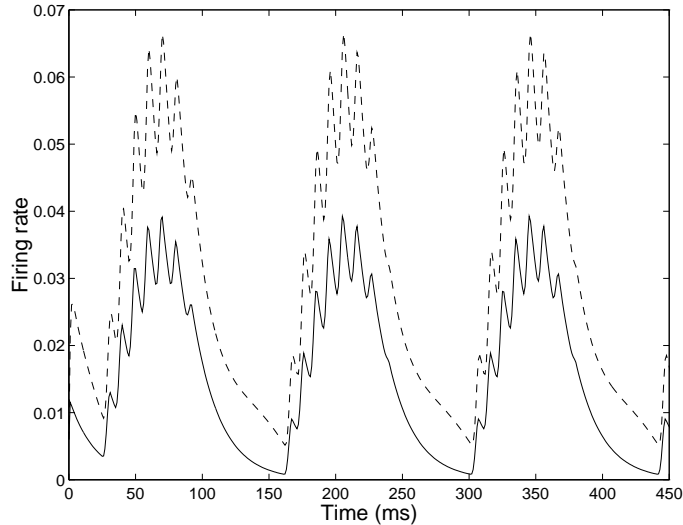


Figure 5: Solution of Equations 1 and 2 with the periodic input to the inhibitory population given by Equation 13 in the parameter regime corresponding to strong excitation. $E(t)$ - Solid line; $I(t)$ - dashed line. The response function was $g(x) = \tanh(x)$. Parameters were $\tau = 20ms$, $\tau' = 10ms$, $J_{ee} = 40$, $J_{ei} = 25$, $J_{ie} = 30$, $J_{ii} = 15$, $e = 0.1$, $i_0 = 0$, $i_1 = 0.1$. Since the slope of the response function is not constant, the condition for the stability of a fixed point is violated in part of a theta cycle, giving rise to oscillations in the gamma range.

Predictions

The forgoing analysis of the network model leads to three main predictions:

- Changes in external input to inhibitory interneurons can cause their activity to be modulated in the direction opposite to the change in the input if the intrinsic excitatory connections are sufficiently strong.
- For oscillatory inputs, the phase difference between the excitatory and inhibitory populations may vary depending on the strengths of internal interactions.
- For oscillatory inputs, the depth of modulation of the inhibitory population should be largest when its phase shift relative to the excitatory population is 0° or 180° .

Comparison with experimental data

A previous study (Skaggs et al., 1996) found that pyramidal cells oscillated in synchrony with each other over the CA1 region of hippocampus when a theta rhythm was present while rats ran for food reward. At the same time, inhibitory interneurons had a broad distribution of phase relations to the pyramidal cell population. According to our model, this suggests that the strength of recurrent excitation was in the transitional region; the variability in the phase shift would then reflect variations in the strength of recurrent excitation in different groups of excitatory cells.

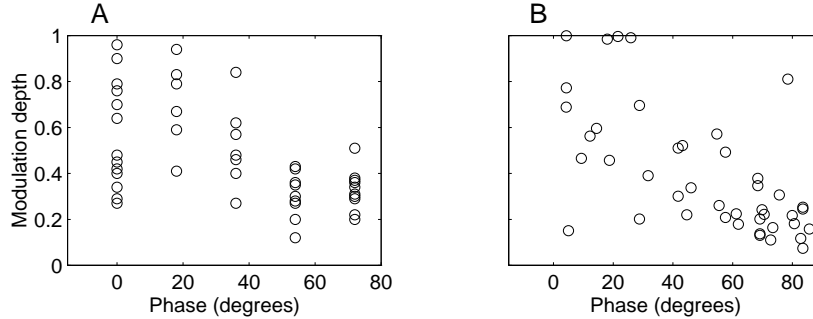


Figure 6: A, Depth of modulation of theta oscillations for a sample of 46 inhibitory interneurons, recorded from 6 rats running for food reward. The horizontal axis is the deviation from their phases from either 0° or 180° , whichever is less. B, Results of simulations presented in the same format. A network representing 50 excitatory and 50 inhibitory subpopulations was simulated. Each of the subpopulations was connected with a random choice of 10 other subpopulations of each type. The strength of all of the connections was randomly chosen from a uniform distribution from zero to twice the average value. The response function was the same as in Fig. 5. Parameters were $\tau = 20ms$, $\tau' = 10ms$, $J_{ee} = 0.2$, $J_{ei} = 0.23$, $J_{ie} = 0.22$, $J_{ii} = 0.21$, $e = 0.2$, $i_0 = 0.2$, $i_1 = 0.1$ (these are the average values for each of the existing connections).

The prediction derived from the present analysis—that interneurons with phase shifts far from 0° or 180° would show relatively weak theta modulation—was tested using a sample of 46 interneurons recorded in or near the CA1 cell body layer of the rat hippocampus. These cells were recorded simultaneously with groups of 50–150 CA1 pyramidal cells, using ensemble recording methods as described by Skaggs et al. (1996). A cell was classified as an interneuron if (1) it had a narrow spike waveform, less than $300 \mu\text{sec}$ from peak to valley, (2) it did not fire in complex-spike bursts, and (3) it had a mean rate greater than 5 Hz averaged across the whole session.

The depth of modulation for each interneuron and its phase relative to the excitatory population were obtained as explained in the methods. In Fig. 6A we plot the normalized depth of modulation for the sample of 46 interneurons against the deviation from their phases from either 0° or 180° , whichever is less. The correlation coefficient between the coordinates is -0.55 and is significantly different from zero ($p = 0.0002$, two-sided correlation test). To demonstrate that observed variability in-phase shifts of interneurons could indeed result from inhomogeneity in the synaptic strengths, we simulated a network consisting of 50 excitatory and 50 inhibitory subpopulations with random set of connections between them. The results of the simulations, presented in Fig. 6B, display the same trend as is seen in Fig. 6A. The results of the recordings are therefore broadly consistent with the predictions of the model.

Discussion

At a general level, perhaps the most important outcome of the analysis presented here is the finding that networks with strong recurrent excitation stabilized by feedback inhibition have

some counterintuitive properties; in particular, the responses of the interneurons to direct manipulation can be completely reversed by the context in which they are embedded. Since networks of this type are very common in the brain, our findings are likely to be widely applicable.

Another point that has perhaps not been appreciated is that in this type of network, inhibitory-inhibitory connections are critical for stability. It is interesting to note that GABAergic neurons in most areas of the brain have numerous GABAergic terminals synapsing on their dendrites. These have often been treated as a mystery and omitted from models: our analysis indicates that they are actually essential to the stability of the network.

There are a number of existing experimental findings that are consistent with the model. In awake, moving animals, when a theta rhythm is present, most hippocampal interneurons fire in-phase with the excitatory population, as reported in Skaggs et al (1996). It has also been reported that, under urethane anesthesia, most interneurons fire out-of-phase with the excitatory population (Fox et al., 1986). Based on our analysis, a shift of this sort could easily be produced by a general reduction in the effectiveness of excitatory synapses due to anesthesia (Moroni et al., 1981), i.e., a reduction in the effective values of J_{ee} and J_{ie} .

In the model, the source of variability in-phase shift among different interneurons was hypothesized to derive from differences in recurrent connectivity among the populations of pyramidal cells they were connected with. It is worth noting that a somewhat different explanation is also possible. Several lines of evidence indicate that hippocampal interneurons fall into a number of distinct classes, with different patterns of connectivity. A straightforward extension of the present analysis indicates that, at a fixed point of the network, each type of interneuron must obey an equation similar to eq. 10; however the “effective” values of the J_{ee} , J_{ie} , and J_{ei} coefficients can be different for different types. This suggests that different types of interneurons can “see” different effective values for the recurrent connectivity of the remainder of the network. A full analysis of this situation, however, taking into account the possibility of differential external modulation of multiple interneuron types, would be quite difficult to carry out in complete generality.

In addition to the evidence presented here in favor of the predicted relationship between the phase shift and depth of modulation for each interneuron, the model makes several other predictions that could be further investigated:

1. The hippocampal theta rhythm is implemented largely via GABAergic projections from the medial septal area that terminate on inhibitory interneurons (Freund and Antal, 1988). We predict that many interneurons driven by this input will fire in-phase with pyramidal cells, and that this will be even more the case in CA3 than in CA1.
2. Serotonergic inputs to the hippocampus are known to terminate largely on inhibitory interneurons, and their synaptic effects are thought to be inhibitory (Freund et al., 1990). We predict that when these inputs are activated, the firing rates of many of the interneurons will increase rather than decrease.
3. Certain anatomical types of GABAergic interneurons in the hippocampus are thought to project specifically to other types of interneurons. We predict, that when these cells are activated, their targets may show increases rather than decreases in firing rate.

More specific predictions from this type of model will only be possible once more detailed data on the connectivity and pharmacological properties of hippocampal neurons becomes available. Regardless of the details, though, it is very likely that the paradoxical responses described here will be seen at some locations in the hippocampal system. The same analysis could also be applied to the neocortex, which has a more complex structure than the hippocampus, but is also dominated by strong recurrent excitation and feedback from local inhibitory interneurons. Blockade of inhibition in the neocortex leads to runaway activity of the excitatory cells, culminating in an epileptic seizure, which suggests that the neocortex may also operate near the edge of instability when there are strong rhythms.

References

- Abeles, M. (1991). *Corticonics*. Cambridge University Press, New York.
- Amaral, D. G. and Witter, M. P. (1995). Hippocampal formation. In Paxinos, G., editor, *The Rat Nervous System*, pages 443–493. Academic Press, San Diego.
- Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., and Buzsáki, G. (1995). Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.*, 15:47–60.
- Braun, M. (1986). *Differential equations and their applications*. Springer-Verlag, New York.
- Fox, S. E., Wolfson, S., and Ranck, Jr., J. B. (1986). Hippocampal theta rhythm and the firing of neurons in walking and urethane anesthetized rats. *Exp. Brain Res.*, 62:495–508.
- Freund, T. F. and Antal, M. (1988). GABA-containing neurons in the septum control inhibitory interneurons in the hippocampus. *Nature*, 336:170–173.
- Freund, T. F., Gulyás, A. I., Acsády, L., Görcs, T., and Tóth, K. (1990). Serotonergic control of the hippocampus via local inhibitory interneurons. *Proc. National Acad. Sci.*, 87:8501–8505.
- Grinvald, A., Frostig, R. D., and Lieke, E. (1988). Optical imaging of neuronal activity. *Physiol. Rev.*, 68:1285–1366.
- Jones, E. G. (1986). Connectivity of the primate sensory-motor cortex. In Jones, E. G. and Peters, A., editors, *Cerebral Cortex*. Plenum Press, New York.
- Moroni, F., Corradetti, R., Casamenti, F., Moneti, G., and Pepeu, G. (1981). The release of endogenous gaba and glutamate from the cerebral cortex in the rat. *Naunyn-Schmiedeberg's Arch. of Pharmacol.*, 316:235–239.
- Sik, A., Penttonen, M., Ylén, A., and Buzsáki, G. (1995). Hippocampal cal interneurons: An in vivo intracellular labeling study. *J. Neurosci.*, 15:6651–6665.
- Skaggs, W. E., Wilson, M. A., and McNaughton, B. L. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6:149–172.

- Traub, R. D., Jefferys, J. G. R., Whittington, M. A., Buzsaki, G., Penltonen, M., and Colling, S. B. (1995). Models of gamma ('40 hz') and the theta rhythm in the rat hippocampus in vitro and in vivo. In *Society for Neuroscience Abstracts*, volume 21. 473.4.
- Traub, R. D. and Miles, R. (1991). *Neuronal Networks of the Hippocampus*. Cambridge University Press, New York.
- Tsodyks, M., Skaggs, W., Sejnowski, T., and McNaughton, B. (1996). Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus*, 6:271–280.
- Vanderwolf, C. H. (1969). Hippocampal electrical activity and voluntary movement in the rat. *EEG & Clin. Neurophysiol.*, 26:407–418.
- Whittington, M. A., Traub, R. D., and Jeffreys, J. G. R. (1995). Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature*, 373:612–615.
- Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. Journal*, 12:1– 24.