Network: Computation in Neural Systems
September 2006; 17(3): 277–297

informa
healthcare

An analysis of the mean theta phase of population activity in a model of hippocampal region CA1

ERIC A. ZILLI[1] & MICHAEL E. HASSELMO[1,2]

[1]*Program in Neuroscience,* [2]*Department of Psychology, Center for Memory and Brain, 2 Cummington Street, Boston University, Boston, MA 02215, USA*

(Received 24 June 2006; accepted 26 September 2006)

Abstract
In a recent study, Manns et al. (2007) showed that CA1 population spiking activity differs in mean phase between conditions of exposure to stimuli of varying levels of familiarity. Here, we provide an analysis of a computational model of the hippocampus from Hasselmo et al. (2002) to examine how an animal's history of exposures to stimuli affects CA1 population activity. We show how the model can reproduce the major findings from the study by Manns et al. Specifically, we show that differences in direction and magnitude of the mean phase of population activity between two stimuli in the model depend on the order of presentation of the stimuli and the duration of exposure during each presentation. The analyses also reveal that the model's learning rule results in stability and normalization of synaptic weights and show how dysfunction of neuromodulatory systems might lead to epileptic or amnestic type conditions.

Keywords: *Theta rhythm, entorhinal cortex, CA3*

Introduction

During many behaviors, a strong oscillatory component of the hippocampal EEG appears in the 3 to 10 Hz range and is referred to as theta rhythm (Buzsaki 2002). Over the course of each cycle of the theta rhythm, experimental data shows that some physiological variables wax and wane in strength, including amplitude of synaptic inputs from CA3 and entorhinal cortex layer III (ECIII) into CA1 (Brankack et al. 1993) and inducibility of plasticity (Pavlides et al. 1988; Hölscher et al. 1997; Hyman et al. 2003). Specifically, Brankack et al. (1993) showed that inputs from CA3 and ECIII to CA1 are strongest at different phases of the cycle, ECIII being strongest on the downslope of theta as recorded at the hippocampal fissure, and CA3 input being strongest on the upslope. Hyman et al. (2003) showed that LTP is most easily induced at the peak of fissure theta, while stimulation at the trough results in LTD or depotentiation. Analyses of a hippocampal model taking these variations into account (Hasselmo et al. 2002; Judge and Hasselmo 2004) showed that the phases reported in the physiological experiments correspond to optimal performance for specific memory performance measures.

Correspondence: Eric A. Zilli, Program in Neuroscience, Center for Memory and Brain, 2 Cummington Street, Boston University, Boston, MA 02215, USA. Tel: 1 617 353 1431. Fax: 1 617 353 1424. E-mail: zilli@bu.edu

ISSN: 0954-898X print / ISSN 1361-6536 online
DOI: 10.1080/09548980601028908

In those models, CA3 input to CA1 was interpreted as reflecting learned associations to sensory stimuli, while ECIII input was thought to reflect the processed sensory input about the stimuli currently present in the environment. Based on this interpretation, experiments were recently performed (Manns et al. 2007) to test the model's predictions regarding the activity of CA1 in response to stimuli of varying degrees of familiarity. Since CA3 and ECIII inputs occur at different phases, familiar versus novel stimuli would be predicted to be associated with different mean phases of CA1 spiking activity. Manns et al. tested rats in two different tasks. In their novel object task, unit activity from CA1 was recorded while rats sampled novel versus familiar objects in an environment. Manns et al. also tested rats in a delayed non-match to odor task, comparing neuronal activity when sampling a trial-repeated odor with activity when sampling a trial-novel odor. The data showed that the magnitude of the difference in mean phases between two conditions varied across animals from as little as 12.3 to as much as 91.7 degrees difference. Further, some animals in that study exhibited a clockwise shift in mean phases between conditions, while others exhibited a counterclockwise shift (i.e., some differences were positive and some negative). Because the model of Hasselmo et al. (2002) made no predictions regarding spiking activity, we analyze that model in more detail to better understand the way the phases of synaptic input and history of stimulus presentations interact to alter the mean phase of CA1 population spiking activity.

Equations describing the model

The model analyzed here is based on that of Hasselmo et al. (2002). As shown in the following equations, the model comprises three regions of medial temporal cortex including entorhinal cortex layer III and hippocampal regions CA3 and CA1. Activity in region CA1 results from the sum of inputs from CA3 and ECIII, modulated by sinusoids that vary over the course of each theta rhythm. Plasticity is modulated similarly, alternating between phases of LTD (when CA3 input is strongest) and LTP (when ECIII input is strongest). More complete details can be found in Hasselmo et al. (2002).

Brankack et al. (1993) demonstrated that the synaptic input from CA3 and ECIII to CA1 are not constant over a theta cycle but rather each input has a range of phases in which it is strongest. Therefore, in the model the strengths of synaptic input to CA1 vary over the course of each theta cycle in the form of sinusoids that are shifted in the positive y direction and scaled to range from 0 to 1. The sinusoids have phase offsets, Φ_{CA3} and Φ_{ECIII}, such that the peaks of the sinusoids match the phases of strongest input reported in Brankack et al. (1993) from current-source density analysis (CSD) (Table I in that paper, $\varphi_{CA3} = 276^\circ$, $\varphi_{ECIII} = 129^\circ$). To translate those peak phases to sine offsets in the model, we subtract 90° from them (because the peak of a sine function is at 90°) so the peaks of the modulatory functions match the CSD data: $\Phi_{CA3} = 186^\circ$, $\Phi_{ECIII} = 39^\circ$. Specifically, the functions are:

$$\Theta_{CA3}(t) = \frac{1}{2} + \frac{1}{2}\sin(t + \Phi_{CA3}) \tag{1}$$

$$\Theta_{ECIII}(t) = \frac{1}{2} + \frac{1}{2}\sin(t + \Phi_{ECIII}) \tag{2}$$

Except for the Schaffer collaterals, which are represented by a matrix W, all synaptic connections in the network are taken to be identity matrices. During a theta cycle, activity patterns in CA3 and ECIII, $a_{CA3}(t)$and $a_{ECIII}(t)$, respectively, remain constant. Input to region CA1 is modeled as a continuous function given by the sum of modulated CA3 and

ECIII activity:

$$a_{CA1}(t) = \Theta_{CA3}(t)Wa_{CA3}(t) + \Theta_{ECIII}(t)a_{ECIII}(t) \quad (3)$$

In the simplest case, one can imagine that W in this equation increases proportional to duration of exposure to a given stimulus. Then one can extract the appropriate values for substitution into Equation 6 below (as done in Equations 8 and 9) and this is sufficient to reproduce the basic results from Manns et al. (2007) regarding the magnitude and direction of CA1 mean spiking phase differences. However, we will later be interested in details regarding the time course of modification of these synapses and so will use the following learning rule from Hasselmo et al. (2002).

The magnitude and direction of plasticity are also modulated by a sinusoid, but this sinusoid varies from -1 (LTD) at the peak of fissure theta to 1 (LTP) at the trough of fissure theta, in agreement with electrophysiological data reported in CA1 in Hyman et al. (2003) and Hölscher et al. (1997) and in dentate gyrus by Pavlides et al. (1988):

$$\Theta_{LTP}(t) = \sin(t) \quad (4)$$

The learning rule used in the model is a variation of a standard Hebb-type rule in which a synaptic weight changes only when both pre- and postsynaptic cells are firing, but with the direction of plasticity controlled by $\Theta_{LTP}(t)$:

$$W(T+1) = W(T) + \int \Theta_{LTP}(t)a_{CA1}(t)a_{CA3}(t)^{T}dt \quad (5)$$

To simplify the following analyses, the actual strengths of synapses do not immediately change but are delayed until the end of each theta cycle indicated by the modification time step T. In the following numerical simulations, the changes in synaptic weights are further multiplied by the length of the time step in each theta cycle (in radians) so that the values match those obtained by numerically evaluating the integral in Equation 5.

The present work does not take into account the effects of inhibitory interneuron firing on CA1 spiking activity, but unpublished analyses and numerical simulations indicate that the main result of such inhibition is to decrease the range of mean CA1 spiking phases. Notably, in this model inhibition does not affect learning, as plasticity has been shown to occur even when cell bodies are hyperpolarized, as long as there is depolarization in the dendrites (Golding et al. 2002; Lisman & Spruston 2005).

Representations of stimuli

For simplicity we will restrict our consideration to the case where two stimuli are presented. Each of the two stimuli will be represented in ECIII by the activity of a set of cells. Some of those active cells will be unique for each stimulus, but we will also allow for the case where some number of cells is shared between the two. If we set the number of unique cells to be the same for each stimulus, we can describe the general case as [s,u] $\equiv$ [number of shared cells, number of unique cells], resulting in a total of s + 2u cells for the two stimuli. We can describe the representations of the stimuli in CA3 in the same form, allowing the two regions to have different representations. Thus for CA3 we can write [sC,uC] and for ECIII, [sE,uE]. To specifically refer to the unique cells associated with odor A we can write uCA or uEA, so, when odor A is present, the activity of the shared cells and the unique cells for that odor will be set to 1 and the others to 0. Because the matrix from ECIII to CA1 is an identity matrix, we can also identify cells in CA1 according to the ECIII cell from which the CA1 cell in question receives its input (e.g., cells receiving input from sE, from uEA, etc.).

The matrix representing the Schaffer collaterals, then, has rows for each CA1 cell labeled by the type of EC input it receives and columns for each CA3 cell labeled similarly, as shown below.

$$\begin{bmatrix} \text{sC, sE} & \text{uCA, sE} & \text{uCB, sE} \\ \text{sC, uEA} & \text{uCA, uEA} & \text{uCB, uEA} \\ \text{sC, uEB} & \text{uCA, uEB} & \text{uCB, uEB} \end{bmatrix} \equiv \begin{bmatrix} \text{Wss} & \text{Wuas} & \text{Wubs} \\ \text{Wsua} & \text{Wuaua} & \text{Wubua} \\ \text{Wsub} & \text{Wuaub} & \text{Wubub} \end{bmatrix}$$

sC, uC, sE, and uE will be used throughout the paper to refer to both the number of such cells and the populations of cells themselves. Context should make clear which meaning is intended. We will also consider sets of synapses between these cells, calling them, e.g., Wss for the synapses from sC cells to CA1 cells receiving input from sE cells. Where unique cells are used, the letter u will be followed by a letter representing the stimulus to which the unique cells correspond, but, unless specifically noted, Wuu will refer to connections between unique cells for the same stimulus in both CA3 and ECIII.

Afferent contributions to CA1 activity

In the model, CA1 activity is produced only by CA3 and ECIII input; the mean phase of CA1 activity is a function of the phases at which the inputs are strongest and the magnitudes of those peak inputs. We assume that the phases of those inputs are always the same (with the values given above, from the CSD data), but the magnitude of the inputs will change with learning, as we will show below. With those phases fixed, we can describe CA1 activity in terms of the magnitude of CA3 and ECIII inputs as {magnitude of CA3, magnitude of ECIII}.

If only CA3 or only ECIII input were present ({1,0} or {0,1}, respectively), the mean phase of CA1 spiking would be identical to the phase at which the single input is strongest. If both inputs were present with equal magnitudes, {1,1}, the mean phase of CA1 activity would be the circular mean of the phases of maximum input of both CA3 and ECIII. We will see, however, that the magnitude of the two inputs can change, so we need a general formula for computing the mean phase of CA1 spiking activity. Intuitively, the mean phase will be closer to the phase of the input with the larger magnitude. As derived in Appendix A, inputs of magnitude a and b, with peaks at phases Φ_A and Φ_B, respectively, produce a mean phase of:

$$\bar{\Theta} = \tan^{-1} \frac{\mathrm{a}\sin\left(\frac{\pi}{2} - \Phi_{\mathrm{A}}\right) + \mathrm{b}\sin\left(\frac{\pi}{2} - \Phi_{\mathrm{B}}\right)}{\mathrm{a}\cos\left(\frac{\pi}{2} - \Phi_{\mathrm{A}}\right) + \mathrm{b}\cos\left(\frac{\pi}{2} - \Phi_{\mathrm{B}}\right)} \tag{6}$$

To find the difference in mean phases of CA1 spiking for two conditions, $\bar{\Theta}_m = \tan^{-1} m$ and $\bar{\Theta}_n = \tan^{-1} n$, where m and n are the fractions for each condition in the equation above, we can simplify the computation by using a difference of inverse tangents formula:

$$\bar{\Theta}_{\mathrm{m}} - \bar{\Theta}_{\mathrm{n}} = \tan^{-1}\left(\frac{\mathrm{m} - \mathrm{n}}{1 + \mathrm{mn}}\right) \tag{7}$$

Delayed non-match to sample task

We simulate the response of region CA1 in a simplified version of a delayed non-match to sample (DNMS) task that consists solely of stimulus presentations, no attempt is made here to simulate behavior or decision making. We will first examine a three-presentation

task comprising an exposure to odor A for one theta cycle, then a subsequent exposure to odors A and B in rapid succession, as real animals do not sample the odors simultaneously. In this task, synaptic weights will only be updated after the initial exposure, but not after presentation of the second two test odors. A second analysis will consider an arbitrary series of stimuli presented one at a time for one theta cycle each with the synaptic weights updating at the end of each cycle. Lastly, we will consider numerical simulations of this model and compare them to the results from the following analyses.

Analysis 1—Magnitude of spiking phase differences

We first consider how the magnitude of differences in CA1 spiking phases is related to memory of stimuli. This analysis is meant to represent a single trial of a DNMS task in which the rat is presented with a single odor in a sample phase, then, after a delay, is presented with the same odor and a different odor. We will restrict the weights of synapses and the activity levels of cells to be only 0 or 1, and will let $sC = uC = sE = uE = 1$, so that there is one cell shared between the two odors and one unique cell for each odor. After the first presentation of odor A, the synapses will be updated so that each of sC and uCA links to CA1 cells receiving both sE and uEA as shown in Figure 1, top. When odor A and odor B are presented subsequently, CA1 will have the same total magnitude of ECIII input for either odor ($sE + uEA = 2$ for odor A and $sE + uEB = 2$ for odor B), though the specific cells activated will be different. The total magnitude of CA3 input, however, will differ between the two cases. When odor A is presented (the match condition), CA1 cells receiving input

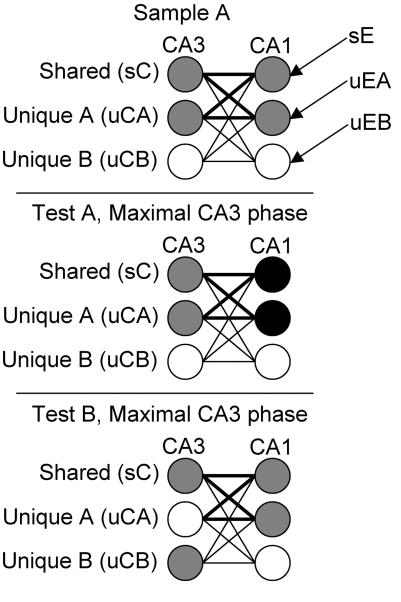

Figure 1. Example of different magnitudes of CA3-driven CA1 activity in a single DNMS trial. Top: During the sample phase, odor A is presented which activates the shared unit and the unique A unit in both CA3 and CA1 through ECIII. LTP links each active CA3 unit to the active CA1 units. Middle: Activity in the test phase of the task at a point where CA3 input to CA1 is strong and ECIII input is weak. Thus, postsynaptic activity (in gray) is due to CA3 input at this phase. When odor A is presented, the shared and unique A units in CA1 each receive two synaptic inputs. Bottom: When odor B is presented, only the connections from the shared CA3 unit are activated and each CA1 cell receives only a single input.

from sE and uEA will each receive input from both sC and uCA cells, so the total number of inputs CA1 receives is 4. When odor B is presented (non-match), on the other hand, CA1 cells receiving input from sE and uEA will only receive one input each from sC, while uCB will cause no activity because it has not been linked to any CA1 cells, for a total of 2 inputs. These cases are shown in the middle and bottom of Figure 1. Using the notation from earlier of {magnitude of CA3, magnitude of ECIII}, we can write CA1 activity for the match condition as {4,2}, while CA1 for the non-match condition will be {2,2}. Using the formula for determining the mean phase given two amplitudes from earlier in the paper, we can compute the mean phase of CA1 population activity for the match and non-match conditions

$$\bar{\Theta}_{\text{match}} = \tan^{-1} \frac{4\sin\left(\frac{\pi}{2} - \Phi_{\text{CA3}}\right) + 2\sin\left(\frac{\pi}{2} - \Phi_{\text{ECIII}}\right)}{4\cos\left(\frac{\pi}{2} - \Phi_{\text{CA3}}\right) + 2\cos\left(\frac{\pi}{2} - \Phi_{\text{ECIII}}\right)} = -71^{\circ} \tag{8}$$

$$\bar{\Theta}_{\text{non - match}} = \tan^{-1} \frac{2\sin(\frac{\pi}{2} - \Phi_{\text{CA3}}) + 2\sin\left(\frac{\pi}{2} - \Phi_{\text{ECIII}}\right)}{2\cos\left(\frac{\pi}{2} - \Phi_{\text{CA3}}\right) + 2\cos\left(\frac{\pi}{2} - \Phi_{\text{ECIII}}\right)} = -23^{\circ} \tag{9}$$

The full range of values produced by mean phase equations of this sort is -96 to 51°.

More generally, in this version of the task, CA1 activity in the match condition will be {(sC + uC) ∗ (sE + uE),(sE + uE)}, while that for the non-match condition will be {sC ∗ (sE + uE),(sE + uE)}. The only difference between these two is the presence of uC in the CA3 magnitude of the match condition. Thus, a higher uC, i.e., a decreased amount of overlap in CA3 representations, causes an increase in the phase difference between the two conditions.

This suggests that one explanation for the difference in CA1 mean phase reported in Manns et al. (2007) is that learning occurring during the sample phase increases the CA3 component for the match stimulus, resulting in a phase difference qualitatively similar to that computed above.

Similarly, if the magnitude of the CA3 input for a stimulus reflects the duration of exposure to it, then from this example it can be seen that, given two stimuli, the difference of population spiking phase in CA1 between the two could be either positive or negative, depending on which has been experienced more. While this fact alone is sufficient to demonstrate the means by which the results of Manns et al. (2007) could occur, the following analyses consider in greater detail the effects of specific stimulus presentation histories on synaptic weights and thereby on population response in CA1. This closer consideration also leads to interesting results on normalization and stability of the present learning rule.

Analysis 2—Direction of phase differences

As stated earlier, experimental data show that, in addition to having varying magnitudes in phase differences, the direction of the difference can also vary between animals, thus we continue the analysis to understand how factors influencing the strength of input can influence phase shift directions. While the previous analysis was restricted to a single trial of a DNMS task, actual experiments using that task present multiple trials and often reuse odors over the course of the task. Further, it is likely that synapses in an animal's brain can take on intermediate weights between 0 and 1, so we proceed with a more detailed analysis of the model under these conditions. We begin by substituting the activity equation for CA1 from above into the learning rule, so that we can better understand how synaptic weights

change over the course of stimulus exposures. The substitution yields:

$$W(T+1) = W(T) + \int \Theta_{LTP}(t)[\Theta_{CA3}(t)W(T)a_{CA3}(t) + \Theta_{ECIII}(t)a_{ECIII}(t)]a_{CA3}(t)^T\,dt \tag{10}$$

We can rewrite this difference equation and solve it explicitly if we restrict the network to have only a single CA3 cell (either shared or unique). By doing so we can consider the weight of, e.g., a synapse from an sC to a CA1 cell receiving input from sE and can avoid the complication of, e.g., a uC also causing postsynaptic activity that is dependent on a different synaptic weight.

Since the integrals in this equation are with respect to time within a theta cycle, t, while the weights themselves change with respect to whole theta cycles, T, we can pull W out of the first integral and rewrite the equation as:

$$W(T+1) = W(T) + W(T)\int \Theta_{LTP}(t)\Theta_{CA3}(t)a_{CA3}(t)^2dt + \int \Theta_{LTP}(t)\Theta_{ECIII}(t)a_{CA3}(t)a_{ECIII}(t)dt \tag{11}$$

Since we're now dealing with only a single synapse, $a_{CA3}(t)$ and $a_{ECIII}(t)$ should be interpreted as referring only to the particular pre- and post-synaptic cells for that synapse. The integrals are effectively constants, so we can substitute:

$$X = \int \Theta_{LTP}(t)\Theta_{CA3}(t)a_{CA3}(t)^2dt \tag{12}$$

$$Y = \int \Theta_{LTP}(t)\Theta_{ECIII}(t)a_{ECIII}(t)a_{CA3}(t)dt \tag{13}$$

and write $W(T+1)$ as a simpler equation:

$$W(T+1) = W(T) + W(T)X + Y \tag{14}$$

For the specific phases shown in the CSD analysis of Brankack et al. (1993), X will be negative or zero and Y will be positive or zero, depending on the activity of the cells. If the presynaptic cell, a_{CA3}, is not firing, both X and Y will be zero and no weight change will occur. If the presynaptic cell fires, but the postsynaptic cell does not, X will be negative and Y will be zero, so LTD will occur. If both pre- and postsynaptic cells fire, X will be negative and Y will be positive, and, as shown below, the weight will increase toward an asymptotic value.

As described in Appendix B, it is reasonable to treat this difference equation as a differential equation for this analysis. The asymptotic synaptic value is then given by solving this differential equation to obtain:

$$W(T) = \frac{-Y}{X}(1 - e^{TX}) + W(0)e^X \tag{15}$$

The behavior of this function depends on the signs of X and Y (but also see Appendix B). If $X > 0$, the function grows exponentially toward infinity (positive infinity when $Y > 0$ and negative infinity when $Y < 0$). If $X < 0$, the function grows to an asymptote of $-Y/X$, and thus toward a positive value when $Y > 0$ and toward a negative value when $Y<0$. As discussed later, we do not allow a change in sign of synaptic strength. The signs of these integrals depend on the difference in phases between $\Theta_{LTP}(t)$ and $\Theta_{CA3}(t)$ for X and $\Theta_{LTP}(t)$

and $\Theta_{ECIII}(t)$ for Y. Specifically, $X > 0$ when $-\frac{\pi}{2} < \Phi_{CA3} - \Phi_{LTP} < \frac{\pi}{2}$, and, likewise, $Y > 0$ when $-\frac{\pi}{2} < \Phi_{ECIII} - \Phi_{LTP} < \frac{\pi}{2}$. For the phases found experimentally in Brankack et al. (1993), the synapses will grow to an asymptote of $-Y/X$, a positive value. This suggests that one function of the distinct oscillatory input phases in relation to oscillatory changes in the induction of plasticity might be to bound the maximum values of synapses, thereby avoiding stability issues that can arise when synaptic weights can grow without an upper limit. Note that, biologically, synaptic weights are bounded by the physiology of the synapses, but, as we will show later, this asymptote actually applies to the sum of all weights of active synapses onto a cell, and thus it can prevent all synapses from saturating at their maximum values, which has been shown to cause instabilities and memory deficits in neural networks (Hasselmo 1994).

We can now consider the effects of a series of stimulus presentations on a synapse in the model. The simplest case is that of a Wss synapse. In this case, the pre- and postsynaptic cells will be active on every presentation. If the weight starts with value 0, the strength of the synapse after T single-cycle presentations will simply be:

$$Wss(T) = \frac{-Y}{X}(1 - e^{TX}) \tag{16}$$

For a synapse from a uCA to a CA1 cell receiving input from sE or uEA, the presynaptic cell will only fire on trials in which the stimulus it represents is present, but on any trial it does fire, the postsynaptic cell will also be firing. Thus, at any point, if the stimulus has been seen for k theta cycles, the weight will be:

$$Wuaua(k) = Wuas(k) = \frac{-Y}{X}(1 - e^{kX}) \tag{17}$$

In this single synapse network, a presynaptic unique cell for one stimulus will never fire on the same trial as a postsynaptic unique cell for a different stimulus, so those weights will always be 0. This leaves the final case of a synapse from an sC to a CA1 cell receiving input from a uEA cell. Here the presynaptic cell will fire on every trial, but the postsynaptic cell will only fire on trials in which a particular stimulus, A, is presented. The weight of this synapse after T theta cycles will thus reflect a series of weight increments when stimulus A is presented and weight decrements on other trials (the decrements are caused by the CA3 cell driving CA1 activity during the phases where Θ_{LTP} is negative). We will represent the series of theta cycles as a string, E, where the ith digit, E_i, is 1 when odor A was present on that cycle and 0 when odor A was not present. Because no weight change will occur before the first presentation, we take $E_1 = 1$. After n theta cycles with history E, we will have a weight of:

$$Wsu(n, E) = \frac{-Y}{X}\left[\sum_{i=1}^{n} E_i e^{(n-i)X}(1 - e^{X})\right] \tag{18}$$

Because this equation allows synapses to decrease in strength under appropriate conditions, it is worth considering whether it has a lower bound in addition to its upper asymptote at $-Y/X$. Indeed, as is clear from the equation $dW/dT = WX + Y$, since $X < 0$ and $Y > 0$, the function can only decrease when W is nonzero, so starting from a zero or positive value, W will never become negative and will simply approach zero as a lower bound. This, however, is only the case for this single synapse network. In a network with multiple synapses, as will be shown below, W here is effectively the sum of active synapses onto a postsynaptic cell, and so the network constrains the sum to be between 0 and $-Y/X$. Without an imposed lower bound at 0, some synapses can decrease toward a bounded, negative value. Because it

is not thought to be biologically plausible to have a synapse change from being excitatory to inhibitory in most conditions, we restrict synapses to be greater than or equal to zero for the rest of the paper.

Now that we have expressions for all of the values in the weight matrix, we can consider CA1 population mean phases in response to two different stimuli, A and B, given the history of trials up to that point.

As before, we will assume the numbers of unique cells for the two stimuli are the same. The contribution of ECIII to CA1 population activity will again be the same when exposed to A or B, and so it is the difference in CA3 contributions from the weight matrix that will cause differences in the mean phase of CA1 population activity. If we assume here that plasticity is delayed until after both A and B have been presented (so that we may compare the CA1 response to each without the first presentation affecting the second), the contribution from the shared CA3 units will be the same and any difference in CA1 activity, D, will come from the two columns in the weight matrix corresponding to the unique CA3 cells. We can write this as:

$$D = uC(Wuaua + Wuas - Wubub - Wubs) \tag{19}$$

As shown above, Wuaua=Wuas and Wubub = Wubs, so we can simplify that to:

$$D = 2uC(Wuaua - Wubub) \tag{20}$$

Then by substituting in the solutions to these synaptic weights, equation (17), and simplifying, we find that the difference in CA3 contributions is:

$$D = \frac{-uC2Y}{X}(e^{jX} - e^{kX}) \tag{21}$$

where j is the number of times A has been presented up to that point and k is the number of presentations of B. Thus, on any particular trial, the phase difference between the spiking activity to A and B can be positive or negative, depending on which stimulus has been experienced more often. Importantly, this is affected by the total duration of the exposure and not simply the number of times a stimulus has been presented. In the Manns et al. (2007) experiment, the duration of odor sampling was not controlled, so even if two rats had identical orders of odor presentation, individual differences in sampling duration could cause differences in phase of CA1 responses.

The actual mean phase computation, however, depends on the specific values of the CA3 contributions, not their difference. The total CA3 contribution for stimulus A is given by:

$$D = sC(Wss + Wsua + Wsub) + uC(Wuaua + Wuas) \tag{22}$$

into which Equations 16 to 18 can be similarly substituted, and likewise for stimulus B. As the number of presentations gets larger, both e^{kX} and e^{jX} approach 0, because $X < 0$, and the phase difference disappears.

Phase differences and performance measure

In Hasselmo et al. (2002), a memory performance measure was used to determine the optimal phase offsets for the oscillatory variables in the model. Here we compute an analogous performance measure for the DNMS task to compare sets of parameters that produce both a large phase offset and a high value of the performance measure. We present a sample phase in which odor A is presented for a single theta cycle, then odor B is presented for a single cycle. We then turn off plasticity and present a test phase of odor A followed by odor B and

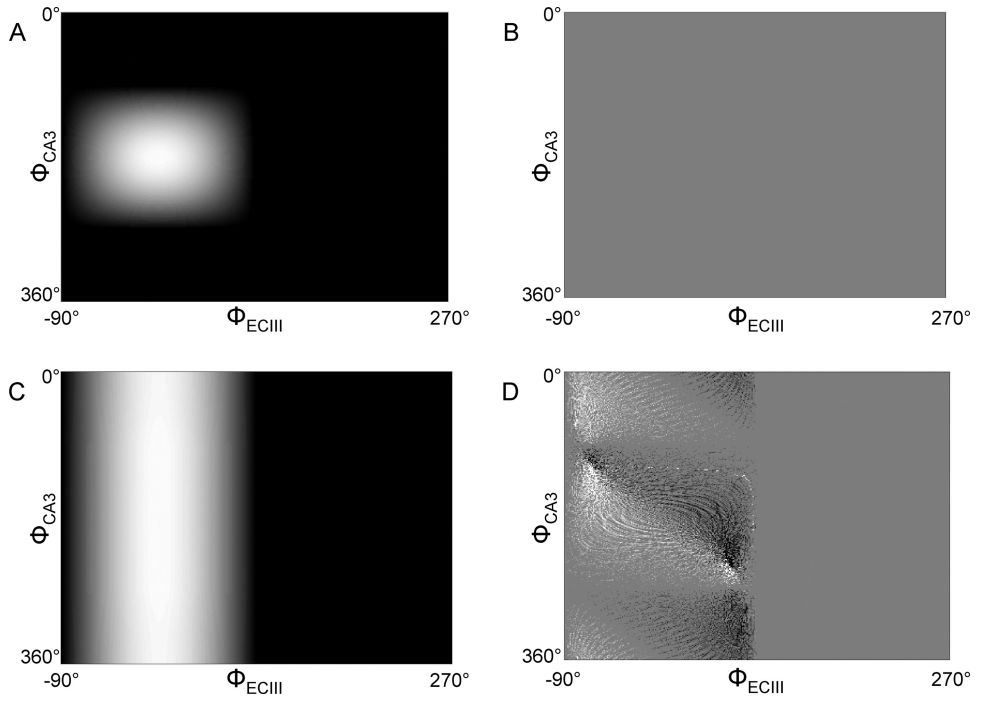

Figure 2. Performance measures and mean phase differences in CA1 spiking activity for two parameter sets. Each point represents network response after a single presentation of each of two odors. The x axes correspond to phases of maximal ECIII input varying from −90 to 270 degrees. The y axes show phases of maximal CA3 input varying from 0 to 360 degrees. A and B demonstrate parameters uC = 0, sC = 1, uE = 1, sE = 0. In A, nonzero performance (positive values in white, negative values in black) is limited to a small region of parameter space centered at $\Phi_{ECIII} = 0$ and $\Phi_{CA3} = 180$. B shows that, for these parameters, the phase difference is always zero (zero phase difference in medium gray shade). In C and D, uC = 100, sC = 1, uE = 1, sE = 1. C shows that, in this case, Φ_{CA3} is less important for nonzero performance because there is strong activity from unique CA3 synapses. With uCA3 > 0, D shows that mean spiking phase differences appear in response to the two odors.

compare the phase difference of CA1 response between the two. Additionally, a behavioral performance measure equivalent to that presented in Hasselmo et al. (2002) is computed:

$$M = (a_{ECIII}(B) - a_{ECIII}(A)) \cdot \int_0^{2\pi} W * \Theta_{CA3}(t) a_{CA3}(B) dt \tag{23}$$

where $a_{ECIII}(A)$ and $a_{ECIII}(B)$ are the ECIII activity patterns corresponding to odors A and B, respectively, and likewise for $a_{CA3}(B)$. This measure is maximized when the retrieved activity for odor B is most similar to the ECIII activity for that odor and least similar to the ECIII activity for odor A. The performance measure and phase difference from the test phase are repeated for all phases of Φ_{CA3} and Φ_{ECIII} from 0 to 2π, with $\Phi_{LTP} = 0$ as above. The results are shown in Figure 2. Figure 2A shows the performance measure and Figure 2B shows the phase difference for uC = 0, sC = 1, uE = 1, sE = 0, which corresponds to the hippocampal network presented in Hasselmo et al. (2002). Figure 2A is identical to Figure 4 in that paper, showing this performance measure is equivalent to the one used therein. The maximal performance occurs when ECIII input is in phase with LTP and CA3 input is antiphase with it. However, Figure 2B shows that, for the conditions of Figure 2A, there is no phase difference between the responses to the two test odors. This occurs because the

CA3 patterns are identical for both test odors (since there are no stimulus-unique neurons) and thus the CA1 activity driven by CA3 is the same. On the other hand, Figure 2C and D show the case of uC = 100 (to emphasize phase differences), sC = 1, uE = 1, sE = 1. In Figure 2C, nonzero performance is still restricted to the range of phases where ECIII input is in phase with LTP, but the phase of CA3 input is less important. When $\Phi_{LTP} = \pi$, CA3 input undergoes LTD and performance is as in Hasselmo et al. (2002). For other CA3 phases, as long as the learning rate is $\ll 1$, CA3-driven cells in CA1 will have less activity than ECIII driven cells (because the CA3 to CA1 weights will have only increased a small amount in the sample trials) so less interference will occur and retrieval during the test trial will more closely resemble B than A. If the learning rate is 1 this will not occur and the nonzero performance will be restricted to the same CA3 phases as in Figure 2A. Notice that Figure 2D shows a wide range of spiking phase differences for synaptic phase values that result in high performance. This demonstrates that spiking phase differences are unrelated to high performance values for this particular performance measure; it is possible that other memory performance measures might have nonzero performance only when spiking phase differences occur.

Numerical simulations

Comparison of the evolution of synaptic weights in a full network simulation with the results of the analysis above reveals further complexities. In the full network, the change in strength of synapses from sC cells to CA1 cells receiving input from sE cells depends on both those strengths and on the strength of synapses from uCs to the same CA1 cells, because both cause activity in the postsynaptic cell (and likewise for synapses onto uE-receiving CA1 cells). Specifically:

$$\frac{dWss}{dT} = X(Wss * sC + Wus * uC) + Y \tag{24}$$

$$\frac{dWus}{dT} = X(Wss * sC + Wus * uC) + Y \tag{25}$$

One need not solve these equations to see that if $Wss * sC + Wus * uC = -Y/X$, then the change in weights will equal 0 (note that this is not matrix multiplication; we are multiplying the value of a single synapse by the number of cells in the respective populations). Hence, when the sum of synapses from active presynaptic cells onto a postsynaptic cell equals $-Y/X$, the weights will stop growing. When there is only a single synapse, the synapse can grow to $-Y/X$ and it reduces to the case in the previous analysis. Note that if the sum of the weights of active synapses is less than $-Y/X$, the weights will all increment by the same amount, $Y + X(Wss * sC + Wus * uC)$. Also, if the sum of the weights exceeds $-Y/X$, then, because X is negative here, the negative term will exceed the positive term, Y, and the synapses will all weaken by the same amount in the direction of $-Y/X$. This means that this learning rule implements a thresholded, subtractive normalization of active synapses. The normalization is subtractive because the amount of decrement of a synapse is the same for all synapses, regardless of their relative strengths.

To determine the actual weights that Wss and Wus grow to, we must consider what happens before normalization occurs, when the sum of the active synapses is less than $-Y/X$. If we consider a series of alternating presentations of stimuli A and B, we will see that, early on, Wss is incremented every presentation, but Wuas and Wubs are alternately incremented, and so, at any time, Wss will be roughly twice the value of either Wuas or Wubs. This will continue to be the case until the sum of active synapses on a trial equals or exceeds the

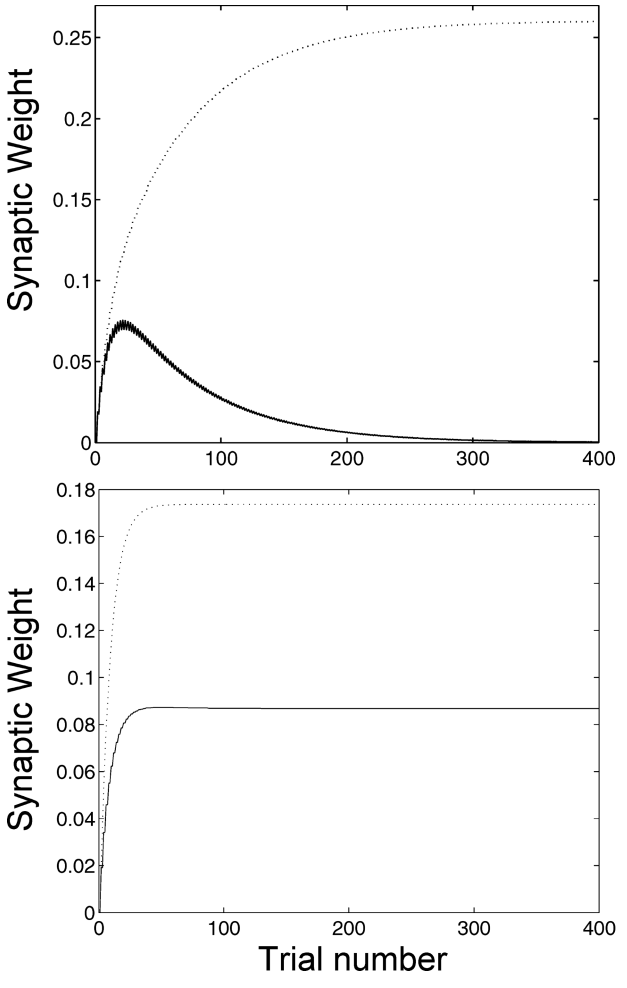

Figure 3. Evolution of synaptic weights onto CA1 cells during alternating presentation of two stimuli. Top: Synapses onto stimulus-unique CA1 cells. The solid line represents a synapse from a shared CA3 cell to a unique CA1 cell. Initially, the weight grows on trials when the CA1-coded stimulus is presented and decreases on trials when it is not presented. Eventually, however, the subtractive normalization described after Equation 27 in the text begins to weaken the synapse and it is driven to 0. The dotted line represents the weight of a synapse from a unique CA3 cell to a unique CA1 cell. Since the cells are unique to a particular stimulus and the two stimuli are alternately presented, the weights only change on every other trial. Because shared to unique synapses eventually have weight 0, the automatic normalization of active synapses allows unique to unique synapses to have higher values before their sum reaches the upper bound. Bottom: synapses onto shared CA1 cells. The solid line represents the weight of a synapse from a unique CA3 cell to a shared CA1 cell. The dotted line represents a connection from a shared CA3 cell to a shared CA1 cell. As described in the text, the shared to shared synapses attain a value twice as large as those of shared to unique synapses because they are incremented on every trial, as opposed to every other trial for shared to unique synapses. (Note: Figures 3–5 have uC = sC = 3, uE = sE = 1.)

threshold of $-Y/X$. At this point, we will have sC Wss synapses and $2 * uC$ Wus synapses in which the sum of the sC Wss synapses and one set of uC Wus synapses equal $-Y/X$, and therefore the value of a single Wss will roughly equal $(-Y/X)*(2/(2sC + uC))$ and that of a Wus will roughly equal $(-Y/X)*(1/(2sC + uC))$. Figure 3, top shows the evolution of Wsu and Wsu toward their asymptotes, while Figure 3, bottom shows the time course of changes in Wus and Wss.

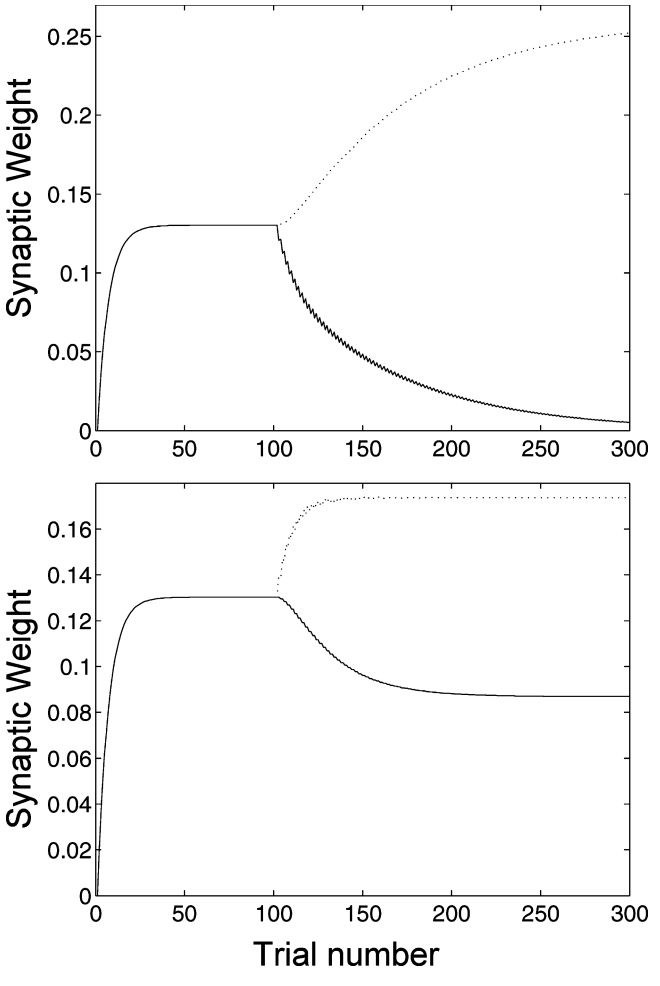

Figure 4. Long-term behavior of synapses does not depend on order of stimulus presentations. Stimulus A was presented 100 times and then stimulus B was introduced, alternating with stimulus A. Top: The values reached by synapses onto unique CA1 cells are identical to those in the top of Figure 3. Bottom: Shared synapses also reach values identical to those shown in the bottom of Figure 3.

These values are not dependent on A and B having been alternately presented. If we consider a large number of repeated presentations of A such that the synapses reach their asymptote and then begin alternately presenting B, we will see that the same values are reached. After the last presentation of A, before B is introduced, Wss and Wuas will have the same value, $(-Y/X)/(sC + uC)$, as Wuas will have increased at the same rate as Wss, while Wubs will be 0 because B has not yet been presented. When B is first presented, only the Wss synapses will be active, and since their sum will be less than $-Y/X$, both Wss and Wubs synapses will increase in value. When A is next presented, Wss will have been increased and the sum of the synapses will exceed $-Y/X$ so all synapses will be depotentiated by the same amount, but Wss will be larger than Wuas. Further presentations of B will increase Wss and Wubs toward their asymptotes, while subsequent presentations of A will continue decreasing Wss and Wuas, until Wss is again twice the value of Wuas and Wubs, as above. This case is shown in Figure 4.

While Wss and Wus behave similarly to the simplified analysis presented earlier, as does Wuu as will be shown, Wsu behaves rather differently. Before the sum of active Wuu and Wsu synapses reach $-Y/X$, they will grow as initially described. When the threshold is reached, however, Wsu begins to decrease to 0 while Wuu approaches an asymptote of $(-Y/X)/uC$. By again considering alternating presentations of A and B and focusing on Wsua, we can see why this happens. Since synaptic weights only change when the presynaptic cell is active, Wuu will only change when A is presented. Wsua, on the other hand, will increase when A is presented, but will decrease on presentations of B. When the sum of the synapses reach $-Y/X$, each presentation of B will weaken Wsua, and each presentation of A will strengthen both, but the difference between Wuu and Wsua will continue to increase until Wsua is 0 and Wuu has taken up all of the available synaptic weight. This reasoning also applies to cases in which there are more than two stimuli being presented. If there are a total of n stimuli in use, Wss and Wus will roughly equal $(-Y/X)*(n/(nsC + uC))$ and $(-Y/X)*(1/(nsC + uC))$, respectively.

This long-term behavior results in an interesting function being performed by the learning rule if we consider a different sort of input to the network. In the network, eventually all synapses will be 0 except those for which the CA3 cell is always active when the ECIII-driven CA1 cell is active. Said another way, a synapse from a CA3 cell to a particular CA1 cell will only be nonzero if the firing of that CA3 cell is fully predictive of the CA1 cell firing. Because the theta rhythm has been recorded in many cortical structures aside from the hippocampus (reviewed in Kahana et al. 2001) and because this learning rule is tied to physiological variables modulated by theta, this function may occur in other brain regions. For instance, it may be useful for learning which qualities of an object correlate with a sweet taste in order to identify ripe fruits. It does, however, have an application in the hippocampus as well. In any network that is continually forming new associations, there must be a mechanism by which older memories can be cleared away when they are no longer needed. In the short term, this learning rule continues to strengthen all synapses involved in experienced stimuli, but as shown in Figure 3, top and Figure 5, the long term behavior drives many synapses back to zero. In this way this rule might aid in recycling synapses that have been frequently used

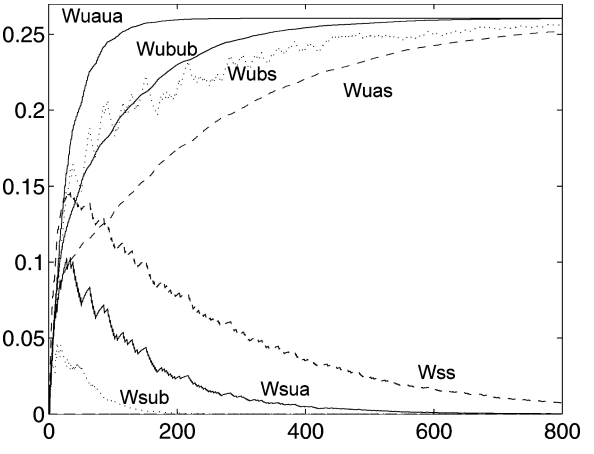

Figure 5. Lack of perfect correlation eventually drives synapses to zero. The simulation is as in Figures 3 and 4, but here the shared CA3 cell was randomly active on only half the presentations of stimulus B. As before, shared to unique synapses Wsua and Wsub approach zero, but now the shared to shared synapses Wss do as well, since the shared presynaptic cell doesn't always predict firing of the shared postsynaptic cell.

(and, therefore, would be thought to represent memories that would have been consolidated in neocortex by that time).

Discussion

Motivated by recent experimental findings reported in Manns et al. (2007), we have shown that the computational model described in Hasselmo et al. (2002) can explain those findings. Specifically, we have demonstrated that the model explains CA1 spiking activity in terms of contributions from sensory input and memory. The memory contribution, and thus the mean phase of CA1 spiking, is dependent on the total duration of exposure to a stimulus and so the difference in mean spiking phase between two stimuli can be positive or negative, depending on the duration of exposure to the two and on the order in which they have been presented.

In the numerical simulations, as in analysis 2, after a sufficient number of presentations of two stimuli, the phase difference in the response to the two will become 0. While many theories stress that the hippocampus is focused on memory for specific episodes and less so on familiarity (Norman & O'Reilly 2003), this model demonstrates the ability to produce familiarity signals for relatively novel stimuli or contexts (where a familiarity signal is simply defined as an aspect of neural activity that correlates with the amount of exposure to a particular stimulus). Though this signal disappears after a relatively small number of presentations, that may be a sufficient amount of time for adjacent cortical structures to form representations of the stimuli and take over the familiarity function. Such a change in response to a novel object was shown in Manns et al. (2007) where, in the novel object task, the mean phase of CA1 activity slowly shifted over nine seconds of exploring a novel object. Further, one efferent of CA1 is perirhinal cortex where Hölscher et al. (2003) showed that, over the course of 10 days, the response of cells to initially novel stimuli slowly became more and more similar to the response to very familiar stimuli, an analogous result to the present one, albeit on a longer timescale.

We have shown that the learning rule first proposed in Hasselmo et al. (2002) can automatically produce upper-bounded synaptic weights and provides subtractive normalization of active synapses. Notably, while many approaches to normalizing synaptic weights require information from all synapses, and are thus non-local, this rule normalizes only active synapses based on the total dendritic activity of the postsynaptic cell, and so normalization here is based only on locally available values (other local rules include, e.g., Oja 1982; Song et al. 2000).

Another difference between this learning rule and other common rules is the separation in time of LTP and LTD such that, with the phases used here, LTP is based on a correlation of pre- and postsynaptic activity (a traditional Hebb rule), while LTD is presynaptically gated and depends on the current strength of the active synapses onto the postsynaptic cell. The rule could be written as:

$$\frac{\mathrm{dW}}{\mathrm{dT}} = pre * post - pre * \mathrm{W} = pre * (post - \mathrm{W})$$

In this form, the rule appears identical to the Instar of Grossberg (1976), or presynaptically-gated, learning rule. The rule presented here is, however, derived entirely from physiological data (LTD/LTP: Hyman et al. 2003, CSD: Brankack et al. 1993) as opposed to a derivation from first principles, as in Grossberg (1976). Further, because the two terms in the learning rule take place at different times, the postsynaptic activity is not the result of activity spreading over the synapses, as would be expected in the Instar and other traditional,

presynaptically-gated rules, but arrives via a separate set of synapses. As a result, while the Instar rule modifies all synapses onto a postsynaptic cell toward the value of the activity of its presynaptic cells, thereby adjusting the synapses so that they are effectively a snapshot of the presynaptic activity, this rule only modifies synapses from active presynaptic cells and adjusts them so that presynaptic cells that are always coactive with the target cell approach a positive bound, while presynaptic cells that are not consistently coactive come to have a strength of zero.

The present learning rule also begs comparison with the oscillating learning algorithm described in Norman et al. (2005). Their algorithm consisted of a learning cycle of fixed network input during which network inhibition first decreases, allowing activation of subthreshold units that are similar to the retrieval cue which undergo LTD to reduce similarity, then inhibition increases, identifying retrieved units that become inactive and thus undergo LTP to increase the internal consistency of the retrieved representation. While lacking a completely biological basis, they demonstrated that this algorithm is a very powerful means of maintaining old memories in the face of new, interfering memories. Despite a similarity in that both rules alternate between LTD and LTP when updating synaptic weights, the two rules perform rather different functions. The oscillating learning algorithm attempts to maintain a complete representation of the memory patterns being stored, while the learning rule analyzed in the present paper produces two stages of function. Initially, the present learning rule updates synaptic weights to reflect the number of times a given stimulus has been experienced, then later, when the weights reach their asymptotes, it maintains nonzero weights only between pairs of cells that are always coactive. This long-term behavior is particularly suited for tasks such as reversal learning or DNMS in which old associations must be unlearned. The two rules are thus both useful, but under distinct circumstances.

This analysis also offers an explanation for the variability in mean phase magnitude and direction reported in Manns et al. (2007) by showing that these two measures are dependent on the animal's history of experienced stimuli. As discussed below Equation 21, the direction of the mean spiking phase difference between two stimuli depends on the total duration of exposure to the two in the model. There are, however, other sources of variability that these analyses didn't consider. One of the assumptions made above is that the number of unique cells for different stimuli are the same (i.e., $uEA = uEB$). In reality, it is likely that different stimuli will have representations of different sizes and that these representations will change over time (e.g., repetition suppression or enhancement shown in Suzuki et al. 1997 and modeled in Sohal & Hasselmo 2002; Bogacz & Brown 2003). It is also possible that the phases of maximal synaptic input themselves vary between animals. The analysis above showed that there is a fairly wide range of input phases for which network performance is stable and if, for example, the two inputs were just less than 180° apart the mean phase would be in one direction, whereas if the inputs were a bit more than 180° apart, the mean phase of CA1 spiking would be in the other direction, while ECIII input alone would have the same phase, thus the difference could be clockwise or counterclockwise. These influences will further affect the mean phase of CA1 population activity.

Although the present analysis has not emphasized the utility of this learning rule and the resulting differences in the mean phase of CA1 population activity on behavior, it is noteworthy that the analyses showed that, given two input patterns to CA1, a greater difference in mean phase response occurs when the two patterns have a lesser amount of overlap. O'Reilly et al. (1994) suggested in a thorough analysis that the anatomy of the hippocampus is conducive to performing either pattern separation or completion, dependent on the similarity of an input pattern to patterns stored in the network. Then it might be the case that, given an array of objects, by making the pattern of novel objects as distinct as possible

from familiar objects, a greater phase difference can result which could act as a signal of novelty.

An interesting feature of this learning rule that this analysis has shown is that the asymptotic synaptic values depend on X and Y, which themselves are reflections of the phases of maximal input to CA1 from CA3 and ECIII, as well as the phase of strongest induction of LTP. The behavior of the network will change depending on those phases. As stated earlier, the bounding behavior of the rule only occurs when X is negative, which is when the phase of maximal LTP induction is more than 90 degrees away from the phase of maximal CA3 input. If those two phases are less than 90 degrees away from each other, X becomes positive and synapses will grow exponentially toward infinity. This happens because maximal CA3 input will occur during a period of LTP in those conditions and the synapses will grow in strength in proportion to their current strength, causing successively larger increments on each theta cycle.

In some neurological disorders, e.g., Alzheimer's disease, neuromodulatory systems (which may underlie the phasic modulations of these physiological variables) become disrupted, which might result in pathological phase relationships which could cause all synapses to saturate at their maximum values (see Hasselmo 1994), potentially causing epileptiform synchronous activity across large populations of cells, or causing an amnesic-like inability to encode or retrieve memories. While the peak phase of induction of plasticity may be difficult to measure in humans, intracranial EEG recordings have been done in preparation for surgery on humans and might allow performance of a CSD analysis across the laminae of the hippocampus in healthy and impaired patients, which could show alterations in the phases of sinks in stratum radiatum and stratum lacunosum-moleculare in CA1. Consistent with this suggestion that normal theta rhythm produces dynamics that are conducive to stability, a number of studies (e.g., Miller et al. 1994; Ferencz et al. 1998) have shown that theta rhythm has seizure-resistant effects in rats. Further, Burchfiel et al. (1979) showed that the kindling model of epilepsy in rats is associated with hippocampal supersensitivity to acetylcholine and Wasterlain et al. (1985) showed that cholinergic agonists in the rat amygdala can induce kindling, suggesting that the protective effects are due to aspects of the theta rhythm beyond simply the release of acetylcholine in the hippocampus.

We have shown that the phase of maximal CA1 population activity reflects the familiarity of a stimulus in the sense that stimuli that have been presented more often produce mean population spiking phases later in the cycle. The question remains as to how, if used as a familiarity signal, this phase change is read out and used to guide behavior. One possibility is that regions receiving output from CA1 may have cells that are sensitive to differences in spike train timing (e.g., the tempotron model of Gutig et al. 2006), converting the temporal code into a population code reflecting the level of familiarity. Alternately, afferent regions of CA1 may be driven by oscillatory inputs of different frequencies or phases. Jensen (2004) showed that an arrangement such as this can produce selective transmission of information at particular phases of oscillations.

Acknowledgements

The authors would like to thank Lisa Giocomo, Randal Koene, Joe Manns, and Matthew McHale for comments on early versions of this manuscript. We are also grateful for the helpful comments offered by the anonymous reviewers. This research was supported by NSF Science of Learning Center SBE 0354378 (CELEST), NIH MH60013, MH61492 and NIDA DA16454 as part of the Collaborative Research in Computational Neuroscience program.

References

Bogacz R, Brown MW. 2003. Comparison of computational models of familiarity discrimination in the perirhinal cortex. Hippocampus 13:494–524.

Brankack J, Stewart M, Fox SE. 1993. Current source density analysis of the hippocampal theta rhythm: Associated sustained potentials and candidate synaptic generators. Brain Res 615:310–327.

Burchfiel JL, Duchowny MS, Duffy FH. 1979. Neuronal supersensitivity to acetylcholine induced by kindling in the rat hippocampus. Science 204:1096–1098.

Buzsaki G. 2002. Theta oscillations in the hippocampus. Neuron 33:325–340.

Ferencz I, Kokaia M, Elmer E, Keep M, Kokaia Z, Lindvall O. 1998. Suppression of kindling epileptogenesis in rats by intrahippocampal cholinergic grafts. Eur J Neurosci 10:213–20.

Golding NL, Staff NP, Spruston N. 2002. Dendritic spikes as a mechanism for cooperative long-term potentiation. Nature 418:326–331.

Grossberg S .1976. Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biol Cybern 23:121–134.

Gutig R, Sompolinsky H. 2006. The tempotron: a neuron that learns spike timing-based decisions. Nat Neurosci 9:420–428.

Hasselmo ME. 1994. Runaway synaptic modification in models of cortex: Implications for Alzheimer[1]s disease. Neural Networks 7:13–40.

Hasselmo ME, Bodelon, C, Wyble BP. 2002. A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. Neural Computation 14:793–817.

Hölscher C, Anwyl R, Rowan MJ. 1997. Stimulation on the positive phase of hippocampal theta rhythm induces long-term potentiation that can be depotentiated by stimulation on the negative phase in area CA1 in vivo. J Neurosci 17:6470–6477.

Hölscher C, Rolls ET, Xiang J. 2003. Perirhinal cortex neuronal activity related to long-term familiarity memory in the macaque. Eur J Neurosci 18:2037–2046.

Hyman JM, Wyble BP, Goyal V, Rossi CA, Hasselmo ME. 2003. Stimulation in hippocampal region CA1 in behaving rats yields LTP when delivered to the peak of theta and LTD when delivered to the trough. J Neurosci 23:11725–11731.

Jensen O. 2004. Computing with oscillations by phase encoding and decoding. Proceedings of the International Joint Conference on Neural Networks; 2004 July 25–29; Budapest.

Judge SJ, Hasselmo ME. 2004. Theta rhythmic stimulation of stratum lacunosum-moleculare in rat hippocampus contributes to associative LTP at a phase offset in stratum radiatum. J Neurophysiol 92:1615–1624.

Kahana MJ, Seelig D, Madsen JR. 2001. Theta returns. Curr Opin Neurobiol 11:739–744.

Lisman J, Spruston N. 2005. Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. Nat Neurosci 8:839–841.

Manns JR, Zilli EA, Ong KC, Hasselmo ME, Eichenbaum HB. 2007. Hippocampal CA1 spiking during encoding and retrieval: relation to theta phase. Neurobiology of Learning and Memory 87:9–20.

Miller JW, Turner GM, Gray BC. 1994. Anticonvulsant effects of the experimental induction of hippocampal theta activity. Epilepsy Res 18:195–204.

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psych Review 110:611–646.

Norman KA, Newman EL, Perotte AJ. 2005. Methods for reducing interference in the Complementary Learning Systems model: oscillating inhibition and autonomous memory rehearsal. Neural Network 18:1212–1228.

Oja E. 1982. A simplified neuron model as a principal component analyzer. J Math Biol 15:267–273.

O'Reilly RC, McClelland JL. 1994. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus 4:661–682.

Pavlides C, Greenstein YJ, Grudman M, Winson J. 1988. Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. Brain Res 439:383–387.

Sohal V, Hasselmo ME. 2000. A model for experience-dependent changes in the responses of inferotemporal neurons. Network: Comp Neural Syst 11:169–190.

Song S, Miller KD, Abbott LF. 2000. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat Neurosci 3:919–926.

Suzuki WA, Miller EK, Desimone R. 1997. Object and place memory in the macaque entorhinal cortex. J Neurophysiol 78:1062–1081.

Wasterlain CG, Farber DB, Fairchild D. 1985. Cholinergic kindling: What has it taught us about epilepsy? J Neural Trans 63:119–132.

Appendix A

The circular mean phase of two positive-shifted sinusoids, $\Theta_A(t) = \frac{a}{2} + \frac{a}{2}\sin(t + \Phi_A)$ and $\Theta_B(t) = \frac{b}{2} + \frac{b}{2}\sin(t + \Phi_B)$, is given by

$$\bar{\Theta} = \tan^{-1}\frac{a\sin\left(\frac{\pi}{2} - \Phi_A\right) + b\sin\left(\frac{\pi}{2} - \Phi_B\right)}{a\cos\left(\frac{\pi}{2} - \Phi_A\right) + b\cos\left(\frac{\pi}{2} - \Phi_B\right)}$$

Proof. We consider a number of vectors from the origin each with angle$0 \leq \Phi_X < 2\pi$ and magnitudes $\Theta_A(\Phi_X)$ and $\Theta_B(\Phi_X)$. The angle of the resultant vector when all of these vectors are summed will be the mean phase of the sum of the two positive-shifted sinusoids. Each vector can be written as a sum of two components: one in the direction of the peak of the sinusoid and one in a direction perpendicular to the peak. Because the sinusoids are symmetric about their peaks, the perpendicular component of any vector will be cancelled out by a vector an equal angle in the other direction from the peak. Then the sum of all the vectors from a particular positive-shifted sinusoid will be a vector in the same direction as the peak. To find the magnitude of that vector, we ignore the phase offset in the sinusoid and, for simplicity, rewrite the sinusoid as a cosine so its peak will be at an angle of 0. For a vector at angle Φ_X, the component of that vector in the direction of the new peak will be $\cos\Phi_X$ times the magnitude of the vector. Thus, we can multiply this positive-shifted cosine by a cosine to find the component of each vector in the direction of angle 0, $R_X = \cos\Phi_X(\frac{a}{2} + \frac{a}{2}\cos\Phi_X)$, where R_X is the magnitude of $\Theta_A(\Phi_X)$ in the direction of angle 0, and likewise for Θ_B. Integrating this equation for each of the two sinusoids over the interval $[0, 2\pi)$ will yield the magnitude of the vector in the direction of the sinusoid's maximum:

$$R = \int_0^{2\pi}\cos\Phi\left(\frac{a}{2} + \frac{a}{2}\cos\Phi\right)d\Phi = \frac{a}{2}\int_0^{2\pi}\cos\Phi d\Phi + \frac{a}{2}\int_0^{2\pi}\cos^2\Phi d\Phi$$

$$= 0 + \frac{a}{2}\Bigg|_0^{2\pi}\left[\frac{\cos\Phi\sin\Phi}{2} + \frac{\Phi}{2}\right] = \frac{a}{4}\Bigg|_0^{2\pi}\left[\cos\Phi\sin\Phi + \Phi\right] = \frac{a\pi}{2}$$

Since the peak of $\Theta_A(t)$ is at $t = \frac{\pi}{2} - \Phi_A$ and likewise for B, we have the magnitude and peak angle of both sinusoids. In polar coordinates these vectors are $(\frac{a\pi}{2}, \frac{\pi}{2} - \Phi_A)$ and $(\frac{b\pi}{2}, \frac{\pi}{2} - \Phi_B)$. To sum them, we will convert to rectangular coordinates, sum the components, then convert back to polar coordinates and will have the mean phase. Converting to rectangular coordinates:

$$\left(\frac{a\pi}{2}\cos\left(\frac{\pi}{2} - \Phi_A\right), \frac{a\pi}{2}\sin\left(\frac{\pi}{2} - \Phi_A\right)\right), \left(\frac{b\pi}{2}\cos\left(\frac{\pi}{2} - \Phi_B\right), \frac{b\pi}{2}\sin\left(\frac{\pi}{2} - \Phi_B\right)\right).$$

Summing the two vectors:

$$\left(\frac{a\pi}{2}\cos\left(\frac{\pi}{2} - \Phi_A\right) + \frac{b\pi}{2}\cos\left(\frac{\pi}{2} - \Phi_B\right), \frac{a\pi}{2}\sin\left(\frac{\pi}{2} - \Phi_A\right) + \frac{b\pi}{2}\sin\left(\frac{\pi}{2} - \Phi_B\right)\right).$$

Finally, the angle of this vector in polar coordinates is:

$$\bar{\Theta} = \tan^{-1}\frac{a\sin\left(\frac{\pi}{2} - \Phi_A\right) + b\sin\left(\frac{\pi}{2} - \Phi_B\right)}{a\cos\left(\frac{\pi}{2} - \Phi_A\right) + b\cos\left(\frac{\pi}{2} - \Phi_B\right)}.$$

■

Appendix B

Because the update of synaptic weights in this model occurs at discrete time points, the learning rule is a difference equation:

$$\mathrm{W(T+1) = W(T) + XW(T) + Y}, \quad \mathrm{W(0) = 0}$$

Solving this recurrence relation yields:

$$\mathrm{W_1(T) = Y \sum_{i=0}^{T-1} \binom{T}{i} X^{T-i-1}}$$

The behavior of W is not clear from this equation, so we wish to find a simpler equation that is equivalent to this one. Euler's forward method for solving differential equations is a method by which a differential equation is solved as if it were a difference equation. The amount of error introduced over h time steps in this method is known to be $O(h^2)$ and thus we can solve the difference equation as a differential equation with a known error bound, the solution being:

$$\mathrm{W_2(T) = \frac{-Y}{X}(1 - e^{TX}) + W(0)e^X}$$

Although the amount of error introduced is potentially large, we can show that, for proper values of X, both solutions have the same asymptote (and thus the error goes to 0) and that the behavior of the solutions is sufficiently similar for the purposes of the analysis in the text.

The original difference equation has six possible asymptotic behaviors:

1. If $\mathrm{Y = 0}$ then W will always be 0.
2. If $\mathrm{X > 0}$ then W will grow to infinity (positive infinity if $\mathrm{Y > 0}$ and negative infinity if $\mathrm{Y < 0}$).
3. If $\mathrm{X = 0}$ then $\mathrm{W(T) = T{*}Y}$.
4. If $\mathrm{-2 < X < 0}$ then W will grow to $\mathrm{-Y/X}$.
5. If $\mathrm{X = -2}$ then $\mathrm{W(1) = Y}$, $\mathrm{W(2) = Y - 2 * Y + Y = 0}$, $\mathrm{W(3) = 0 + -2 * 0 + Y = Y}$, etc, and so W alternates between 0 and Y.
6. If $\mathrm{X < -2}$ then W will grow to infinity (positive infinity if $\mathrm{Y < 0}$ and negative infinity if $\mathrm{Y > 0}$).

The differential solution has four asymptotic behaviors:

1. If $\mathrm{Y = 0}$ then W will always be 0.
2. If $\mathrm{X > 0}$ then W will grow to infinity (positive infinity if $\mathrm{Y > 0}$ and negative infinity if $\mathrm{Y < 0}$).
3. If $\mathrm{X = 0}$ then W(T) will be undefined.
4. If $\mathrm{X < 0}$ then W will grow to $\mathrm{-Y/X}$.

Thus if we restrict $\mathrm{0 < |X| < 2}$, the two solutions have the same asymptotic behavior. For the sine offsets used in this paper, X can only be 0 when the presynaptic cell for the synapse in question is inactive, but the analysis where this is used specifically takes that case into account by altering the differential solution, and thus this is not a problem. To ensure that the absolute value of X is less than 2, one must simply set a learning rate α to an appropriately low value, as when the learning rule is written to include a learning rate:

$$\mathrm{W(T+1) = W(T) + \alpha(XW(T) + Y) = W(T) + \alpha XW(T) + \alpha Y}$$

and so the effective value of X can be scaled appropriately.

To check that the functions behave similarly, we note that the only time the actual value of a synapse is used in the analysis is in equation (21). That result depends only on the fact that, for the phases used in this paper, $W(a) > W(b)$ if $a > b$ and $W(a)$ and $W(b)$ are less than W's asymptote, or, more simply, $W(a + 1) > W(a)$. Since here we have $Y > 0$, If $-2 < X < -1$ then $-Y/X$ will be less than Y and since $W(1) = Y$, the evolution of W will be a damped oscillation around $-Y/X$, contradicting our assumption that $W(a)$ and $W(b)$ will necessarily be less than $-Y/X$, so we will restrict further consideration to the case $-1 < X < 0$. When we rewrite the difference equation as:

$$W(a + 1) - W(a) = XW(a) + Y$$

Since $Y > 0$, if $W(a+1) - W(a) > 0$ then it must be that $XW(a) < Y$, but since we require that $W(a) < \frac{-Y}{X}$ here, clearly that holds and we have $W(a + 1) > W(a)$.

For the differential solution, we similarly require that $W_2(a + 1) > W_2(a)$. We start with:

$$\begin{aligned} W_2(a + 1) - W_2(a) &= \frac{-Y}{X}(1 - e^{X(a+1)}) - \frac{-Y}{X}(1 - e^{aX}) \\ &= \frac{-Y}{X}(1 - e^{X(a+1)} - 1 + e^{aX}) \\ &= \frac{-Y}{X}(e^{aX} - e^{X(a+1)}) \\ &= \frac{-Y}{X}(e^{aX} - e^{aX}e^{X}) \\ &= \frac{-Y}{X}(e^{aX}(1 - e^{X})) \end{aligned}$$

Because $X < 0$, $0 < e^X < 1$ and so $(1 - e^X) > 0$. Also, since $X < 0$, $-Y/X$ is positive and thus the whole expression is positive, so $W_2(a + 1) > W_2(a)$.