Interactions of Excitatory and Inhibitory Feedback Topologies in Facilitating Pattern Separation and Retrieval

Jane X. Wang
1
1 Applied Physics Program, University of Michigan, Ann Arbor, MI 48105, U.S.A.

Michal Zochowski
1, 2, 3
1 Department of Physics, University of Michigan, Ann Arbor, MI 48105, U.S.A.
2 Biophysics Program, University of Michigan, Ann Arbor, MI 48105, U.S.A.
3 Michigan Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48105, U.S.A.

Keywords: memory, hippocampus, inhibitory topology, network, pattern separation

Abstract

Within the brain, the interplay between connectivity patterns of neurons and their spatiotemporal dynamics is believed to be intricately linked to the bases of behavior, such as the process of storing, consolidating, and retrieving memory traces. Memory is believed to be stored in the synaptic patterns of anatomical circuitry in the form of increased connectivity densities within subpopulations of neurons. At the same time, memory recall is thought to correspond to activation of discrete areas of the brain corresponding to those memories. Such regional subpopulations can selectively activate
during memory recall or retrieval, signifying the process of accessing a single memory or concept. It has been shown previously that recovery of single memory activity patterns is mediated by global neuromodulation signifying transition into different cognitive states such as sleep or awake exploration. We examine how underlying topology can affect memory awake activation and sleep reactivation when such memories share increasing proportions of neurons. The results show that while single memory activation is diminished with increased overlap, pattern separation can be recovered by offsetting excitatory associations between two memories with targeted and heterogeneous inhibitory feedback. Such findings point to the importance of excitatory-to-inhibitory current balance at both the global and local level in the context of memory retrieval and replay, and highlight the role of network topology in memory management processes.

1 Introduction

The hippocampus is thought to be both the initial store and regulator of episodic memory, with the CA3 region in particular computationally functioning as a content-addressable memory with an ability to both store and complete patterns (Marr, 1971; Rolls, 1996). Replay during sleep of activity patterns that were experienced during active exploration within the hippocampus plays an important role in memory consolidation, the process of forming hippocampus-independent memory schemas within the cortex (Louie & Wilson, 2001; Gais et al., 2006). The hippocampal CA3 subregion is known to be highly recurrent and is therefore thought to be an associative memory store (Marr, 1971; Treves & Rolls, 1994). This framework is based on the assumption that subsets of neurons which are more highly connected to each other would result in the stable persistent firing of the entire ensemble, corresponding to representations of memories and information encoded through local variations in the densities of synapses and synaptic strengths (Dubnau et al., 2003). Heterogeneity in synaptic patterns introduce spatially varying yet temporally stable patterns of firing, reflecting the underlying connectivity and which represent attractor states of the network (Treves & Rolls, 1994). Working and short-term memory in particular are thought to function through persistent activity of subpopulations of neurons which are either topographically or functionally associated with each other (Durstewitz et al., 2000; Wang, 1999). Previous models
have shown that recurrent networks underlying working and short-term memory allow for focal persistent discharges, representing remembered information, as a result of the network being bistable between a resting state and a persistent memory activation state (Amit & Brunel, 1997; Roxin et al., 2004; Jablonski et al., 2007; Wang et al., 2008).

In dealing with complex episodic memories which can comprise many hierarchies of substructure and modularity, an important issue in the storage and retrieval process concerns how to deal with representations which are overlapped. Correlation or overlap between patterns tends to degrade pattern separation and the selective retrieval of single memories, and several theoretical frameworks have been formulated to address this issue.

It has been suggested that mossy fiber recruitment of inhibitory interneurons (Mori et al., 2007) serves as a mechanism for pattern separation within the CA3 (Myers & Scharfman, 2009; Leutgeb et al., 2007). Experimental evidence exists which shows that excitation is frequently balanced on both a global and local level by inhibition, such as in the local dendritic branches of hippocampal neurons (Liu, 2004) and in neocortical dynamics (Haider & McCormick, 2009). Theoretical models have implemented inhibitory feedback as a mechanism of threshold control and stabilization of activity (Hasselmo et al., 1995). Others have shown detailed excitatory-inhibitory current balance to be vital to signal gating (Vogels & Abbott, 2009). Inhibitory feedback therefore increases the dynamic range of the responsiveness of the network and allows for optimal information processing under a variety of sensory input levels. Loss of inhibitory interneurons has been linked to brain disorders such as epilepsy (Epsztein et al., 2006) and schizophrenia (Benes & Berretta, 2001), and inhibitory cells hold significance in spatiotemporal organization involved in binding or memory tasks (Buzsáki, 2006).

Concurrently, neuromodulation, mediated by differential levels of acetylcholine present during different sleep stages and wake-vs-sleep behavior (Hasselmo et al., 1996), has been suggested as a method of preventing runaway learning by separating the encoding phase from the retrieval phase during consolidation (Hasselmo, 1995). Theoretical and experimental investigations have supported the notion of varying input drives throughout consolidation supporting memory encoding (Booth & Poe, 2006; Poe et al., 2000). Amit et al., 1997, conducted a mathematical analysis of spontaneous and stimulated states of attractor networks containing previously learned memory patterns,
but did not consider overlapping pattern representations or the possibility of structured feedback inhibition. More generally, little is known about how specifically different excitatory and inhibitory topologies interact with global top-down signaling to affect pattern retrieval of hierarchically stored memories.

These concepts motivate our exploration of inhibitory feedback and global drives as mechanisms of pattern separation within a multiple, hierarchical memory model. In this paper, we focus on how overlapping concepts with complex substructure can be optimally stored within a neural network and examine the role of topologically structured inhibition in mediating competitive memory retrieval and activation. In particular, we focus on the link between pattern separation and the balance of inhibitory and excitatory current on both a global as well as a spatially local scale. We define three distinct functional states which arise due to interaction of heterogeneous connectivity with global signaling: regime 1 – low-frequency, noisy firing, regime 2 – selective activation of single population subsets reflecting the underlying structural heterogeneities, and regime 3 – global, synchronized bursting behavior (Jablonski et al., 2007; Wang et al., 2008). These three regimes are controlled by a global excitability parameter representing neuromodulatory changes during different sleep and exploratory states. Effects of this parameter on memory storage and retrieval in different network topologies are assessed. We examine the network mechanisms controlling both awake activation and sleep reactivation of stored memory patterns and determine processes which promote the quick pattern separation evident in episodic memory retrieval (O’Reilly & McClelland, 1994). By examining the robustness of various network topologies with regard to the global excitability term, we can infer the stability and importance of various connectivity schemes and how they interact with global modulation.

We first examine how increased distribution and increased overlap between two memories embedded within an excitatory-inhibitory network degrades pattern separation in different global modulatory conditions. Then we characterize the abilities of different inhibitory-to-excitatory connection topologies in counteracting this degradation, highlighting the role of inhibitory interneurons in being able to regulate selective activations of stored memories. Memory retrieval dynamics are further investigated for different network and memory sizes. A key finding is that the amount of pattern separation upon retrieval is heavily dependent upon the balance of excitatory
and inhibitory associative connections between two overlapped memory structures. We show that structure in the pattern of inhibitory-excitatory connections which matches excitatory structure optimizes this current balance and therefore single memory replay. Through relatively simple rules of targeted inhibition, hypothetically possible through spike-timing dependent plasticity (STDP) (Hebb, 1949; Song et al., 2000) and synaptic scaling mechanisms (Turrigiano & Nelson, 2004), networks are able to achieve selective memory replay even when memories are highly associated with each other. This study suggests that inhibitory networks working in close conjunction with excitatory networks can support complex memory retrieval functions, indicating the importance of inhibitory topological structure and excitatory-inhibitory current matching to cognition.

2 Methods

2.1 Network structure and dynamics

The network is composed of two subpopulations, an excitatory subnetwork and an inhibitory subnetwork, each composed of recurrently connected neuronal units as depicted in Figure 1A. As an approximation to complex feature representation, we institute within the $N_e = 1500$ cells of the excitatory network three hierarchical layers of organization. At the bottom level, cells are grouped into small cell assemblies, or clusters, which represent basic, simple features. Cell assemblies (Harris, 2005) are subpopulations of neurons which can activate and display persistent coordinated activity (i.e. reverberation) without direct stimulation, and are therefore ideal for the modeling of memory activation. Persistent network activity is promoted in dynamic small-world networks (Roxin et al., 2004) due to the existence of both local, nearest neighbor connections and longer-range random shortcuts. Neurons within each cluster are therefore coupled in small-world fashion consistent with the Watts-Stogatz model (Watts & Strogatz, 1998), with the probability of a particular link being a random shortcut set at $p_{ex} = 0.1$. In determining locality and nearest neighbors, excitatory neurons are considered to exist on a one-dimensional periodic lattice, but in general we do not attempt to represent relative spatial distances in this model. The total number of links exist-
ing between all cells in a cluster is set at 8% of the total possible (also referred to as “connectivity density” or simply “connectivity”).

At the next level, clusters are grouped into \( M \) memory structures to represent the binding of low-level features into complex higher level concepts or memories. Assignment of clusters to memories is done in a stochastic fashion. Each memory is composed of \( C \) clusters of size \( g = N_M/C \), where \( N_M \) is the total number of neurons within a memory. Clusters can exist in more than one memory to represent common features shared by multiple memories, and the amount of overlap \( L \) is defined as the percentage of total number of neurons in a memory structure \( N_M \) which are also shared with another memory. Figure 1C qualitatively illustrates an example of clusters grouped into \( M = 2 \) memories in just this fashion within the excitatory network. These two subgroups represent memory structures formed through long term potentiation (LTP) processes thought to underlie learning during exploration of novel environments (Mehta et al., 1997; Ekstrom et al., 2001; Davis et al., 2004). The connectivity density between clusters is set at 4% of the total possible number of connections; these links are added in addition to within-cluster connections.

Memory structures become more distributed as the number of neurons in each cluster, \( g \), decreases and the number of clusters per memory increases. For the majority of this paper we will be considering \( M = 2 \) memories, but explore briefly networks with \( M = 3 \) and 4 as well.

At the highest level we have the entire excitatory network, which is linked with an additional uniform background connection probability of 0.5% among all excitatory cells. This background connectivity effectively introduces noise interference which is independent from the hierarchical memory structure. Our model therefore accounts for multiple levels of memory storage and representation by implementing different degrees of association and connection density at different levels.

The inhibitory neuronal subnetwork is composed of \( N_i = 300 \) cells connected in random fashion; each inhibitory neuron is connected to a fraction \( S_i = 10\% \) of the other inhibitory neurons. The excitatory subnetwork is connected to the inhibitory subnetwork in a local fashion, so that each inhibitory neuron receives connections from 15 nearest excitatory neurons (\( S_{ei} = 1\% \) of total possible connections). It should be noted that in this model, “local” does not correspond to anatomically correct or Euclidean
Figure 1: Schematic of model. A) The network is composed of an excitatory subpopulation and a smaller inhibitory subpopulation activated locally by the excitatory network, which is itself inhibited with connections from the inhibitory layer. The excitatory network is composed of clusters each of size $g$ neurons which are connected in small-world fashion with $p = 0.1$ and 8% connectivity density. Clusters within a memory are randomly coupled with a 4% probability. Inhibitory neurons are randomly coupled to each other. B) Formation of inhibitory synaptic connections onto excitatory neurons are determined by a Gaussian probability function with the center $g$ excitatory neurons taken out so that surrounding clusters are suppressed. C) Conceptual illustration of the excitatory network embedded with two memory structures (localized increases in connectivity density), with no overlap (top) and $L=50\%$ overlap (bottom). Memory structures are composed of $C$ clusters, and increased overlap occurs when memories share increased numbers of clusters. D) Targeted, additional inhibitory feedback is created in the form of random inhibitory-to-excitatory connections both from memory structures to the rest of the network as well as vice versa. Inhibitory cells not depicted. Amount of targeted inhibition is tuned by the parameter $S_a$, which is the fraction of total inhibitory connections possible between the memory structure and the rest of the network.
positions, but rather to a specific, preserved ordering of the units in determining their coupling. For instance, locally connected neurons preferentially connect to a specific set of other neurons which are determined to be “local,” whereas randomly connected neurons have no preference. Local coupling from the excitatory layer to the inhibitory layer therefore means that certain inhibitory cells are preferentially associated with and innervated by certain excitatory cells.

Different conditions of inhibitory-to-excitatory layer connections are tested to determine effects of inhibitory topology on replay performance. The network size ratios and connection densities used are chosen to grossly reflect biological distributions and connectivity patterns in the hippocampal CA1 and CA3 regions. Total network sizes ($N_e$ and $N_i$) are varied in a later section to determine the effects of changing memory and network size on dynamics.

We use leaky-integrate-and-fire neurons given by

$$\tau_m \frac{dV^j_{i/e}}{dt} = -\alpha_j V^j_{i/e} + I^j_{i/e} + \sum_k w^{jk} I^k_{syn} + I^j_d$$

(1)

to represent the reduced dynamics of the network elements. The $i/e$ denotes either an inhibitory or excitatory neuron, $V^j_{i/e}$ is the membrane voltage of the $j$-th neuron, $\alpha_j$ is the membrane leak rate constant randomly distributed such that $\alpha_j \in [1, 1.3]$, $I^k_{syn}$ is the incoming current to the $j$-th neuron from the $k$-th neuron, and $w^{jk}$ is the connection strength between neurons $j$ and $k$. $\tau_m = 30$ ms is the membrane time constant controlling the rate at which membrane potential increases as a result of integrated incoming current. The excitatory subnetwork is connected with connection strengths of $w_e = 2$ and the inhibitory subnetwork with $w_i = 2$; the excitatory-to-inhibitory network connections are of strength $w_{ei} = 4$ and inhibitory-to-excitatory network connections $w_{ie} = 3$. Synaptic strengths are chosen to inversely scale with the number of incoming connections so that total possible inhibitory current is roughly balanced with excitatory current. The external current $I^j_{i/e}$ is uniform over the entire inhibitory/excitatory network. $I_i$ is held constant at 5; $I_e$ is allowed to vary in the range 0-10 and functions as a global modulatory parameter that controls network response transitions from low-frequency random activity, to spontaneous activation of discrete network regions, and finally to global bursting. To represent input from other cortical regions or external sensory information, neurons can be additionally stimulated with additional current $I_d$. 
Only neurons within existing, stored memories are driven to represent stimulation of familiar concepts or object representations. Initially we explore the network dynamics of pattern separation, or the ability of primary, activated memories to suppress activation of associated overlapped memories, by stimulating all neurons within one memory with varying input drive $I_d$ and then examining activity patterns of all other memories. We later examine pattern completion dynamics by driving a constant number of neurons while varying the size of memories.

Excitability therefore comprises the two input drives $I_{i/e}$ and $I_d$. Biologically, excitability represents the responsiveness of neurons within the network to both recurrent activity and external input. Such global neuromodulation within the hippocampus has been posited to be mediated by differential levels of acetylcholine present during sleep and awake stages (Hasselmo et al., 1996). Within a certain range of excitability, regions with high local density of connections activate selectively while suppressing the rest of the network. This is due to the fact that heterogeneities allow for additional synaptic transmission within a localized region, amounting to higher effective excitability. This network architecture promotes regional inhibition driven by focal excitation that creates selective, persistent reactivation patterns. For a detailed description, refer to (Jablonski et al., 2007; Wang et al., 2008).

When the membrane potential of a given cell assumes a maximum value of $V_t = 1$, the neuron emits an action potential, its membrane potential is reset to $V_{\text{rest}} = 0$, and the neuron enters a refractory period for $\tau_{\text{refr}} = 10\,\text{ms}$. The synaptic current emitted by spiking neuron $k$ is of the form

$$I_{\text{syn}}^k(t) = \exp\left(-\frac{(t - t_{\text{spike}}^k)}{\tau_s}\right) - \exp\left(-\frac{(t - t_{\text{spike}}^k)}{\tau_f}\right),$$

(2)

where $(t - t_{\text{spike}}^k)$ is the time since neuron $k$ last spiked, $\tau_s = 3\,\text{ms}$ is the slow time constant, and $\tau_f = 0.3\,\text{ms}$ is the fast time constant. Aside from the deterministic input drive received from other cells, all neurons have a $p_{\text{fire}} = 10^{-3}$ probability of firing spontaneously per millisecond, representing nonspecified and stochastic sources of noise not explicitly accounted for in our model, such as variable input from other brain regions or activity due to internal neuronal dynamics not modeled by the integrate-and-fire neuron.
2.2 Topology of inhibitory-to-excitatory connectivity

The inhibitory network is connected to the excitatory network using one of four connectivity conditions: 1) purely random with connectivity $S_r$, 2) purely local with connectivity $S_{le}$, 3) part local and part random inhibition, and 4) local inhibition coupled with additional targeted connections with connectivity $S_a$. Connectivities $S_{le}$ and $S_r$ are calculated as a percentage of the total possible number of inhibitory-to-excitatory connections ($N_e \times N_i$). Local inhibition consists of lateral suppression of locally placed clusters similar to Mexican hat inhibition in visual (Müller et al., 2005), auditory (Wehr & Zador, 2003), and other cortical network models (Wang et al., 2004). Physiologically, lateral inhibition allows for selective tuning of receptor cells, and layer-specific lateral inhibition has been shown to be important to cortical organization (Adesnik & Scanziani, 2010). The excitatory cluster of $g$ neurons most local to and therefore associated with active inhibitory neurons is not inhibited while nearby clusters of excitatory neurons which are most likely associated with other memory structures are suppressed (see Figure 1A). The connection probability between an inhibitory and excitatory cell is represented by a Gaussian of standard deviation $\sigma = g$ excitatory neurons, modified with the center $g$ neurons taken out (Figure 1B), where $g$ is the number of neurons per cluster. For the purely local inhibition condition, we also consider a range of $\sigma$ between $g$ and $7 \times g$ in order to account for intermediate topologies between completely local and completely random. Fully random connectivity assigns connections randomly between inhibitory and excitatory cells according to a uniform probability.

Mixed connectivity topologies consist of both local and nonlocal connections. On top of a baseline level of local connectivity $S_{le} = 10\%$, networks also include either additional random inhibitory-excitatory synapses (at connectivity $S_r$) or additional targeted synapses (at connectivity $S_a$) to test for the effect of selective inhibition between memory structures. The main difference between local and targeted inhibition is that targeted inhibition is selective for excitatory cells according to memory structure, whereas local inhibition is simply selective for clusters of excitatory cells which are deemed “local” according to some specific ordering of cells. Targeted inhibitory feedback effectively counteracts increased association which exists when two memories are overlapped and thus share neurons. These links are added such that inhibitory neurons associated with each memory structure (i.e. locally activated by excitatory
neurons within these memory structures) will synapse randomly onto all other areas of the excitatory network (Figure 1D), so that activation of a single memory will globally suppress the rest of the network and other memories through these inhibitory feedback pathways. Reciprocally, excitatory cells of each memory structure receive the same number of (but not identical) inhibitory connections from non-associated areas of the inhibitory network. We test various amounts of added inhibitory connections, $S_a = 0\%, 2\%, 4\%,$ and $6\%$, where $S_a$ is the percentage of total possible connections between a single memory structure and the rest of the network. The total number of possible connections is given by $N_M \times (N_e - N_M)$.

Comparisons are made between runs which have the same total amount of inhibitory connections, the only difference being the topology of the connections. All simulations are run with a time step of 0.1 ms.

### 2.3 Measures

**Activity overlap** $A$. For a pair of memories $u$ and $v$, we can define a memory activation overlap $A_{u,v}$ from the activity traces $S_{u/v}(i)$ of the nonoverlapped neurons (belonging to only one memory) in $u$ and $v$. In other words, $S_{u/v}(i)$ is the summed and binned activity time trace of each memory less the activity contributed by neurons which are common to both memories as well as the baseline noise:

$$A_{u,v} = 2 \times (1 - \frac{1}{T} \sum_i^T \frac{\max\{S_u(i), S_v(i)\}}{S_u(i) + S_v(i)})$$

where

$$S_x(i) = \sum_{n \in x, n \notin y} (s(n, i) - b * p_{fire}),$$

for $x, y \in \{u, v\}$. The total number of spikes $s(n, i)$ fired by neuron $n$ at time bin $i$ is calculated with a sliding window of $b = 20$ ms with 50% overlap between successive time bins, and $T$ is the total number of time bins for the simulation.

$A_{u,v}$ has been normalized to a range of $[0,1]$, where 0 denotes minimal overlap between two memories. In the case of more than two memories, $A$ is defined as the average calculated from all possible pairs of memories. $A = 0$ would signify that the network activity is entirely due to the activity of one memory, while $A = 1$ denotes that all memories are activating simultaneously at the same amplitude. Activity overlap
measure gives us an idea of how exclusively the network is within one attractor state, as it is a measure of the degree of coactivation of embedded memories. Figure 2B presents five simulation runs and depicts how the value $A$ for this parameter set relates to memory co-activation as excitability is increased. The dip in activity overlap is due to the network being in a noise regime at low excitability, giving an activity overlap of 0.4, a value dependent on the noise level $p_{\text{fire}}$ and the size of the window used to calculate activity levels. As excitability increases, activity overlap dips drastically for the driven network, and less so for the reactivation cases, until reaching full network bursting in which all memories are co-activated and thus give an $A$ of 1. It should be noted that in Figure 2B, the peak in $A$ between regimes 1 and 2 is only observed for certain parameter sets and regimes. For illustration purposes we wanted to include memory activation time traces for all possible dynamical states, even though they are not all typically observed for any given network topology.

**Integrated activity overlap $\mu$.** In the interest of studying the effects on memory retrieval of network topology, we focused on quantifying the selective memory activation regime (regime 2), which exists when global excitability levels are such that only discrete subpopulations of the network are active. This regime changes for different network topologies, and could in fact disappear altogether, so that the network transitions directly from a uniform low activity state to a uniform global bursting state as global excitability, $I_e$, is varied. The robustness of this regime can be studied by examining the variation of $A$ with changes in global excitability. We calculate a measure called integrated activity overlap $\mu$ which represents the area of the $A$ versus $I_e$ curve under 0.4 (illustrated in Figure 2C):

$$
\mu = \sum_{I_e} (0.4 - A(I_e)) \cdot H(0.4 - A(I_e)),
$$

where $H(x)$ is the Heaviside step function, which is 0 for negative $x$ and 1 for positive $x$, and $A(I_e)$ is the activity overlap value for the network at excitability $I_e$. The higher $\mu$ is, the more robust selective memory reactivation is for a particular network configuration. Excitability ranges are chosen so that the network displays both regime 1 and regime 3 dynamics at either end of this range. Other measures were also investigated, such as the absolute minimum $A$ value achieved, or the total range of excitabilities for which $A$
Table 1: List of key parameters and measures.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>Value range</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of excitatory neurons</td>
<td>$N_e$</td>
<td>200 – 2000</td>
<td>1500</td>
</tr>
<tr>
<td>Number of inhibitory neurons</td>
<td>$N_i$</td>
<td>40 – 40</td>
<td>300</td>
</tr>
<tr>
<td>Number of neurons per memory</td>
<td>$N_M$</td>
<td>100 – 1000</td>
<td>750</td>
</tr>
<tr>
<td>Number of memories</td>
<td>$M$</td>
<td>2 – 4</td>
<td>2</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>$C$</td>
<td>5 – 50</td>
<td>5</td>
</tr>
<tr>
<td>Number of neurons per cluster</td>
<td>$g$</td>
<td>15 – 150</td>
<td>150</td>
</tr>
<tr>
<td>Probability of random shortcut within a cluster</td>
<td>$p_{ex}$</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Range of local inhibition</td>
<td>$\sigma$</td>
<td>$g - 7 \times g$</td>
<td>$g$</td>
</tr>
<tr>
<td>Ratio of memory size to excitatory network size</td>
<td>$f$</td>
<td>0.05 – 0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Memory overlap</td>
<td>$L$</td>
<td>0 – 70%</td>
<td>n/a</td>
</tr>
<tr>
<td>Excitability</td>
<td>$I_e$</td>
<td>0 – 10</td>
<td>n/a</td>
</tr>
<tr>
<td>Driving current to one memory</td>
<td>$I_d$</td>
<td>0 – 5</td>
<td>0, 2</td>
</tr>
<tr>
<td>Local inhibitory-to-excitatory connectivity</td>
<td>$S_{ie}$</td>
<td>0 – 16%</td>
<td>10%</td>
</tr>
<tr>
<td>Random inhibitory-to-excitatory connectivity</td>
<td>$S_r$</td>
<td>0 – 16%</td>
<td>0%</td>
</tr>
<tr>
<td>Targeted inhibitory-to-excitatory connectivity</td>
<td>$S_a$</td>
<td>0 – 6%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Variable</th>
<th>Value range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity overlap</td>
<td>$A$</td>
<td>0 – 1</td>
</tr>
<tr>
<td>Integrated activity overlap</td>
<td>$\mu$</td>
<td>n/a</td>
</tr>
<tr>
<td>Associativity</td>
<td>$\rho$</td>
<td>n/a</td>
</tr>
</tbody>
</table>
is below 0.4, but little additional information was obtained since the integrated activity overlap $\mu$ combines both of these factors and thus is fairly well correlated with them.

**Associativity $\rho$.** In order to quantify the net structural association between memories $u$ and $v$, we calculate the associativity $\rho$:

$$\rho = \frac{1}{N_e} \sum_{j \in u} \sum_{k \in v} w_{jk}^e - \frac{1}{N_e} \sum_{j \in u} \sum_{k \in v} w_{jk}^i.$$  \hspace{1cm} (6)

In words, this measure sums up the total weight of connections from inhibitory neurons associated with one memory to excitatory neurons of another memory and subtracts this from the total weight of excitatory connections from one to the other. This allows us to quantify the net possible interaction between two memories and make comparisons between their structure and dynamics.

### 3 Results

Previous work has highlighted the role of recurrent synaptic connections on reverberating patterns of activity (Roxin et al., 2004), as well as the effect of global neuronal excitability on network states. A recurrent network with regional heterogeneities, or localized increases in connectivity densities, displays three regimes of behavior depending on a global control parameter, excitability $I_e$ (Jablonski et al., 2007). Figure 2A illustrates three regimes of behavior as excitability changes from low to high: 1) low-frequency, noisy firing, 2) discrete activation of population subsets, and 3) globally synchronous bursting behavior. Regime 2 is linked with memory reactivation during sleep or memory activation during exploration (with stimulation). In comparison with reactivation, external stimulation of one memory biases the network toward one attractor state, allowing for much more robust regime 2 dynamics, as depicted in Figure 2C. While it may seem obvious that single pattern retrieval is better in the presence of stimulation, our results show that, due to highly recurrent connections, network activity in response to biases in input is distinctly nonlinear.

We implement a simple neural network memory model and explore the limits of the network’s ability to manage overlapped and distributed memories in both the offline and the active waking state. We examine various inhibitory feedback topologies and self-
Figure 2: Dynamics and network regimes. A) Top: Raster plots of the three network regimes - noise (Regime 1, left), single/selective memory activation (Regime 2, middle), and global bursting (Regime 3, right). Two memories with $C = 5$ clusters each are embedded in a network with clusters located in alternating fashion for visualization purposes. A) Bottom: Raster plots of networks for memories stored in $C = 10$ clusters with no overlap (left) and $C = 5$ clusters with $L = 40\%$ overlap (right). Green and blue colors represent spiking activity of memory 1 and memory 2, respectively; gray color represents both memories; black dots, neither. B) Example activity-overlap-excitability curve given for the reactivation condition with random inhibitory feedback topology. Insets: Memory dynamics over time are displayed for characteristic values along the activity overlap curve, showing distinct regimes of activity. Green: memory 1, blue: memory 2. C) Example activity overlap curves during reactivation (no external stimulation) and driving of one memory (current $I_d = 2$), for the case of local inhibitory feedback. Integrated activity overlap $\mu$ quantifies robustness of regime 2 (single memory activation) to variation in excitability $I_e$. All examples given are with $C = 5$ clusters, overlap $L = 0\%$, and density of inhibitory feedback $S_{ie} = 10\%$. 
regulating mechanisms in optimizing memory reactivation and recognition. Our main finding is that the topology of inhibitory feedback is important for maintaining pattern separation when memories are overlapped and that this is due to detailed current balance on a spatial scale in interaction with global modulatory effects.

3.1 Interplay between memory storage and inhibitory feedback topology

In characterizing inhibitory feedback topology, we examine various conditions of memory storage, including increased distribution (smaller clusters $g$ or increased number of clusters $C$) and increased overlap in memory representation ($L$). Four types of inhibitory feedback are tested: 1) local, modified-Gaussian (Figure 1B), 2) random, 3) local combined with random, and 4) local combined with targeted inhibition.

In the case of purely local inhibition, larger clusters tended to promote more single memory activation, as shown in Figure 3 in the left column. For very distributed memories ($C = 50$), the activity overlap versus excitability curve displays a prominent rise before descending to the low values of regime 2 dynamics. This peak occurs as a result of a uniform rise in activity level across the whole network (see Figure 2B) due to either its inability to fall into a single memory attractor state or its remaining in a global activation attractor state (Wong & Wang, 2006). Such dynamics can be attributed to the diffuse nature of excitatory connections as the number of clusters increases and size of the clusters decreases. As excitatory connectivity becomes increasingly global due to increased memory distribution, for moderate levels of excitability $I_e$, current is unable to concentrate within any one discrete area and contributes to a uniform rise in activity, most likely because of the inability of such small clusters to maintain persistent activity (Riecke et al., 2007).

Randomness in inhibitory topology can also induce such dynamics, regardless of the size of clusters (see Figure 3, right column). For the random inhibitory feedback case, this peak in activity overlap between regime 1 and regime 2 increases as cluster size decreases, reaffirming the notion that random topology tends to impede the network from falling into selective memory attractor states.

The ability to recover a single memory, as quantified by integrated overlap $\mu$, dis-
Figure 3: Comparison of network performance between local inhibitory-excitatory topology (left) and random, long-range inhibitory-excitatory topology (right). Activity overlap plotted versus excitability for networks with $C = 3, 5, 15,$ and $50$ clusters in the cases of A) inhibitory connectivity $S_{ie} = 10\%$ (left), $S_r = 10\%$ (right), and B) $S_{ie} = 16\%$ (left), $S_r = 16\%$ (right). C) Plots of integrated activity overlap $\mu$ versus $S_{ie}$ and $S_r$ for different $C$, showing dependence of regime 2 robustness on the strength of inhibition and distribution of patterns stored. Error bars represent standard error of the mean, as for all subsequent figures.
plays different functional relationships with the amount of inhibitory feedback for the two cases, as shown in Figure 3C. For local inhibition, as the number of inhibitory connections increases, $\mu$ generally increases, representing an increase in robustness of regime 2 (single pattern recovery) dynamics across different levels of excitability. However for highly random topologies, either excitatory ($C = 50$ cell clusters) or inhibitory, as with the random inhibition case, $\mu$ begins to fall for higher values of inhibitory-excitatory connectivity, as dynamics are dominated by uniform global increases in activity (see Figure 2B for a typical activity time trace) and the range of excitability for which single memories can activate disappears. This effect is especially striking with highly fragmented or distributed memories, $C = 50$, in combination with random inhibitory feedback.

Next we examined the effects of increased overlap between memories in the case of sleep reactivation (input drive $I_d = 0$) and external driving ($I_d = 2$). Without external stimulation, for both local and random inhibition, the network transitions directly from noise to bursting dynamics without displaying single memory activity for all but 0% overlap, as illustrated in Figure 4 in the left column. It’s clear that as overlap increases, co-activation of stored memories rises sharply, driven by activation of shared neurons. However, with the addition of a relatively small stimulation to one subpopulation of neurons (9% bias in input current), representing stimulation of a single pattern, the network is able to display single memory activation for $L = 20\%$ overlap (see Figure 4, right column). In both driven ($I_d = 2$) and nondriven ($I_d = 0$) cases, as memory patterns become increasingly overlapped, purely random or local inhibitory feedback becomes inadequate to maintain pattern separation, especially in the absence of external driving. In fact, extrapolating from local to random topology by increasing the range of inhibitory-to-excitatory connections defined by $\sigma$ results in little change in the robustness of single memory activation, as shown in Figure 5B.

We next examine more complex connectivities: $S_{le} = 10\%$ local connectivity combined with $S_r = 6\%$ random, and $S_{le} = 10\%$ local connectivity combined with $S_a = 6\%$ targeted inhibitory connectivity (see Figure 1D). As shown in Figure 5, targeted inhibition provides the best single memory activation (low $A$) for all tested ranges of pattern overlap ($L = 0 - 40\%$). In the non-overlapped case, although random inhibition allows for a lower minimum activity overlap $A$, the large peak between the noise
Figure 4: Activity overlap plotted versus excitability $I_e$ for different amounts of memory overlap $L = 0\%$, 20\%, 40\%, and 60\% with A) the local inhibitory feedback and B) random inhibitory feedback. Reactivation dynamics are shown in the left column; driving dynamics with $I_d = 2$ shown in the right column. All simulations run with $C = 5$ clusters and inhibitory connectivity $S_{ie/r} = 10\%$. 
and single memory regime results in a decrease the robustness of regime 2 dynamics. For the high overlap case, only targeted inhibition combined with baseline local topology is able to recover single memory replay.

In order to examine more mechanistically how targeted inhibition enhances pattern separation, we sought to quantify the topology of links between memory patterns. Specifically, the associativity $\rho$ calculates the net weight of excitatory and inhibitory links between two overlapped memories, giving the total potential for association between neurons in one memory with neurons in the other (see Methods Section 2.3). Figure 6A illustrates how associativity varies as the overlap $L$ and percentage of targeted inhibition $S_a$ between two memories change. In general, increasing targeted inhibition lowers associativity, but this effect increasingly diminishes for larger $S_a$ and $L$. This can be understood by considering that the amount of cells in one memory but not in both, which are therefore available for targeted inhibition, decreases as overlap increases, since this inhibition is directed between a memory and the entire rest of the network. Comparing integrated activity overlap $\mu$ with this quantification of topological linkage between stored patterns $\rho$, we find a very clear dependence of the robustness of single memory activation on the storage topology. In Figure 6B, two memories are stored in networks with purely random or local inhibitory topologies at varying levels of connectivity $S_{ie/r}$. We find that as overlap increases, the associativity between them increases because connectivity within a cluster (8%) is significantly higher than connectivity within a memory (4%), and shared clusters would thus contribute large amounts of linked excitation.

When the strength of inhibitory links between two regions is greater than the strength of excitatory links by roughly $10^4$, $\mu$ sharply increases for all connectivity conditions. In general, as the connections between memories are increasingly dominated by inhibition, they display increased separation in their activity, tending to activate in a mutually exclusive manner and exhibit spontaneous symmetry breaking.

Similar dependencies on associativity $\rho$ by integrated activity overlap $\mu$ exist when the number of memories is scaled up to 3 and 4. In Figure 6C, $\rho$ is plotted against $\mu$ for varying numbers of memories stored as well as different amounts of overlap $L$ and targeted inhibition $S_a$. Overlap is defined as the average pairwise overlap between all pairs of memories. We see a similar trend of increased $\mu$ as associativity decreases (i.e.
Figure 5: Addition of targeted connections from the inhibitory subnetwork to the excitatory subnetwork in addition to baseline local inhibition creates optimal separation between memory patterns. A) Comparison of four different inhibitory-excitatory connectivity patterns for reactivation case: 1) local inhibitory connectivity $S_{ie} = 16\%$, 2) random connectivity $S_r = 16\%$, 3) local ($S_{ie} = 10\%$) combined with random ($S_r = 6\%$), and 4) local ($S_{ie} = 10\%$) combined with targeted ($S_a = 6\%$). Left graph is for no overlap, $L = 0\%$, and right graph is with $L = 40\%$ overlap. All four network topologies contain the same number of inhibitory-to-excitatory connections. B) Integrated activity overlap $\mu$ values for each case. For the purely inhibitory connectivity case, different values of $\sigma$ are tested; for mixed connectivities, $\sigma = g$. 
Figure 6: A) Change in associativity $\rho$ with varying levels of overlap $L$ and targeted inhibition $S_a$. $\rho$ is calculated by subtracting the summed weight of inhibitory-to-excitatory connections from the summed weight of excitatory connections between two memories. B-C) Associativity $\rho$ versus integrated activity overlap $\mu$ for various networks. B) Inhibitory-to-excitatory connectivity is either purely local or purely random (data pooled). Two memories stored, $N_e = 1500$ neurons in excitatory network, $N_i = 300$ neurons in inhibitory network, memory overlap $L=0-70\%$, and inhibitory-to-excitatory connectivity density $S_{ie/r}=10-16\%$. Data points are differentiated according the amount of overlap, showing that as pattern overlap increases, integrated activity overlap decreases and associativity increases. C) Inhibitory-excitatory connectivity is local combined with targeted links. Two, three, and four memories stored; $N_e = 2000$; $N_i = 400$; $S_{ie}=10\%$; $L=0-70\%$; and targeted inhibitory-to-excitatory connectivity density $S_a=0-5\%$. Data is plotted three different ways: 1) top left – differentiating between number of memories stored, 2) top right – differentiating between amount of targeted inhibition, and 3) bottom – differentiating between different amounts of memory overlap.
Figure 7: Calculation of integrated activity $\mu$ for various parameters to show dependence of single memory activity (regime 2 dynamics) robustness on A) overlap $L$, B) targeted inhibition connectivity density $S_a$, and C) amount of driving current $I_d$. Unless being explicitly varied, default parameter values are $S_a = 4\%$, $L = 40\%$, and number of clusters per memory $C = 10$.

gets more negative), which holds for 2, 3, and 4 memories stored, although $\mu$ appears to reach lower maximum values for more memories. However, independent of how many memories are stored, there remains a strong dependence of single memory dynamics on associativity. This was verified by conducting a linear regression analysis, which showed that the associativity accounted for 45.2\% of the variance in the integrated activity overlap ($p < 0.001$), while the number of memories only contributed a further 0.3\% explanation of the variance ($p > 0.05$, not statistically significant).

Changing the amount of targeted inhibition results in similar $\rho$ vs. $\mu$ curves, but with different rise times and offsets. Overall single memory activity increases significantly for associativity below $-10^4$, as well as for targeted inhibitory connectivity greater than $S_a = 2\%$, which tends to increase $\mu$ even for higher or positive associativity.

Aside from the connectivity topology of inhibitory feedback, the three main factors determining the range of excitability for which regime 2 dynamics occurs are 1) density of inhibitory feedback connections, 2) amount of overlap between the two memory representations, and 3) amount of external sensory driving, all of which display interactions with each other. Figure 7 summarizes the multivariate effects of different sets of these parameters on single memory activation dynamics. Increased overlap between pattern representations decreases the extent of single memory replay (Figure 7A), but increased
inhibitory feedback counteracts this effect by creating anti-correlations in memory activation and thus resurrecting or broadening the single memory regime, illustrated in Figure 7B. Similarly, increased external driving introduces a bias to one of the memories which allows for more robust single memory activation than during reactivation. However, the ability to separate activation of associated memories is limited in the absence of additional targeted inhibition even with external driving (see Figure 7C); a driving current of $I_d = 3$ is required to recover single memory activation with no additional inhibition, but with targeted inhibition $S_a = 2\%$ only small driving current is required, while with $S_a = 4\%$ no driving current is needed at all.

### 3.2 Varying memory size and pattern completion

We next investigated the effect that the size of a memory $N_M$ has on pattern retrieval dynamics and completion. However, in changing memory size, various other parameters are also affected, such as the fraction of the total network the memory comprises $f = N_M/N_e$ (and subsequently the total size of the excitatory network $N_e$), the size of individual clusters $g$, or number of clusters $C$. Three simulation conditions were run to assess the effects of changing memory size when two parameters are held constant while varying $N_M$: 1) fraction $f = 0.5$ and cluster size $g = 50$ held constant with varying total network size $N_e$ and cluster number $C$, 2) $f = 0.5$ and $C = 5$ held constant with varying $N_e$ and $g$, and 3) $N_e = 2000$ and $g = 50$ held constant with varying $f$ and $C$. External stimulation was held constant for different memory sizes by randomly stimulating 50 neurons in one memory with driving current $I_d = 2$, allowing for the ability to investigate the relative contributions of internally generated activity and external driving on pattern completion dynamics.

As shown in Figure 8A (left), with overlap $L = 0\%$, targeted inhibitory connectivity $S_a = 0\%$, and driving current $I_d = 2$, robustness of single memory activation $R$ initially decreases with memory size and then rises. This initial decrease is due to the fact that for larger memories, the number of stimulated neurons becomes a smaller fraction of the total memory. However, at a certain size, the external stimulation results in generation of self-sustaining internal activity through excitatory feedback loops. This is shown in Figure 8A (right), which depicts the mean ratio $R$ of activation per neuron...
Figure 8: Effects of different memory sizes and ratio of memory to total network size on network performance and single memory activation due to partial stimulation. A) Single memory activation and reactivation as measured by integrated activity overlap $\mu$ and ratio of internal activity to stimulated activity $R$ shows different functional dependence on memory size for three cases: constant cluster number $C = 5$ and ratio $f = N_M/N_e = 0.5$, constant cluster size $g = 50$ neurons and $f = 0.5$, and constant $g = 50$ neurons and $N_e = 2000$ neurons. Integrated activity overlap $\mu$ (left) and ratio $R$ (right) as a function of memory size $N_M$ for overlap $L = 0\%$, targeted inhibition $S_a = 0\%$, and external current $I_d = 2$. B) Comparison of effects of external driving and targeted inhibition on $\mu$ (left) and $R$ (right) for overlap $L = 40\%$, $f = N_M/N_e = 0.5$, and $g = 50$ neurons per cluster. Driving current is fed to 50 randomly chosen neurons in one memory.
generated internally (nondriven neurons) to activation due to stimulated activity (driven neurons) when the network is displaying regime 2 dynamics (defined by activity overlap $A < 0.3$). We see that for memories of size $N_M = 500$ and more, all neurons within a memory are able to be activated (high $R$) despite direct stimulation of only a fraction of the memory, signifying pattern completion. For memories of this critical size or larger, the fraction of the network that the memory comprises $f$ significantly affects the value of integrated activity overlap $\mu$. Specifically, as the fraction $f$ increases and approaches 0.5 (light gray dashed line), $\mu$ rises and approaches the curves corresponding to the constant $f = 0.5$ cases (solid and dotted black lines). For memories of size 500 and $N_e=1500$ (light gray dashed line), there are simultaneously low $\mu$ and high $R$ values, corresponding memory completion as well as coactivation of both memories. Smaller memories tend to have better separation but worse completion, and larger memories have both better separation and higher completion and additionally benefit more much from targeted inhibition.

For all parameter combinations tested, there is little significant difference between holding the cluster size $g$ constant and holding cluster number $C$ constant, indicating that the overall size of the memory and total network size both contribute more to single memory activation than distribution or cluster size individually.

In order to examine the individual effects of external driving and addition of targeted inhibitory connections, we simulated memories overlapped with $L = 40\%$, making up $f = 0.5$ fraction of the excitatory network, and composed of $g = 50$ neurons. As shown in Figure 8B, when 50 neurons are driven in the absence of targeted inhibition, single memory activity now decreases as memories get larger (black solid line) due to coactivation of linked nonstimulated memories. However, in the presence of targeted inhibition but no driving (dotted dark gray line), larger memories perform better. When both are combined (light gray dashed line), external driving appears to dominate activity for smaller memories, while internally generated activity within a single memory boosted by targeted inhibition is prominent for larger memories (see Figure 8B, right). This shows that different mechanisms can contribute to single memory recall or activation, and in varying degrees as a function of memory size and other topological characteristics.
4 Discussion

We have modeled a hierarchical memory storage and retrieval system loosely based on the functions of the hippocampal CA3 region thought to be linked to episodic memory encoding (Rolls, 2007). More complex concepts or memories will tend to share common features, which would be encoded by the same sets of cell assemblies. With standard autoassociative memory models such as those of Hopfield (Hopfield, 1982) or Hebb (Hebb, 1949), this poses a problem because activation of one representation also activates all associated representations simultaneously, resulting in the retrieval of an amalgam of many memories instead of only one. Recurrent networks also frequently suffer runaway synaptic modification due to uncontrolled activity because novel inputs tend to also activate old memories and become co-encoded (Hasselmo et al., 1995). Therefore, additional inhibitory or competitive feedback drives must be in place to facilitate pattern separation and single memory recall. Our results highlight the importance of inhibitory feedback topology in counteracting coactivation of overlapped and hierarchically stored memory patterns.

This study explored retrieval dynamics in the absence of learning, as a first step in understanding the complex role network topology plays in system dynamics. We utilized global excitability as a control parameter to probe the robustness of retrieval to changes in and transitions between global network states such as sleep or receptive, awake states. In the absence of external stimulation, stored memory traces activate spontaneously offline, as has been observed experimentally to happen within the hippocampus during sleep or quiet waking states after awake learning (Louie & Wilson, 2001). Similar dynamics are observed within previously studied attractor state networks (Treves & Rolls, 1994; Wang, 1999; Ben-Yishai et al., 1995). During active exploration, it’s possible to encounter familiar stimuli and preferentially activate the relevant memory pattern in a process of recognition. In our model, this waking state is represented by biasing input drive to one pattern over another, activating inhibitory feedback pathways which suppress the nonstimulated pattern even in cases of higher pattern overlap.

Heterogeneous storage of memory patterns creates localized changes in excitability, and storage of multiple memory patterns creates problems on retrieval because of pattern overlap. Overlapped memories become essentially linked through their common
cell assemblies, which become an effective source of excitation due to their increased numbers of incoming and outgoing excitatory connections. The key finding was that network structure in the inhibitory-excitatory feedback network matching the structure of the excitatory layer served to counteract excitatory association between linked but distinct patterns. Matching incoming inhibition to this increased excitability within these common neurons skews the balance of total current toward inhibition and helps to enhance single pattern recall. This is consistent with findings of excitatory-inhibitory current balance as key to cortical (Vogels & Abbott, 2009; Haider & McCormick, 2009) and hippocampal dynamics Liu (2004). Intuitively, we would expect that only outgoing inhibition from the memory to rest of the network is necessary for selective memory activation and reactivation, but further simulations show that networks with only this type of inhibitory-to-excitatory connectivity tend to consistently activate common neurons across wide ranges of excitability, independently of noncommon neuronal activity. This is due to incoming current within the overlapped portions being heavily skewed toward excitability, while the rest of the network is more suppressed than usual. We posit that maintaining current balance across the entire network is important to selectively retrieving whole patterns. Overall, these results indicate that the dynamics of memory activation are heavily dependent upon both global network excitability as well as interactions between excitatory and inhibitory topology.

More specifically, we show that targeted inhibition is important to maintaining pattern separation by increasing basins of attraction for single memories, while purely random nontargeted inhibition tends to promote uniform reinjection of current which prevents the network from falling into existing attractor states. We investigated the effects of network architecture on dynamics by studying static states involving memory retrieval independently of previous or subsequent learning and information processing. However, it’s important to ask how targeted inhibitory connections can be created on the timescale of memory formation in a biological context. The targeted inhibition presented in this paper is simple enough to be implemented via straightforward activity-dependent learning processes such as STDP (Hebb, 1949; Song et al., 2000) and homeostatic plasticity in the form of synaptic scaling (Turrigiano & Nelson, 2004) of the inhibitory-to-excitatory connections. For instance, upon activation of a subset of excitatory neurons due to external sensory stimulation, two simultaneous hypothetical
processes of plasticity are occurring: 1) downstream activation of coupled inhibitory cells causes depression of local inhibitory-excitatory synapses since these activated inhibitory cells are firing after excitatory cells, and 2) synaptic scaling within the activated excitatory neurons compensates for the overall reduction in inhibition by strengthening all incoming inhibitory synaptic connections. Although synaptic scaling has primarily been studied in excitatory synapses, evidence exists for regulation of GABA receptors and inhibitory synaptic density by manipulation of neuronal activity within hippocampal slice cultures (Marty et al., 2000) and dissociated cortical cultures (Kilman et al., 2002). Through the course of learning multiple memories, preferential strengthening of interneuronal synapses leads to pattern separation of the final embedded memory traces, allowing for the creation of a functional hierarchical memory structure from relatively simple plasticity rules.

It was not the goal to explore network capacity but rather how the network can manage increased overlap between pattern representations. Although our analyses focus mainly on 2-memory networks, the results generalize to larger numbers of memories because we relate pattern retrieval to the total amount of interference from other stored memory patterns. Interference in the form of external excitatory associations can originate from either storage of more memories or larger amounts of overlap from one other memory. We showed that the relationship between associativity and single memory dynamics remain qualitatively the same for 3- and 4-memory networks, highlighting that it is the balance of negative and positive associations between pairs of memories which is important for pattern separation. That there is little effect on performance by the total number of memories indicates that the extent and robustness of regime 2 dynamics, when memories mutually inhibit each other, can be primarily explained by pairwise associative interactions between memories rather than higher order effects involving more than two memories. It’s well known that in associative networks, more stored patterns lead to more overlap and therefore increased interference. Our results show that it is not the amount of overlap, but rather the balance of inhibition and excitation between two memory structures which determines retrieval dynamics.

This current balance, in conjunction with an interplay between recurrent and external current, also plays a role in pattern completion. We characterized the impact of memory size and targeted inhibition on single memory retrieval due to partial stimu-
lation. It was seen that a certain critical memory size was required for internally generated activity and pattern completion, and that the fraction of the entire network that the memory comprises can also affect selective activation. These dynamics therefore do not scale trivially with size, especially in the absence of targeted inhibition, as there exist complex size-dependent recurrent dynamics. However, the addition of targeted inhibition helps to maintain both pattern completion and pattern separation as memories grow larger.

Further, we systematically explored the nonlinear relationship between pattern driving and ensuing network dynamics and show that it depends on many network characteristics such as memory size and targeted inhibition. The highly recurrent connectivity structure of our network is key to pattern completion, but also significantly contributes to the activation of nonstimulated memories. Recurrent activity is highly affected by small biases in input drives to separate memories, but the addition of targeted, recurrent inhibition allows for significant pattern separation with any amount of input driving. Even small amounts of driving and input bias results in substantially better pattern separation and selective memory activation than with no external input, and these benefits are enhanced with targeted inhibition. Cumulatively, these results indicate two important mechanisms for accurate and specific memory retrieval, mechanisms which are synergistic when combined but could also act independently, possibly in different brain states.

While our model is motivated by hippocampal memory dynamics, it is not meant to accurately reproduce the complex dynamics and timescales of memory reactivation during sleep (Louie & Wilson, 2001), but is instead focused on examining the contributions of network topology to associative memory activation. The net structural inhibition between two overlapped memories is highly correlated with pattern separation, indicating an important role of inhibitory feedback morphology in efficient associative memory encoding in the hippocampus. We also suggest that simple rules of STDP and synaptic scaling can be combined to form more optimal learning and encoding strategies which can lead to networks exhibiting complex network topologies. Only global tuning of excitation in the absence of dynamic plasticity was considered in this report. Nevertheless, we are able to characterize complex interactions between both global and spatially patterned sources of current in the context of pattern recovery. Further studies
incorporating plasticity are needed to gain an even fuller understanding of the roles of structure and excitatory-inhibitory current balance in information processing and pattern retrieval.

Acknowledgments

This work was supported by funding from the NIH Molecular Biophysics Training Grant No. T32 GM008270 at the University of Michigan and by NIH Grant No. EB008163 (NIBIB). The authors would like to acknowledge and thank both the Center for the Study of Complex Systems as well as the Center for Advanced Computing at the University of Michigan for use of their computing resources.

References


O’Reilly, R. C. & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage,

Poe, G. R., Nitz, D. A., McNaughton, B. L., & Barnes, C. A. (2000). Experience-
dependent phase-reversal of hippocampal neuron firing during REM sleep. *Brain

Riecke, H., Roxin, A., Madruga, S., & Solla, S. A. (2007). Multiple attractors, long
chaotic transients, and failure in small-world networks of excitable neurons. *Chaos,
17*, 026110.

601–620.

Rolls, E. T. (2007). An attractor network in the hippocampus: theory and neurophysi-


Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: the importance of
NMDA receptors to working memory. *J. Neurosci.*, 19, 9587–9603.


